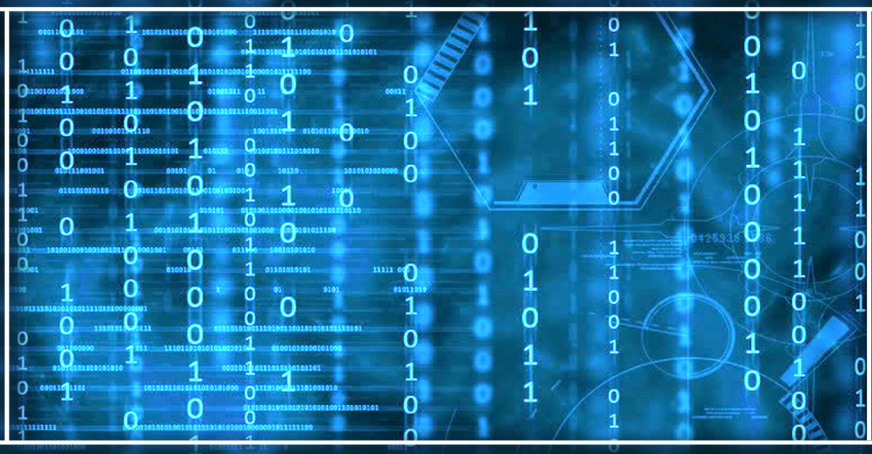
IJACSA
WHERE WISDOM SHARES
International Journal of Advanced Computer Science and Applications

Volume 15 Issue 8
August 2024



ISSN 2156-5570(Online)
ISSN 2158-107X(Print)



www.ijacsa.thesai.org

Editorial Preface

From the Desk of Managing Editor…

It may be difficult to imagine that almost half a century ago we used computers far less sophisticated than current home desktop computers to put a man on the moon. In that 50 year span, the field of computer science has exploded.

Computer science has opened new avenues for thought and experimentation. What began as a way to simplify the calculation process has given birth to technology once only imagined by the human mind. The ability to communicate and share ideas even though collaborators are half a world away and exploration of not just the stars above but the internal workings of the human genome are some of the ways that this field has moved at an exponential pace.

At the International Journal of Advanced Computer Science and Applications it is our mission to provide an outlet for quality research. We want to promote universal access and opportunities for the international scientific community to share and disseminate scientific and technical information.

We believe in spreading knowledge of computer science and its applications to all classes of audiences. That is why we deliver up-to-date, authoritative coverage and offer open access of all our articles. Our archives have served as a place to provoke philosophical, theoretical, and empirical ideas from some of the finest minds in the field.

We utilize the talents and experience of editor and reviewers working at Universities and Institutions from around the world. We would like to express our gratitude to all authors, whose research results have been published in our journal, as well as our referees for their in-depth evaluations. Our high standards are maintained through a double blind review process.

We hope that this edition of IJACSA inspires and entices you to submit your own contributions in upcoming issues. Thank you for sharing wisdom.

Thank you for Sharing Wisdom!

Kohei Arai
Editor-in-Chief
IJACSA
Volume 15 Issue 8 August 2024
ISSN 2156-5570 (Online)
ISSN 2158-107X (Print)

Editorial Board

Editor-in-Chief

Dr. Kohei Arai - Saga University
Domains of Research: Technology Trends, Computer Vision, Decision Making, Information Retrieval, Networking, Simulation

Associate Editors

Alaa Sheta
Southern Connecticut State University
Domain of Research: Artificial Neural Networks, Computer Vision, Image Processing, Neural Networks, Neuro-Fuzzy Systems

Arun D Kulkarni
University of Texas at Tyler
Domain of Research: Machine Vision, Artificial Intelligence, Computer Vision, Data Mining, Image Processing, Machine Learning, Neural Networks, Neuro-Fuzzy Systems

Domenico Ciuonzo
University of Naples, Federico II, Italy
Domain of Research: Artificial Intelligence, Communication, Security, Big Data, Cloud Computing, Computer Networks, Internet of Things

Dorota Kaminska
Lodz University of Technology
Domain of Research: Artificial Intelligence, Virtual Reality

Dr Ronak AL-Haddad
Anglia Ruskin University / Cambridge
Domain of Research : Technology Trends,Communication,Security,Software Engineering and Quality,Computer Networks,Cyber Security,Green Computing,Multimedia Communication,Network Security,Quality of Service

Elena Scutelnicu
"Dunarea de Jos" University of Galati
Domain of Research: e-Learning, e-Learning Tools, Simulation

In Soo Lee
Kyungpook National University
Domain of Research: Intelligent Systems, Artificial Neural Networks, Computational Intelligence, Neural Networks, Perception and Learning

Krassen Stefanov
Professor at Sofia University St. Kliment Ohridski
Domain of Research: e-Learning, Agents and Multi-agent Systems, Artificial Intelligence, e-Learning Tools, Educational Systems Design

Renato De Leone
Università di Camerino
Domain of Research: Mathematical Programming, Large-Scale Parallel Optimization, Transportation problems, Classification problems, Linear and Integer Programming

Xiao-Zhi Gao
University of Eastern Finland
Domain of Research: Artificial Intelligence, Genetic Algorithms

CONTENTS

Paper 1: Ensemble Learning with Sleep Mode Management to Enhance Anomaly Detection in IoT Environment
Authors: Khawlah Harahsheh, Rami Al-Naimat, Malek Alzaqebah, Salam Shreem, Esraa Aldreabi, Chung-Hao Chen
PAGE 1 – 9

Paper 2: Optimizing Low-Resource Zero-Shot Event Argument Classification with Flash-Attention and Global Constraints Enhanced ALBERT Model
Authors: Tongyue Sun, Jiayi Xiao
PAGE 10 – 17

Paper 3: A Personalized Hybrid Tourist Destination Recommendation System: An Integration of Emotion and Sentiment Approach
Authors: Suphitcha Chanrueang, Sotarat Thammaboosadee, Hongnian Yu
PAGE 18 – 27

Paper 4: Automatic Identification and Evaluation of Rural Landscape Features Based on U-net
Authors: Ling Sun, Jun Liu, Yi Qu, Jiashun Jiang, Bin Huang
PAGE 28 – 38

Paper 5: Analysis of Customer Behavior Characteristics and Optimization of Online Advertising Based on Deep Reinforcement Learning
Authors: Zhenyan Shang, Bi Ge
PAGE 39 – 49

Paper 6: Logistics Transportation Vehicle Monitoring and Scheduling Based on the Internet of Things and Cloud Computing
Authors: Kang Wang, Xin Wang
PAGE 50 – 61

Paper 7: Performance and Accuracy Research of the Large Language Models
Authors: Nicoleta Cristina GAITAN
PAGE 62 – 69

Paper 8: Image Generation Using StyleVGG19-NST Generative Adversarial Networks
Authors: Dorcas Oladayo Esan, Pius Adewale Owolawi, Chunling Tu
PAGE 70 – 80

Paper 9: Applied to Art and Design Scene Visual Comprehension and Recognition Algorithm Research
Authors: Yuxin Shi
PAGE 81 – 90

Paper 10: Quantitative Measurement and Preference Research of Urban Landscape Environmental Image Based on Computer Vision
Authors: Yan Wang
PAGE 91 – 100

Paper 11: Clustering Algorithms to Analyse Smart City Traffic Data
Authors: Praveena Kumari M K, Manjaiah D H, Ashwini K M
PAGE 101 – 107

Paper 12: Prediction of Outpatient No-Show Appointments Using Machine Learning Algorithms for Pediatric Patients in Saudi Arabia
Authors: Abdulwahhab Alshammari, Fahad Alotaibi, Sana Alnafrani
PAGE 108 – 116

Paper 13: Performance Optimization of Support Vector Machine with Adversarial Grasshopper Optimization for Heart Disease Diagnosis and Feature Selection
Authors: Nan Tang, Lele Wang, Kangming Li, Zhen Liu, Yanan Dai, Ji Hao, Qingdui Zhang, Huamei Sun, Chunmei Qi
PAGE 117 – 128

Paper 14: Sleep Disorder Diagnosis Through Complex-Morlet-Wavelet Representation Using Bi-GRU and Self-Attention
Authors: Mubarak Albathan
PAGE 129 – 144

Paper 15: Digital Landscape Architecture Design Combining 3D Image Reconstruction Technology
Authors: Chen Chen
PAGE 145 – 154

Paper 16: Software Systems Documentation: A Systematic Review
Authors: Abdullah A H Alzahrani
PAGE 155 – 162

Paper 17: Optimizing Data Security in Computer-Assisted Test Applications Through the Advanced Encryption Standard 256-Bit Cipher Block Chaining
Authors: M. Afridon, Agus Tedyyana, Fajar Ratnawati, Afis Julianto, M. Nur Faizi
PAGE 163 – 170

Paper 18: Diabetes Prediction Using Machine Learning with Feature Engineering and Hyperparameter Tuning
Authors: Hakim El Massari, Noreddine Gherabi, Fatima Qanouni, Sajida Mhammedi
PAGE 171 – 179

Paper 19: Computational Modeling of the Thermally Stressed State of a Partially Insulated Variable Cross-Section Rod
Authors: Zhuldyz Tashenova, Elmira Nurlybaeva, Zhanat Abdugulova, Shirin Amanzholova, Nazira Zharaskhan, Aigerim Sambetova, Anarbay Kudaykulov
PAGE 180 – 189

Paper 20: Performance Analysis of a Hyperledger-Based Medical Record Data Management Using Amazon Web Services
Authors: Mohammed K Elghoul, Sayed F. Bahgat, Ashraf S. Hussein, Safwat H. Hamad
PAGE 190 – 196

Paper 21: Enhancing Business Intelligence with Hybrid Transformers and Automated Annotation for Arabic Sentiment Analysis
Authors: Wael M.S. Yafooz
PAGE 197 – 207

Paper 22: A Method by Utilizing Deep Learning to Identify Malware Within Numerous Industrial Sensors on IoTs
Authors: Ronghua MA
PAGE 208 – 217

Paper 23: Quantifying the Effects of Homogeneous Interference on Coverage Quality in Wireless Sensor Networks
Authors: Qingmiao Liu, Qiang Liu, Minhuan Wang
PAGE 218 – 231

Paper 24: Lightweight and Efficient High-Resolution Network for Human Pose Estimation
Authors: Jiarui Liu, Xiugang Gong, Qun Guo
PAGE 232 – 240

Paper 25: Enhanced Resume Screening for Smart Hiring Using Sentence-Bidirectional Encoder Representations from Transformers (S-BERT)
Authors: Asmita Deshmukh, Anjali Raut
PAGE 241 – 249

Paper 26: Machine Learning Techniques for Protecting Intelligent Vehicles in Intelligent Transport Systems
Authors: Yuan Chen
PAGE 250 – 258

Paper 27: Automation of Book Categorisation Based on Network Centric Quality Management System
Authors: Tingting Liu, Qiyuan Liu, Linya Fu
PAGE 259 – 268

Paper 28: Optimization of Distribution Routes in Agricultural Product Supply Chain Decision Management Based on Improved ALNS Algorithm
Authors: Liling Liu, Yang Chen, Ao Li
PAGE 269 – 278

Paper 29: Harnessing Technology to Achieve the Highest Quality in the Academic Program of University Studies
Authors: Rania Aboalela
PAGE 279 – 292

Paper 30: Enhancing Digital Financial Security with LSTM and Blockchain Technology
Authors: Thanyah Aldaham, Hedi HAMDI
PAGE 293 – 304

Paper 31: Sketch and Size Orient Malicious Activity Monitoring for Efficient Video Surveillance Using CNN
Authors: K. Lokesh, M. Baskar
PAGE 305 – 311

Paper 32: Enhancing Arabic Phishing Email Detection: A Hybrid Machine Learning Based on Genetic Algorithm Feature Selection
Authors: Amjad A. Alsuwaylimi
PAGE 312 – 325

Paper 33: A Feature Interaction Based Neural Network Approach: Predicting Job Turnover in Early Career Graduates in South Korea
Authors: Haewon Byeon
PAGE 326 – 335

Paper 34: A Systematic Review of Virtual Commerce Solutions for the Metaverse
Authors: Ghazala Bilquise, Khaled Shaalan, Manar Alkhatib
PAGE 336 – 346

Paper 35: DIAUTIS III: A Fuzzy and Affective Platform for Obtaining Autism Mental Models and Learning Aids
Authors: Mohamed El Alami, Sara El khabbazi, Fernando de Arriaga
PAGE 347 – 362

Paper 36: Stock Price Forecasting with Optimized Long Short-Term Memory Network with Manta Ray Foraging Optimization
Authors: Zhongpo Gao, Junwen Jing
PAGE 363 – 379

Paper 37: Modified TOPSIS Method for Neutrosophic Cubic Number: Multi-Attribute Decision-Making and Applications to Music Composition Effectiveness Evaluation of Film and Television
Authors: Liang Yang, Jun Zhao
PAGE 380 – 388

Paper 38: Heuristic Intelligent Algorithm-Based Approach for In-Depth Development and Application Analysis of Micro- and Nanoembedded Systems
Authors: Buzhong Liu
PAGE 389 – 398

Paper 39: Optimization of Knitting Path of Flat Knitting Machine Based on Reinforcement Learning
Authors: Tianqi Yang
PAGE 399 – 409

Paper 40: Design and Research of Cross-Border E-Commerce Short Video Recommendation System Based on Multi-Modal Fusion Transformer Model
Authors: Yiran Hu
PAGE 410 – 419

Paper 41: A Hidden Markov Model-Based Performance Recognition System for Marching Wind Bands
Authors: Wei Jiang
PAGE 420 – 431

Paper 42: Fitness Equipment Design Based on Web User Text Mining
Authors: Jinyang Xu, Xuedong Zhang, Xinlian Li, Shun Yu, Yanming Chen
PAGE 432 – 440

Paper 43: Evaluating the Impact of Fuzzy Logic Controllers on the Efficiency of FCCUs: Simulation-Based Analysis
Authors: Harsh Pagare, Kushagra Mishra, Kanhaiya Sharma, Sandeep Singh Rawat, Shailaja Salagrama
PAGE 441 – 449

Paper 44: Hybrid Machine Learning Approach for Real-Time Malicious URL Detection Using SOM-RMO and RBFN with Tabu Search
Authors: Swetha T, Seshaiah M, Hemalatha K L, Murthy S V N, Manjunatha Kumar BH
PAGE 450 – 458

Paper 45: Missing Value Imputation in Data MCAR for Classification of Type 2 Diabetes Mellitus and its Complications
Authors: Anik Andriani, Sri Hartati, Afiahayati, Cornelia Wahyu Danawati
PAGE 459 – 466

Paper 46: Optimizing Dance Training Programs Using Deep Learning: Exploring Motion Feedback Mechanisms Based on Pose Recognition and Prediction
Authors: Yuting Jiao
PAGE 467 – 475

Paper 47: Improved Decision Support System for Alzheimer's Diagnosis Using a Hybrid Machine Learning Approach with Structural MRI Brain Scans
Authors: Niranjan Kumar Parvatham, Lakshmana Phaneendra Maguluri
PAGE 476 – 484

Paper 48: Innovative Melanoma Diagnosis: Harnessing VI Transformer Architecture
Authors: Sreelakshmi Jayasankar, T. Brindha
PAGE 485 – 493

Paper 49: Enhancing Safety for High Ceiling Emergency Light Monitoring
Authors: G. X. Jun, M. Batumalay, C. Batumalai, Prabadevi B
PAGE 494 – 499

Paper 50: Data-Driven Approaches to Energy Utilization Efficiency Enhancement in Intelligent Logistics
Authors: Xuan Long
PAGE 500 – 508

Paper 51: Design and Application of the DPC-K-Means Clustering Algorithm for Evaluation of English Teaching Proficiency
Authors: Mei Niu
PAGE 509 – 518

Paper 52: Enhancing Indonesian Text Summarization with Latent Dirichlet Allocation and Maximum Marginal Relevance
Authors: Muhammad Faisal, Bima Hamdani Mawaridi, Ashri Shabrina Afrah, Supriyono, Yunifa Miftachul Arif, Abdul Aziz, Linda Wijayanti, Melisa Mulyadi
PAGE 519 – 528

Paper 53: Research on Traffic Flow Prediction Using the MSTA-GNet Model Based on the PeMS Dataset
Authors: Deng Cong
PAGE 529 – 539

Paper 54: Laboratory Abnormal Behavior Recognition Method Based on Skeletal Features
Authors: Dawei Zhang
PAGE 540 – 550

Paper 55: Twitter Truth: Advanced Multi-Model Embedding for Fake News Detection
Authors: Yasmine LAHLOU, Sanaa El FKIHI, Rdouan FAIZI
PAGE 551 – 560

Paper 56: Protein-Coding sORFs Prediction Based on U-Net and Coordinate Attention with Hybrid Encoding
Authors: Ziling Wang, Wenxi Yang, Zhijian Qu
PAGE 561 – 572

Paper 57: An Improved YOLOv8 Method for Measuring the Body Size of Xinjiang Bactrian Camels
Authors: Yue Peng, Alifu Kurban, Mengmei Sang
PAGE 573 – 580

Paper 58: Towards Secure Internet of Things-Enabled Healthcare: Integrating Elliptic Curve Digital Signatures and Rivest Cipher Encryption
Authors: Longyang Du, Tian Xie
PAGE 581 – 589

Paper 59: Dose Archiving and Communication System in Moroccan Healthcare: A Unified Approach to X-Ray Dose Management and Analysis
Authors: Lhoucine Ben Youssef, Abdelmajid Bybi, Hilal Drissi, El Ayachi Chater
PAGE 590 – 601

Paper 60: UTAUT Model for Digital Mental Health Interventions: Factors Influencing User Adoption
Authors: Mohammed Alojail
PAGE 602 – 610

Paper 61: ResNet50 and GRU: A Synergistic Model for Accurate Facial Emotion Recognition
Authors: Shanimol. A, J Charles
PAGE 611 – 620

Paper 62: Efficient Parallel Algorithm for Extracting Fuzzy-Crisp Formal Concepts
Authors: Ebtesam Shemis, Arabi Keshk, Ammar Mohammed, Gamal Elhady
PAGE 621 – 630

Paper 63: Automatic Plant Disease Detection System Using Advanced Convolutional Neural Network-Based Algorithm
Authors: Sai Krishna Gudepu, Vijay Kumar Burugari
PAGE 631 – 638

Paper 64: Towards Secure Cloud-Enabled Wireless Ad-Hoc Networks: A Novel Cross-Layer Validation Mechanism
Authors: Zhenguo LIU
PAGE 639 – 646

Paper 65: UWB Printed MIMO Antennas for Satellite Sensing System (SRSS) Applications
Authors: Wyssem Fathallah, Chafai Abdelhamid, Chokri Baccouch, Alsharef Mohammad, Khalil Jouili, Hedi Sakli
PAGE 647 – 656

Paper 66: Novel Data-Driven Machine Learning Models for Heating Load Prediction: Single and Optimized Naive Bayes
Authors: Fangyuan Li
PAGE 657 – 668

Paper 67: Facial Expression Real-Time Animation Simulation Technology for General Mobile Platforms Based on OpenGL
Authors: Mingzhe Cao
PAGE 669 – 678

Paper 68: Noise Reduction Techniques in Adas Sensor Data Management: Methods and Comparative Analysis
Authors: Ahmed Alami, Fouad Belmajdoub
PAGE 679 – 695

Paper 69: DMMFnet: A Dual-Branch Multimodal Medical Image Fusion Network Using Super Token and Channel-Spatial Attention
Authors: Yukun Zhang, Lei Wang, Muhammad Tahir, Zizhen Huang, Yaolong Han, Shanliang Yang, Shilong Liu, Muhammad Imran Saeed
PAGE 696 – 705

Paper 70: Deep Learning and Computer Vision-Based System for Detecting and Separating Abnormal Bags in Automatic Bagging Machines
Authors: Trung Dung Nguyen, Thanh Quyen Ngo, Chi Kien Ha
PAGE 706 – 719

Paper 71: Implementing Optimization Methods into Practice to Enhance the Performance of Solar Power Systems
Authors: Luçiana Toti, Alma Stana, Alma Golgota, Eno Toti
PAGE 720 – 728

Paper 72: Combined Framework for Type-2 Neutrosophic Number Multiple-Attribute Decision-Making and Applications to Quality Evaluation of Digital Agriculture Park Information System
Authors: Wei Ji, Ning Sun, Botao Cao, Xichan Mu
PAGE 729 – 742

Paper 73: Using Pretrained VGG19 Model and Image Segmentation for Rice Leaf Disease Classification
Authors: Gulbakhram Beissenova, Almira Madiyarova, Akbayan Aliyeva, Gulsara Mambetaliyeva, Yerzhan Koshkarov, Nagima Sarsenbiyeva, Marzhan Chazhabayeva, Gulnara Seidaliyeva
PAGE 743 – 752

Paper 74: Blockchain-Based Vaccination Record Tracking System
Authors: Shwetha G K, Jayantkumar A Rathod, Naveen G, Mounesh Arkachari, Pushparani M K
PAGE 753 – 759

Paper 75: Enhance the Security of the Cloud Using a Hybrid Optimization-Based Proxy Re-Encryption Technique Considered Blockchain
Authors: Ahmed I. Alutaibi
PAGE 760 – 771

Paper 76: Malicious Website Detection Using Random Forest and Pearson Correlation for Effective Feature Selection
Authors: Esha Sangra, Renuka Agrawal, Pravin Ramesh Gundalwar, Kanhaiya Sharma, Divyansh Bangri, Debadrita Nandi
PAGE 772 – 780

Paper 77: Enhancing Orchard Cultivation Through Drone Technology and Deep Stream Algorithms in Precision Agriculture
Authors: P. Srinivasa Rao, Anantha Raman G R, Madira Siva Sankara Rao, K. Radha, Rabie Ahmed
PAGE 781 – 795

Paper 78: Comparative Analysis of Small and Medium-Sized Enterprises Cybersecurity Program Assessment Model
Authors: Wan Nur Eliana Wan Mohd Ludin, Masnizah Mohd, Wan Fariza Paizi@Fauzi
PAGE 796 – 804

Paper 79: Real-Time Robotic Force Control for Automation of Ultrasound Scanning
Authors: Ungku Muhammad Zuhairi Ungku Zakaria, Seri Mastura Mustaza, Mohd Hairi Mohd Zaman, Ashrani Aizzuddin Abd Rahni
PAGE 805 – 812

Paper 80: Deep Learning Model for Enhancing Automated Recycling Machine with Incentive Mechanisms
Authors: Razali Tomari, Aeslina Abdul Kadir, Wan Nurshazwani Wan Zakaria, Dipankar Das, Muhamad Bakhtiar Azni
PAGE 813 – 819

Paper 81: A Hybrid of Extreme Learning Machine and Cellular Neural Network Segmentation in Mangrove Fruit Classification
Authors: Romi Fadillah Rahmat, Opim Salim Sitompul, Maya Silvi Lydia, Fahmi, Shifani Adriani Ch, Pauzi Ibrahim Nainggolan, Riza Sulaiman
PAGE 820 – 828

Paper 82: Evaluating the Impact of Yoga Practices to Improve Chronic Venous Insufficiency Symptoms: A Classification by Gaussian Process
Authors: Feng Yun Gou
PAGE 829 – 840

Paper 83: A Semantic Segmentation Method for Road Scene Images Based on Improved DeeplabV3+ Network
Authors: Lihua Bi, Xiangfei Zhang, Shihao Li, Canlin Li
PAGE 841 – 849

Paper 84: Enhancing Tuberculosis Diagnosis and Treatment Outcomes: A Stacked Loopy Decision Tree Approach Empowered by Moth Search Algorithm Optimization
Authors: Huma Khan, Mithun D'Souza, K. Suresh Babu, Janjhyam Venkata Naga Ramesh, K. R. Praneeth, Pinapati Lakshmana Rao
PAGE 850 – 859

Paper 85: Complex Environmental Localization of Scenic Spots by Integrating LANDMARC Localization System and Traditional Location Fingerprint Localization
Authors: Shasha Song, Cong Li
PAGE 860 – 870

Paper 86: Development of a 5G-Optimized MIMO Antenna with Enhanced Isolation Using Neutralization Line and SRRs Metamaterials
Authors: Chaker Essid, Linda Chouikhi, Alsharef Mohammad, Bassem Ben Salah, Hedi Sakli
PAGE 871 – 883

Paper 87: Deep Learning Fusion for Intracranial Hemorrhage Classification in Brain CT Imaging
Authors: Padma Priya S. Babu, T. Brindha
PAGE 884 – 894

Paper 88: Romanian Sign Language and Mime-Gesture Recognition
Authors: Enachi Andrei, Turcu Cornel, George Culea, Sghera Bogdan Constantin, Ungureanu Andrei Gabriel
PAGE 895 – 902

Paper 89: Multiclass Osteoporosis Detection: Enhancing Accuracy with Woodpecker-Optimized CNN-XGBoost
Authors: Mithun D'Souza, Divya Nimma, Kiran Sree Pokkuluri, Janjhyam Venkata Naga Ramesh, Suresh Babu Kondaveeti, Lavanya Kongala
PAGE 903 – 914

Paper 90: Attention-Based Joint Learning for Intent Detection and Slot Filling Using Bidirectional Long Short-Term Memory and Convolutional Neural Networks
Authors: Yusuf Idris Muhammad, Naomie Salim, Sharin Hazlin Huspi, Anazida Zainal
PAGE 915 – 922

Paper 91: Attention-Based Deep Learning Approach for Pedestrian Detection in Self-Driving Cars
Authors: Wael Ahmad AlZoubi, Girish Bhagwant Desale, Sweety Bakyarani E, Uma Kumari C R, Divya Nimma, K Swetha, B Kiran Bala
PAGE 923 – 932

Paper 92: Cryptographic Techniques in Digital Media Security: Current Practices and Future Directions
Authors: Gongling ZHANG
PAGE 933 – 941

Paper 93: Detecting Online Gambling Promotions on Indonesian Twitter Using Text Mining Algorithm
Authors: Reza Bayu Perdana, Ardin, Indra Budi, Aris Budi Santoso, Amanah Ramadiah, Prabu Kresna Putra
PAGE 942 – 949

Paper 94: Securing RPL Networks with Enhanced Routing Efficiency with Congestion Prediction and Load Balancing Strategy
Authors: Saumya Raj, Rajesh R
PAGE 950 – 961

Paper 95: Optimizing Hyperparameters in Machine Learning Models for Accurate Fitness Activity Classification in School-Aged Children
Authors: Britsel Calluchi Arocutipa, Magaly Villegas Cahuana, Vanessa Huanca Hilachoque, Marco Cossio Bolaños
PAGE 962 – 972

Paper 96: Modeling Micro Traffic Flow Phenomena Based on Vehicle Types and Driver Characteristics Using Cellular Automata and Monte Carlo
Authors: Tri Harsono, Kohei Arai
PAGE 973 – 985

Paper 97: Convolutional Neural Network Model for Cacao Phytophthora Palmivora Disease Recognition
Authors: Jude B. Rola, Jomari Joseph A. Barrera, Maricel V. Calhoun, Jonah Flor Oraño – Maaghop, Magdalene C. Unajan, Joshua Mhel Boncalon, Elizabeth T. Sebios, Joy S. Espinosa
PAGE 986 – 990

Paper 98: Leveraging Mechanomyography Signal for Quantitative Muscle Spasticity Assessment of Upper Limb in Neurological Disorders Using Machine Learning
Authors: Muhamad Aliff Imran Daud, Asmarani Ahmad Puzi, Shahrul Na'im Sidek, Ahmad Anwar Zainuddin, Ismail Mohd Khairuddin, Mohd Azri Abdul Mutalib
PAGE 991 – 1002

Paper 99: Interactive Color Design Based on AR Virtual Implantation Technology Between Users and Artificial Intelligence
Authors: Jun Ma, Ying Chen
PAGE 1003 – 1012

Paper 100: A Comprehensive Authentication Taxonomy and Lightweight Considerations in the Internet-of-Medical-Things (IoMT)
Authors: Azlina binti Ahmadi Julaihi, Md Asri Ngadi, Raja Zahilah binti Raja Mohd Radzi
PAGE 1013 – 1025

Paper 101: Dynamic Simulation and Forecasting of Spatial Expansion in Small and Medium-Sized Cities Using ANN-CA-Markov Models
Authors: Chengquan Gao
PAGE 1026 – 1039

Paper 102: Advanced IoT-Enabled Indoor Thermal Comfort Prediction Using SVM and Random Forest Models
Authors: Nurtileu Assymkhan, Amandyk Kartbayev
PAGE 1040 – 1050

Paper 103: Exploration of Deep Semantic Analysis and Application of Video Images in Visual Communication Design Based on Multimodal Feature Fusion Algorithm
Authors: Yanlin Chen, Xiwen Chen
PAGE 1051 – 1061

Paper 104: A Deep Reinforcement Learning (DRL) Based Approach to SFC Request Scheduling in Computer Networks
Authors: Eesha Nagireddy
PAGE 1062 – 1065

Paper 105: Improving Automatic Short Answer Scoring Task Through a Hybrid Deep Learning Framework
Authors: Soumia Ikiss, Najima Daoudi, Manar Abourezq, Mostafa Bellafkih
PAGE 1066 – 1073

Paper 106: BlockChain and Deep Learning with Dynamic Pattern Features for Lung Cancer Diagnosis
Authors: A. Angel Mary, K. K. Thanammal
PAGE 1074 – 1083

Paper 107: Application of Improved CSA Algorithm-Based Fuzzy Logic in Computer Network Control Systems
Authors: Jianxi Yu
PAGE 1084 – 1094

Paper 108: Application of Sanda-Assisted Teaching System Integrating VR Technology from a 5G Perspective
Authors: Zhaoquan Zhang, Yong Ding
PAGE 1095 – 1107

Paper 109: Data Collection Method Based on Data Perception and Positioning Technology in the Context of Artificial Intelligence and the Internet of Things
Authors: Xinbo Zhao, Fei Fei
PAGE 1108 – 1118

Paper 110: Hyperparameter Optimization in Transfer Learning for Improved Pathogen and Abiotic Plant Disease Classification
Authors: Asha Rani K P, Gowrishankar S
PAGE 1119 – 1140

Paper 111: Design and Application of Intelligent Visual Communication System for User Experience
Authors: Chao Peng
PAGE 1141 – 1148

Paper 112: Synchronous Update and Optimization Method for Large-Scale Image 3D Reconstruction Technology Under Cloud-Edge Fusion Architecture
Authors: Jian Zhang, Jingbin Luo, Yilong Chen
PAGE 1149 – 1160

Paper 113: The Application of Anti-Collision Algorithms in University Records Management
Authors: Ying Wang, Ying Mi
PAGE 1161 – 1171

Paper 114: Advancements in Deep Learning Architectures for Image Recognition and Semantic Segmentation
Authors: Divya Nimma, Arjun Uddagiri
PAGE 1172 – 1185

Paper 115: Optimized Retrieval and Secured Cloud Storage for Medical Surgery Videos Using Deep Learning
Authors: Megala G, Swarnalatha P
PAGE 1186 – 1195

Paper 116: Rolling Bearing Reliability Prediction Based on Signal Noise Reduction and RHA-MKRVM
Authors: Yifan Yu
PAGE 1196 – 1205

Paper 117: EiAiMSPS: Edge Inspired Artificial Intelligence-based Multi Stakeholders Personalized Security Mechanism in iCPS for PCS
Authors: Swati Devliyal, Sachin Sharma, Himanshu Rai Goyal
PAGE 1206 – 1217

Paper 118: Integrated IoT-Driven System with Fuzzy Logic and V2X Communication for Real-Time Speed Monitoring and Accident Prevention in Urban Traffic
Authors: Khadiza Tul Kubra, Tajim Md. Niamat Ullah Akhund, Waleed M. Al-Nuwaiser, Md Assaduzzaman, Md. Suhag Ali, M. Mesbahuddin Sarker
PAGE 1218 – 1226

Paper 119: TGMoE: A Text Guided Mixture-of-Experts Model for Multimodal Sentiment Analysis
Authors: Xueliang Zhao, Mingyang Wang, Yingchun Tan, Xianjie Wang
PAGE 1227 – 1234

Paper 120: A Simple and Efficient Approach for Extracting Object Hierarchy in Image Data
Authors: Saravit Soeng, Vungsovanreach Kong, Munirot Thon, Wan-Sup Cho, Tae-Kyung Kim
PAGE 1235 – 1242

Paper 121: Priority-Based Service Provision Using Blockchain, Caching, Reputation and Duplication in Edge-Cloud Environments
Authors: Tarik CHANYOUR, Seddiq EL KASMI ALAOUI, Mohamed EL GHMARY
PAGE 1243 – 1257

Paper 122: A Data Augmentation Approach to Sentiment Analysis of MOOC Reviews
Authors: Guangmin Li, Long Zhou, Qiang Tong, Yi Ding, Xiaolin Qi, Hang Liu
PAGE 1258 – 1264

Paper 123: Design and Implementation of Style-Transfer Operations in a Game Engine
Authors: Haechan Park, Nakhoon Baek
PAGE 1265 – 1273

Paper 124: Under Sampling Techniques for Handling Unbalanced Data with Various Imbalance Rates: A Comparative Study
Authors: Esraa Abu Elsoud, Mohamad Hassan, Omar Alidmat, Esraa Al Henawi, Nawaf Alshdaifat, Mosab Igtait, Ayman Ghaben, Anwar Katrawi, Mohmmad Dmour
PAGE 1274 – 1284

Paper 125: Multiclass Chest Disease Classification Using Deep CNNs with Bayesian Optimization
Authors: Maneet Kaur Bohmrah, Harjot Kaur
PAGE 1285 – 1300

Paper 126: Preprocessing Techniques for Clustering Arabic Text: Challenges and Future Directions
Authors: Tahani Almutairi, Shireen Saifuddin, Reem Alotaibi, Shahendah Sarhan, Sarah Nassif
PAGE 1301 – 1314

Paper 127: Impact of Emojis Exclusion on the Performance of Arabic Sarcasm Detection Models
Authors: Ghalyah Aleryani, Wael Deabes, Khaled Albishre, Alaa E. Abdel-Hakim
PAGE 1315 – 1322

Paper 128: A Configurable Framework for High-Performance Graph Storage and Mutation
Authors: Soukaina Firmli, Dalila Chiadmi, Kawtar Younsi Dahbi
PAGE 1323 – 1331

Paper 129: Detecting Malware on Windows OS Using AI Classification of Extracted Behavioral Features from Images
Authors: Nooraldeen Alhamedi, Kang Dongshik
PAGE 1332 – 1339

Paper 130: The Impact of Virtual Collaboration Tools on 21st-Century Skills, Scientific Process Skills and Scientific Creativity in STEM
Authors: Nur Atiqah Jalaludin, Mohamad Hidir Mhd Salim, Mohamad Sattar Rasul, Athirah Farhana Muhammad Amin, Mohd Aizuddin Saari
PAGE 1340 – 1347

Paper 131: The Impact of E-Commerce Drivers on the Innovativeness in Organizational Practices
Authors: Abdulghader Abu Reemah A Abdullah, Ibrahim Mohamed, Nurhizam Safie Mohd Satar
PAGE 1348 – 1355

Ensemble Learning with Sleep Mode Management to Enhance Anomaly Detection in IoT Environment

Khawlah Harahsheh[1], Rami Al-Naimat[2], Malek Alzaqebah[3], Salam Shreem[4], Esraa Aldreabi[5], Chung-Hao Chen[6]
Ph.D. Student, Department of Electrical and Computer Engineering, Old Dominion University, Norfolk, VA, 23529 USA[1]
Independent Scholar, Karak, Jordan[2]
Department of Mathematics-College of Science, Imam Abdulrahman Bin Faisal University, Dammam, Saudi Arabia[3]
Basic and Applied Scientific Research Center, Imam Abdulrahman Bin Faisal University, Dammam, Saudi Arabia[3]
Independent Scholar, Chicago, USA[4]
Independent Scholar, New York, USA[5]
Department of Electrical and Computer Engineering, Old Dominion University, Norfolk, VA, 23529 USA[6]

***Abstract*—The rapid proliferation of Internet of Things (IoT) devices has underscored the critical need for energy-efficient cybersecurity measures. This presents the dual challenge of maintaining robust security while minimizing power consumption. Thus, this paper proposes enhancing the machine learning performance through Ensemble Techniques with Sleep Mode Management (ELSM) approach for IoT Intrusion Detection Systems (IDS). The main challenge lies in the high-power consumption attributed to continuous monitoring in traditional IDS setups. ELSM addresses this challenge by introducing a sophisticated sleep-awake mechanism, activating the IDS system only during anomaly detection events, effectively minimizing energy expenditure during periods of normal network operation. By strategically managing the sleep modes of IoT devices, ELSM significantly conserves energy without compromising security vigilance. Moreover, achieving high detection accuracy with limited computational resources poses another problem in IoT security. To overcome this challenge, ELSM employs ensemble learning techniques with a novel voting mechanism. This mechanism integrates the outputs of six different anomaly detection algorithms, using their collective intelligence to enhance prediction accuracy and overall system performance. By combining the strengths of multiple algorithms, ELSM adapts dynamically to evolving threat landscapes and diverse IoT environments. The efficacy of the proposed ELSM model is rigorously evaluated using the IoT Botnets Attack Detection Dataset, a benchmark dataset representing real-world IoT security scenarios, where it achieves an impressive 99.97% accuracy in detecting intrusions while efficiently managing power consumption.**

Keywords*—*IoT; IDS; machine learning; ensemble technique; sleep-awake cycle; cybersecurity; anomaly detection

I. Introduction

The Internet of Things (IoT) encompasses many devices, from small sensors to larger, more complex machines. These devices often have specific characteristics and limitations concerning power capacity and time consumption, which are crucial to consider in their design and deployment. Examples of these limitations include power capacity limitations, time-consuming processes, and capacity constraints [1]. Integrating IoT devices into daily life has raised concerns about power consumption and security. Indeed, many IoT devices run outdated firmware, rely on old legacy protocols, and have constrained computational resources. These devices are prone to failure and are vulnerable to a wide spectrum of anomalies, such as malicious attacks, traffic congestion, connectivity problems, and flash crowds [2].

Nevertheless, the high resource requirements of complex and heavy-weight conventional security mechanisms cannot be afforded by (a) the resource-constrained IoMT edge devices with limited processing power, storage capacity, and battery life, and/or (b) the constrained environment in which The IoMT devices are deployed and interconnected using lightweight communication protocols [3]. Traditional security measures often require continuous device operation, leading to excessive power usage. A solution that allows IoT devices to remain in sleep mode until necessary can significantly reduce energy consumption while maintaining security.

IoT has revolutionized the way people interact with technology, connecting a myriad of devices to the Internet. However, this interconnectivity poses significant security challenges. IoT devices are often resource-constrained and become easy targets for cyber-attacks, making robust security mechanisms crucial for maintaining system integrity and user privacy. The wide range of different communication technologies (e.g., WLANs, Bluetooth, Zigbee) and types of IoMT devices (e.g., medical sensors, actuators) incorporated in IoMT edge networks are vulnerable to various types of security threats. This raises many security and privacy challenges for such networks, as well as for the healthcare systems relying on these networks [4].

Software-defined networking (SDN) is a modern paradigm that enhances network management through its dynamic and programmable architecture [5]. However, SDN lacks inherent security features, and one significant issue that may impede its widespread adoption is the potential for novel assaults [6]. Characteristics such as network programmability and centralized control introduce new faults and vulnerabilities, opening the door to threats that did not previously exist [7, 8]. The literature identifies seven potential attack vectors against SDNs, with three specifics to SDN networks [9]. A key countermeasure to secure SDN is the deployment of Intrusion Detection Systems (IDS), which can identify and mitigate

malicious activities in real-time by monitoring network traffic [10]. IDS plays a pivotal role in identifying and mitigating cyber threats in IoT environments [11]. By monitoring network traffic and analyzing system behavior, IDS can detect potential threats before they cause harm. However, traditional IDS solutions often require significant computational resources, leading to high power consumption and latency, which are impractical for many IoT devices [12].

Traditional IDS solutions are often characterized by their high-power consumption, primarily due to the intensive computational processes involved in monitoring and analyzing network traffic [13]. This poses a significant challenge for IoT environments, where devices are designed to operate with minimal energy use. The deployment of power-intensive IDS can lead to rapid battery depletion, reducing the operational lifespan of IoT devices. This creates a critical trade-off situation: ensuring robust security through comprehensive IDS capabilities versus maintaining the energy efficiency critical to IoT device functionality.

Given the resource constraints of IoT devices, developing Intrusion Detection Systems (IDS) that efficiently manage time and power consumption is critical for practical deployment. This entails ensuring prolonged battery life and quick response times, striking a balance between robust security measures and minimal resource usage. In various domains like risk control, fraud detection, and decision-making, there's a significant focus on identifying unexpected events or patterns from datasets. Both static IoT systems (e.g., smart homes, smart buildings) and dynamic IoT networks (e.g., Vehicular Ad-hoc Networks) incorporate numerous lightweight, resource-constrained devices. Thus, efforts to develop IDS for IoT environments must optimize both intrusion detection effectiveness and resource efficiency to seamlessly integrate and widely adopt IDS within IoT ecosystems [14].

Machine learning (ML) offers promising avenues for enhancing the efficiency of IDS in IoT. By using ML algorithms, IDS can adapt to new threats more efficiently and reduce the time and power required for data processing and threat detection. This paper aims to explore the implementation of machine learning techniques in improving time and power efficiency in IoT intrusion detection systems. Energy efficiency is increased when a device has a sleep-wake mechanism, which enables it to go into sleep mode while not in use. The sleep-wake cycle in IoT devices is a critical feature designed to balance energy consumption with operational functionality. This cycle allows devices to conserve power by entering a low power 'sleep' mode when active operation is not needed and 'waking up' to a fully operational state when required to perform tasks. Additionally, the module integrates an isolation forest with an autoencoder, combining the best features of both techniques to provide enhanced anomaly detection in Internet of Things devices, even in sleep mode.

In this paper, a novel hybrid model designed to conserve power within IoT environments is introduced. This model incorporates a mechanism to keep warning devices inactive until abnormal traffic is detected, thereby reducing unnecessary power consumption. Our proposed module leverages ensemble learning and a strategic sleep-wake protocol to optimize power usage while maintaining robust anomaly detection capabilities.

By integrating multiple machine learning models, such as Logistic Regression (LR), Random Forest (RF), Decision Tree (DT), Support Vector Machine (SVM), K-Nearest Neighbors (KNN), and XGBoost, and using an ensemble learning and voting mechanism, the module enhances detection accuracy. The sleep-wake mechanism ensures that IoT devices remain in a low-power state until an anomaly is detected, significantly conserving energy. The main contributions of this work can be summarized as follows:

- Enhanced anomaly detection with hybrid ensemble learning: A novel hybrid model integrating multiple machine learning models with a voting mechanism is introduced to improve the accuracy of anomaly detection in IoT systems.
- Optimized power consumption with sleep-wake mechanism integration: This allows devices to conserve energy by entering sleep states when not actively needed, thereby extending battery life and device operation time.
- Improved computational efficiency with caching: A caching mechanism for the fastest-performing model within the ensemble is introduced to reduce processing time for future detections and improve real-time application performance.

The paper is structured as follows: Section II provides an overview of related work and compares the main differences between our work and other existing research. In Section III, the research methodology details the experimental configuration, and the system flowchart is presented. The experimental results and analysis are thoroughly described in Section IV. Future work is articulated in Section V, followed by a conclusion in Section VI.

II. Related Work

In recent years, significant advancements have been made in developing Intrusion Detection Systems (IDS) for secure IoT environments. The integration of machine learning techniques has greatly enhanced the ability to detect and mitigate cyber threats. Various approaches have been proposed to improve the detection accuracy and efficiency of IDS, especially using ensemble learning methods.

An ensemble learning approach that integrates logistic regression, decision trees, and gradient boosting to enhance the performance of IDS was proposed in study [15]. This method addresses challenges in accuracy, speed, false alarms, and unknown attack detection. Using the CSE-CIC-IDS2018 dataset and Spearman's rank correlation, 23 out of 80 features were selected. The proposed model achieved 98.8% accuracy, demonstrating its effectiveness in improving data security.

Another ensemble-based intrusion detection model was proposed in study [16] using multiple machine learning techniques, such as Decision Trees, J48, and SVM. Particle swarm optimization was used to select the nine most relevant

features in the KDD99 dataset, resulting in a model with 90% accuracy.

Additionally, a hybrid IDS model based on Naive Bayes and SVM was presented in study [17]. The study normalized and preprocessed a real-time historical log dataset, enhancing the model to achieve 95% accuracy and precision. The addition of session-based features further increased classifier performance.

Performance analysis of multiple classical machine learning algorithms on several ID-based datasets, including CICIDS2018, UNSW-NB15, ISCX2012, NSLKDD, and CIDDS001, was conducted in study [18]. Techniques such as SVM, k-nearest Neighbors, and Decision Trees were deployed, with Decision Trees outperforming other classifiers, achieving detection accuracy rates between 99% and 100% for all datasets.

A lightweight IDS developed using SVM aimed to detect unknown and misuse attempts in IoT networks [19]. Experiments on different functions, such as linear, polynomial, and radial basis, showed reduced processing time and complexity due to selected features. However, the algorithm struggled to detect intrusions without affecting traffic flow rates.

A framework for botnet attack detection using a sequential detection architecture was introduced, reducing processing resource demand through relevant feature selection [20]. The N-BaIoT dataset was used, achieving 99% detection performance with Decision Trees (DT), Naive Bayes (NB), and Artificial Neural Networks (ANN). Hybrid classification was used in each sub-engine to enhance accuracy.

Alzaqebah et al. [21] developed a Network Intrusion Detection System (NIDS) using a modified Grey Wolf Optimization algorithm to enhance the efficiency of IDS. This approach smartly grouped wrapper and filter methods to include informative features in every iteration, combined with an Extreme Learning Machine (ELM) for faster classification.

Additionally, Ugendhar et al. [22] proposed an IDS that uses a deep multilayer classification approach. This system incorporates an autoencoder with a reconstruction feature to perform dimensionality reduction. The developed deep multilayer classification approach uses the autoencoder to reduce the dimensionality of the reconstruction feature, enhancing the efficiency and accuracy of the IDS.

In study [23], Grey Wolf Optimization (GWO) was proposed for feature selection, combined with Particle Swarm Optimization (PSO) to optimize the updating process for each grey wolf position. This hybrid approach harnesses PSO's ability to preserve the individual's best position information, which helps prevent GWO from falling into local optima. The performance of this technique is verified using the NSL-KDD dataset. Classification is conducted using k-means and SVM algorithms, and performance is measured in terms of accuracy, detection rate, false alarm rate, number of features, and execution time.

In another study, Samriya and Kumar [24] used a fuzzy-based Artificial Neural Network (ANN) for developing an IDS. They employed the Spider Monkey Optimization algorithm for dimensionality reduction and tested the model with the NSL-KDD dataset. Moreover, an ensemble approach combining multiple machine learning models, such as Decision Tree (DT), Naive Bayes (NB), and Support Vector Machine (SVM), has been shown to enhance the detection capabilities of IDS. This method leverages the strengths of individual classifiers to improve overall system performance.

An Ensemble learning-based method was proposed in study [25], combining Isolation Forest and Pearson's Correlation Coefficient to reduce computational cost and prediction time. The Random Forest classifier was used to enhance performance. Evaluations on Bot-IoT and NF-UNSW-NB15-v2 datasets showed RF-PCCIF and RF-IFPCC achieving up to 99.99% accuracy and prediction times as low as 6.18 seconds, demonstrating superior performance.

A study proposed in [26] developed an IDS for CAN bus networks using ensemble techniques and the Kappa Architecture. The IDS combines multiple machine learning classifiers to enhance real-time attack detection. Supervised models were developed and improved with ensemble methods. Evaluation of common CAN bus attacks showed the stacking ensemble technique achieving an accuracy of 98.5%.

A novel approach to enhance intrusion detection was proposed in study [27]. It begins with denoising data to address the imbalance, followed by employing the enhanced Crow search algorithm for feature selection. An ensemble of four classifiers then classifies intrusions. Evaluation of NSL-KDD and UNSW-NB15 datasets shows accuracy rates of 99.4% and 99.2%, respectively, highlighting superior performance compared to existing methods.

Table I provides a detailed analysis of how researchers have used machine learning algorithms for intrusion detection across different datasets. It explores the specific techniques employed in each study, the datasets leveraged for training and evaluation, and any noteworthy remarks about the methodology or findings.

While many studies employ classic machine learning techniques, there is a notable lack of approaches addressing power efficiency in IoT environments, which is critical given the resource constraints of IoT devices. Moreover, these methods often fail to incorporate mechanisms for real-time intrusion detection and efficient computational performance, which is crucial for practical deployment in dynamic IoT networks.

The proposed Ensemble Learning with Sleep Mode Management (ELSM) aims to fill these gaps by using a more recent IoT botnet dataset, offering a contemporary perspective on current security challenges. The proposed method achieves a high detection rate by employing an ensemble learning technique with a novel voting mechanism while enhancing robustness and efficiency. This is particularly important in IoT environments where power conservation is crucial; Hence, integrating sleep-awake management optimizes power usage. Additionally, our approach accelerates the detection process by caching the fastest module for subsequent iterations, thereby improving computational efficiency.

TABLE I. A COMPREHENSIVE REVIEW OF MACHINE LEARNING TECHNIQUES FOR INTRUSION DETECTION IN VARIOUS DATASETS

Reference	Dataset	Techniques Used	Acc	Remarks
[15]	CSE-CIC-IDS2018	Logistic Regression, Decision Trees, Gradient Boosting	98.8%	Uses Spearman's rank correlation.
[16]	KDD99	Decision Trees, J48, SVM	90%	Employs Particle Swarm Optimization for feature selection.
[17]	Real-time historical log	Naive Bayes, SVM	95%	Data normalized and preprocessed before applying machine learning algorithms.
[18]	CICIDS2018, UNSW-NB15, ISCX2012, NSL-KDD, CIDDS001	SVM, k-Nearest Neighbors, Decision Trees	between 99% and 100%	Evaluates multiple classical machine learning algorithms on various datasets.
[19]	CICIDS2017	SVM (linear, polynomial, radial basis functions)	98.03%	Focuses on detecting denial-of-service attacks using simple features (e.g., packet arrival rates).
[20]	N-BaIoT	Artificial Neural Network, J48 Decision Tree, Naive Bayes	99%	Reduces processing demands by selecting relevant features.
[21]	NSL-KDD	Grey Wolf Optimization, ELM (Extreme Learning Machine)	99.12%	Focuses on achieving fast classification using ELM.
[22]	NSL-KDD	A deep multilayer classification approach	96.7%	Autoencoder Employed for dimensionality reduction.
[23]	NSL-KDD	The classification is done using the k-means and SVM algorithms	-	Hybrid Grey Wolf Optimizer with Particle Swarm Optimization for feature selection.
[24]	NSL-KDD	Fuzzy ANN (Artificial Neural Network), Spider Monkey Optimization	98.70%	Utilizes fuzzy logic for potentially enhanced accuracy.
[25]	Bot-IoT, NF-UNSW-NB15-v2	Isolation Forest, Pearson's Correlation Coefficient, Random Forest	99.99%	Utilizes Isolation Forest for anomaly detection and feature selection with Random Forest for classification.
[26]	CAN bus attack datasets	Random Forest, Decision Tree, XGBoost	98.5%.	Uses ensemble learning techniques with Kappa Architecture for improved performance.
[27]	NSL-KDD, UNSW-NB15	SVM, k-nearest Neighbors, Random Forest, Long Short-Term Memory (LSTM)	99.4% and 99.2%	Applies data denoising and Crow search algorithm for feature selection and explores LSTM for deep learning.
[28]	CIC-IDS 2017	Improved Golden Jackal Optimizer, ANN	98.60%	Applies deep learning and optimization techniques for intrusion detection in smart city environments.
[29]	NSL-KDD	Random Forest, Naive Bayes, J48	99.10%	Investigates the effectiveness of hybrid classification methods.

III. PROPOSED METHOD

Power management is critical in the world of IoT devices, with a special focus on leveraging sleep mode to enhance energy efficiency. This is particularly crucial for devices operating on limited resources such as batteries. Sleep-wake cycles are essential to prolong operational life and increase longevity by reducing active mode duration, thus lowering maintenance demands and costs. In networked IoT environments, timely activation and communication can mitigate network congestion and minimize communication overhead. Key design considerations include optimizing wake-up frequency to align with specific application needs and power constraints, minimizing wake-up duration to the minimum for task completion before returning to sleep mode, and ensuring device responsiveness and reliability to guarantee timely wakeups.

Anomaly detection in IoT frameworks is crucial for the early identification of issues ranging from minor system faults to major cybersecurity vulnerabilities, thereby enhancing system reliability, security, and operational efficiency. Machine learning is essential in this context, leveraging historical data to detect patterns and anomalies. In IoT, anomalies manifest as deviations from expected behaviors, signaling potential system failures, security breaches, or environmental changes. Anomalies can be categorized into point anomalies, which are significant deviations in individual data points, context anomalies specific to conditions, and collective anomalies, where seemingly normal data points together indicate suspicious activity. Each type requires distinct detection methodologies, with machine learning playing a pivotal role in their identification and resolution.

The proposed hybrid model aims to reduce power consumption in IoT environments by optimizing usage through a strategic sleep-wake protocol activated exclusively during anomaly detection phases. This methodology employs a composite model merging an autoencoder with an isolation forest, offering dual benefits of increased energy efficiency and enhanced detection precision. The model's construction, illustrated in Fig. 1, includes steps designed to conserve power and enhance IDS compatibility with IoT device limitations. Moreover, deploying this model at the edge is preferable to using cloud resources, as the primary goal is power preservation, while cloud data transmission consumes additional energy. Cloud computing may not be entirely suitable for provisioning IoT applications [29], primarily due to connectivity challenges between cloud resources in the core network and edge devices [30].

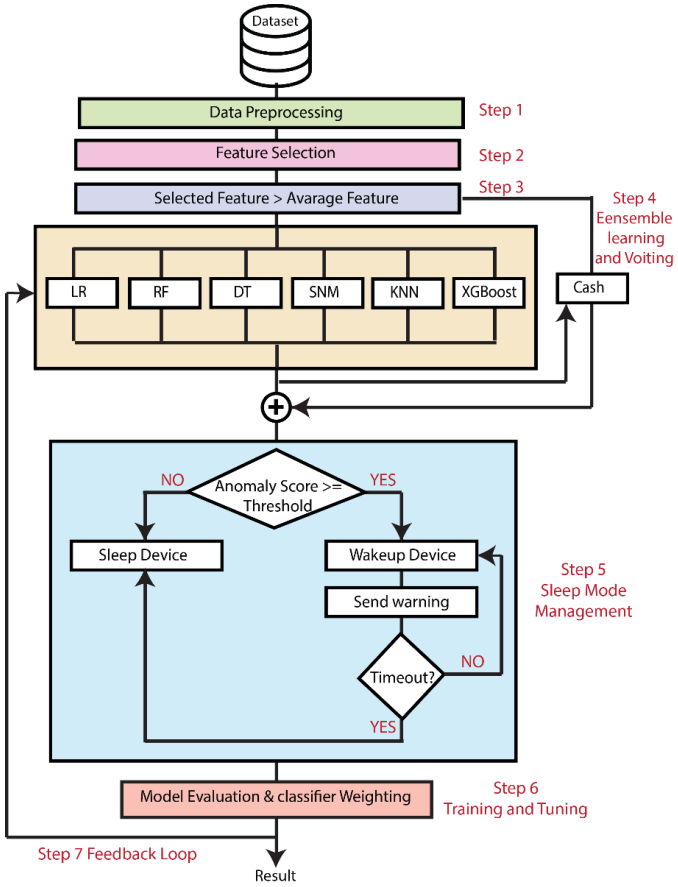


Fig. 1. The flowchart of the ELSM module.

The flowchart in Fig. 1 illustrates a comprehensive methodology for enhancing anomaly detection in IoT environments by leveraging ensemble learning and incorporating a sleep mode management mechanism. The process is divided into several key steps, each contributing to optimizing power efficiency and improving detection accuracy. The steps for Ensemble Learning with Sleep Mode Management (ELSM) are as follows:

1) Data preprocessing: Data collection and preprocessing from IoT sensors and devices are critical steps to ensure the suitability of data for machine learning analysis. Initially, data is gathered from sensors embedded in IoT devices, capturing real-time information about the environment or the device itself. This raw data then undergoes preprocessing, which involves cleaning, filtering, and transforming it into a format suitable for analysis. Tasks in preprocessing may include removing noise, missing handling values, normalizing data, and extracting features. Additionally, data may be aggregated or sampled to reduce dimensionality and enhance computational efficiency. By meticulously preparing the data and ensuring its quality and relevance, subsequent machine learning algorithms can effectively derive meaningful insights and support informed decision-making in IoT applications.

2) Feature selection: In the proposed methodology, a filter method is employed to enhance the performance of the Intrusion Detection System (IDS). Improving speed is crucial for minimizing power consumption, reducing the reliance on extensive computational resources, and achieving faster processing times. Filter methods are noted for their swift execution compared to other feature selection techniques [32].

3) Feature voting: Selected features that surpass a specified average threshold are advanced to the next phase. This ensures that only the most significant features are used, thereby maintaining system efficiency and ensuring optimal feature selection for further analysis and performance optimization. The Feature Voting process involves ranking features based on their contribution to the classification task. Features with important scores above a predefined average threshold are considered significant and selected for further analysis. This reduction is crucial for maintaining high accuracy while ensuring computational efficiency, particularly in resource-constrained IoT environments.

4) Ensemble learning and voting: This step involves training multiple machine-learning models, including Logistic Regression (LR), Random Forest (RF), Decision Tree (DT), Support Vector Machine (SVM), K-Nearest Neighbors (KNN), and XGBoost. These models are combined using ensemble learning techniques to enhance overall prediction accuracy. The ensemble method typically employs a voting mechanism, where each model votes on the presence of an anomaly, and the final decision is based on the majority vote. Each model was evaluated based on two criteria: accuracy and prediction time. For each model, the training time and prediction time per instance are calculated as in Eq. (1), Eq. (2), and Eq. (3) respectively.

$$t_{\text{train},i} = \text{time_end}\, d_{\text{train},i} - \text{time_start}_{\text{train},i} \quad (1)$$

$$t_{\text{pred},i} = \frac{\text{time_end}_{\text{pred},i} - \text{time_start}_{\text{pred},i}}{n_{\text{test}}} \quad (2)$$

$$Accuracy_i = \frac{\sum_{j=1}^{n_{\text{tot}}} \mathbf{1}(\hat{y}_j = y_j)}{n_{\text{test}}} \quad (3)$$

where, 1 is the indicator function that returns 1 if the predicted label $\hat{y}_j$ matches the true label yj, and $n$test is the number of test instances. The weight and the normalized weight for each module were then calculated separately, as shown in Eq. (4) and Eq. (5), respectively.

$$w_i = \frac{Accuracy}{t_{pred,i}} \quad (4)$$

$$w_i' = \frac{w_i}{\sum_{k=1}^{6} w_k} \quad (5)$$

Afterwards, the ensemble model which combines the predictions of all individual models using a weighted voting mechanism was calculated. For a given instance x, the ensemble prediction $\hat{y}_{\text{ensemble}}$ as in Eq. (6):

$$\hat{y}_{\text{ensemble}} = arg\max_{c} \sum_{i=1}^{6} w_i' \cdot P_i(y = c \mid x) \quad (6)$$

Where $(y=c|x)$ is the probability predicted by model mi that the instance x belongs to class c. Finally, the accuracy is calculated for the ensemble techniques as in Eq. (7).

$$Accuracy_{ensemble} = \frac{\sum_{j=1}^{n_{tet}} \mathbf{1}\left(\hat{y}_{ensemble,j}=y_j\right)}{n_{test}} \quad (7)$$

5) Sleep mode management: The sleep-wake cycle plays a crucial role in managing energy consumption in devices, particularly in battery-powered or energy-constrained environments. During sleep mode, the device conserves energy by shutting down non-essential functions, maintaining only vital components in a significantly reduced power state. Triggers such as scheduled timers, external signals such as motion detection, or internal alerts such as critical battery levels prompt the device to transition from sleep to wake mode. In wake mode, the device resumes full functionality, enabling tasks such as sensing, data processing, and communication. Once these tasks are completed, the device can return to sleep mode, ensuring efficient power preservation. This cycle optimizes energy usage, thus enhancing device efficiency and longevity.

6) Training and tuning: Testing and validation of the system under real-world conditions are essential to confirm its effectiveness in anomaly detection and power efficiency. During testing, the system's ability to accurately identify anomalies across various scenarios and environmental conditions is evaluated, providing insights into its robustness and reliability. Additionally, efforts are focused on assessing the system's power efficiency to ensure optimal operation within IoT device constraints, while minimizing energy consumption without compromising detection accuracy. By rigorously testing performance metrics and conducting thorough validation procedures, the system's suitability for deployment in practical applications is affirmed, instilling confidence in its ability to enhance security and efficiency in IoT environments.

7) The feedback loop: The feedback loop for continuous improvement in the system is pivotal for enhancing its performance over time. This iterative process ensures that the system remains agile and responsive to changes in its environment, thereby improving accuracy in anomaly detection. Additionally, the feedback mechanism facilitates self-optimization based on performance metrics and energy consumption data, allowing for adjustments to algorithms and configurations as needed. This iterative feedback loop enhances the system's overall effectiveness while promoting efficient energy utilization, ensuring sustainable operation in IoT environments.

IV. Experimental Results and Discussion

This section presents the experimental results and analysis of the proposed model, which aims to enhance anomaly detection in IoT environments through ensemble learning and a strategic sleep-wake mechanism. Additionally, the process crucial for maintaining high accuracy while ensuring efficiency in the resource-constrained nature of IoT environments is demonstrated.

A. Dataset

The dataset utilized in this research is known as the InSDN dataset, published in 2020 by M. Elsayd et al. [31]. The primary objective of creating the InSDN dataset was to reduce its size compared to other IDS datasets while providing a realistic representation of traffic in an SDN environment. Unlike NSL-KDD and KDD99, InSDN includes real-world network traffic collected from an SDN environment and categorizes it based on the presence of DDoS attacks [32].

The dataset comprises a total of 343,889 records with 84 features. Of these, 127,828 records correspond to normal traffic, while 216,061 records represent attack traffic in both OSV and metasploitable files. InSDN encompasses various attack scenarios, including SYN, TCP, UDP, and ICMP floods, as well as Slowloris attacks. The distribution of samples within the InSDN dataset is detailed in Fig. 2.

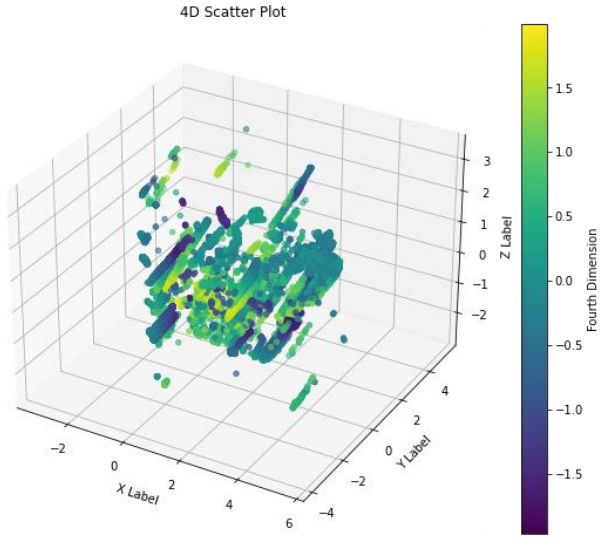


Fig. 2. 4D scatter plot for records on InSDN dataset.

B. Experimental Result and Analysis

The study employed a Random Forest classifier trained on the dataset to evaluate feature importance and classification accuracy. The dataset was split into training and testing sets, achieving an impressive accuracy of 99.97%. Through analysis, the top features contributing to classification were identified, ranked by importance in descending order. Fig. 3 presents a bar chart visualizing the importance of each feature, offering valuable insights into which features are most critical for the model's decisions.

Features with an average importance score above 0.0556 were selected for further analysis, as depicted in Fig. 4, effectively reducing the dataset's dimensionality. This focused approach enabled us to focus on the most relevant features, optimizing the model's performance and computational efficiency. Such selection is crucial for maintaining high accuracy while ensuring the system remains efficient and suitable for the resource-constrained nature of IoT environments.

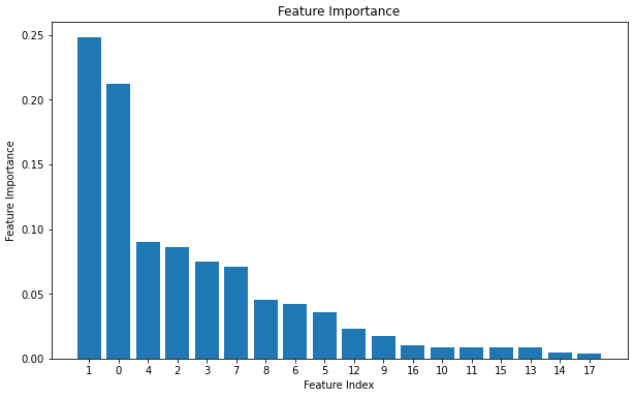


Fig. 3. The importance of the selected features.

After feature selection, an ensemble model was developed by combining Logistic Regression, Decision Tree, and SVM to leverage the strengths of multiple classifiers, as explained in Section III, Step 4.

As a result, Logistic Regression achieved an accuracy of 99.99% with a prediction time of 1.88e-7 seconds, weighted at 871.5946. The Decision Tree model exhibited 99.96% accuracy with a prediction time of 11.16e-7 seconds, having the highest weight of 5789.0159. SVM achieved an accuracy of 99.98% with a prediction time of 0.00011 seconds, weighted at 167.8950. Details of other classifier results are provided in Table II.

As visualized in Fig. 5, ELSM emerges as a robust performer across multiple key metrics. It achieved an accuracy of 99.97%, closely aligning with top-performing models such as LR, SVM, and XGBoost, which scored 99.99%. ELSM's precision of 94.97% underscores its ability to accurately identify positive instances. Moreover, ELSM achieved a recall of 93.93%, demonstrating its effectiveness in capturing the most actual positive instances. Its F1 score of 94.31% strikes a balance between precision and recall, ensuring robust overall performance.

C. Discussion

ELSM excels with a perfect ROC AUC of 100.00%, indicating its exceptional ability to differentiate between classes with high confidence, which is crucial for minimizing false positives and false negatives in intrusion detection scenarios. This comprehensive evaluation positions ELSM favorably, demonstrating its efficacy in maintaining high accuracy while optimizing precision, recall, F1 score, and ROC AUC. The visual representation in the figure provides a clear comparative analysis, illustrating ELSM's strong performance across these critical metrics and substantiating its suitability for deployment in real-world applications that require reliable and efficient intrusion detection systems.

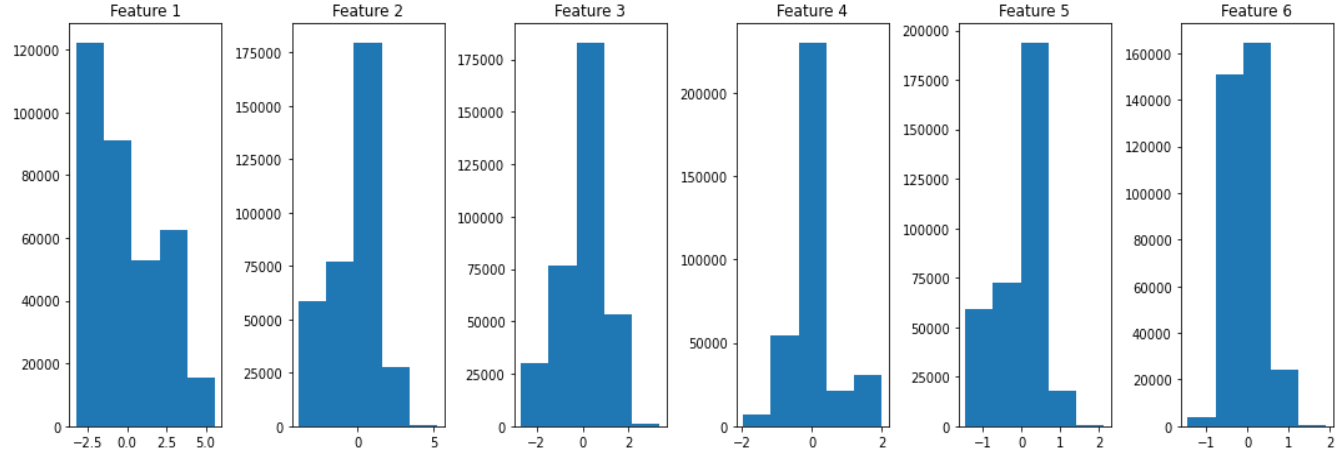


Fig. 4. The selected features with average importance above 0.0556.

TABLE II. CLASSIFIERS RESULTS AND WEIGHT

Model	Accuracy (%)	Precision (%)	Recall (%)	F1 Score (%)	ROC AUC (%)	Training time (Sec)	Prediction time (Sec)	Weight	Normalized Weight
Logistic Regression	99.99	99.07	94.01	95.59	99.57	18.91	1.88e-7	5295085.502	0.3552
Decision Tree	99.97	95.06	96.78	95.82	98.39	6.77	11.16e-7	8589122.0416	0.5762
SVM	99.98	99.51	93.9	95.77	100	223.49	0.00011	9100.85	0.0006
Random Forest	99.99	99.51	91.24	92.87	99.79	114.69	8.36e-5	119585.259	0.0080
K-Nearest Neighbors	99.97	94.97	93.93	94.31	96.97	0.087	0.00018	5446.28	0.0003
XGBoost	99.99	99.51	94.01	95.83	100	6.54	1.13e-6	887210.96	0.0595
Proposed ELSM	99.97	94.97	93.93	94.31	100				

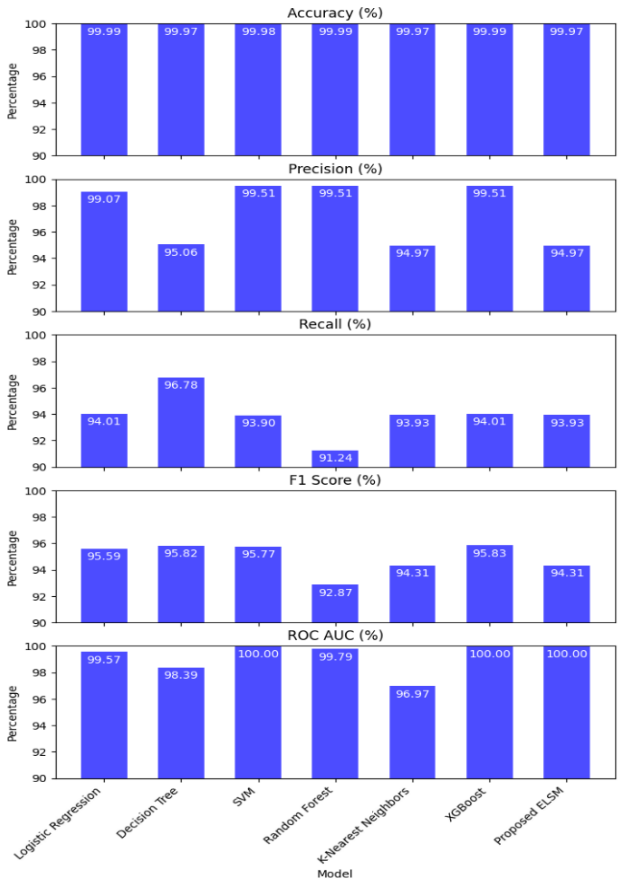


Fig. 5. Evaluation metrics for different models.

By combining these individual models using a weighted voting mechanism, the ensemble model ELSM achieved an overall accuracy of 99.97%. This high accuracy was attained by leveraging the complementary strengths of different classifiers, thereby enhancing the robustness of predictions and ensuring more reliable anomaly detection. In contrast, other methods achieved lower accuracy for the same dataset, as shown in Table I. This detailed comparison will effectively demonstrate the superiority of our approach in improving anomaly detection performance and reliability.

Subsequently, all previous stage results were integrated into the sleep-wake cycle management. The ensemble model's anomaly score determines whether the device remains in sleep mode or wakes up to handle potential threats. If the anomaly score meets or exceeds a predefined threshold, the device wakes up and sends a warning. If no anomalies are detected, the device remains in sleep mode to conserve energy. This strategic activation ensures that IoT devices are active only when necessary, optimizing power consumption while maintaining high-security standards. This approach effectively addresses the dual challenges of maintaining high-security standards and reducing power consumption, making it ideal for the resource-constrained nature of IoT environments.

V. Future Work and Challenges

Our research presents a promising approach to enhancing IoT security through ensemble learning and a strategic sleep-wake mechanism. However, several areas require further exploration. Future work will focus on scaling the anomaly detection algorithms to handle the increasing volume of IoT data while ensuring real-time processing capabilities. Additionally, the aim is to develop adaptive algorithms capable of dynamically adjusting to diverse IoT environments and data types, explore advanced feature selection techniques, and integrate the model with an edge computing framework to reduce latency.

Despite the improvements demonstrated, several challenges remain. Balancing energy conservation with high performance, addressing data heterogeneity, achieving real-time detection without compromising accuracy, and keeping up with the dynamic threat landscape are significant hurdles.

VI. Conclusion

This paper introduces a novel hybrid model designed to enhance anomaly detection in IoT environments by leveraging ensemble learning and a strategic sleep-wake mechanism, significantly improving the efficiency and effectiveness of Intrusion Detection Systems (IDS) in these settings. Our approach addresses the critical challenges of maintaining high-security standards and reducing power consumption, which are essential for the resource-constrained nature of IoT devices. The integration of multiple machine learning models using ensemble learning techniques significantly improved the overall prediction accuracy.

The experimental results demonstrated that our model achieved an impressive accuracy of 99.97% on the IoT Botnets Attack Detection Dataset. The use of feature importance assessment through Random Forest allowed us to reduce the dataset's dimensionality, focusing on the most relevant features and optimizing the model's performance and computational efficiency. By employing a weighted voting mechanism, the strengths of individual classifiers were effectively combined, enhancing the robustness and reliability of anomaly detection. Additionally, the integration of the sleep-wake mechanism ensured that IoT devices remained in a low-power state until an anomaly was detected, thereby conserving energy.

References

[1] H. Allioui and Y. Mourdi, "Exploring the full potentials of IoT for better financial growth and stability: A comprehensive survey," *Sensors*, vol. 23, no. 19, pp. 8015, 2023.

[2] N. Najari, S. Berlemont, G. Lefebvre, S. Duffner, and C. Garcia, "Radon: Robust autoencoder for unsupervised anomaly detection," in *Proc. 2021 14th Int. Conf. Security of Information and Networks (SIN)*, vol. 1, pp. 1-8. IEEE. Dec. 2021.

[3] I. S. Essop, J. C. Ribeiro, M. Papaioannou, G. Zachos, G. Mantas, and J. Rodriguez, "Generating datasets for anomaly-based intrusion detection systems in IoT and industrial IoT networks," *Sensors*, vol. 21, no. 4, pp. 1528, 2021.

[4] G. Zachos, G. Mantas, I. S. Essop, K. Porfyrakis, J. C. Ribeiro, and J. Rodriguez, "Prototyping an anomaly-based intrusion detection system for Internet of Medical Things Networks," in *Proc. 2022 IEEE 27th Int. Workshop on Computer Aided Modeling and Design of Communication Links and Networks (CAMAD)*, pp. 179-183, Nov. 2022.

[5] Z. Shu, J. Wan, D. Li, J. Lin, A. V. Vasilakos, and M. Imran, "Security in software-defined networking: Threats and countermeasures," *Mobile Networks and Applications*, vol. 21, pp. 764-776, 2016.

[6] O. E. Tayfour and M. N. Marsono, "Collaborative detection and mitigation of DDoS in software-defined networks," *The Journal of Supercomputing*, vol. 77, no. 11, pp. 13166-13190, 2021.

[7] T. A. Tang, D. McLernon, L. Mhamdi, S. A. R. Zaidi, and M. Ghogho, "Intrusion detection in SDN-based networks: Deep recurrent neural network approach," in *Deep Learning Applications for Cyber Security*, pp. 175-195, 2019.

[8] H. Y. Ibrahim, P. M. Ismael, A. A. Albabawat, and A. B. Al-Khalil, "A secure mechanism to prevent ARP spoofing and ARP broadcasting in SDN," in *Proc. 2020 Int. Conf. Computer Science and Software Engineering (CSASE)*, pp. 13-19, Apr. 2020.

[9] D. Kreutz, F. M. Ramos, P. E. Verissimo, C. E. Rothenberg, S. Azodolmolky, and S. Uhlig, "Software-defined networking: A comprehensive survey," *Proc. IEEE*, vol. 103, no. 1, pp. 14-76, Jan. 2014.

[10] M. Said Elsayed, N. A. Le-Khac, S. Dev, and A. D. Jurcut, "Network anomaly detection using LSTM based autoencoder," in *Proc. 16th ACM Symposium on QoS and Security for Wireless and Mobile Networks*, pp. 37-45, Nov. 2020.

[11] J. C. S. Sicato, S. K. Singh, S. Rathore, and J. H. Park, "A comprehensive analysis of intrusion detection system for IoT environment," *Journal of Information Processing Systems*, vol. 16, no. 4, pp. 975-990, 2020.

[12] M. Latah and L. Toker, "An efficient flow-based multi-level hybrid intrusion detection system for software-defined networks," *CCF Transactions on Networking*, vol. 3, no. 3, pp. 261-271, 2020.

[13] P. Dini, A. Elhanashi, A. Begni, S. Saponara, Q. Zheng, and K. Gasmi, "Overview on Intrusion Detection Systems Design Exploiting Machine Learning for Networking Cybersecurity," *Applied Sciences*, vol. 13, no. 13, pp. 7507, 2023.

[14] X. W. Wu, Y. Cao, and R. Dankwa, "Accuracy vs Efficiency: Machine Learning Enabled Anomaly Detection on the Internet of Things," in *Proc. 2022 IEEE Int. Conf. Internet of Things and Intelligence Systems (IoTaIS)*, pp. 245-251, Nov. 2022.

[15] Q. R. S. Fitni and K. Ramli, "Implementation of ensemble learning and feature selection for performance improvements in anomaly-based intrusion detection systems," in *Proc. 2020 IEEE Int. Conf. Industry 4.0, Artificial Intelligence, and Communications Technology (IAICT)*, pp. 118-124, Jul. 2020.

[16] A. Kumari and A. K. Mehta, "A hybrid intrusion detection system based on decision tree and support vector machine," in *Proc. 2020 IEEE 5th Int. Conf. Computing Communication and Automation (ICCCA)*, pp. 396-400, Oct. 2020.

[17] P. Pokharel, R. Pokhrel, and S. Sigdel, "Intrusion detection system based on hybrid classifier and user profile enhancement techniques," in *Proc. 2020 Int. Workshop on Big Data and Information Security (IWBIS)*, pp. 137-144, Oct. 2020.

[18] I. F. Kilincer, F. Ertam, and A. Sengur, "Machine learning methods for cyber security intrusion detection: Datasets and comparative study," *Computer Networks*, vol. 188, p. 107840, 2021.

[19] S. U. Jan, S. Ahmed, V. Shakhov, and I. Koo, "Toward a lightweight intrusion detection system for the Internet of Things," *IEEE Access*, vol. 7, pp. 42450-42471, 2019.

[20] Y. N. Soe, Y. Feng, P. I. Santosa, R. Hartanto, and K. Sakurai, "Machine learning-based IoT-botnet attack detection with sequential architecture," *Sensors*, vol. 20, no. 16, pp. 4372, 2020.

[21] A. Alzaqebah, I. Aljarah, O. Al-Kadi, and R. Damaševičius, "A modified grey wolf optimization algorithm for an intrusion detection system," *Mathematics*, vol. 10, no. 6, p. 999, 2022.

[22] A. Ugendhar, B. Illuri, S. R. Vulapula, M. Radha, K. S., F. Alenezi, et al., "A Novel Intelligent-Based Intrusion Detection System Approach Using Deep Multilayer Classification," *Mathematical Problems in Engineering*, vol. 2022, no. 1, p. 8030510, 2022.

[23] M. Otair, O. T. Ibrahim, L. Abualigah, M. Altalhi, and P. Sumari, "An enhanced grey wolf optimizer based particle swarm optimizer for intrusion detection system in wireless sensor networks," *Wireless Networks*, vol. 28, pp. 721-744, 2022.

[24] J. K. Samriya and N. Kumar, "A novel intrusion detection system using hybrid clustering-optimization approach in cloud computing," *Materials Today: Proceedings*, vol. 2, no. 1, pp. 23-54, 2020.

[25] M. Mohy-Eddine, A. Guezzaz, S. Benkirane, M. Azrour, and Y. Farhaoui, "An ensemble learning based intrusion detection model for industrial IoT security," *Big Data Mining and Analytics*, vol. 6, no. 3, pp. 273-287, 2023.

[26] E. Alalwany and I. Mahgoub, "An Effective Ensemble Learning-Based Real-Time Intrusion Detection Scheme for an In-Vehicle Network," *Electronics*, vol. 13, no. 5, p. 919, 2024.

[27] D. Jayalatchumy, R. Ramalingam, A. Balakrishnan, M. Safran, and S. Alfarhood, "Improved Crow Search-based Feature Selection and Ensemble Learning for IoT Intrusion Detection," *IEEE Access*, vol. 12, p. 32554, 2024.

[28] R. Chinnasamy, M. Subramanian, and N. Sengupta, "Devising Network Intrusion Detection System for Smart City with an Ensemble of Optimization and Deep Learning Techniques," in *Proc. 2023 Int. Conf. Modeling & E-Information Research, Artificial Learning and Digital Applications (ICMERALDA)*, pp. 179-184, Nov. 2023.

[29] S. Yangui, "A panorama of cloud platforms for IoT applications across industries," *Sensors*, vol. 20, no. 9, p. 2701, 2020.

[30] A. Yahyaoui, T. Abdellatif, S. Yangui, and R. Attia, "READ-IoT: Reliable event and anomaly detection framework for the Internet of Things," *IEEE Access*, vol. 9, pp. 24168-24186, 2021.

[31] M. S. Elsayed, N. A. Le-Khac, and A. D. Jurcut, "InSDN: A novel SDN intrusion dataset," *IEEE Access*, vol. 8, pp. 165263-165284, 2020.

[32] K. Harahsheh, R. Al-Naimat, and C. H. Chen, "Using Feature Selection Enhancement to Evaluate Attack Detection in the Internet of Things Environment," *Electronics*, vol. 13, no. 9, p. 1678, 2024.

Optimizing Low-Resource Zero-Shot Event Argument Classification with Flash-Attention and Global Constraints Enhanced ALBERT Model

Tongyue Sun[1], Jiayi Xiao[2]
School of Engineering and Informatics, University of Sussex, Brighton, UK[1]
International Business School Suzhou, Xi'an Jiaotong-Liverpool University, Suzhou, China[2]
Management School, University of Liverpool, Liverpool, UK[2]

***Abstract*—Event Argument Classification (EAC) is an essential subtask of event extraction. Most previous supervised models rely on costly annotations, and reducing the demand for computational and data resources in resource-constrained environments is a significant challenge within the field. We propose a Zero-Shot EAC model, ALBERT-F, which leverages the efficiency of the ALBERT architecture combined with the Flash-Attention mechanism. This novel integration aims to address the limitations of traditional EAC methods, which often require extensive manual annotations and significant computational resources. The ALBERT-F model simplifies the design by factorizing embedding parameters, while Flash-Attention enhances computational speed and reduces memory access overhead. With the addition of global constraints and prompting, ALBERT-F improves the generalizability of the model to unseen events. Our experiments on the ACE dataset show that ALBERT-F outperforms the Zero-shot BERT baseline by achieving at least a 3.4% increase in F1 score. Moreover, the model demonstrates a substantial reduction in GPU memory consumption by 75.1% and processing time by 33.3%, underscoring its suitability for environments with constrained resources.**

***Keywords*—Artificial intelligence; natural language processing; event argument classification; zero-shot learning; flash-Attention; global constraints; low-resource**

I. Introduction

Event Argument Classification (EAC) is a crucial part of event understanding and event argument extraction, embodying the complexity and importance of this interdisciplinary field [1, 2]. This domain, which integrates natural language processing (NLP) and knowledge representation, is dedicated to converting narrative event descriptions and their relational dynamics into a structured form of knowledge. As shown in Fig. 1, for a trigger word "paid" in a "Transfer-Money" event, it has several argument spans (e.g., "O'neal"). By determining the roles of these arguments (e.g., identifying "O'neal" as the "Giver"), this structured knowledge enables us to better understand events and use them for knowledge reasoning and automated decision support, benefiting applications such as biomedical research and question answering recommendation systems.

In the domain of event argument classification, a prevalent strategy has been the manual annotation of domains and patterns. Although effective, this approach necessitates significant labeling efforts for model training. This method also presents challenges in transferring knowledge across different application domains and scaling to new datasets. The laborious nature of annotation incurs significant costs. To mitigate this, some EAC models have turned to few-shot learning [3–6], which, despite its potential, is sensitive to the selection of examples and requires costly, task-specific training, limiting its practicality.In contrast, zero-shot EAC models have been introduced, leveraging label semantic understanding or prompt learning strategies [7–10]. Although existing methods perform well when dealing with events similar to the training data, they may not achieve the expected results when faced with significantly different new events. Some studies have attempted to improve performance in zero-shot and few-shot learning scenarios by integrating Large Language Models (LLMs) [2, 11], but there is still a considerable gap compared to models that have been specifically fine-tuned. Moreover, the operation of LLMs requires a significant amount of computational resources, which may limit their potential for application in resource-constrained environments. Therefore, tackling the efficiency constraints inherent to Zero-Shot EAC tasks in resource-scarce environments has become a formidable obstacle.

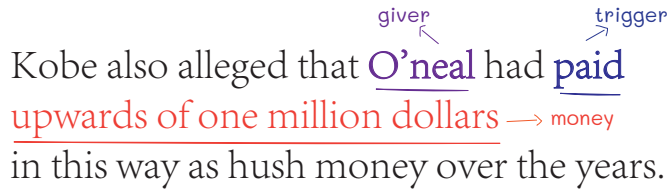


Fig. 1. An example of EAC. The arrows indicate the trigger and argument types respectively.

To address the challenges in the field of event argument classification, we propose a Zero-Shot model tailored for low-resource scenarios. This model integrates an ALBERT architecture [12] optimized with Flash-Attention [13] and is enhanced by global constraints with prompting, aiming to improve the performance of zero-shot EAC tasks.

The ALBERT model mitigates the issues of excessive parameters and inefficiency by simplifying its design in the BERT [12, 14]. Furthermore, global constraints provide critical supervisory guidance to our model, as a manifestation of

domain knowledge [2]. This guidance is particularly crucial in zero-shot learning environments with a scarcity of fully annotated data, as it enables the model to better understand and generalize the relations between event arguments. To further accommodate the constrained resources in low-resource scenarios, we propose ALBERT-F, a solution that optimizes the ALBERT model using a Flash Attention module. Flash Attention leverages efficient upper-level storage computational units to reduce access to the slower lower-level storage, thereby maintaining performance while significantly enhancing the model's resource utilization efficiency [13].

Through a series of experiments, we have validated the effectiveness of our proposed method. Specifically, our approach achieved at least a 3.4% increase in F1 score on the ACE dataset compared to the Zero-shot BERT baseline model, with a 75.1% reduction in GPU memory consumption and a 33.3% reduction in processing time. Furthermore, the introduction of Flash Attention resulted in a further 5.1% reduction in GPU memory consumption and an 11.1% decrease in processing time compared to the original ALBERT model. These results not only demonstrate the significant advantage of our method in reducing resource consumption but also confirm its effectiveness in enhancing performance.

Subsequent sections present our experimental setup, results, and a comparative analysis with existing models. We conclude with a discussion on the implications of our findings, limitations and avenues for future research.

II. Related Works

A. Event Extraction

Event Extraction (EE) is one of the most fundamental tasks in information extraction, which can be further divided into four subtasks: trigger identification, trigger classification, argument identification, and argument classification [1, 15–18].

Traditional event extraction relies heavily on feature engineering, which poses its central challenge [1, 18]. However, these methods often encounter limitations when dealing with deep or complex nonlinear patterns. In recent years, some advanced works based on supervised learning have attracted attention due to their two main advantages: first, the applicability of their embedding representations to large-scale datasets; second, the combination of automated feature extraction with specific deep architectures, which effectively captures more intricate nonlinear patterns [19–23].

In the task of information extraction, models can identify the actions in sentences and their corresponding participants by defining constraints [24]. One of the applications of constraint modeling in NLP is in syntactic analysis, where it is used to represent that an object must satisfy general or very specific properties to exclude those that do not belong to the structure of the languag [25]. Particularly in zero-shot scenarios, constraint modeling can provide useful indirect supervision to the model, thereby further improving its performance [26].

Nevertheless, the inherent limitations of supervised learning may impact the model's generalization capabilities across different domains. Moreover, the demand for computational resources and specialized skills (including constraint modeling), along with the reliance on a substantial amount of manually annotated data, become bottlenecks in their practical application.

B. Few-Shot Learning for EE

Few-shot learning methods have garnered widespread attention in the domain of event extraction, and the majority of current research is concentrated on the task of event identification within the context of Few-shot Event Detection (FSED) [1]. These approaches are dedicated to achieving accurate predictions for specific tasks with minimal training samples, such as one-shot, five-shot, etc. By leveraging prior knowledge, transfer learning, or meta-learning strategies, few-shot learning endeavors to surmount the challenge of data scarcity and enhance the model's generalization capability on novel tasks [3–6].

The DEGREE model [6] excels at synthesizing events from a text segment into coherent, naturally constructed sentences that conform to a pre-established template, aided by manually curated prompts. By integrating the semantic essence of labels with the collective intelligence across sub-tasks, DEGREE discerns interdependencies among entities, thereby reducing the volume of training data required. Many previous works on event extraction (EE) necessitate extensive annotations for model training [6, 8, 23], which incurs high costs due to the labor-intensive nature of annotation and poses challenges in scaling to new domains. While DEGREE refines a pre-existing generative language model [27], the output it generates may reflect the characteristics of the corpus from which it was trained. Although infrequent, there is a possibility that the model might produce sentences that are malevolent, mendacious, or prejudiced, thus raising ethical concerns [28, 29].

For classification tasks, LoLoss [4] is employed for training few-shot learning models based on the matching information of examples within the support set.

$$L(x, S) = L_{\text{query}}(x, S) + \lambda \cdot L_{\text{aux}}(S) \tag{1}$$

The components of this equation are as follows: $L(x, S)$ represents the total loss, which is contingent upon the model parameters x and the training samples S. $L_{\text{query}}(x, S)$ refers to the query loss, which assesses the discrepancy between the model's predictions for the query set and their corresponding true labels. λ is a hyperparameter that modulates the trade-off between the query loss and the auxiliary loss within the total loss calculation. $L_{\text{aux}}(S)$ is the auxiliary loss, which leverages the internal matching information of examples within the support set to provide additional training signals. It not only matches the query examples with those in the support set but also further matches the examples among themselves within the support set, thereby providing additional training signals for the model.

The scarcity of samples in long-tail categories increases the complexity of classification in few-shot learning tasks. To overcome this challenge, the Multi-Level Matching and Aggregation Network (MLMAN) [3] employs a hierarchical matching and aggregation strategy. This strategy comprehensively analyzes the support vectors and query vectors at different levels, capturin local features integrating global contextual information, thereby enhancing the classification accuracy of

long-tail category samples. Adaptive Attentional Network for Few-Shot Knowledge Graph Completion (FAAN) [5] employs a minimal set of reference samples to adeptly predict and discern connections and relations. These reference relation triples are adaptively encoded within a transformer network through the application of embeddings and an attention mechanism, ensuring precise alignment with the query. FAAN's adaptive encoding of entities and reference pairs significantly enhances the performance of traditional knowledge graph embedding methods, particularly for long-tail relations that are characterized by a paucity of samples.

Nonetheless, constrained by a limited sample size, the model is susceptible to overfitting, with an exacerbated risk in scenarios characterized by class imbalance. This propensity may compromise the model's capacity for robust generalization. Furthermore, the necessity for supplementary computational resources or the adoption of intricate model architectures could potentially restrict the practical applicability of these models.

C. Zero-shot Learning for EE

In the context of lacking prior knowledge and labeled data, existing research tends to adopt preset event information frameworks or experience-based strategies to achieve effective classification of unknown event types [7–10]. Similarly, the zero-shot contrastive learning strategy also emphasizes the use of unlabeled data during the training phase to cultivate features that can distinguish between different categories [2]. Although these methods still have a significant gap compared to supervised methods, they offer an insightful perspective and suggest possible directions for improvement in event extraction under resource-constrained environments.

The event extraction task was conceptualized by Huang et al. [7] as a "grounding" problem, wherein it is encapsulated within a structured ontology that delineates event mentions and their respective types. Semantic similarity measures are harnessed for the purpose of prediction. A transferable neural architecture was proposed by Huang et al., one that capitalizes on manually annotated event patterns alongside a modest subset of previously encountered types. This architecture was adept at transferring knowledge from known types to the extraction of novel types, thereby enhancing the scalability of event extraction and conserving human resources. Further exploration into transfer learning methodologies for novel events was undertaken by Lyu et al. [9] They reframed the event extraction challenge within the contexts of textual entailment (TE) and question answering (QA), advocating for the direct application of pre-trained TE/QA models.

Although these models have demonstrated exceptional performance on standard benchmark tests, they have not yet realized the anticipated generalization effect when applied to the event extraction dataset. Nonetheless, they offer an insightful vision and suggest a possible direction for improvement in event extraction within very low-resource environments.

Lin et al. [2] proposed a Global Constraint Regularization Module that standardizes predictions through three types of global constraints: cross-task constraints, cross-parameter constraints, and cross-event constraints. They utilized a method that combines global constraints with prompting, employing the large language model GPT-J [30], which makes it possible to effectively perform event parameter classification without any annotations or task-specific training. Chen et al. [11], also employing large language models for research, utilized a large language model as an expert annotator for event extraction. Strategically incorporating sample data from the training dataset into the prompts, the researchers ensure that the generated samples from the language model align with the data distribution of the benchmark dataset. This enables the creation of an augmented dataset to supplement the existing benchmarks, alleviating challenges of data imbalance and scarcity, thus enhancing the performance of fine-tuned models. However, existing open-source large language models often require expensive hardware configurations and substantial computational resources [31]. Furthermore, the utility of these models is limited by the fact that most current hardware was developed prior to the emergence of large-scale models, potentially rendering it inadequate for the computational demands of such models during inference. This limitation is particularly pronounced in low-resource settings, where specialized hardware is required to facilitate efficient inference processes for large models [32].

To address low-resource scenarios, we propose a Zero-shot EAC model that incorporates global constraints and prompt, coupled with ALBERT-F. This approach aims to enhance the performance of Zero-shot EAC tasks in resource-constrained environments.

III. Methodology

Our model comprises two distinct modules. As shown in Fig. 2, the first is the prompting module, which is tasked with the generation of several new passages and the subsequent evaluation of their quality. During this creation process, the model integrates candidate role with prefix prompts that contain information regarding the event type and trigger. These candidates are connected to the target parameter range by embedding them within the passages through a cloze prompt. Subsequently, the model employs an ALBERT model optimized with Flash-Attention (ALBERT-F) to score the newly generated passages. Without the need for manual annotation, the initial prediction is the role with the highest prompting score. The second module is the global constraint regularization module, wherein the model regularizes the predictions through three types of global constraints. These are based on domain knowledge related to inter-task, inter-parameter, and inter-event relationships within the event-related context.

A. ALBERT-F

Before delving into the two primary modules, we provide a overview of ALBERT-F. By substituting the attention mechanisms across all modules, the primary structure of our ALBERT network is depicted in Fig. 3.

Flash-attention is designed to expedite the computation of attention mechanisms and curtail memory usage [13]. It leverages the knowledge of the memory hierarchy of underlying hardware, such as the memory architecture of GPUs, to enhance computational speed and reduce the overhead of memory access. By using statistical measures and altering

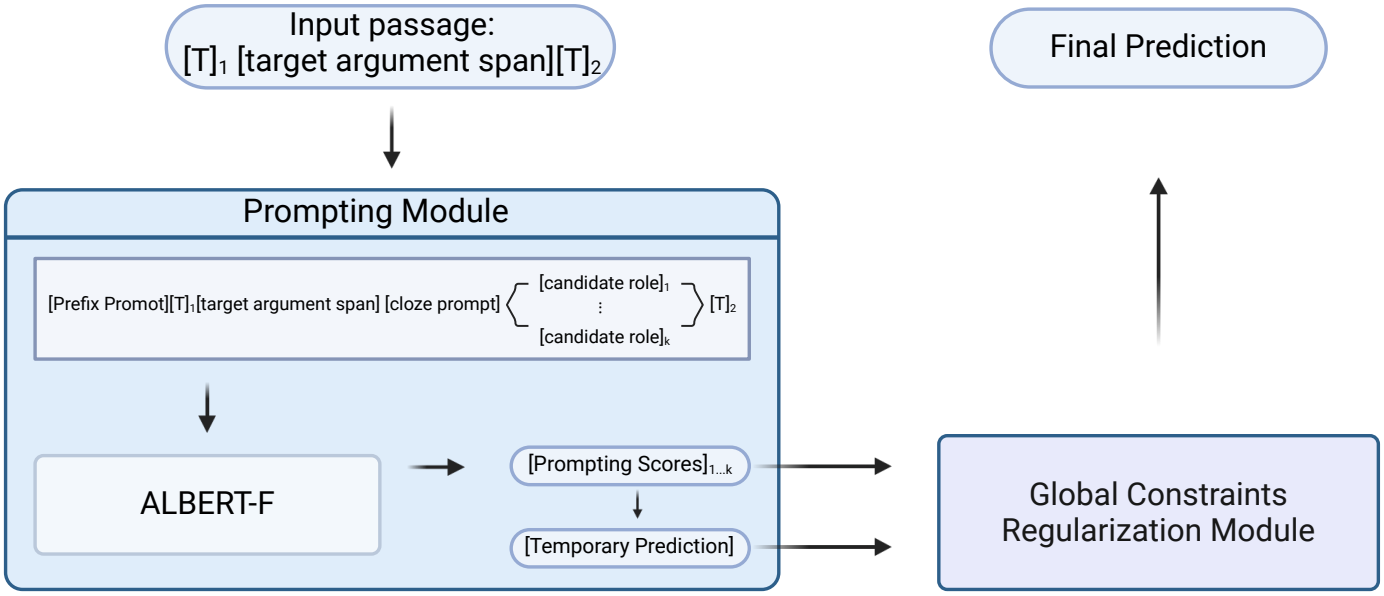


Fig. 2. Model summary, illustrated with the prediction of a single argument span. $[T]_1$ represents the segment of the input text preceding the span, while $[T]_2$ denotes the segment that follows. The variable k signifies the total count of potential roles associated with the event type.

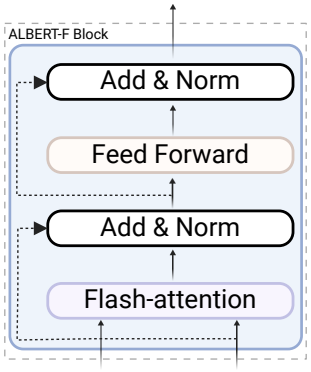


Fig. 3. The main structure of the ALBERT-F module. Enhancing model efficiency by introducing flash-attention to reduce computational resource consumption.

the computation sequence of the attention mechanism, Flash-attention computes in chunks rather than approximates, effectively reducing complexity. The outcomes of Flash-attention are entirely equivalent to those of the native attention mechanism [13, 33].

Increasing the size of a Pre-trained Language Model (PLM) typically enhances its inferential capabilities; however, once the model reaches a certain magnitude, it encounters limitations imposed by the memory capacity of GPUs/TPUs [33]. Consequently, ALBERT implements factorized embedding parameterization, which decomposes the embedding matrix. Instead of directly projecting one-hot encoded vectors into a hidden vector space of dimension H, the vectors are first projected into a lower-dimensional embedding vector space and then into the hidden vector space. This decomposition significantly reduces the number of embedding parameters and results in a more uniform distribution.

In ALBERT, the parameters of the fully connected layers and the attention layers are shared, meaning that ALBERT retains the deep multi-layer connections, but the parameters between layers are identical [12]. Consequently, ALBERT-F optimizes ALBERT using Flash-attention to reduce the model's computational resource consumption, yielding more satisfactory results in low-resource scenarios where computation and memory are constrained.

B. Prompting Module

In this section, we primarily elucidate the prompting module. Given a passage, we initially append a prefix prompt at the onset, which encapsulates information pertaining to the event type and the scope of the trigger. Such a prompt serves to guide ALBERT-F in: (1) accurately capturing the correlation between the input text and the event-related associations; (2) possessing a clear trigger awareness capability. In accordance with the definitions of events and triggers [17], we have formulated the following prefix prompt:

- "This is a $[P_1]$ event whose occurrence is most clearly expressed by $[P_2]$."

Where the first and second pairs of square brackets are placeholders for the event type(P_1) and trigger span (P_2), respectively.

For each candidate role, the module inserts a cloze test prompt subsequent to the target parameter range, with the role filling the slot of the prompt. The cloze prompt employs a hypernym extraction pattern "M and any other []", wherein "M" denotes the parameter range, and the square brackets serve as placeholders for the candidate roles. Such prompts harness the linguistic and common-sense knowledge stored within the ALBERT-F to assist in identifying which candidate role is the most plausible [2, 34].

For each novel passage, we apply ALBERT-F to compute the language modeling loss. The prompting score for the corresponding paragraph is determined by the negative value of the loss, where a more negative loss indicates a higher score, reflecting greater plausibility as assessed by ALBERT-F. Since the scoring process for each candidate role is independent of other candidate roles, we implement the steps for different candidate roles in parallel. This parallel implementation significantly enhances the efficiency of our model.

C. Global Constraints Regularization Module

This module regularizes predictions through global constraints. We refer to and leverage the constraint strategy proposed by Lin et al. [2], employing Event Argument Entity Typing (EAET) as an auxiliary task, which aims to categorize arguments into their contextually relevant entity types. These constraints provide our model with a global understanding of event arguments.

By utilizing the label dependencies between EAC and the auxiliary task, our model can glean global information about event arguments from the auxiliary task. Concurrently, to adapt to applications in low-resource scenarios, our model limits the number of specific arguments for certain or all events.

IV. Experiments

We initially present the experimental setup, the baselines for comparison, and certain implementation details. Subsequently, we demonstrate and analyze the results of the experiments. We then conduct an analysis of computational resources and processing duration. Finally, we perform an error analysis.

A. Settings

We utilize the ACE (2005-E +) dataset [23, 35] as the basis for our experiments. The ACE dataset encompasses a total of 33 event types and 22 roles. We preprocess all events, as is done in the work of Lin et al. [23], to retain only the event subtypes when applicable. Since our approach is zero-shot, for each dataset, we consolidate all splits into a single test set, following the preprocessing in the study by Lyu et al. [9].

We evaluate using the F1 score, as proposed by Ji and Grishman [36], employing the ALBERT-F model as the foundational model for our module implementation. We run our experiments on a single NVIDIA RTX 4000 Ada GPU.

B. Main Results

We report results that are compared with several existing zero-shot methods, including those by Huang et al. (2018) [7], Liu et al. (2020) [8], Zhang et al. (2021) [10], and the current state-of-the-art zero-shot approach by Lin et al. (2023) [2]. In our comparisons, we also evaluated the work of Lin et al., where we compared two PLM (Pre-trained Language Model) bases: Bert-large-uncased [12] with a parameter count of 330 million, and GPT-6J [30] with a parameter count of 6 billion.

From Table I, we have the following observations: Compared to all zero-shot baselines, our model has demonstrated superior performance in the Settings category. Specifically, our model has achieved an F1 score on the ACE dataset that surpasses the best non-large model zero-shot baseline by 3.4% (Lin et al., 2023 [2]). This represents a significant gap. Such substantial performance improvement can be attributed to several factors: (1) the prefix and cloze prompts effectively guide the PLM to capture the input's event-related perspectives and triggers; (2) the global constraint regularization incorporates global information and domain knowledge into the inference process; (3) our model has effectively enhanced the inferential capabilities of the EAC (Event Argument Classification) task.

TABLE I. Comparison F1 Score of Different Models on the ACE 2005 E+ Dataset, the Best Performance of Non-LLM is Marked in Bold Font

Model	Year	ACE 2005 E+
Lin et al. [2] (GPT-6J)	2023	66.1
Liu et al. [8]	2020	46.1
Lyu et al. [9]	2021	47.8
Zhang et al. [10]	2021	53.6
Lin et al. [2] (BERT-Large)	2023	58.2
Ours	2024	**61.6**

Compared to the state-of-the-art (SOTA) results based on large models, there remains a significant performance gap for our model. Specifically, Lin et al. achieved a 4.5% higher score on the ACE dataset than our model. The advantage of models with ample resources over our zero-shot method is even more pronounced. This may be due to the fact that our model's parameter count (60M) is only 1% of the SOTA model's parameter count (6B). Based on the theoretical knowledge presented in Section III-A, there is still a partial performance gap between our EAC model and those utilizing Large Language Models (LLMs).

C. Comparison Between Different Prefix Prompts

In this section, we conduct experiments on the ACE (2005 E+) dataset to compare the effectiveness of using different prefix prompts within the model. We compare the following prefix prompts with those mentioned in Section III-B:

1) "[P1] most accurately represents the occurrence of this [P2]."
2) "The event type is [P1] and the trigger is [P2]."

TABLE II. Performance of Different Prefix Prompts

Prefix Prompt	F1 Score
Prefix(0)	61.6
Prefix(1)	61.3
Prefix(2)	60.8

From Table II, it can be observed that the prompts described in Section III-B are the most effective, which may be attributed to the fact that the prefix prompts are not only based on the definitions of events and triggers [17], but also possess a naturally fluent expression [2].

D. Computational Resource Analysis

The state-of-the-art (SOTA) model based on GPT-J, with its substantial parameter count of 6 billion, excels in resource-intensive tasks but also implies a significant demand for computational resources, making it generally unsuitable for low-resource scenarios. Therefore, in this section, we primarily compare the version implemented within the framework of Lin et al. [2] using BERT-Large with our model.

In contrast, the BERT-Large model used by Lin et al. has a parameter count of 334 million, whereas our model has a parameter count of only 60 million, significantly reducing the model's storage and computational requirements. Runtime and GPU memory usage are key indicators for gauging the feasibility of models in practical applications. We independently ran each model five times to calculate their average resource consumption.

TABLE III. THE RESOURCE CONSUMPTION OF EACH MODEL, WITH NUMERICAL VALUES REPRESENTING THE AVERAGE DURATION AND GPU MEMORY USAGE OVER FIVE INDEPENDENT RUNS

Model	Paramaters	Run Time	GPU Memory
Lin et al. (BERT-Large)	334M	1.2h	2151MiB
Ours. (w/o Flash-att)	60M	0.9h	563MiB
Ours. (Flash-att)	60M	0.8h	534MiB

As shown in Table III, the model of Lin et al. requires 1.2 hours to complete training or inference, while our model (without Flash-Attention) and (with Flash-Attention) only requires 0.9 hours and 0.8 hours, respectively. Compared to the BERT-Large-based model and the model without Flash-Attention, the time consumption is reduced by 33.3% and 11.1%, respectively, indicating that our model can provide faster processing speeds while maintaining a smaller parameter size.

The model of Lin et al. (BERT-Lager) [2] requires 2151 MiB of GPU memory, whereas our model significantly reduces this demand, with the version without Flash-Attention requiring 563 MiB and the Flash-Attention version further reducing to 534 MiB. This indicates that our model, while maintaining a smaller parameter size, has reduced GPU memory usage by 75.1% and 5.1%, respectively, making it more suitable for operation in resource-constrained environments.

The results indicate a substantial improvement in F1 score and a significant reduction in resource consumption. We attribute these improvements to the synergistic effect of Flash-Attention and global constraints within our model. However, we also acknowledge potential limitations, such as the model's generalizability to other domains and the need for further adaption to enhance its robustness.

E. Discussion

The ALBERT-F model, without the implementation of Flash-Attention, exhibits an average runtime of 0.9 hours, which is further reduced to 0.8 hours with the integration of Flash-Attention. This is a significant reduction compared to the 1.2 hours required by the BERT-Large model. Concurrently, the GPU memory consumption is markedly decreased from 2151 MiB for the BERT-Large to 563 MiB for the ALBERT-F model without Flash-Attention, and an additional reduction to 534 MiB is achieved with the utilization of Flash-Attention. These results indicate that the ALBERT-F model substantially diminishes resource consumption while maintaining performance, making it particularly suitable for scenarios with limited computational resources.

The fusion of global constraints and prompting strategies enhances the model's generalizability to unknown events, rendering it more competitive in zero-shot learning tasks. This characteristic implies that in practice, the model can make reasonable predictions for new event types even without specific training data, which is invaluable in situations where data is scarce or difficult to annotate. However, despite the ALBERT-F model demonstrating advantages in multiple aspects, its limitations in handling complex event structures and long-distance dependencies remain a subject worthy of investigation. Complex events often involve multi-layered nested semantic relationships, and long-distance dependencies require the model to capture associations between words that are distant in the text. The model's performance may be compromised in such cases, as traditional attention mechanisms may not effectively span long sequences to capture crucial information. Therefore, future research could focus on developing more advanced attention mechanisms or model architectures to strengthen the model's comprehension of complex events.

V. CONCLUSION

In conclusion, we propose a ALBERT-F model for zero-shot EAC that employs global constraints and prompting. Compared to previous works, our model has a significantly lower parameter count, which not only reduces storage requirements but also potentially mitigates the risk of model overfitting. Additionally, it offers advantages in terms of run time, implying faster iterations and adaptation to new data. In terms of GPU memory usage, our model is substantially suitable for operation on devices with limited memory. These advantages make our model particularly appealing in resource-constrained environments.

VI. LIMITATIONS

In this section, we summarize the limitations of our work as follows:

1) Expressiveness: Although the ALBERT-F model demonstrates exceptional resource efficiency, it may not match the robust expressiveness of large language models in certain complex natural language understanding tasks. Large language models typically excel in handling intricate linguistic structures and long-distance dependencies due to their substantial parameter count and deeper network architectures.

2) Domain-specific performance: In certain domains or tasks, large language models may exhibit superior performance due to exposure to a more diverse range of texts during their pre-training phase. While the ALBERT-F model possesses strong zero-shot learning capabilities, it may require additional domain adaptation to achieve optimal results with specialized terminology and concepts in specific fields.

3) Scalability: Although the ALBERT-F model shows significant optimization in resource consumption, whether it can maintain these advantages when dealing with larger datasets or more complex tasks, or if further adjustments to the model structure and parameters are needed, remains a subject that necessitates further research and validation.

ACKNOWLEDGMENT

We sincerely appreciate the collaboration and insightful discussions with all authors, which greatly contributed to this work. Our thanks also extend to the anonymous reviewers for their thorough and thoughtful evaluations, which have helped us refine our approach and presentation of the findings.

REFERENCES

[1] Q. Li, J. Li, J. Sheng, S. Cui, J. Wu, Y. Hei, H. Peng, S. Guo, L. Wang, A. Beheshti *et al.*, "A survey on deep learning event extraction: Approaches and applications," *IEEE Transactions on Neural Networks and Learning Systems*, 2022.

[2] Z. Lin, H. Zhang, and Y. Song, "Global constraints with prompting for zero-shot event argument classification," in *Findings of the Association for Computational Linguistics: EACL 2023*, 2023, pp. 2527–2538.

[3] Z.-X. Ye and Z.-H. Ling, "Multi-level matching and aggregation network for few-shot relation classification," in *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, 2019, pp. 2872–2881.

[4] V. D. Lai, F. Dernoncourt, and T. H. Nguyen, "Exploiting the matching information in the support set for few shot event classification," in *Advances in Knowledge Discovery and Data Mining: 24th Pacific-Asia Conference, PAKDD 2020, Singapore, May 11–14, 2020, Proceedings, Part II 24*. Springer, 2020, pp. 233–245.

[5] J. Sheng, S. Guo, Z. Chen, J. Yue, L. Wang, T. Liu, and H. Xu, "Adaptive attentional network for few-shot knowledge graph completion," *arXiv preprint arXiv:2010.09638*, 2020.

[6] I.-H. Hsu, K.-H. Huang, E. Boschee, S. Miller, P. Natarajan, K.-W. Chang, and N. Peng, "Degree: A data-efficient generation-based event extraction model," in *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 2022, pp. 1890–1908.

[7] L. Huang, H. Ji, K. Cho, I. Dagan, S. Riedel, and C. Voss, "Zero-shot transfer learning for event extraction," in *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 2018, pp. 2160–2170.

[8] J. Liu, Y. Chen, K. Liu, W. Bi, and X. Liu, "Event extraction as machine reading comprehension," in *Proceedings of the 2020 conference on empirical methods in natural language processing (EMNLP)*, 2020, pp. 1641–1651.

[9] Q. Lyu, H. Zhang, E. Sulem, and D. Roth, "Zero-shot event extraction via transfer learning: Challenges and insights," in *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*, 2021, pp. 322–332.

[10] H. Zhang, H. Wang, and D. Roth, "Zero-shot label-aware event trigger and argument classification," in *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*, 2021, pp. 1331–1340.

[11] R. Chen, C. Qin, W. Jiang, and D. Choi, "Is a large language model a good annotator for event extraction?" in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 38, no. 16, 2024, pp. 17 772–17 780.

[12] Z. Lan, M. Chen, S. Goodman, K. Gimpel, P. Sharma, and R. Soricut, "Albert: A lite bert for self-supervised learning of language representations," in *International Conference on Learning Representations*, 2019.

[13] T. Dao, D. Fu, S. Ermon, A. Rudra, and C. Ré, "Flashattention: Fast and memory-efficient exact attention with io-awareness," *Advances in Neural Information Processing Systems*, vol. 35, pp. 16 344–16 359, 2022.

[14] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," *arXiv preprint arXiv:1810.04805*, 2018.

[15] B. M. Sundheim, "Overview of the fourth message understanding evaluation and conference," in *Proceedings of the 4th conference on Message understanding - MUC4 '92*, Jan 1992.

[16] R. Grishman and B. Sundheim, "Message understanding conference-6," in *Proceedings of the 16th conference on Computational linguistics*, Jan 1996.

[17] R. Grishman, D. Westbrook, and A. Meyers, "Nyu's english ace 2005 system description," *Ace*, vol. 5, no. 2, 2005.

[18] W. Xiang and B. Wang, "A survey of event extraction from text," *IEEE Access*, vol. 7, pp. 173 111–173 137, 2019.

[19] Y. Chen, L. Xu, K. Liu, D. Zeng, and J. Zhao, "Event extraction via dynamic multi-pooling convolutional neural networks," in *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, 2015, pp. 167–176.

[20] Y. Chen, S. Liu, S. He, K. Liu, and J. Zhao, "Event extraction via bidirectional long short-term memory tensor neural networks," in *Chinese Computational Linguistics and Natural Language Processing Based on Naturally Annotated Big Data: 15th China National Conference, CCL 2016, and 4th International Symposium, NLP-NABD 2016, Yantai, China, October 15-16, 2016, Proceedings 4*. Springer, 2016, pp. 190–203.

[21] L. Sha, F. Qian, B. Chang, and Z. Sui, "Jointly extracting event triggers and arguments by dependency-bridge rnn and tensor-based argument interaction," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 32, no. 1, 2018.

[22] D. Wadden, U. Wennberg, Y. Luan, and H. Hajishirzi, "Entity. relation, and event extraction with contextualized span representations," in *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, Jan 2019.

[23] Y. Lin, H. Ji, F. Huang, and L. Wu, "A joint neural model for information extraction with global features," in *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, Jan 2020.

[24] H. Wang, M. Chen, H. Zhang, and D. Roth, "Joint constrained learning for event-event relation extraction," *arXiv preprint arXiv:2010.06727*, 2020.

[25] P. Blache, "Constraints, linguistic theories, and natural language processing," in *International Conference on Natural Language Processing*. Springer, 2000, pp. 221–232.

[26] K. Ganchev, J. Graça, J. Gillenwater, and B. Taskar, "Posterior regularization for structured latent variable models," *The Journal of Machine Learning Research*, vol. 11, pp. 2001–2049, 2010.

[27] T. Hagendorff, "Mapping the ethics of generative ai: A comprehensive scoping review," *arXiv preprint arXiv:2402.08323*, 2024.

[28] X. Fang, S. Che, M. Mao, H. Zhang, M. Zhao, and X. Zhao, "Bias of ai-generated content: an examination of news produced by large language models," *Scientific Reports*, vol. 14, no. 1, p. 5224, 2024.

[29] L. Acion, M. Rajngewerc, G. Randall, and L. Etcheverry, "Generative ai poses ethical challenges for open science," *Nature Human Behaviour*, vol. 7, no. 11, pp. 1800–1801, 2023.

[30] B. Wang and A. Komatsuzaki, "GPT-J-6B: A 6 Billion Parameter Autoregressive Language Model," https://github.com/kingoflolz/mesh-transformer-jax, May 2021.

[31] G. Bai, Z. Chai, C. Ling, S. Wang, J. Lu, N. Zhang, T. Shi, Z. Yu, M. Zhu, Y. Zhang *et al.*, "Beyond efficiency: A systematic survey of resource-efficient large language models," *arXiv preprint arXiv:2401.00625*, 2024.

[32] S. Zeng, J. Liu, G. Dai, X. Yang, T. Fu, H. Wang, W. Ma, H. Sun, S. Li, Z. Huang *et al.*, "Flightllm: Efficient large language model inference with a complete mapping flow on fpgas," in *Proceedings of the 2024 ACM/SIGDA International Symposium on Field Programmable Gate Arrays*, 2024, pp. 223–234.

[33] J. Kaddour, J. Harris, M. Mozes, H. Bradley, R. Raileanu, and R. McHardy, "Challenges and applications of large language models," *arXiv preprint arXiv:2307.10169*, 2023.

[34] H. Dai, Y. Song, and H. Wang, "Ultra-fine entity typing with weak supervision from a masked language model," in *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, Jan 2021. [Online]. Available: http://dx.doi.org/10.18653/v1/2021.acl-long.141

[35] G. Doddington, A. Mitchell, M. Przybocki, L. Ramshaw, S. Strassel, and R. Weischedel, "The automatic content extraction (ace) program tasks, data, and evaluation," *Language Resources and Evaluation,Language Resources and Evaluation*, May 2004.

[36] H. Ji and R. Grishman, "Refining event extraction through cross-document inference," *Meeting of the Association for Computational Linguistics,Meeting of the Association for Computational Linguistics*, Dec 2008.

A Personalized Hybrid Tourist Destination Recommendation System: An Integration of Emotion and Sentiment Approach

Suphitcha Chanrueang[1], Sotarat Thammaboosadee[2]*, Hongnian Yu[3]
Faculty of Engineering, Mahidol University, Nakhon Pathom, Thailand[1, 2]
School of Computing Engineering and the Built Environment, Edinburgh Napier University, Edinburgh, United Kingdom[3]

***Abstract*—This research introduces a personalized hybrid tourist destination recommendation system tailored for the growing trend of independent travel, which leverages social media data for trip planning. The system sets itself apart from traditional models by incorporating both emotional and sentiment data from social platforms to create customized travel experiences. The proposed approach utilizes Machine Learning techniques to improve recommendation accuracy, employing Collaborative Filtering for emotional pattern recognition and Content-based Filtering for sentiment-driven destination analysis. This integration results in a sophisticated weighted hybrid model that effectively balances the strengths of both filtering techniques. Empirical evaluations produced RMSE, MAE, and MSE scores of 0.301, 0.317, and 0.311, respectively, indicating the system's superior performance in predicting user preferences and interpreting emotional data. These findings highlight a significant advancement over previous recommendation systems, demonstrating how the integration of emotional and sentiment analysis can not only improve accuracy but also enhance user satisfaction by providing more personalized and contextually relevant travel suggestions. Furthermore, this study underscores the broader implications of such analysis in various industries, opening new avenues for future research and practical implementation in fields where personalized recommendations are crucial for enhancing user experience and engagement.**

Keywords*—Recommendations; hybrid recommendation system; Collaborative Filtering; Content-based Filtering; social media data; travel planning

I. Introduction

Tourism significantly boosts the economy, creates jobs, and reduces poverty through spending, investments, and government backing [1]. Effective government policies are crucial for maximizing tourism's benefits and fostering overall economic growth. Social media , especially Facebook, plays a pivotal role in influencing travel choices and providing essential information to travelers [2]. Many travelers primarily rely on blogs and vlogs [3] for travel inspiration.

Emotions expressed on Facebook, such as anger, sadness, fear, joy, and love, can be identified through advanced textual analysis techniques [4]. Among these, anger, sadness, and fear are the most common, while love and amazement top the reactions [5]. Sentiment analysis, powered by sophisticated machine learning algorithms, evaluates sentiments on Facebook by analyzing data from the Top Page [6]. Understanding user perceptions and engagement is crucial for informed decision-making on social media platforms. This technique can also be applied to other digital platforms to assess customer feedback sentiment. Automating this process facilitates the examination of large datasets, effectively addressing the inherent challenges in sentiment analysis [7].

Recommender systems are instrumental in helping tourists discover attractions that match their personal preferences and needs. These systems utilize various techniques, including social and Bayesian networks, Collaborative Filtering (CF) algorithms, and deep learning models, to analyze user behavior, textual sentiment, and similarities among options. By considering user characteristics, behaviors, social network connections, and search contexts, these systems can make precise recommendations for tourist destinations that align with individual interests [8].

In this study, we introduce a personalized hybrid tourist destination recommendation system that leverages emotional and sentiment data from social media platforms. Unlike conventional models, our system integrates these emotional cues to provide more nuanced and accurate travel suggestions. Our system combines CF and Content-based Filtering (CB) with sentiment analysis techniques. CF is used to recognize emotional patterns, while CB analyzes sentiment-driven data, resulting [9] in a sophisticated weighted hybrid model [10]. This approach ensures recommendations are finely tuned to capture the nuanced emotional responses of users.

By integrating emotional and sentiment analysis, our system enhances user satisfaction by adapting to changes in user preferences over time, providing dynamic and contextually aware recommendations. This leads to a more engaging and satisfying user experience. The key advancements of our model include dynamic weighting of user data, enhanced emotional resonance, and adaptability. These improvements make our model more accurate and personalized compared to traditional methods.

Overall, the sentiment and emotion-based Weighted Hybrid technique not only improves the technical robustness of recommender systems but also significantly elevates their practical application by delivering a more personalized, accurate, and emotionally resonant travel experience. This advancement represents a substantial leap forward in the field of tourism recommendation systems, setting a new standard for personalized travel planning.

In terms of technical aspects, the proposed model combines CF and CB with sentiment analysis techniques. CF is used to recognize emotional patterns, while CB analyzes sentiment-driven data, resulting in a sophisticated weighted hybrid model. This approach ensures recommendations are finely tuned to capture the nuanced emotional responses of users. To the best of our knowledge, no previous research has utilized emotions derived from social media reactions in their recommendation systems.

Regarding practical applications, by integrating emotional and sentiment analysis, the developed approach enhances user satisfaction by adapting to changes in user preferences over time, providing dynamic and contextually aware recommendations. This leads to a more engaging and satisfying user experience. The key advancements of this model include dynamic weighting of user data, enhanced emotional resonance, and adaptability. These improvements make the proposed model more accurate and personalized compared to traditional methods.

Therefore, the objectives of the presented paper are to:

1) Extract and analyze data focusing on the emotions and sentiments in posts, comments, and reactions about tourist spots to discern their influence on travel decisions.

2) Develop a state-of-the-art hybrid recommender system that combines CB and CF with sentiment analysis from Facebook data, offering personalized suggestions for tourist destinations.

The paper is structured into six sections, including a comprehensive review of related works in Section II, system design and methods in Section III, experimental results in Section IV, a discussion of these findings in Section V, and conclusions and future directions in Section VI.

II. Related Works

A. Recommender Systems

Recommendation systems have become indispensable across numerous sectors, including e-commerce, entertainment, news, and social networking, by facilitating access to tailored information and resources. These systems streamline the search process, allowing users to find resources suitable to their needs by providing individualized suggestions or guiding them to relevant resources within a large data space. They simplify the process of finding information and solutions, making it easier for customers and project providers to identify and receive projects and other services [11]. In the tourism sector, these systems assist users in locating resources that match their specific requirements by offering personalized recommendations or directing them towards pertinent resources within a vast data environment [12]. By analyzing user preferences and behavior, they filter and present tailored options, significantly reducing the time and effort needed to find relevant information or items in an otherwise overwhelming data landscape [13]. They also play a crucial role in guiding customers throughout their shopping journey, presenting the most relevant products without the need for explicit searches [14]. By incorporating ontologies and machine-learning algorithms, recommender systems enhance accuracy and efficiency [15], addressing challenges and improving business productivity.

Recommendation systems can be categorized into several types, each with a unique approach to providing personalized recommendations. CF is one of the most common methods [16] and can be divided into user-based and item-based approaches. User-based Collaborative Filtering recommends items based on the preferences of users who have similar tastes, while Item-based Collaborative Filtering suggests items that are similar to those the user has previously liked or interacted with. CB focuses on recommending items that share similar attributes or features with those the user has shown interest in. Hybrid Systems (HS) combine multiple recommendation techniques, such as CB and CF, to enhance the overall accuracy and relevance of the recommendations. Context-aware Recommender Systems consider contextual factors such as time, location, or current activity to tailor recommendations more closely to the user's present situation. Demographic Recommender Systems provide suggestions based on demographic data, such as age, gender, or education level.

The research methodologies employed in tourism recommendation systems exhibit considerable diversity. Some studies focus on analyzing and quantifying user sentiment toward tourism destinations based on text reviews [17], integrating these sentiments [18] into the recommendation model. Others employ hybrid methods that combine CB and CF, utilizing preprocessed data from websites for recommendation. Additional approaches include probabilistic topic modeling and custom day itinerary models to analyze tourist travel patterns and preferences. While some studies emphasize recommending points of interest within a tourist attraction based on visitor interests, others offer broader recommendations spanning entire countries.

In summary, personalized hybrid recommendation systems across various domains leverage individualized suggestions and advanced techniques like opinion mining and hybrid filtering (HF) to enhance accuracy [19] and user experience. Despite their effectiveness in simplifying information discovery and improving the customer journey, these systems encounter challenges related to relevance computation, personalization, and the integration of specific user interests within large data spaces.

B. Factors Influencing Personal Travel Destination Choices

The decision-making process regarding travel destinations is influenced by a complex interplay of factors that vary significantly among individuals based on their preferences and circumstances, timing of the travel, and the quality of infrastructure and traffic conditions, which collectively shape the feasibility [20] and appeal of a destination [21]. The broad availability of information from different channels like online platforms, print media, and travel agencies plays a crucial role in informing potential travelers about their options, thereby significantly influencing their destination choices [22]. Moreover, the operational efficiency and overall attractiveness of a tourism destination, determined by factors like labor quality, capital investment, technological advancement, environmental sustainability, financial expenditure, generated revenue, and the potential length of stay, are critical in swaying personal travel

destination choices [23]. These considerations encompass a range of practical, economic, and subjective factors that contribute to the appeal and competitiveness of a destination, highlighting the multifaceted nature of travel decision-making. In essence, the choice of a travel destination emerges from a dynamic balance of these practical considerations, individual preferences, and the intrinsic attributes of the destination itself, underscoring the complexity of travel planning and the importance of understanding these factors for stakeholders in the tourism industry.

C. Sentiment and Emotion as New Factors for Recommendation Systems in the Tourism Domain

Sentiment and emotion play a crucial role in enhancing recommendation systems [24], [25] in the tourism domain. Incorporating sentiment analysis from user-generated content like reviews can significantly improve the accuracy [26] and quality of recommendations. By utilizing sentiment and emotion scores derived from user reviews, tourism recommendation systems can better capture user preferences and generate personalized recommendation lists based on semantic similarity [27]. Additionally, aspect-based sentiment analysis models can extract sentiment polarity from reviews, providing insights into tourists' evaluations and enhancing service and product upgrades [28].

These sentiment-driven approaches not only help in understanding tourists' emotions but also assist tourism organizations in their decision-making processes, ultimately leading to more effective recommendations [29] and greater customer satisfaction.

D. Weight Hybrid Recommendation System

A weighted hybrid model in recommendation systems combines multiple techniques, such as CF and CB, by assigning different weights to each technique based on their effectiveness in predicting user preferences. This approach leverages the strengths of each method to enhance recommendation accuracy and relevance [30]. The model integrates CF, which identifies patterns based on user interactions, and CB, which recommends items with similar attributes to those the user likes. Each method generates a recommendation score, and their contributions are weighted differently, with the weights determined [31] through experiments or data characteristics. One of the key advantages of a weighted hybrid model is its flexibility. The weights can be adjusted dynamically based on the recommendation context, user behavior, or data changes. This adaptability helps improve recommendation relevance over time, addressing limitations like the cold start problem and data sparsity [32], and providing more accurate and diverse suggestions.

In summary, a weighted hybrid model strategically combines multiple recommendation techniques with assigned weights to enhance accuracy and personalization. This method leverages the strengths of different techniques, adapts to changing data and user behaviors, and provides a robust and personalized recommendation experience.

III. System Design and Methods

The model aims to develop an innovative personalized tourist attraction recommendation system, showcasing a novel architecture that incorporates advanced HF techniques, as illustrated in Fig. 1. This system leverages three distinct types of information, marking a significant advancement in tourist recommendation technologies.

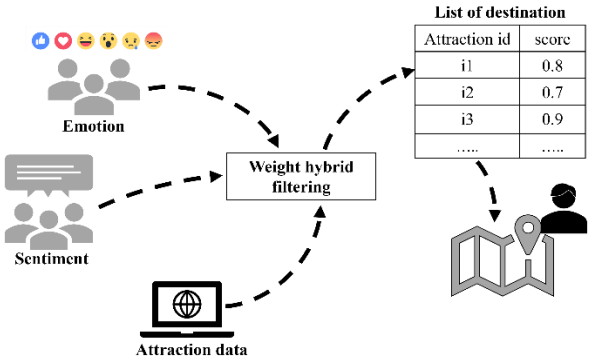


Fig. 1. Proposed architecture for a tourism recommender system.

A. Data Collection and Authorization

In collaboration with five Thai Facebook fan pages, we secured authorization to extract a wealth of data, including comments from followers, their emotional reactions via the 'reaction' button, and detailed information on various tourist attractions. This comprehensive dataset, accumulated two years, offering deep insights into tourist preferences and behaviors.

The architecture of a Personalized Hybrid Tourist Destination Recommendation System, as depicted in Fig. 1, employs a hybrid approach that integrates both CB and CF models. An in-depth explanation of the main segments and their respective processes is provided in multiple sections, as shown in Fig. 2.

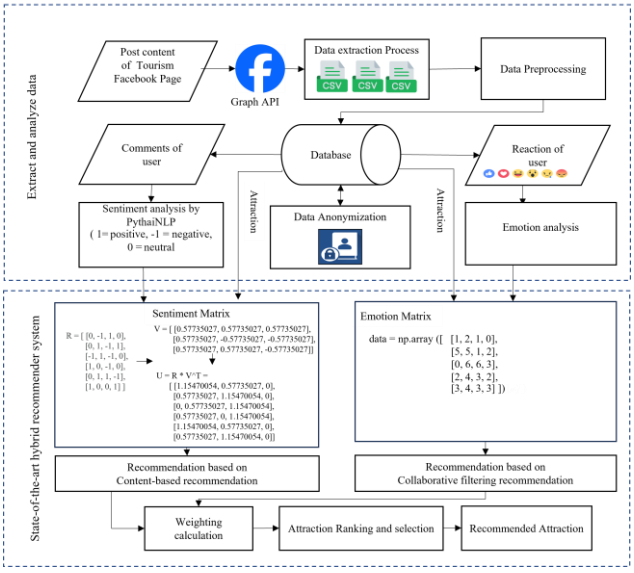


Fig. 2. Architecture of a weighted hybrid recommendation system.

B. Data Collection Process

As illustrated in Fig. 2, the architecture of a WHF begins with the data collection process, which gathers information from various sources. One such source is Facebook, where data is collected using the Graph API. This data, which ranges from

user demographics to their interactions with content, is typically stored in a CSV format to facilitate handling and analysis.

The initial stage of data preprocessing involves extracting and refining the gathered information through a series of steps. This phase focuses on ensuring data quality by cleaning inaccuracies, making necessary corrections, and identifying essential information to enable accurate recommendations. In respect of user privacy, data anonymization is implemented. During this process, data is stripped of personal identifiers, ensuring user privacy protection while still allowing for personalized content suggestions.

The system performs sentiment analysis by categorizing user comments as positive, neutral, or negative. Additionally, it evaluates user reactions, such as likes or emojis, by assigning numerical values from one to six to quantify user engagement. The processed data is organized into a matrix format, where users are listed alongside the items they interact with, creating a comprehensive map of interactions.

C. Data Selection Criteria

Selecting data for a tourism recommendation system is a complex endeavor due to the sheer number of tourism attractions and the overwhelming volume of information available online and across social media platforms. Existing recommender systems encounter challenges in delivering precise recommendations, as they must contend with variations in users' interests, the ever-changing contexts, and the sequential patterns of travel [33]. The lack of sufficient historical user data in the tourism sector further complicates matters, leading to difficulties such as cold starts and data sparsity, which hinder the delivery of accurate and reliable recommendations [34]. Furthermore, the infrequent browsing and purchasing of travel products, along with the influence of factors such as departure, destination, and price, adds another layer of complexity to recommending travel products [35].

Our method for selecting sentiment and emotion data follows strict criteria to ensure its relevance, accuracy, and diversity, while maintaining privacy and ethical standards. We identify key emotional data like user comments and reactions, verify their accuracy, and source them from various platforms, including social media. This data must be scalable and comprehensive to support reliable analysis and improve the system's ability to offer accurate and trustworthy recommendations, thereby enhancing its effectiveness.

D. Data Extraction Process

In social media data extraction, using Facebook's Graph API is crucial for researchers and technologists retrieving data from fan pages. Facebook Graph API allows developers to access and interact with Facebook data, such as user profiles, posts, and photos, using HTTP requests. It requires authentication via access tokens for secure data access and supports CRUD operations. This API enables the integration of Facebook data into applications for social media management, analytics, and personalized content delivery. Our study utilizes Facebook's Graph API [36] as a key tool, following a systematic method that values user privacy and adheres to privacy regulations , as shown in Fig. 3.



Fig. 3. Data preparation process.

The process starts by configuring an application on Facebook Developer Console to obtain an App ID and App Secret, enabling user consent through OAuth 2.0 for an access token. The access token allows fetching data from fan pages using Graph API, which is crucial for analysis.

Data retrieval involves accessing posts, reactions, and comments from fan pages through Graph API to gather raw data for analysis and recommendations. Anonymization techniques are applied to protect user data, including removing identifiers, randomizing sensitive data, and auditing the process regularly. The system securely stores anonymized reactions and comments in a privacy-compliant database and manages them through an ETL pipeline to maintain anonymization. Utilizing Facebook's Graph API involves access setup, data retrieval, and strict anonymization, laying the groundwork for further data processing in the Data Preparation phase.

E. Data Preparation

The data preparation phase involves handling three types of raw data crucial for constructing the dataset.

1) Social network data: Provides attraction names from social networks to identify and categorize tourist destinations.

2) Sentiment data transformation: Categorizes opinions from user reviews into three sentiment categories using the PyThaiNLP library [38] for sentiment analysis tailored to the Thai language. This library is vital for tasks specific to Thai, such as word tokenization for interpreting user sentiments accurately.

3) Model for identifying and scoring user emotions: This model identifies core emotions such as anger, disgust, fear, happiness, sadness, and surprise. The proposed model is consistent with Ekman's framework and is widely acknowledged in the field of emotion recognition. Emotions are scored to reflect user engagement [39]: Love = 6, Like = 5, Haha = 3, Wow = 4, Sad = 2, and Angry = 1, providing insights into the subtleties of user emotional feedback. The process is illustrated in Fig. 4.

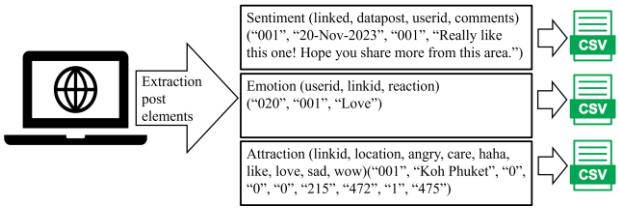


Fig. 4. Data preparation process.

Social network and user review data hold implicit details, requiring specific feature extraction techniques. Social computing retrieves social information from social network data, whereas sentiment analysis uncovers emotional cues from user feedback. Data collection and integration procedures prioritize safeguarding user privacy by anonymizing sensitive

information. These privacy steps are vital for upholding trust and efficiency in our recommendation system, offering personalized suggestions while protecting user privacy.

F. Creation of Recommendation System

The creation of the recommendation system involves the integration of CF and CB models. Below, we detail the methodologies used for each model:

1) Collaborative Filtering model: The development of a tourist attraction recommendation system using CF with the Singular Value Decomposition (SVD) algorithm is achieved through the following steps, with conceptual underpinnings and practical implementation in Python. Installation and Setup: Begin by installing the Surprise package using a package manager like pip, and import necessary classes such as Dataset, Reader, SVD, and accuracy functions from the library.

Conceptual Framework of SVD: For a matrix A with dimensions $m \times n$, SVD decomposes A into three matrices (1):

$$R = U \cdot \Sigma \cdot V^T \tag{1}$$

U: User ser features matrix, where rows represent users and columns represent hidden characteristics.

Σ: Diagonal matrix of singular values indicating the importance of the latent features.

V^T: Item-feature matrix, with row for items (attractions) and columns for latent features.

The procedure includes constructing matrices for User factors and Item factors, as illustrated in Fig. 5. Based on the example data provided, this step entails training a model using SVD.

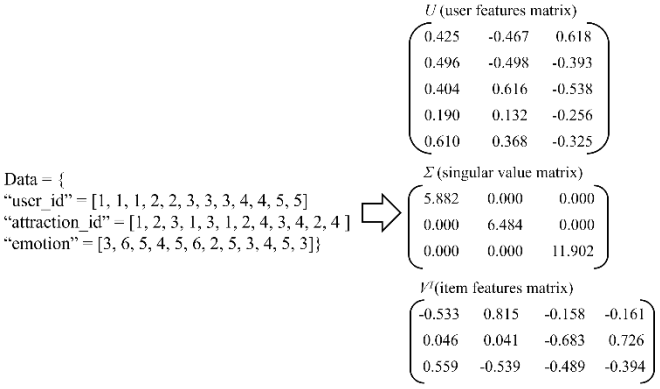


Fig. 5. Matrices are generated for user factors and item factors.

Data Preparation: Prepare the data, including 'User_id', 'Attraction_id', and 'emotion' scores ranging from 1 to 6. An example data structure is provided.

Model Training: Define the range of 'emotion' scores using the Reader class, setting the minimum and maximum values. Load the data using the Dataset module, formatted according to the Reader specifications. Split the data into a training set for training the model and a test set for evaluating its performance. Instantiate the SVD algorithm and fit it to the training dataset.

Prediction and Evaluation: The system makes predictions for unseen user-attraction combinations in the test set. For example, it predicts the 'emotion' score for a given user-attraction pair, such as 'User123' (user_ID) and 'Attraction456' (attraction_ID). If the actual emotion score given by the user is 4 and the score estimated by the SVD model is 3.8, this indicates the model's performance. The details section, which shows 'was impossible': False, confirms that the prediction was successfully computed. To evaluate the model's accuracy, metrics such as Root Mean Square Error (RMSE) and Mean Absolute Error (MAE) are used [40]. These metrics compare the predicted scores against the actual scores in the test set. RMSE measures the square root of the average squared differences between the predicted and actual values, while MAE measures the average of the absolute differences between the predicted and actual values.

Practical Considerations: To protect user privacy, data anonymization is crucial. This involves removing or obscuring personal identifiers from the data to prevent individual users from being easily identified. Regular audits of the anonymization process are essential to ensure the data remains secure and to minimize the risk of re-identification, maintaining user trust and compliance with privacy regulations.

2) Content-based filtering model: Implementing a CB model for travel recommendations with Support Vector Machine (SVM) involves a structured approach, focusing on harnessing powerful capabilities in handling complex data patterns. The key steps in this process include data preparation, model selection and tuning, training, and evaluation.

Data Preparation: Transform sentiment data into numerical scores of (-1, 0, 1) to align with the SVM models' requirements. This step converts subjective sentiments into objective data points. Convert characteristics of tourist attractions, such as type and location, into numerical forms labeled as 'Attraction_type'. This numerical transformation is crucial for machine learning algorithms to process and learn from the data.

Model Selection and Tuning: Choosing and tuning the model is crucial as we opt for SVM due to its proficiency in classification tasks, especially its effectiveness in high-dimensional spaces and its ability to handle non-linear data separation through kernel methods. Fine-tune parameters like C (regularization), `kernel type`, and `gamma` (kernel coefficient) to optimize the model for the dataset. In classification tasks, the primary objective is to find an optimal hyperplane that best separates the classes in the given dataset.

The SVM algorithm seeks to find an optimal hyperplane that best separates different classes in the dataset, as expressed by the hyperplane Eq. (2):

$$w^T x + b = 0 \tag{2}$$

where w represents the weight vector, x is the vector of data points, and b signifies the bias. Alongside the hyperplane equation, SVM involves an optimization problem, which is geared towards maximizing the margin between the data classes, as shown in Eq. (3):

$$min(w, b) \left(\frac{1}{2}||w||^2\right) \tag{3}$$

Subject to the constraints for each data point *i*:

$$y_i(w^T x_i + b) \geq 1, \forall i \quad (4)$$

In this context: $||w||2$ is the norm of the weight vector, and minimizing it is key to maximizing the margin. The labels of the data points are denoted by y_i and x_i represents each data point.

Training and Evaluation: During model training, we divide our dataset into training and test sets to both train the model and evaluate its predictive performance accurately. After training, we assess the model's performance using metrics such as accuracy, precision, recall, and F1-score. These metrics provide a comprehensive understanding of the model's effectiveness in classification tasks.

Implementation: Once the model proves its effectiveness, it can be utilized to forecast user preferences for different tourist attractions and provide suitable recommendations. Utilizing SVM enables us to adjust and explore the model for optimal performance on our dataset. This methodology is especially beneficial for tasks demanding nuanced data analysis.

Visualize the sentiment feature matrix in SVM as shown in Table I to understand the data's distribution and how the model determines decision boundaries. Once the model is optimized and validated, it can accurately predict user preferences for various tourist attractions, providing personalized recommendations based on these analyzed features.

1) Weight hybrid recommendation: Integrating CB with CF in a WHS involves a structured approach to leverage the advantages of both methods. The workflow is as follows*:*

TABLE I. SENTIMENT FEATURE MATRIX OF SVM

Seq	Attraction_id	User_id	Sentiment
0	101	201	-1
1	102	202	0
3	104	204	-1
4	105	201	-1
5	105	202	0

a) Data processing and score calculation: Initially, both CB and CF systems process their respective datasets to compute scores for tourist destinations. These scores are based on each system's unique algorithms and the data provided.

b) Blending scores: The critical step of blending involves merging the scores from both systems using a predefined formula. This formula assigns specific weights to the scores from each system, balancing their contributions. For instance, the hybrid score can be calculated as:

$$\Sigma\chi o\rho\varepsilon_{H\psi\beta\rho\iota\delta} = \alpha \times \Sigma\chi o\rho\varepsilon_{\chi B} + (1 - \alpha) \times \Sigma\chi o\rho\varepsilon_{\chi o\lambda\lambda\alpha\beta o\rho\alpha\tau\iota\varpi\varepsilon} \quad (5)$$

Here, α represents the weight assigned to the score from the CB Filtering system.

1) Optimizing the weight parameter (α): Fine-tuning α is crucial for balancing CB and CF systems. Experimenting with various α values, calculating blended scores, and analyzing outcomes enhances recommendation accuracy and diversity. K-grid tuning optimizes model parameters by adjusting the K value for cross-validation groups. Selecting an optimal K value, evaluating model performance, and refining based on results analysis ensures models align with data characteristics and task requirements.

2) Performance evaluation: After the hybrid recommendation system is in place, its performance should be evaluated to ensure it provides relevant and accurate suggestions. Utilizing feedback from users and analyzing performance metrics such as Accuracy, Precision, Recall, and F1-Score are integral to this phase. These evaluations facilitate ongoing refinement, improving the system's capability to effectively cater to user preferences.

This methodology underscores the importance of a balanced integration of CB and CF techniques, ensuring that the recommendations are not only accurate but also varied, catering to the diverse interests of users.

IV. EXPERIMENTAL RESULTS

The analysis of recommendation results derived from a comprehensive dataset provides valuable insights for the development and evaluation of a tourist attraction recommendation system. The following discussion presents a structured analysis based on the dataset results.

A. Dataset Results

The data extraction process leveraged a fan page dedicated to tourist attraction reviews, a valuable resource for assessing recommendation models. The dataset's composition and its implications for the recommendation system are as follows:

1) Dataset overview: The dataset contains 252,568 records split into two segments: 151,541 records for training and 101,027 for testing. This division ensures a robust framework for both developing and validating the recommendation model.

2) User participation and attraction diversity: A total of 38,739 users have contributed to the dataset, reviewing 508 different attractions. Examples of the data can be found in Tables II to IV. This level of participation and variety underscores the dataset's richness and diversity, providing a solid foundation for generating nuanced and wide-ranging recommendations.

3) Sentiment classification: Sentiments extracted from the dataset are categorized into three categories: positive, negative, and neutral, as indicated in Table II. This classification facilitates a detailed understanding of user preferences and emotions regarding various attractions.

TABLE II. EXAMPLE OF SENTIMENT DATA

User Id	Attraction		Sentiment	
	Id	Type	Type	Score
1	157	1	Neutral	0.5
1	122	5	Positive	0.8
2	28	1	Positive	0.7
3	89	6	Positive	0.6
3	210	2	Neutral	0.5

4) Emotion ratings: User ratings are based on an emotional scale from 1 to 6, where each number corresponds to a specific emotion (e.g., love = 6, like = 5, wow = 4, haha = 3, sad = 2 and angry = 1), as illustrated in Table III.

TABLE III. EXAMPLE OF USER EMOTION DATA

User ID	Attraction		Emotion	
	Id	Type	Word	Value
1	627	2	wow	4
1	781	5	like	5
3	783	6	wow	4
3	210	2	like	5
7	157	1	love	6

The dataset shown in Table IV includes a variety of attraction types, illustrating the diverse interests of the users. Understanding the range of attractions is crucial for tailoring recommendations to suit individual user preferences effectively.

The following table, Table V, lists different types of attractions along with their corresponding type names.

TABLE IV. EXAMPLE OF ATTRACTION DATA

Attraction ID	Name		Location	Type
1	Khun Dan Prakarn Chon Dam		14.7994° N, 98.5969° E	4
2	Sai Yok National Park		14.417778° N, 98.747222°E	4
3	Singha Historical Park		14.03583° N, 99.23972°E	3
4	Phra Pathom Chedi		13° 49' 6.59" N, 100° 03' 22.20" E	6

TABLE V. EXAMPLE OF ATTRACTION TYPE DATA

Attraction Type	Type Name
1	Eco tourism
2	Arts and sciences educational attraction
3	Historical attraction
4	Natural attraction
5	Recreational attraction
6	Cultural attraction

B. Recommendation Results

The recommendation results in Table VI provide a comparative analysis of the performance of three models: CF, CB, and HF using Precision, Recall, and F1-score metrics. The HF model outperforms both CF and CB in all three metrics. This superior performance can be attributed to its ability to combine the strengths of both CF and CB techniques, along with additional enhancements such as sentiment and emotion analysis. This integration allows the HF model to provide more accurate and reliable recommendations, better predicting user preferences and enhancing the overall user experience in tourism settings. In conclusion, the HF model is the most effective approach for recommending tourist attractions, as it achieves the highest precision, recall, and F1-score, significantly improving recommendation accuracy compared to the CF and CB models.

TABLE VI. RESULT OF PRECISION, RECALL AND F1- SCORE

Model	Precision	Recall	F1-score
CF	0.780	0.740	0.760
CB	0.660	0.690	0.670
HF	0.850	0.830	0.840

This analysis compares the efficacy of three recommendation models: CF, CB Filtering, and HF, using Precision, Recall, and F1-score as performance metrics. Fig. 6 displays the trends of these metrics as the value of K changes, illustrating how different parameter settings affect model performance.

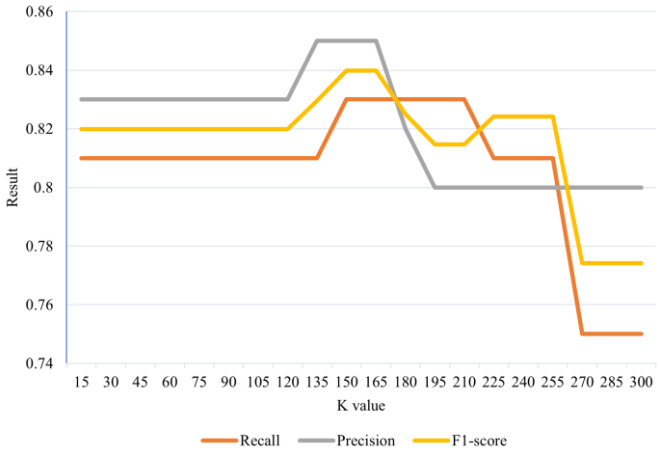


Fig. 6. Model discrimination score by K-value.

Content-Based Filtering vs. Collaborative Filtering: CF demonstrates robust performance with a well-balanced trade-off between Precision (0.78) and Recall (0.74), resulting in an F1-score of 0.76. In contrast, CB Filtering, while slightly less effective, demonstrated modest performance with Precision (0.66) and Recall (0.69), and an F1-score of 0.67. This suggests a modest decline in performance compared to CF. The HF model, which combines CB and CF, demonstrated superior performance across all metrics. It achieved the highest Precision (0.85), Recall (0.83), and F1-score (0.84), signifying its effectiveness in providing accurate and comprehensive recommendations.

The accuracy of a recommendation system, which measures how precisely it predicts user preferences, can be assessed using metrics like the RMSE or MAE [37]. These metrics, along with the MSE, provide insight into the system's performance and can be calculated using the functionalities available in Scikit-learn [41]. The results of these calculations are presented in Fig. 7.

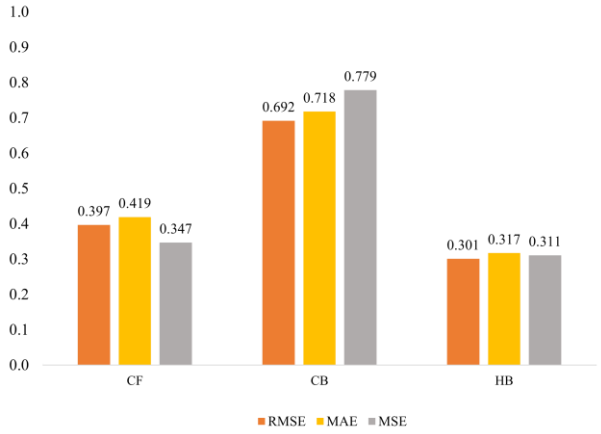


Fig. 7. Model performance comparison.

Error Rate Analysis: The analysis of RMSE, MAE, and MSE metrics presented in Fig. 7 revealed that the CB model exhibited the highest error rates, suggesting limitations in its predictive accuracy. The CF model showed moderate error values, indicating a balanced yet not optimal level of accuracy. The Hybrid model outperformed both with the lowest error values in all metrics, indicating its superior accuracy in predicting user preferences. This summary indicates that the HF model, which combines elements of both Collaborative and CB methodologies, offers the most effective approach for accurate user preference prediction.

C. Comparison of Methods, Items, and Values Results

The proposed method demonstrates the highest efficiency in recommendation systems, exhibiting the lowest error rates across all evaluated metrics, including RMSE, MAE, and MSE. This method integrates sentiment and emotion analysis with CF and CB techniques, leading to a significant improvement in recommendation accuracy. The integration of these emotional and sentimental data allows the system to better understand and predict user preferences, resulting in more personalized and precise recommendations.

In comparison, other methods such as those referenced in, [21], [30], and [42], utilize various combinations of CF and CB techniques, sometimes incorporating additional methods like SVD and SVM. Despite these efforts, they still exhibit higher error rates. For instance, the method in [21], which uses a hybrid approach with SVD and weighted techniques, shows better performance than some but still falls short compared to the proposed method. The methods in [33] and [42], although incorporating diverse techniques, demonstrate even higher error rates, indicating less accuracy in their recommendations. All the data is detailed in Table VII.

Overall, the proposed method's ability to integrate emotional and sentiment analysis with traditional filtering techniques sets it apart, achieving superior performance and underscoring the importance of these factors in enhancing recommendation systems.

TABLE VII. COMPARISONS WITH RECENT METHODS

Ref	Method	Technique	Result		
			RMSE	MAE	MSE
[21]	Hybrid	SVD Weighted	0.500	0.414	0.254
[30]	CB+CF	Weighted	0.880	0.670	-
[42]	Hybrid	Cosine, SVD, SVM	0.864	0.666	-
Proposed Method	Baseline+ CB+CF	SVD, SVM, Sentiment, Emotion	0.301	0.317	0.311

V. DISCUSSION

This study demonstrates the potential of a Hybrid Filtering method for enhancing efficiency and accuracy in tourist attraction recommendations. By intelligently combining CB and CF techniques with adjustable method weights, the proposed approach delivers highly personalized recommendations that align closely with individual user preferences. Previous research has highlighted the strengths of CB and CF methods individually, but the integration of these techniques with customizable weights offers a novel approach that addresses limitations in prior studies [8], [10]. Despite these advancements, the methodology encounters significant challenges in sentiment analysis and data extraction from social media platforms, particularly Facebook.

A. Sentiment Analysis Challenges

The study acknowledges the inherent complexities in interpreting emotions expressed on social media, consistent with findings from existing literature [4], [6]. Social media users often present idealized versions of their emotions, which may not accurately reflect their true sentiments [20]. Additionally, the diversity of content types (text, images, videos) and nuanced language used on these platforms further complicate sentiment interpretation [27]. These challenges underscore the need for advanced sentiment analysis tools capable of understanding diverse expressions and cultural contexts [26]. Future research could build on recent advancements in sentiment analysis techniques to improve interpretation accuracy [18].

B. Data Extraction Complexities

Relying on Facebook's Graph API for data retrieval introduces significant challenges, a problem well-documented in the literature [4]. Researchers must navigate strict personal data access restrictions, frequent API changes, data request limits, and complex verification processes, all while managing privacy risks. The complexity of ensuring compliance with Facebook’s policies adds another layer of difficulty, requiring careful data transformation to protect personal information—a process that is often time-consuming. While the WHF method demonstrates superior performance by effectively leveraging data from multiple sources, it still faces substantial hurdles in sentiment analysis and social media data extraction. These findings are consistent with earlier studies that have identified similar challenges in working with social media data. These challenges highlight critical areas for future research, emphasizing the need for:

1) Advanced sentiment analysis tools: The development of tools capable of accurately interpreting complex emotions expressed through various content types and linguistic nuances on social media.

2) Improved data extraction techniques: The exploration of efficient methods adaptable to the dynamic nature of social media platforms and APIs, while ensuring user privacy and data compliance [16], [23].

The discussion section highlights the efficacy of the WHS method in providing accurate and personalized tourist attraction recommendations. However, it also underscores the challenges in sentiment analysis and data extraction. Future research should focus on developing advanced sentiment analysis tools and improving data extraction techniques to further enhance the performance and reliability of hybrid recommendation systems.

VI. Conclusions and Future Directions

This section provides a summary of the research conclusions and suggests future directions, highlighting key findings, limitations, implications, and areas for further investigation.

A. Conclusions

The WHF model demonstrates significant potential for personalized recommendations in tourism, outperforming CF and CB models in Precision, Recall, and F1-score. Its ability to align recommendations with users' emotions and preferences highlights the model's superiority. This success has broader implications for recommendation systems across various sectors, where aligning with user emotions and preferences can enhance satisfaction and engagement through personalized experiences.

B. Limitations

The extraction of large volumes of data from social media is time-consuming and requires careful handling, slowing the process. Additionally, sentiment analysis faces challenges when dealing with abbreviations and slang, complicating accurate interpretation.

C. Future Directions

To further enhance the model's capabilities and expand its applications, the following strategies are proposed:

1) Advanced hybrid data preprocessing techniques: Implementing sophisticated hybrid data preprocessing methods to improve model efficiency and performance across various databases. This will facilitate more accurate comparisons and refinements, leading to superior recommendation accuracy.

2) Image-based preference analysis: Utilizing image-based approaches to analyze user preferences more accurately for travel products. Integrating visual data will enable the recommendation system to better understand and predict user interests.

By adopting these strategies, the WHF model can address current challenges in sentiment analysis and data extraction, thereby advancing its personalization and accuracy. These enhancements have the potential to transform recommendation systems not only in tourism but also across other sectors. Ongoing development and the integration of advanced techniques will ensure that recommendation systems continue to evolve, providing increasingly personalized and effective solutions to meet diverse user needs.

Acknowledgment

I sincerely thank the Ministry's studentship division for their generous scholarship, which has been vital to my doctoral research.

References

[1] S. Naseem, "The Role of Tourism in Economic Growth: Empirical Evidence from Saudi Arabia," Economies, vol. 9, no. 3, Art. no. 3, Sep. 2021, doi: 10.3390/economies9030117.

[2] L. S. Budovich, "The impact of religious tourism on the economy and tourism industry," HTS Teol. Stud. Theol. Stud., vol. 79, no. 1, May 2023, doi: 10.4102/hts.v79i1.8607.

[3] Shreeraksha Shankar and Hampesh K S, "Impact of Travel Blogs and Vlogs in Social Media on Tourism," Int. J. Eng. Technol. Manag. Sci., vol. 7, no. 1, pp. 16–22, Feb. 2023, doi: 10.46647/ijetms.2023.v07i01.004.

[4] M. Alonazi, "Analyzing Sentiment in Terms of Online Feedback on Top of Users' Experiences," Int. J. Adv. Comput. Sci. Appl., vol. 14, no. 11, 2023, doi: 10.14569/IJACSA.2023.0141114.

[5] V. Balakrishnan, V. Govindan, N. I. Arshad, L. Shuib, and E. Cachia, "Facebook user reactions and emotion: an analysis of their relationships among the online diabetes community," Malays. J. Comput. Sci., pp. 87–97, Dec. 2019, doi: 10.22452/mjcs.sp2019no3.6.

[6] J. Kim and C. Stavrositu, "Feelings on Facebook and their correlates with psychological well-being: The moderating role of culture," Comput. Hum. Behav., vol. 89, pp. 79–87, Dec. 2018, doi: 10.1016/j.chb.2018.07.024.

[7] D. D. Albesta, M. L. Jonathan, M. Jawad, O. Hardiawan, and D. Suhartono, "The impact of sentiment analysis from user on Facebook to enhanced the service quality," Int. J. Electr. Comput. Eng., vol. 11, no. 4, Art. no. 4, Aug. 2021, doi: 10.11591/ijece.v11i4.pp3424-3433.

[8] R. N. Mule and S. S. Mulik, "TRS – A rule based personalized tourism recommender system," Int. J. Res. Appl. Sci. Eng. Technol., vol. 10, no. 10, pp. 1280–1285, Oct. 2022, doi: 10.22214/ijraset.2022.47172.

[9] W. Xia, "Digital Transformation of Tourism Industry and Smart Tourism Recommendation Algorithm Based on 5G Background," Mob. Inf. Syst., vol. 2022, p. e4021706, Sep. 2022, doi: 10.1155/2022/4021706.

[10] D. Roy and M. Dutta, "A systematic review and research perspective on recommender systems," J. Big Data, vol. 9, no. 1, p. 59, May 2022, doi: 10.1186/s40537-022-00592-5.

[11] Z. Zhang et al., "Scholarly recommendation systems: a literature survey," Knowl. Inf. Syst., vol. 65, no. 11, pp. 4433–4478, Nov. 2023, doi: 10.1007/s10115-023-01901-x.

[12] G. Hui, C. Mang, Z. LiQing, and X. ShiKun, "Research on Personalized Recommendation Algorithms Based on User Profile," Int. J. Adv. Comput. Sci. Appl., vol. 15, no. 3, 2024, doi: 10.14569/IJACSA.2024.0150330.

[13] S. Bhattacharya, D. Sarkar, D. K. Kole, and P. Jana, "Chapter 9 - Recent trends in recommendation systems and sentiment analysis," in Advanced Data Mining Tools and Methods for Social Computing, S. De, S. Dey, S. Bhattacharyya, and S. Bhatia, Eds., in Hybrid Computational Intelligence for Pattern Analysis. , Academic Press, 2022, pp. 163–175. doi: 10.1016/B978-0-32-385708-6.00016-3.

[14] M. Rhanoui, M. Mikram, S. Yousfi, A. Kasmi, and N. Zoubeidi, "A hybrid recommender system for patron driven library acquisition and weeding," J. King Saud Univ. - Comput. Inf. Sci., vol. 34, no. 6, pp. 2809–2819, Jun. 2022, doi: 10.1016/j.jksuci.2020.10.017.

[15] M. Kshour, M. Ebrahimi, S. Goliaee, and R. Tawil, "New recommender system evaluation approaches based on user selections factor," Heliyon, vol. 7, no. 7, p. e07397, Jul. 2021, doi: 10.1016/j.heliyon.2021.e07397.

[16] F. Ricci, D. Massimo, and A. De Angeli, "Challenges for Recommender Systems Evaluation," in CHItaly 2021: 14th Biannual Conference of the

Italian SIGCHI Chapter, in CHItaly '21. New York, NY, USA: Association for Computing Machinery, Jul. 2021, pp. 1–5. doi: 10.1145/3464385.3464733.

[17] M. Hu, H. Li, H. Song, X. Li, and R. Law, "Tourism demand forecasting using tourist-generated online review data," Tour. Manag., vol. 90, p. 104490, Jun. 2022, doi: 10.1016/j.tourman.2022.104490.

[18] A. R. Alaei, S. Becken, and B. Stantic, "Sentiment Analysis in Tourism: Capitalizing on Big Data," J. Travel Res., vol. 58, no. 2, pp. 175–191, Feb. 2019, doi: 10.1177/0047287517747753.

[19] I. Mazlan, N. Abdullah, and N. Ahmad, "Exploring the Impact of Hybrid Recommender Systems on Personalized Mental Health Recommendations," Int. J. Adv. Comput. Sci. Appl., vol. 14, no. 6, 2023, doi: 10.14569/IJACSA.2023.0140699.

[20] X. Zheng, Y. Luo, L. Sun, J. Zhang, and F. Chen, "A tourism destination recommender system using users' sentiment and temporal dynamics," J. Intell. Inf. Syst., vol. 51, no. 3, pp. 557–578, Dec. 2018, doi: 10.1007/s10844-018-0496-5.

[21] Y. A. Akbar, Z. A. Baizal, and A. T. Wibowo, "Tourism Recommender System using Weighted Parallel Hybrid Method with Singular Value Decomposition," Indones. J. Comput. Indo-JC, vol. Vol. 6 No. 2, pp. 53-64 Pages, Sep. 2021, doi: 10.34818/INDOJC.2021.6.2.579.

[22] K. X. Li, M. Jin, and W. Shi, "Tourism as an important impetus to promoting economic growth: A critical review," Tour. Manag. Perspect., vol. 26, pp. 135–142, Apr. 2018, doi: 10.1016/j.tmp.2017.10.002.

[23] Y. M. Arif, D. D. Putra, D. Wardani, S. M. S. Nugroho, and M. Hariadi, "Decentralized recommender system for ambient intelligence of tourism destinations serious game using known and unknown rating approach," Heliyon, vol. 9, no. 3, p. e14267, Mar. 2023, doi: 10.1016/j.heliyon.2023.e14267.

[24] Z. Abbasi-Moud, S. Hosseinabadi, M. Kelarestaghi, and F. Eshghi, "CAFOB: Context-aware fuzzy-ontology-based tourism recommendation system," Expert Syst. Appl., vol. 199, p. 116877, Aug. 2022, doi: 10.1016/j.eswa.2022.116877.

[25] M. Chu, Y. Chen, L. Yang, and J. Wang, "Language interpretation in travel guidance platform: Text mining and sentiment analysis of TripAdvisor reviews," Front. Psychol., vol. 13, Oct. 2022, doi: 10.3389/fpsyg.2022.1029945.

[26] N. A. K. M. Haris, S. Mutalib, A. M. A. Malik, S. Abdul-Rahman, and S. N. K. Kamarudin, "Sentiment classification from reviews for tourism analytics," Int. J. Adv. Intell. Inform., vol. 9, no. 1, Art. no. 1, Mar. 2023, doi: 10.26555/ijain.v9i1.1077.

[27] S. Huang, X. Wu, X. Wu, and K. Wang, "Sentiment analysis algorithm using contrastive learning and adversarial training for POI recommendation," Soc. Netw. Anal. Min., vol. 13, no. 1, p. 75, Apr. 2023, doi: 10.1007/s13278-023-01076-x.

[28] Z. Yuan, "Big Data Recommendation Research Based on Travel Consumer Sentiment Analysis," Front. Psychol., vol. 13, Feb. 2022, doi: 10.3389/fpsyg.2022.857292.

[29] M. Balfaqih, "A Hybrid Movies Recommendation System Based on Demographics and Facial Expression Analysis using Machine Learning," Int. J. Adv. Comput. Sci. Appl., vol. 14, no. 11, 2023, doi: 10.14569/IJACSA.2023.0141177.

[30] R. C K and S. K C, "Weighted hybrid model for improving predictive performance of recommendation systems using ensemble learning," Indian J. Comput. Sci. Eng., vol. 13, no. 2, pp. 513–524, Apr. 2022, doi: 10.21817/indjcse/2022/v13i2/221302133.

[31] S. Shah, Y. Raisinghani, and N. Gandhi, "Weighted Hybrid Recommendation System Using Singular Value Decomposition and Cosine Similarity," in Soft Computing and its Engineering Applications, K. K. Patel, G. Doctor, A. Patel, and P. Lingras, Eds., in Communications in Computer and Information Science. Cham: Springer International Publishing, 2022, pp. 367–381. doi: 10.1007/978-3-031-05767-0_29.

[32] M. Elahi, D. Khosh Kholgh, M. S. Kiarostami, M. Oussalah, and S. Saghari, "Hybrid recommendation by incorporating the sentiment of product reviews," Inf. Sci., vol. 625, pp. 738–756, May 2023, doi: 10.1016/j.ins.2023.01.051.

[33] L. Chen, J. Cao, W. Liang, J. Wu, and Q. Ye, "Keywords-enhanced Deep Reinforcement Learning Model for Travel Recommendation," ACM Trans. Web, vol. 17, no. 1, p. 5:1-5:21, Dec. 2022, doi: 10.1145/3570959.

[34] M. Kolahkaj, A. Harounabadi, A. Nikravanshalmani, and R. Chinipardaz, "A hybrid context-aware approach for e-tourism package recommendation based on asymmetric similarity measurement and sequential pattern mining," Electron. Commer. Res. Appl., vol. 42, p. 100978, Jul. 2020, doi: 10.1016/j.elerap.2020.100978.

[35] L. Chen, J. Cao, Y. Wang, W. Liang, and G. Zhu, "Multi-view Graph Attention Network for Travel Recommendation," Expert Syst. Appl., vol. 191, p. 116234, Apr. 2022, doi: 10.1016/j.eswa.2021.116234.

[36] "Graph API - Documentation," Meta for Developers. Accessed: Nov. 12, 2023. [Online]. Available: https://developers.facebook.com/docs/graph-api/.

[37] M. Burkhardt, A. Helmond, T. Seitz, and F. van der Vlist, "The evolution of Facebook's GRAPH API," AoIR Sel. Pap. Internet Res., Oct. 2020, doi: 10.5210/spir.v2020i0.11185.

[38] W. Phatthiyaphaibun et al., "PyThaiNLP: Thai Natural Language Processing in Python," in Proceedings of the 3rd Workshop for Natural Language Processing Open Source Software (NLP-OSS 2023), L. Tan, D. Milajevs, G. Chauhan, J. Gwinnup, and E. Rippeth, Eds., Singapore: Association for Computational Linguistics, Dec. 2023, pp. 25–36. doi: 10.18653/v1/2023.nlposs-1.4.

[39] Y. Qian, Y. Zhang, X. Ma, H. Yu, and L. Peng, "EARS: Emotion-aware recommender system based on hybrid information fusion," Inf. Fusion, vol. 46, pp. 141–146, Mar. 2019, doi: 10.1016/j.inffus.2018.06.004.

[40] T. O. Hodson, "Root-mean-square error (RMSE) or mean absolute error (MAE): when to use them or not," Geosci. Model Dev., vol. 15, no. 14, pp. 5481–5487, Jul. 2022, doi: 10.5194/gmd-15-5481-2022.

[41] F. Pedregosa et al., "Scikit-learn: Machine Learning in Python," Mach. Learn. PYTHON, vol. 12, pp. 2825–2830, 2011.

[42] A. Pramarta and A. Baizal, "Hybrid recommender system using singular value decomposition and support vector machine in bali tourism," JIPI J. Ilm. Penelit. Dan Pembelajaran Inform., vol. 7, no. 2, Art. no. 2, May 2022, doi: 10.29100/jipi.v7i2.2770.

Automatic Identification and Evaluation of Rural Landscape Features Based on U-net

Ling Sun[1], Jun Liu[2*], Yi Qu[3], Jiashun Jiang[4], Bin Huang[5]
School of Architecture and Engineering, Zhanjiang University of Science and Technology, Zhanjiang 524088, China[1, 3]
Department of Information Science, Zhanjiang Preschool Education College. Zhanjiang 524084, China[2]
Guangdong Urban and Rural Planning and Design Institute Technology Group Co., Ltd, Guangzhou 510220, China[4]
China Mobile Communications Group Guangdong Co., Ltd. Zhanjiang Branch, Zhanjiang 524057, China[5]

Abstract*—The study delves into the landscape feature identification method and its application in Xijingyu Village, investigating landscape composition elements. Analyzing rural landscape structure holistically aids in dividing landscape characteristic zoning maps, essential for guiding rural landscape and territorial spatial planning. By utilizing GIS software for superposition analysis based on topography, geology, vegetation cover, and land use, the village range of west well valley undergoes further refinement. To address the inefficiencies of common foreground extraction algorithms relying heavily on rural landscape images, a novel approach is introduced. This new algorithm focuses on directly extracting foreground areas from rural landscape interference images by leveraging stripe sinusoidal characteristics. An adaptive gray scale mask is established to capture the sinusoidal changes in interference stripes, facilitating the direct extraction of foreground areas through a calculated blend of masks. In evaluating the results, the newly proposed algorithm demonstrates significant improvements in operation efficiency while maintaining accuracy. Specific enhancements include classifying pixel gray values into intervals and recalibrating them to enhance analysis metrics. Compared to traditional methods, the algorithm showcases advantageous enhancements across various parameters, such as PRI, GCE, and VOI. Moreover, to address challenges in unwrapping low-quality rural landscape phase areas, a ResU-net convolutional neural network is employed for phase unwrapping. By constructing image datasets of interference stripe wrapping and unwrapping alongside noise simulations for model training, the network structure's feasibility is verified. The study's innovative methodologies aim to optimize rural landscape analysis and planning processes by enhancing accuracy and efficiency in landscape feature identification, foreground area extraction, and phase unwrapping of rural landscapes. These advancements offer substantial improvements in quality and precision for territorial spatial planning and rural landscape management practices.

Keywords*—Rural landscape; foreground area extraction; deep learning; phase unwrapping; ResU-net

I. Introduction

In recent years, it has been faced with many ecological problems in its development. First, a large number of spontaneous rural construction and rural reconstruction under the lack of overall planning, many of the arable land has been converted into construction land, Forest land is changed to cultivated land [1, 2]. China's original harmonious rural ecological environment has been destroyed; Second, excessive exploitation and deforestation, Causing large-scale soil erosion and desertification phenomenon, Such as rural farmland sand and stone accumulation, The decrease in soil fertility, water loss and soil erosion [3, 4]. This also further affects the development of the rural economy and aggravates the deterioration of the ecological environment, Reduce the villagers' production income, Reducing the quality of rural life; Third, the discharge of harmful waste and waste liquid into the nature in the production process, Seriously polluting the environment, Industrial production for the discharge of solid waste containing sulfur dioxide and fuel dust, Disrupted the ecological balance of the countryside, These algorithms use variability between different image features to set thresholds and detect different image features by classification to identify foreground and background regions [5, 6]. The image extraction algorithm based on gray threshold is mainly used for images with target and background occupying different ranges of gray level. Among them, the maximum inter-class variance method proposed by Otsu is a common threshold calculation method. In addition, the best entropy threshold method proposed by Kaptur et al. The main work of image extraction by the threshold method is to select the appropriate threshold value [7]. The cluster-based extraction method enables the extraction of various pixels by classifying pixels with the same or similar features, which can be divided into classification clusters and block clustering methods [8]. The pixel-based extraction algorithm has the advantages of high efficiency, stable algorithm and simple operation, but the extraction effect largely depends on the selection of threshold and the prior knowledge of the operator. In addition, the pixel-based algorithm only analyzes a single pixel attribute, and does not consider the internal characteristics of the image frequency domain [9]. The algorithm is relatively shallow, which makes the image extraction result very sensitive to noise.

In the image foreground extraction algorithm based on the model features, the deformation model has two categories: geometric deformation model and parameter deformation model. The deformation model is robust the definition of the energy function, and the setting of the termination conditions [10]. Therefore, the selection and setting of these factors need to be carefully considered carefully when using the algorithm to improve the stability of the algorithm and the accuracy of segmentation results. With the development of algorithms, model-based extraction methods change to feature extraction methods, which include unsupervised feature extraction algorithms, such as principal component analysis [11, 12]. By

finding a set of orthogonal transformations, and minimizes the sample reconstruction errors generated during the process. Also include supervised learning algorithms, such as linear discriminant analysis, typically used to divide data into two or more categories to find a low-dimensional linear space to maximize variability between categories and minimize variability within the same category [13, 14]. However, these algorithms prefer to target the images with obvious features, while the rural landscape interference stripe belongs to a kind of image without regular change, so it is difficult to effectively extract the foreground area through such methods. The extraction method based on the image edge feature is to find the boundary of the target area in the image according to the assumption of the boundary, and then segment the image along the boundary [15, 16]. Most of the algorithm to image gray gradient change direction and in x direction and y direction partial derivative trend data analysis, through the change rule set threshold of the image pixels, finally to different area pixel properties and feature recognition to complete the effective area edge segmentation, the algorithm field appeared many famous operators, based on edge extraction methods such as Canny operator, Sobel operator and Laplacian operator in extracting the overall target area in the image. But in the study of the interference stripe image stripe light and dark distribution, there are very obvious connected domains between the stripes, so through this kind of image edge algorithm will foreground area into many strips area, cannot realize the correlation between the connected domain, it is difficult to accurately obtain the edge of the foreground area. Therefore, such methods are not applicable in the foreground area of rural landscape interference images in this study [17].

The region growth-based image extraction method divides the image into multiple small regions and gradually merges pixels with similarity into a single region through a series of iterative processes. Each iteration will form a new area until the entire image is completely covered [18]. On the regional growth rules and order, and on setting the regional growth termination conditions. To solve the problem of initial seed point selection, developed an algorithm without seed point can realize automatic image extraction, fuzzy theory and optimization algorithm applied to regional growth algorithm, regional growth method combined with anisotropic filtering technology and algorithm, and add adaptive parameters in the regional growth algorithm, realize the automatic extraction of medical images [19, 20].

II. Data Processing Methods for Laser Interferometry in Rural Landscape

A. Typical Method for Identifying the Foreground Regions of the Interferograms

Large-angle oblique incidence of irradiation. Two identical optical wedges are placed symmetrically on both sides. The front wedge deflects the light at a small Angle, and the light incident to the measured at a large Angle. As shown in Eq. (1) and Eq. (2), I_m is the measured light wave intensity, I_r is the reference light wave intensity, the reflected light is deflected by the small angle of the back light wedge, and the main direction of the deflection light is the same direction as the light before the front light wedge.

$$I_i(i,j) = I_m(x,y) + I_r(x,y) + 2\sqrt{I_m(x,y) I_r(x,y)} \cos\left[\Delta\phi(x,y) + \delta_i\right] \quad (1)$$

$$I_i(x,y) = a_0 + a_1 \cos\delta_i + a_2 \sin\delta_i \quad (2)$$

Using the above large Angle oblique incidence scheme, and for the convenience of experimental operation, such as Eq. (3), Eq. (4), I is the sequence number for introducing phase modulation, E is the grayscale distribution value, the interferometer designed the interferometric measurement system for the model. The light in the measured light path and phase shift in the reference light path.

$$\Delta\phi(x,y) = -\arctan\frac{a_2}{a_1} \quad (3)$$

$$E = \sum_{i=1}^{N} [a_0 + a_1 \cos\delta_i + a_2 \sin\delta_i - I_i(x,y)]^2 \quad (4)$$

The point light emitted by the laser becomes a flat wave beam with uniform light intensity distribution. The beam meets the polarization spectroscopic prism, and the s-polarized path, and the p-polarized light transmitted perpendicular to the long axis of the incident surface enters the reference light path. As shown in Eq. (5), and Eq. (6), P_{RI} is the refractive fiber value, $I()$ is the measured optical path function, in the measured optical path, s polarized light uses the double optical wedge combination to complete the measured large angle oblique incidence, and then enters the imaging co-optical path through the semi-reverse and semi-lens.

$$P_{\mathrm{RI}}(S,\{G_N\}) = \frac{1}{C_n^2}\sum_{i<j}[c_{ij}p_{ij} + (1-c_{ij})(1-p_{ij})] \quad (5)$$

$$I(S,S_{\text{test}}) = \sum_{k=1}^{N}\sum_{k'=1}^{N'} P(k,k')\log[P(k,k')/P(k)/P'(k')] \quad (6)$$

The semi-reverse lens to make it into the imaging common path. As shown in Eq. (7), Eq. (8), $VI()$ is the imaging common path function, $L_{RE}()$ is the interference optical path function, the measurement light and the reference light meet here and produce interference on the photosensitive surface of the CCD camera after the imaging lens. The computer controls the camera to collect and control the PZT in the reference light path to realize phase shift.

$$VI(S,S_{\text{test}}) = H(S) + H(S_{\text{test}}) - 2I(S,S_{\text{test}}) \quad (7)$$

$$L_{\mathrm{RE}}(S,S_{\text{test}},X_i) = \frac{\left|P(S,X_i)/P(S',X_i)\right|}{\left|P(S,X_i)\right|} \quad (8)$$

Relying on rural landscape non-interferometric image mask indirectly extract rural landscape interference image foreground information, cannot achieve directly for rural landscape interference image extraction method and change of measurement conditions will lead to great difference in threshold value, even if the same group phase interference stripe threshold extraction is not interlinked, such as Eq. (9), Eq. (10), $G_{CE}()$ is the threshold calculation function, and m is the modulation ratio calculation formula, and the increase of

the measurement link will not guarantee the consistency of the extraction results, at the same time, the measurement light path has certain limitations, can only be used for non-common light measurement system.

$$G_{\mathrm{CE}}(S,S^{'})=\frac{1}{n}min\left\{\sum L_{\mathrm{RE}}(S,S^{'},X_i),\sum L_{\mathrm{RE}}(S^{'},S,X_i)\right\} \tag{9}$$

$$m=\frac{m_1+m_2}{2}+m_1 \tag{10}$$

B. Rapid Identification Method for the Foreground Area of the Rural Landscape Interference Image

After the helium-neon laser, there is the polarization direction and the optical axis of the polarization spectroscopic prism, as shown in Eq. (11) and Eq. (12), *n* is the angle optical axis selection ratio, and *E()* is the grayscale calculation function, thus changing the light intensity ratio between the p light entering the measured light path and the s light entering the reference light path.

$$n=\frac{n_{min}+n_{max}}{2}+n_{min} \tag{11}$$

$$E(X)=\frac{Gray_L+Gray_R}{2} \tag{12}$$

Even if the light intensity loss in the measured light path is different from the reference light in the reference light path, the measured light intensity can still be adjusted by the rotating half-wave sheet to ensure that the resulting interference stripes have the best contrast. As shown in Eq. (13) and Eq. (14), *m* and *n* is the size of the interference image, *x* and *y* are the coordinates of the interference point, the interference fringe image contains the phase difference caused by the surface topography of the tested rural landscape and is an important way of presenting the measurement information. Through a series of calculations, we can recover the physical properties contained in the measured object, using the characteristics of the light and shade changes in the interference stripe image.

$$\sigma^2=\frac{\left(\varphi_i(x,y)-E(X)\right)^2+\sum_{i=1}^{m\times n-1}\left(\varphi_i(x,y)-E(X)\right)^2}{m\times n} \tag{13}$$

$$\phi(x,y)=\varphi(x,y)+2n\pi \tag{14}$$

The interference phenomenon of light is an important feature of light volatility. It is a phenomenon that meets the interference conditions overlap each other when they meet in space, which always strengthen or weaken in some areas, forming a stable strong and weak distribution. As shown in Eq. (15) and Eq. (16), *J* is the momentum of the light wave, *U* and *V* are the directions of the light wave vector, respectively, only the vibration direction of the coherent light source is the same, the frequency of the two light waves is the same and the phase difference is constant.

$$J=\varepsilon^p=\sum_{i=0}^{M-2}\sum_{j=0}^{N-1}|\phi_{i+1,j}-\phi_{i,j}-\Delta_{i,j}^{x}|^p+\sum_{i=0}^{M-1}\sum_{j=0}^{N-2}|\phi_{i,j+1}-\phi_{i,j}-\Delta_{i,j}^{y}|^p \tag{15}$$

$$\left(\phi_{i,j}-\phi_{i-1,j}-\Delta_{i-1,j}^{x}\right)U(i-1,j)-\left(\phi_{i,j}-\phi_{i,j-1}-\Delta_{i,j-1}^{y}\right)V(i,j-1)=0 \tag{16}$$

The overall process of the foreground area recognition and extraction algorithm is to obtain the processed mask the measured image, as shown in Eq. (17), Eq. (18), *M* and *N* represents the number of refractions, *U(i,j)* represents the mask image threshold, and then the smooth mask image is obtained by smoothly adjusting the image-by-image morphology processing or linear fitting. Finally, the mask image treats the measured image to obtain the image containing only the effective information in the foreground area.

$$J=\varepsilon^2=\sum_{i=0}^{M-2}\sum_{i=0}^{N-1}\left|\phi_{i+1,j}-\phi_{i,j}-\Delta_{i,j}^{x}\right|^0+\sum_{i=0}^{M-1}\sum_{i=0}^{N-2}\left|\phi_{i,j+1}-\phi_{i,j}-\Delta_{i,j}^{y}\right|^0 \tag{17}$$

$$U(i,j)=\left|\phi_{i+1,j}-\phi_{i,j}-\Delta_{i,j}^{x}\right|^{-2} \tag{18}$$

At present, there are many prospect area recognition algorithms in the field of image algorithm, but in the highly targeted research field, the processing results and performance of the algorithm are not satisfactory. Due to the large number of connected areas in the foreground area of the rural landscape interference image and the irregular noise shadow of the gradient in the background area. As shown in Eq. (19) and Eq. (20), *V(i,j)* is the edge algorithm value, *i* and *j* are the cyclic factors, and the results of Canny algorithm cannot connect a large number of connected domains in the rural landscape interference image.

$$V(i,j)=\left|\phi_{i,j+1}-\phi_{i,j}-\Delta_{i,j}^{y}\right|^{-2} \tag{19}$$

$$\rho_{i,j}=\left(\Delta_{i,j}^{x}-\Delta_{i-1,j}^{x}\right)+\left(\Delta_{i,j}^{y}-\Delta_{i,j-1}^{y}\right) \tag{20}$$

III. Design Algorithm for Automatic Recognition Of Fore Landscape Image of Landscape Image

A. Adaptive Threshold Foreground Region Extraction Algorithm Based on Object Image Grayscale

In a set of phase-shift interference maps, the relative modulation regime of pixels refers to the ratio of the amplitude of the AC component to the amplitude of the DC component. When high-precision data or data processing tasks are required, the image quality can be evaluated directly through quantitative analysis, and the relative modulation system can be used to identify the foreground region and the background region in the distinguishing interferogram [21, 22]. The basic basis of the judgment method is: the proportional operation of the AC component amplitude and DC component amplitude can be realized by adjusting the system algorithm. The gray value information of each pixel can be mapped within the range of [0, 1] by the two amplitude proportions [23, 24]. The larger the amplitude of the AC component, the more effective information contained in the pixel, and the higher the regulation value obtained by this algorithm. Therefore, the foreground area can be distinguished from the background area, which can be distinguished from the background area by threshold setting [25, 26]. Fig. 1 is the detailed diagram of the U-net model architecture, and the relative modulation system value can be filtered. Because the path tracking method can be separated from the phase discontinuous integral, the local phase deviation can be prevented from spanning the whole integral

region. The operation speed is fast, and more accurate results can be obtained in the region under the influence of low noise [27, 28]. Therefore, the foreground area of rural landscape interference image can be extracted by evaluating the phase gradient continuity of rural landscape interference image. In the interference field, the interference intensity changes sinusoidal with space, forming an interference stripe between light and dark.

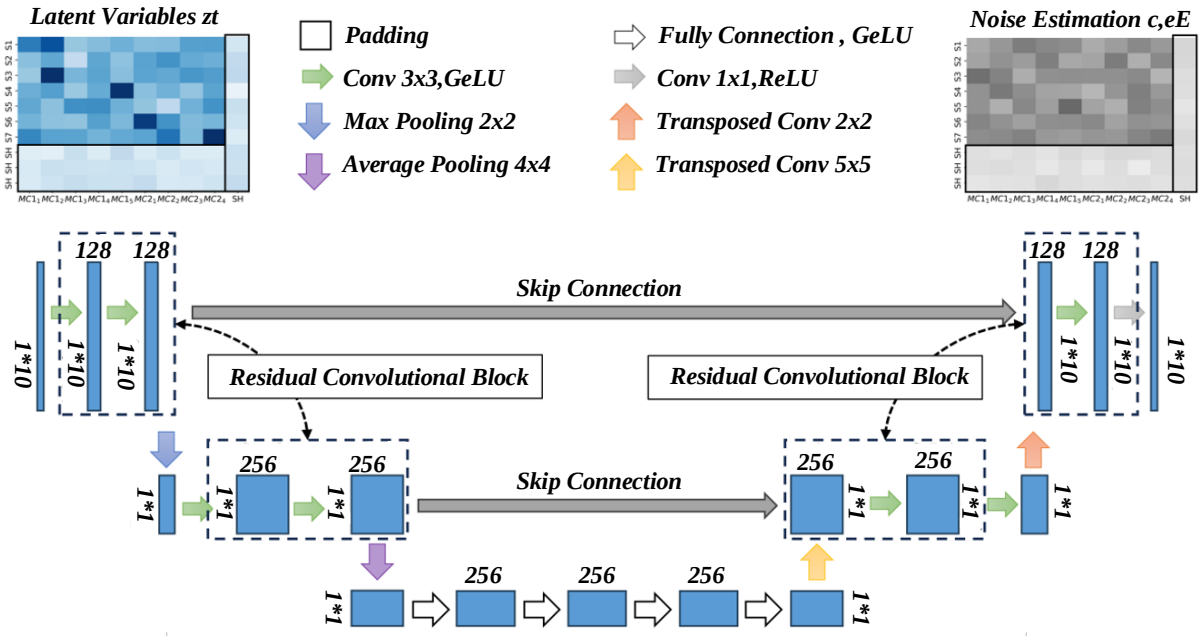


Fig. 1. Detailed diagram of the U-net model architecture.

Therefore, if a single interference stripe image is filtered by gray-scale thresholding, the dark stripes in the foreground region will also be filtered, although the background region can be removed. According to the optical wave shift phase interference formula, the interference intensity of the fixed position in the interference field will change according to the sinusoidal law as the phase shift goes on. This shows that the brightness of any pixel in the interference region changes alternately in a set of phase-shifting interference images. This means that, in a set of phase-shifting interference images taken, the pixel has at least one chance to be captured in its higher brightness state. Considering the above two aspects, a method based on the common filtering of similar interferograms is proposed to distinguish between regions and non-regions in the measured interferograms. Fig. 2 is the graph of feature extraction and fusion strategy. In the common filtering method of the same group of interference graphs, because it is necessary to control the two variables to extract the foreground region simultaneously, so the algorithm extraction results with thresholds of 30,50 and 70 when the flicker times are 1, 2 and 3 are selected. When the flicker number is more than 1 and the threshold is set to 50, the extraction results of the algorithm are still relatively normal, but the right part of the gray value and the background region are not well extracted, the extraction effect of the right part is getting worse and worse, so the algorithm also needs to adjust the parameters for the target image for a long time.

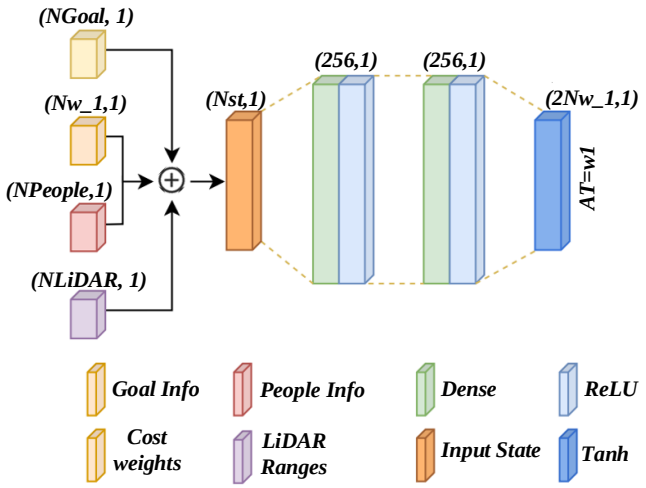


Fig. 2. Feature extraction and fusion strategy plots.

B. Direct Extraction Algorithm for the Foreground Area of Rural Landscape Interference Image Based on Stripe Sinusoidal Characteristics

Unlike the idea of guiding the unwrapping path through the branch cut line, the mass graph guidance method. The algorithm utilizes a tool called a "mass graph" to guide the choice of the mass graph guidance method is the need to define the neighboring solutions. The mass map is a two-dimensional image, where each pixel represents a combination of the search direction and the step size, while the color and direction can be selectively adjusted to achieve a more efficient search process. The idea of path tracking of mass graph guidance method is

similar to that of diffuse water filling method, Starting point selection first performed by phase quality evaluation parameters, The starting point is the point with the highest pixel quality in the parcel phase map, then establish the neighborhood window and its growth ring, Where the neighborhood window can be a cross or a field shape, Common neighborhood window size is 33, Unpackage the adjacent pixels in each direction, Fig. 3 shows the evaluation diagram of the training set and test set, And the phase mass of each point in the neighborhood window is stored in the growth ring from high to low, The highest mass point in the adjacent to each pixel. Cycle this operation until the image completes phase unwrapping. Application of this method need to pay attention to growth ring memory, if the package image size is larger, in the process of cycle package growth ring internal pixel number will increase rapidly, this will lead to the algorithm of package efficiency decreased significantly, the need to pass the algorithm of growth ring processing, on the one hand, to ensure that the pixels solution package order does not change, on the other hand to control the size of the growth ring, said this process for growth ring dressing program.

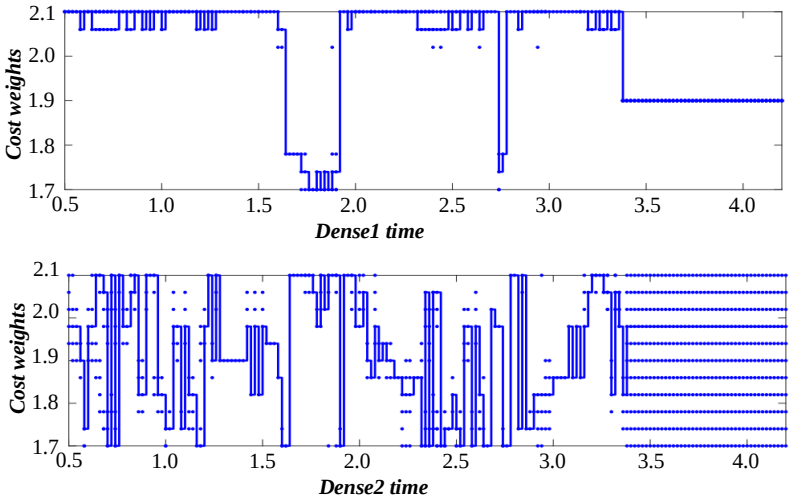


Fig. 3. Evaluation plot of the training set and the test set division.

In order to control the size of the growth ring, the value size needs to be pre-set. For MN size images, the growth ring size can be controlled by the size of MN. Next, the growth ring dressing procedure will work with the number of the pixels in the growth ring. When the number of points exceeds the preset value, the point with the mass value in the growth ring will be removed, and the lowest quality value in the remaining points will be set as the new mass threshold of the growth ring. The removed points were marked with special signs indicating that they were delayed points and would back into the growth ring in the subsequent process. In the next step, only those points with mass values above the threshold are deposited into the growth ring, while those below the threshold are labeled as deferred points. Points in the growth ring are constantly extracted for unwrapping until the growth ring is empty, when the growth ring threshold needs to be lowered (take 0) and the delayed points are put back into the growth ring for unwrapping. Table I shows the GSA pixel matching accuracy evaluation, and the above process is repeated until all points are unwrapped. There are three main parameters reflecting the phase quality: pseudo-coherence coefficient (PSD), phase derivative deviation (PDV) and maximum phase gradient (MPG). At present, the most widely used and most effective is the phase derivative deviation. The main influencing factor of the quality guide method for solving the package result is the image quality of the package image (including the accuracy of the foreground extraction technology and the stability of the shooting equipment), and the image quality will directly affect the accuracy of the final quality map guide. With good image quality, the phase solution of the mass graph guidance method will be better than the other traditional path tracking algorithms. In the mass graph guidance method, there is also a technique called "taboo search", by recording and avoiding the previously searched solutions.

TABLE I. EVALUATION OF PIXEL MATCHING ACCURACY IN GSA

Appraise	Consult	The GSA algorithm results	Results of the method in this paper
Picture	Bear fruit	Pixel number	Accuracy
A group	408000	387681	73.67%
B group	408000	384975	72.98%
C group	745608	724625	75.81%

The mask cutting method, also known as the mask-based cutting algorithm, is a segmentation method based on image edges. The core idea is to divide the image into different regions according to the edge information in the image. The method can be seen as a combination of the pruning method and the mass graph guidance algorithm. It requires searching for residual debris and placing the branch cut line, but the difference from the branch cut method is that the branch cut line placement is guided by the mass map. The brief step is as follows: Select a mask to traverse the entire image, usually select a square or circular mask; move the mask along the edge of the image and calculate the difference of pixel value in the mask; divide the image according to the obtained boundary position, such as algorithms based on the threshold value and

area growth. The process of creating the mask cut line is similar to the pixel diffusion in the higher mass area, but the difference is that the process starts with a residual miss without equilibrium and then gradually spreads into the surrounding low mass area. Fig. 4 evaluates the graph as the loss function changes with the number of iterations. This process continues until the connected residue almost reaches equilibrium or reaches the image boundary. In fact, the process of generating the mask cutting line is a process of regional growth, so the generated mask cut line is rough and needs further refinement. Whether the point on the mask cut line needs to be removed depends on two points: whether the point is close to the handicap; whether the point removal affects the connectivity of the mask cut line. The detection of near disability is simple, while the detection of connectivity needs to be conducted in the 33-neighborhood centered on this point. If the coordinates of this point are (i, j), the point can be removed from the mask cut line if any of the following cases exist.

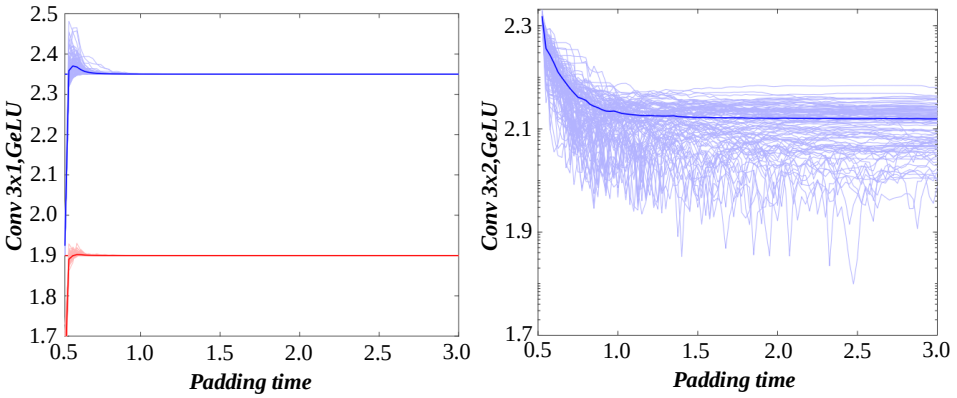


Fig. 4. The loss function changes with the number of iterations.

IV. Automatic Identification and Evaluation of Rural Landscape Features Based on U-net

Phase solution package is optical interferometry structure light projection measurement is a common problem in the field of contactless measurement, phase solution package algorithm in the 80s to 90s in the rapid development stage, this stage appeared a lot of for all kinds of phase solution algorithm, for the subsequent further development provides an effective theoretical basis, phase solution package algorithm basically can be attributed to algorithm reduces the unwrapping problem to the problem of optimal path selection, which uses the correlation of adjacent pixels in space. The classical algorithms based on path tracking mainly include: branch cutting method, mass graph guidance algorithm, mask cutting method and minimum discontinuity algorithm. Fig. 5 is the confusion matrix evaluation graph, which connects the detected positive and negative residues into branch tangents and then bypasses them to unwrapping. The difficulty lies in the setting of the branch, doing a lot of work on the connection strategy of the branch, and proposing many improvement methods, and the effect is obvious. The quality map guide algorithm does not identify the residual or set the branch, but guides the solution path through the phase mass map. The mask cutting method combines the first two methods to guide the setting of branches by mass diagram, without rigorous process and detailed algorithm, which is the first feasible algorithm: minimum discontinuity algorithm. This method draws on the relevant theories of computer graphics, clever conception, solution accuracy and algorithm stability; moreover, proposes the regional growth algorithm, using the phase information of the surrounding pixels to predict the solution results and make consistency test, thus selecting the optimal solution path, achieving success in processing complex SAR phase data.

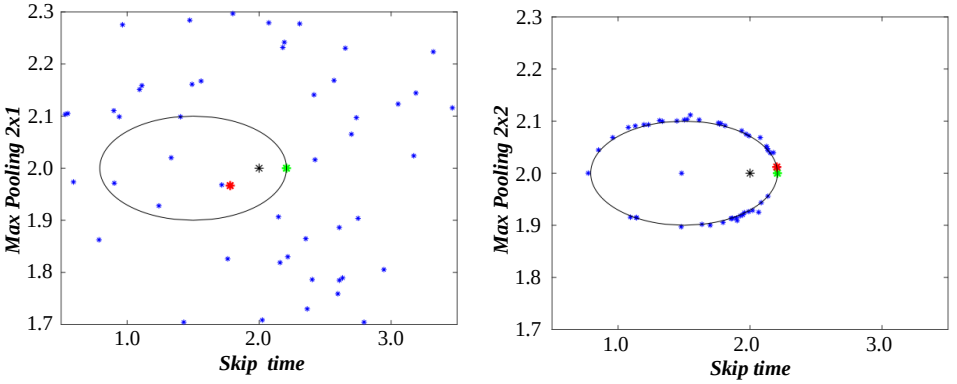


Fig. 5. Confusion matrix evaluation.

The problem of finding the minimum by calculating the minimum of the p power of the difference between the synthetic phase gradient and the wrapping phase gradient. Initially, this method was limited to the least squares method,

divided into weighted least squares and no right least squares method. The concept of minimum norm is proposed, the least squares method is incorporated into this concept, and the significance of different norm values is clarified from the mathematical perspective, and the specific algorithm steps are given, which greatly enriches the understanding of parcel theory. After entering the new century, phase disassembly has entered the stage of diversified development. On the one hand, AI algorithms such as deep learning, genetic algorithms and ant colony algorithms were introduced into the branch method. Table II shows the RMA pixel matching accuracy evaluation to optimize the placement of branches. Compared with previous methods, the artificial intelligence algorithm can better place the branch line, but the characteristics of the branch method are difficult to work in the residual close density area. On the other hand, the network planning algorithm is introduced into the minimum norm method to try to solve this optimization problem through this optimization algorithm. The network planning method has better stability in the accuracy of these algorithms is still not high, and the efficiency is also general. In addition, filtering the parcel phase diagram is also a new development direction. Before unsolving the parcel, the parcel phase diagram is filtered first. While fully retaining the useful information, the noise is filtered out as much as possible to obtain the nearly perfect parcel phase diagram, so as to reduce the difficulty of unwrapping. The effect of this method is very significant on the images with severe noise interference. Different application fields have different requirements for unwrapping results. General algorithms have not appeared in the field of rural landscape interference image phase unwrapping. However, with the gradual iterative progress of neural network structure.

Traditional interferometry obtains useful information by directly analyzing interference stripes, but due by noise interference, stripe density and manual operation error. Phase shift technology is applied to optical interferometry. This new technique analyzes the phase difference to obtain the information of the measured surface. The measurement accuracy is higher. In phase shift interferometry, the inverse triangle function is used to obtain the phase. Due to the nature of the inverse triangle function, the recovered phase information is limited to a fixed interval, so the phase data discontinuity caused by the restriction problem should be eliminated to obtain the real data. Therefore, it is necessary to find the discontinuous cutoff point in the parcel image through the algorithm and restore it to continuous phase information. This process is called phase unwrapping. After decades of optimization, the basic phase unwrapping algorithm is specifically divided into two categories: path tracking method and minimum norm method. Fig. 6 is the ROC assessment diagram, where the path tracking method is a local algorithm. The core idea of this algorithm is to search for another target pixel by evaluating the polarity of the pixel data and determining the polarity, and establish a connection during the search process. If the two pixels have opposite polarity, stop the search; if the polarity is the same, expand the search range and build connections based on their data quality and polarity until the two pixels have the same polarity or the search range reaches the maximum. Then, repeat to override all unconnected pixels throughout the image until the target pixels are searched. The path tracking method mainly includes branch cutting method, mass map guide method, mask cutting method and minimum discontinuity method.

TABLE II. EVALUATION OF PIXEL MATCHING ACCURACY OF RMA

Appraise	Consult	The RMA algorithm results		Results of the method in this paper	
Picture	Bear fruit	Pixel number	Accuracy	Pixel number	Accuracy
A group	475000	392478	92.21%	348272	98.14%
B group	464000	382577	97.53%	347864	97.25%
C group	737808	747816	96.42%	747884	98.31%

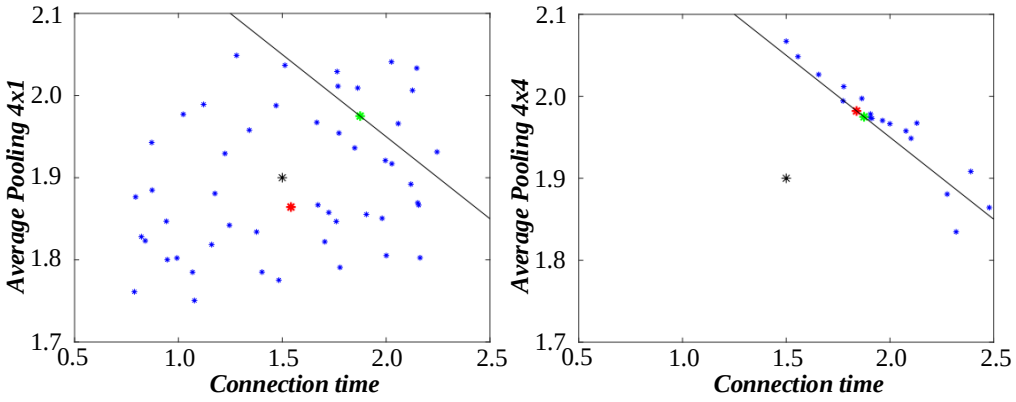


Fig. 6. ROC assessment.

Boundary function to cut down the subspace that is impossible to reach the optimal solution, so as to improve the search efficiency. First, determine the specific functions and constraints of the branch processing target, and then initialize the search queue and the current optimal branch point. The search queue contains the branch point to be expanded, and the current optimal branch point can be set to positive infinity. Secondly, for each unexpended node, the upper and lower bounds corresponding to the node are calculated and sorted according to this bound value, and then each node is expanded in the sorted order. Again, when extending a node, first check if a better solution appears, if so, update the current optimal

solution and compute the bound value of all possible extended nodes of that node and insert it into the search queue. Finally, the above steps are repeated until the wrapping image completes the branch tangent arrangement and the phase unwrapping. After sorting all the residues, the first round of search was started from the first residue in the sequence. Set a search window of size 3 around the remnant close, and traverse the pixel information at the edge of the window. The position of the edge was determined by searching for the residual handicap. First, two adjacent residues are found and connected to form an approximate branch tangent. Fig. 7 evaluates the graph of the feature importance ranking, and then, determines. If so, the branch is considered to be equilibrium, you can stop the search; if not, find a new stump in the adjacent area and connect it to the existing branch and continue the search. When the whole search window is traversed, if still did not reach balance, expand the size of the search window, and on the edge of the search, until find the line balance conditions, size every 2 until to set the maximum size value, after completing the maximum size of the search, whether to achieve balance, will end the round of search. Furthermore, if the edge of the image is reached during the search, the search round also ends.

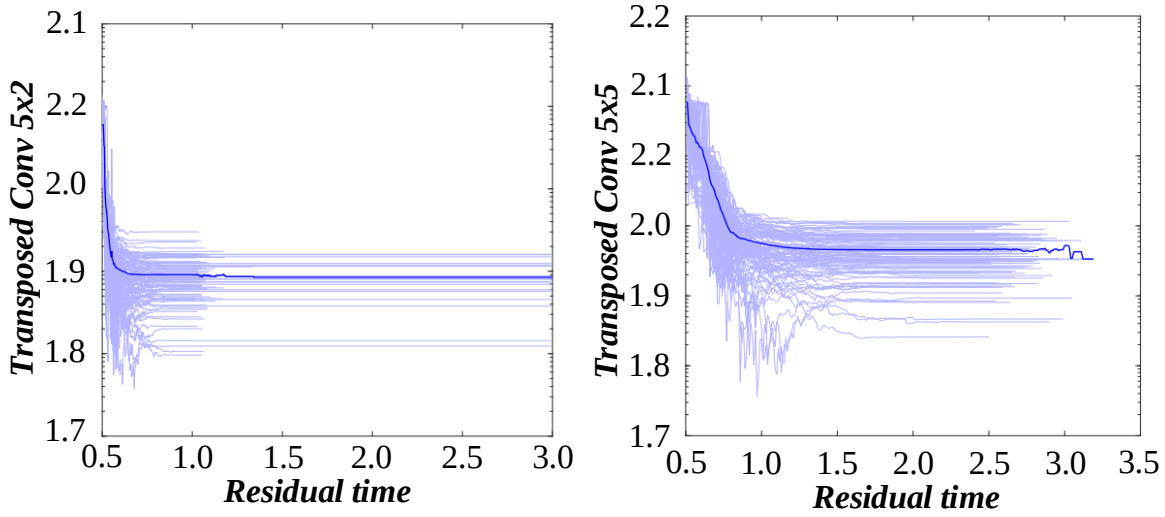


Fig. 7. Feature importance ranking evaluation.

In the phase unwrapping algorithm, starting from a point on the non-branched, unwrapping to the periphery. Before performing unwrapping, a series of judgments are required, including checking whether the point is at the edge of the image or the mask point and whether it has been unwrapped by another path. If none of the above conditions hold and the point is not branch, the solution can be performed. From the point where the package is completed as the starting point, judge again, and choose the appropriate path step method. Tra-walk through the entire image iteratively and untangle its normal parcel phase. An "island" is generated when the branch tangent forms a closed loop. These points are distributed on the branches and need to check if there is phase information that has been successfully unwrapped. If so, use this information to disentangle; otherwise, these points are "exceptions". By this way, we get the real phase distribution map, which effectively solves the error transfer problem caused by noise. The biggest advantage of branch cutting is it's very computationally efficient. In the case of less disability loss, the solution results are relatively credible. However, when the residue is dense, the algorithm will produce a large number of unreasonable branches and lines, which is easy to produce closed areas that cannot unwrap, and can easily lead to errors in the unwrapping results. In addition, this method only uses residual missing information and ignores other information, so the placement of branch tangents lacks convincing criteria.

V. Experimental Analysis

In the rural landscape interferometry, the phase is moved with a fixed step size, so the gray scale of the phase shift interference stripe shows a certain periodic change. Under ideal conditions, to simulate the phase-shift interference stripe image, it can be seen that when the phase shift is different, the difference of the pixel gray value at the same position reaches the maximum. Fig. 8 evaluates the model performance comparison, which has very obvious sinusoidal change characteristics. Simulated phase-shift interference stripe image after adding noise. It can be seen that under the influence of Gaussian noise such as noise, salt and salt noise (PSNR=13.3035) and the phase error of the interference stripe caused by the common light path, the interference stripe still has obvious sinusoidal change characteristics, which ensures that the algorithm logic can be stable in real situations.

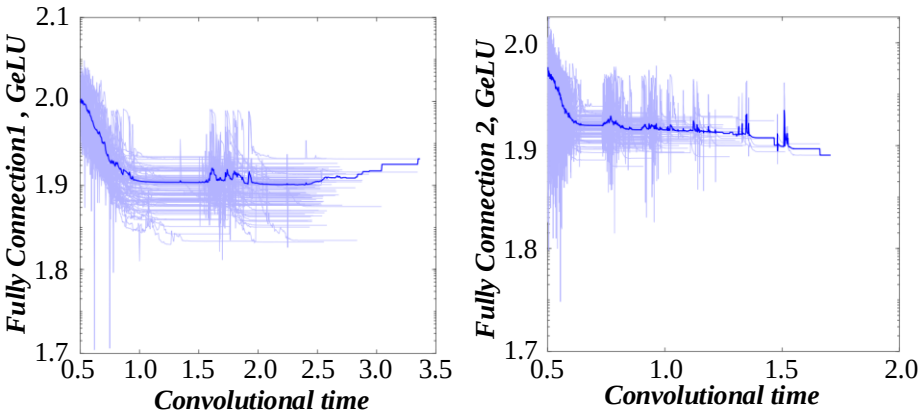


Fig. 8. Model performance comparison evaluation.

Rural landscape interference stripe image groups of rows 150 of all pixels and observe the change in gray scale. In the measured rural landscape interference image, the sinusoidal variation of the interference fringe. The gray value of false foreground pixels in the background area does not change much by phase shift. Fig. 9 shows the overfitting detection evaluation diagram, so the characteristics of the pixels at different phases can be used to eliminate the wrong foreground information, while connecting the connected domain and improving the. The treatment is divided into grayscale mask M1 and repair mask M2 stages. The target image needs to undergo preprocessing, difference operation, gray scale set allocation, threshold extraction, neighborhood local variance analysis and other steps, to realize the direct extraction of the foreground area of the final rural landscape interference image, to avoid the blur of image details.

To obtain the gray value of the processed image to extract the foreground area and take the image segmentation task. For example, the threshold needs to be set to extract the foreground area information of the rural landscape interference image. Mask cutting method is suitable for various types such as target detection, face recognition, medical image analysis, etc. Fig. 10 is the evaluation diagram of IoU index. Although the algorithm is simple and easy to understand, its dependence on edge information, the segmentation effect may be disturbed for some complex images, which needs to be optimized in combination with other image processing methods.

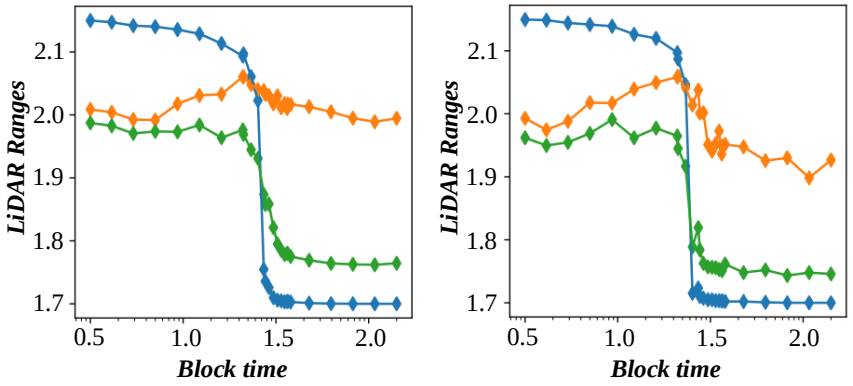


Fig. 9. Overfitting detection assessment.

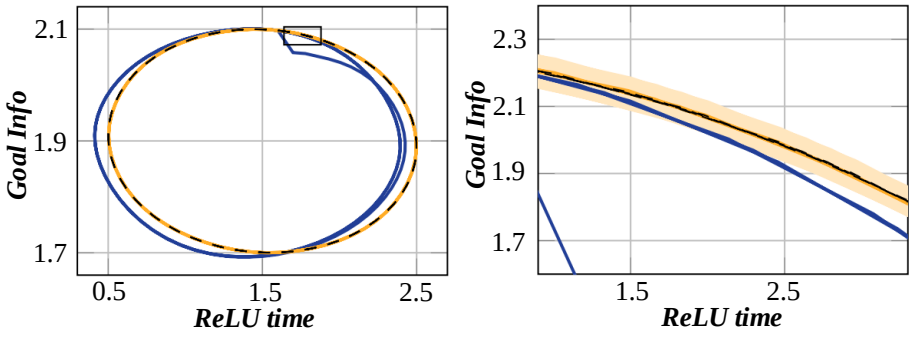


Fig. 10. IoU index evaluation.

VI. Conclusion

In the process of feature identification, we should not only pay attention to the decomposition of landscape feature elements, but also the grasp of the overall landscape structure. Using landscape perception, field research detailed records of local survey information, combined with residents' interviews, increase a lot of subjective observation, this method can more accurately identify and understand the landscape features, help people understand the landscape characteristics of different landscape area, to manage the change of landscape characteristics, help local policy for its landscape change. In the path tracking method, the phase reciprocal deviation is used between the mass map and the weight parameters. The black pixel area appearing in each image represents the absolute value of its phase difference, which belongs to the discontinuous breakpoint appearing when the algorithm phase unwrapped the rural landscape interference image. Such points will have some impact on the restoration accuracy when the final rural landscape morphology error is restored. Longlong branch lines appeared in some areas, The reason for this kind of branch line is the failure to form a good closed ring or some area noise when calculating the phase difference value of adjacent points, The branch tangent has not formed the closure of the wrong package; The mass map guidance method is the good algorithm of phase unwrapping in the path tracking method, Unsuccessful discontinuous points in the unpacking results, However, since the phase quality evaluation parameters can be adjusted appropriately, Make the solution quality of the algorithm in the rural landscape interference image is better than the branch method; The image unwrapping quality of mask cut line method and mass map guidance method is similar, Although part of the branch fails to undergo good phase unwrapping, However, you can see the integral traces of the algorithm; The unwrapping mass of the minimum weighted discontinuity method is similar to the above method, Failure to solve the discontinuity problem in the unpacking process.

In order to further analyze the prospect extraction results, the segmentation results of all the pixels in each group of images are extracted, and the reference segmentation results of the three groups, the comparison algorithm segmentation results and the proposed algorithm segmentation results. The number of pixels in groups A and B was 408000 and 745608. Image similarity reached 97.84% in group A, 96.21% in group B, and 99.33% in group C. The accuracy of GSA algorithm is improved by 2.82%, 5.18% over ICF algorithm, 1.54% over RMA algorithm; GSA algorithm by 1.86%, 4.92% over ICF algorithm, 2.76% from RMA algorithm; GSA algorithm by 2.14%, 2.77% compared with ICF algorithm, and 1.27% over RMA algorithm. By comparison, the maximum accuracy of the algorithm is 2.82% and the minimum increase is 1.86%. Compared with the ICF algorithm, the maximum accuracy increased by 5.18% and 2.77%, the maximum accuracy increased by 2.76% and the minimum improvement by 1.27%.

Acknowledgment

This work was sponsored in part by 2023 Guangdong Provincial Department of Education Special Innovation Projects for Ordinary Universities (2023KTSCX378).

References

[1] Balaha, H. M., Ali, H. A., & Badawy, M. (2021). Automatic recognition of handwritten Arabic characters: a comprehensive review. Neural Computing & Applications, 33(7), 3011-3034.

[2] Chen, Z. Q., Deng, J. H., Zhu, Q. Q., Wang, H. L., & Chen, Y. (2022). A Systematic Review of Machine-Vision-Based Leather Surface Defect Inspection. Electronics, 11(15), 28.

[3] Chisti, M. K. M., Kumar, S. S., & Prasad, G. (2023). Defects Identification, Localization, and Classification Approaches: A Review. Iete Journal of Research, 69(7), 4323-4336.

[4] Francisco, M., Ribeiro, F., Metrolho, J., & Dionisio, R. (2023). Algorithms and Models for Automatic Detection and Classification of Diseases and Pests in Agricultural Crops: A Systematic Review. Applied Sciences-Basel, 13(8), 16.

[5] Vizzari, M., & Sigura, M. (2015). Landscape sequences along the urban–rural–natural gradient: A novel geospatial approach for identification and analysis.Landscape and Urban Planning,140, 42-55.

[6] Čurović, Ž., Čurović, M., Spalević, V., Janic, M., Sestras, P., & Popović, S. G. (2019). Identification and evaluation of landscape as a precondition for planning revitalization and development of mediterranean rural settlements—Case study: Mrkovi Village, Bay of Kotor, Montenegro.Sustainability,11(7), 2039.

[7] Zakariya, K., Ibrahim, P. H., & Wahab, N. A. A. (2019). Conceptual framework of rural landscape character assessment to guide tourism development in rural areas. Journal of Construction in Developing Countries, 24(1), 85-99.

[8] Khan, M. H., Farid, M. S., & Grzegorzek, M. (2021). Vision-based approaches towards person identification using gait. Computer Science Review, 42, 49.

[9] Liu, C. L., Du, Y. C., Yue, G. H., Li, Y. S., Wu, D. F., & Li, F. (2024). Advances in automatic identification of road subsurface distress using ground penetrating radar: State of the art and future trends. Automation in Construction, 158, 21.

[10] Lubna, Mufti, N., & Shah, S. A. A. (2021). Automatic Number Plate Recognition:A Detailed Survey of Relevant Algorithms. Sensors, 21(9), 35.

[11] Mahmud, M. S., Zahid, A., Das, A. K., Muzammil, M., & Khan, M. U. (2021). A systematic literature review on deep learning applications for precision cattle farming. Computers and Electronics in Agriculture, 187, 16.

[12] Manavalan, R. (2020). Automatic identification of diseases in grains crops through computational approaches: A review. Computers and Electronics in Agriculture, 178, 24.

[13] Wilkosz-Mamcarczyk, M., Olczak, B., & Prus, B. (2020). Urban features in rural landscape: A case study of the municipality of Skawina.Sustainability,12(11), 4638.

[14] Plutino, A., Barricelli, B. R., Casiraghi, E., & Rizzi, A. (2021). Scoping review on automatic color equalization algorithm. Journal of Electronic Imaging, 30(2), 32.

[15] Pushpanathan, K., Hanafi, M., Mashohor, S., & Ilahi, W. F. F. (2021). Machine learning in medicinal plants recognition: a review. Artificial Intelligence Review, 54(1), 305-327.

[16] Sachar, S., & Kumar, A. (2021). Survey of feature extraction and classification techniques to identify plant through leaves. Expert Systems with Applications, 167, 14.

[17] Ahmed, S. U., Shuja, J., & Tahir, M. A. (2023). Leaf classification on Flavia dataset: A detailed review. Sustainable Computing-Informatics & Systems, 40, 19.

[18] Attri, I., Awasthi, L. K., & Sharma, T. P. (2024). Machine learning in agriculture: a review of crop management applications. Multimedia Tools and Applications, 83(5), 12875-12915.

[19] Elfferich, J. F., Dodou, D., & Della Santina, C. (2022). Soft Robotic Grippers for Crop Handling or Harvesting: A Review. Ieee Access, 10, 75428-75443.

[20] Elli, G., Hamed, S., Petrelli, M., Ibba, P., Ciocca, M., Lugli, P., & Petti, L. (2022). Field-Effect Transistor-Based Biosensors for Environmental and Agricultural Monitoring. Sensors, 22(11), 38.

[21] Kastelan, N., Vujovic, I., Krcum, M., & Assani, N. (2022). Switchgear Digitalization-Research Path, Status, and Future Work. Sensors, 22(20), 15.

[22] Khan, U., Khan, M. K., Latif, M. A., Naveed, M., Alam, M. M., Khan, S. A., & Su'ud, M. M. (2024). A Systematic Literature Review of Machine Learning and Deep Learning Approaches for Spectral Image Classification in Agricultural Applications Using Aerial Photography. Cmc-Computers Materials & Continua, 78(3), 2967-3000.

[23] Liang, H., Xing, L. Y., & Lin, J. H. (2020). Application and Algorithm of Ground-Penetrating Radar for Plant Root Detection: A Review. Sensors, 20(10), 18.

[24] Morchid, A., Marhoun, M., El Alami, R., & Boukili, B. (2024). Intelligent detection for sustainable agriculture: A review of IoT-based embedded systems, cloud platforms, DL, and ML for plant disease detection. Multimedia Tools and Applications, 40.

[25] Zhang, S. W., & Zhang, C. L. (2023). Modified U-Net for plant diseased leaf image segmentation. Computers and Electronics in Agriculture, 204, 10.

[26] Sachar, S., & Kumar, A. (2021). Survey of feature extraction and classification techniques to identify plant through leaves. Expert Systems with Applications, 167, 14.

[27] Thakur, P. S., Khanna, P., Sheorey, T., & Ojha, A. (2022). Trends in vision-based machine learning techniques for plant disease identification: A systematic review. Expert Systems with Applications, 208, 30.

[28] Wu, Z. N., Chen, Y. J., Zhao, B., Kang, X. B., & Ding, Y. Y. (2021). Review of Weed Detection Methods Based on Computer Vision. Sensors, 21(11), 23.

Analysis of Customer Behavior Characteristics and Optimization of Online Advertising Based on Deep Reinforcement Learning

Zhenyan Shang, Bi Ge
Chongqing College of International Business and Economics, Chongqing 401520, China

***Abstract*—With the shift from traditional media to online advertising, real-time strategies have become crucial, evolving to meet contemporary demands. Advertisers strive to succeed in online advertising evaluations by demand-side platforms to secure display opportunities. Discrepancies in information evaluation can impact click-through rates, emphasizing the need for precise prediction models in asymmetric contexts. Time dynamics significantly influence online ad click-through rates, with rest hours outperforming working hours. This study introduces the ARMA model to refine click predictions by preprocessing hits and employing a single XGBoost model. Furthermore, a reinforcement learning model is developed to explore online advertising strategies amidst information imbalances. Data is segmented into training (70%), validation (15%), and test sets (15%), with model parameters optimized using the DQN algorithm over 48 hours. Validation and testing on separate datasets comprising 15,000 entries each yield model accuracies of 0.85 and recall rates of 0.82. The incorporation of regret minimization algorithms enhances reward functions in deep reinforcement learning. Leveraging Tencent data, a comparative analysis evaluates advertisers' click rates as overrated, underrated, or accurately predicted by DSPs. Findings indicate that smart customer behavior characteristics outperform DQN, converging swiftly to optimal solutions under complete information. Smart characteristics exhibit stability and flexibility, with human-machine collaboration circumventing the drawbacks of random exploration. Transfer Learning amalgamates experimentation with real-world insights, bolstering algorithm adaptability for intelligent decision-making tools in enterprises.**

Keywords*—Real-time online advertising; ARMA-XGBoost model; information asymmetry; deep reinforcement learning decision-making behavior; Transfer Learning

I. Introduction

The publishing platforms and advertising types of online advertising showed a positive trend of development. Online advertising with the Internet as the media, provides advertisers with a high-yield and low-cost way of delivery [1, 2]. Compared with expensive display signs and paper advertisements, online advertising, through big data matching technology based on the characteristics of the audience, can deliver the most accurate revenue for advertisers. At the same time, through the identification of user intention, online advertising can provide users with more interesting advertisements, achieving a win-win situation for users and advertisers [3, 4]. At present, online advertising is generally divided into three categories: sponsored search advertising, general display advertising and real-time online advertising. In general, sponsored search ads are ads displayed in search results after users query their keywords on search engines such as browsers [5]. The general display advertisement will pop up when the user brow the website information, or when the user uses the mobile App, in the open screen animation of the software or the rotation map at the top of the home page [6]. The principle of RTB advertising is that advertisers design online advertising strategies, and through the demand side platform, online advertising on the web or the mobile App, to realize the advertisers to choose the corresponding advertising audience [7, 8].

RTB advertising has experienced explosive growth since its birth. In 2011 internationally, 88% of North American advertisers switched to RTB ads when they bought online ads. The RTB market is expected to grow to $9 billion in 2023, or 40% of the total advertising budget. In China, the RTB market first started with the TANX system launched by Taobao in 2011. By 2013, the number of RTB AD requests in China had reached five billion, and the RTB investment budget for advertisers had increased by 300% to $83 million [9]. RTB advertising is widely used in the era of mobile Internet. Compared with sponsored search advertising and general display advertising in the PC era, RTB advertising has changed the pattern of online advertising to a large extent. Every time in online advertising, you can accurately target the potential user audience, and select the most appropriate ads from the advertising library [10]. RTB iterates on data that is more relevant to user habits than focusing only on context keywords. Advertisers urgently need to make a more accurate prediction and evaluation of the display effect of advertising and design an appropriate online advertising model according to the estimated display effect. The display effect of advertisements can be evaluated by the conversion rate. The better the estimated display effect, the higher the cost of online advertising [11, 12]. Whether we can accurately predict the click rate is the key to the accuracy and effectiveness of the online advertising model. At the same time, in the actual online advertising process, multiple advertisers may participate in the online advertising auction. DSP ranks shot ads based on the effective click cost, and the highest ranked advertisers pay according to the broad second online advertising mechanism [13, 14].

In real, online advertising auctions, the price of online advertising needs to meet the cost constraints of the budget. How to accurately estimate the AD click-through rate, combined with the estimated click-through rate, participate in online advertising within the limited budget, win the DSP online advertising

evaluation has great research significance. In the background, the prediction and bidding model of real-time online advertising advertisements under asymmetric information [15]. After users see the ads they are interested in, they may click, download, register, buy and a series of behaviors, that is, the advertising effect has been transformed. Whether users respond to ads is one of the most concerned issues on the demand side. Users' download, registration and purchase behavior may cause longer delays, so scholars pay more attention to click-through prediction models, both in industrial applications and in academia [16]. In general, the click-through rate prediction model not only predicts the probability of users clicking after seeing an AD, but also describes how much the user is interested in the AD. Early AD click-through rate prediction models can be roughly divided into two categories: feature-based click-through rate prediction models and maximum-likelihood-based click-through rate prediction models. In feature-based methods, the prediction models are constructed based on page display features [17]. These features may include the text of the AD, the picture content, the location on the page, etc. Feature-based methods usually utilize logistic regression models. It describes the click and non-click behavior of ads as a dichotomy problem and is solved with a logistic regression model. Through the experiment, the model has a good prediction effect on the repeatedly displayed advertisements, and the accuracy of the prediction has increased steadily with the increase of the number of repeated displays [18, 19]. However, LR model is only effective for first-order features, and the ability to learn sparse features is poor. It relies on manual preprocessing of combined features, which should meet the shortage of complex data in practical applications.

II. Information Asymmetry in Real-Time Online Advertising

A. Information Asymmetry Analysis in Real-Time Online Advertising

Maximum-likelihood-based methods attempt to smooth the response estimates using the advertised exposure and hits, such as the Gamma-Poisson model. As shown in Eq. (1) and Eq. (2), these methods are all based only on simple linear models and cannot capture the associations between the data. In this case, a hybrid method is proposed by combining the hierarchical information of advertisements and the display information through matrix decomposition using the display features of advertisements.

$$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + \cdots + \phi_p x_{t-p} + \xi_t \tag{1}$$

$$z^p - \phi_1 z^{p-1} - \phi_2 z^{p-2} - \cdots - \phi_p = 0 \tag{2}$$

This method uses MF to learn a set of latent features from the data while correcting the prediction results with a feature-based approach. However, the MF used in this method is limited to binary relationships and does not satisfy higher-order relationships. As shown in Eq. (3) and Eq. (4), in the factorization machine model, the implicit vector of the first-order combined features is calculated to obtain the weight of the second-order combined features.

$$\mathrm{Cov}(\varepsilon_t, \varepsilon_s) = \sigma^2 \delta_{t-s} = \begin{cases} \sigma^2, & t = s, \\ 0, & t \neq s, \end{cases} \tag{3}$$

$$\hat{\rho}_k = \frac{\sum_{t=1}^{N-k} (x_t - \bar{x}_N)(x_{t+k} - \bar{x}_N)}{\Sigma_{t=1}^{N} (x_t - \bar{x}_N)^2} \tag{4}$$

The combination problem under sparse features is further solved. The concept of "field" is proposed on the FM model, classifying the features of the same properties into the same field and learning the hidden vector for each field. When learning different combinations of features, the internal product, as shown in Eq. (5) and Eq. (6), as the weight of the combined features. The experiment proves that FMM model can more accurately estimate the click rate of online ads compared with FM model.

$$LB = n(n+2) \sum_{k=1}^{m} \left(\frac{\hat{\rho}_k^2}{n-k} \right) \tag{5}$$

$$X_t = \phi_0 + \phi_1 X_{t-1} + \cdots + \phi_p X_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q} \tag{6}$$

Although FM and FMM models can solve the combined characteristics well, their theoretical essence is still the second-order combined characteristics stage, so they are not much used in the industry. As shown in Eq. (7) and Eq. (8), when the application of the click-through rate prediction algorithm in the industry occurs, the following problems are encountered: First, the user's response to advertising is a dynamic process, which will change over time. The prediction algorithm of click-through rate needs to take into account the time factor when applying it to landing.

$$\hat{\rho}_k = \frac{\sum_{t=1}^{n-k} (x_t - \bar{x})(x_{t+k} - \bar{x})}{\sum_{t=1}^{n} (x_t - \bar{x})^2}, \forall 0 \leq k \leq n \tag{7}$$

$$\hat{\phi}_{kk} = \frac{\hat{D}_k}{\hat{D}}, \forall 0 < k < n \tag{8}$$

Because mobile applications usually use cold start as a startup in order to save memory, most ads have little history in pages, or limited history. In order to solve the historical exposure rate of advertising or new advertising estimates, a hierarchical importance perception factor machine was developed, HIFM provides an effective general framework, as Eq. (9), Eq. (10), the framework combines importance weight and hierarchical learning, after experimental data validation, HIFM better than the existing FMM model in terms of time sensitivity, and the importance of HIFM perception and hierarchical learning plays a significant role in improving the cold start scenario.

$$min\ AIC = nIn\hat{\sigma}_\varepsilon^2 + 2(p+q+1) \tag{9}$$

$$x_t = \sum_{i=1}^{p} \varphi_i x_{t-p} + \mu_t + \sum_{j=1}^{q} \theta_j \mu_{t-q} \tag{10}$$

B. The Impact of Information Asymmetry on Real-Time Online Advertising

In order to solve the problem of click rate prediction in industrial applications, a method of screening and combining features with gradient promotion decision tree is proposed by generating a discrete feature vector and taking it as the input parameter of the LR model. GBDT is essentially an integrated learner, as shown in Eq. (11) and Eq. (12), the combination of multiple decision trees can describe the differentiated feature combinations more accurately; compared with the single decision tree, the combination method of GBDT and time series is often applied in the recommendation system.

$$\text{Gini}(D)=\sum_{k=1}^{K}p_k(1-p_k)=1-\sum_{k=1}^{K}p_k^2 \tag{11}$$

$$\text{Gini}(D,A)=\frac{|D_1|}{|D|}\text{Gini}(D_1)+\frac{|D_2|}{|D|}\text{Gini}(D_2) \tag{12}$$

Also in the industrial applications, the improvement model is put forward. The model adopts the idea of partition and treatment, group and slice the samples, and LR model is used to predict the partition samples, as shown in Eq. (13) and Eq. (14). Finally, weighted regression is used to combine the results of the partition. The model is proven to fit complex nonlinear functions, and the LS-PLM model using L1 and L2 regulars has good sparsity, which improves the online prediction ability.

$$\Delta D_A=\text{Gini}(D)-\text{Gini}_A(D) \tag{13}$$

$$y=\sum_{k=1}^{K}f_k(x),f_k\subset\Gamma \tag{14}$$

With the in-depth research of deep learning technology in the advertising industry and the breakthrough results, deep neural networks have been realized to fit high-order combination characteristics through nonlinear functions, as shown in Eq. (15) and Eq. (16), so DNN technology is gradually used in the click-through rate prediction model. Based on the DNN model, the deep neural network model of the factorization machine is proposed. This model uses the hidden vector and its weight obtained by FM pre-training as the initial value of the neural network, and then provides the weight of DNN to learn higher-order features, and predicts the click rate of online ads.

$$Obj=\sum_{i=1}^{n}l(y_i,\bar{y}_i)+\sum_{k=1}^{K}\Omega(h_t) \tag{15}$$

$$\Omega(h_t)=\gamma J+\frac{\lambda}{2}\Sigma_{j=1}^{J}\omega_{tj}^2 \tag{16}$$

However, because the DNN model relies on the pre-training of the FM model, it affects the model performance. An improvement is proposed on the DNN model. As shown in Eq. (17) and Eq. (18), the improved PNN model can significantly improve the expression ability of the combined features. The DNN-based model predicts the click rate of online ads by solving the weights of higher-order combined features. We find the importance of low-order features for the prediction of online AD hits and propose Wide and Deep models for the combination of low-order and high-order features.

$$L_t=\sum_{i=1}^{m}L(y_i,f_{t-1}(x_i)+h_t(x_i))+\gamma J+\frac{\lambda}{2}\sum_{j=1}^{J}w_{tj}^2 \tag{17}$$

$$g_{ti}=\frac{\partial L(y_i,f_{t-1}(x_i))}{\partial f_{t-1}(x_i)},h_{ti}=\frac{\partial^2 L(y_i,f_{t-1}(x_i))}{\partial f_{t-1}^2(x_i)} \tag{18}$$

The model is divided into two parts, where the Wide part consists of a generalized linear model, and the Deep part is composed of a DNN model with three hidden layers. In past studies, as shown in Eq. (19) and Eq. (20), click rate prediction and bid optimization are usually carried out in order, firstly minimizing the error between the prediction result and the user response in the real situation by establishing the model. After obtaining the user response prediction result, it is used as input to optimize bids based on other factors such as activity budget, market price, etc.

$$L_t=\sum_{j=1}^{J}[G_{tj}w_{tj}+\frac{1}{2}(H_{tj}+\lambda)w_{tj}^2]+\gamma J \tag{19}$$

$$score=max(score,\frac{1}{2}\frac{G_L^2}{H_L+\lambda}+\frac{1}{2}\frac{G_R^2}{H_R+\lambda}-\frac{1}{2}\frac{(G_L+G_R)^2}{H_L+H_R+\lambda}-\lambda) \tag{20}$$

III. Online Advertising Model Based on Reinforcement Learning

A. CNN Incomplete Information Game and Regret Minimization Algorithm

When a user views the web page or opens a new page in the mobile App, one or more RTB AD spots will appear on the page. The reserved AD space on the page will launch an online AD display request to the supplier platform that provides the advertising agent for the website through the pre-written script code [20, 21]. After receiving the online AD request by clicking the AD display on the user page, the supplier platform SSP will send the AD space information and the web context information to ADX, the AD trading platform that performs the online advertising RTB [22, 23]. AD trading platform ADX receives online AD requests and publishes them to the DSP. In this process, the same online AD request may be sent to multiple different DSPs simultaneously. Fig. 1 shows an overview of the deep reinforcement learning framework. Each demand-side platform publishes details of online advertising requests to advertisers and carries out the first round of online advertising within the demand-side platform. Advertisers can query users' basic personal information, such as education, occupation, gender, etc. After query information, make a decision whether to participate in online advertising auction according to the established online advertising strategy [24, 25]. If you participate in online advertising, you will return your bid to the DSP. After the demand-side platform DSP receives the bid request of each advertiser, it will first rank the bid of each advertiser from the highest to the lowest level.

The advertisers with the highest bids will win the first round of online advertising. The DSP returns the graphic or video information of the top-ranked advertisers together with the bidding price determined by the broad second-highest price mechanism. Each demand side carries the highest-ranked advertising information and bidding price to participate in the second round of online advertising in ADX [26, 27]. ADX ranks

the bids of each demand side, and the highest-ranked demand side wins the second auction. Table I shows the values of AIC, BIC and HQIC, which won the first round of auction by the demand side. The advertising trading platform sends the information of the advertisement to the supplier platform. The supplier platform SSP transmits the advertising text, advertising links or advertising video information of the online advertising winner among the advertisers to the page viewed by the user, and the data will be displayed to the user after rendering through the page [28, 29]. After seeing the advertisement, users may click on the advertisement because of their own interests or be attracted by low discounts, or log in to register the website in the advertisement to complete the transformation, or they may feel that the content of the advertisement is novel and interesting, so they share the advertisement [30]. The complete RTB process describes a two-stage auction. That is, in each DSP internal, the advertisers carry out online advertising auctions. The green box mark part is the second stage of the auction, where each DSP with the highest ads in the first round of online advertising ranking, the second round of online advertising in ADX, competing for the display opportunity of advertising.

Online advertising with accurate positioning of users, and efficient online advertising mechanism, significantly improve users' use experience and delivery efficiency. A top DSP company such as Byte Dance can handle Cookie data from more than 570 million Internet users and use 3,155 attribute tags to represent each Cookie. The DSP sells more than three billion AD displays a day, with each AD display being auctioned off within 50ms. With this Cookies-based audience targeting technology, the market efficiency and effectiveness of RTB advertising have been increased by 50%. It is the Internet big data analysis technology that makes RTB advertising more accurate, controllable and efficient, and also makes RTB become the standard business model of the future online advertising market. The link of RTB is very complex and has a lot of uncertain factors. Fig. 2 shows the updated flow chart of the state-action value function. Therefore, it has certain practical significance to design an applicable online advertising model. When the settlement is CPM, the risk of advertising is estimated and controlled by the demand side. At the same time, because there is no quantification of the user's subsequent behavior, it is difficult to calculate and analyze the advertising effect. When the settlement is made by CPC, the supplier platform SSP can obtain a relatively accurate click rate estimate through the historical user data, and since the subsequent conversion is conducted in the AD demand side site, the AD demand side can make more accurate click value estimation. When the settlement is made by CPA, there is no risk of loss to the demand side, and the supply side is more difficult to operate. So now, in the form of CPA the settlement of online advertising, is gradually decreasing. Through comparison and synthesis, the settlement method of settlement by the click of advertising is the most beneficial to give full play to the advantages of the advertising supply side and the advertising demand side, so the CPC settlement method is widely recognized and accepted in the advertising market. From the perspective of CPC settlement, the online advertising model is discussed.

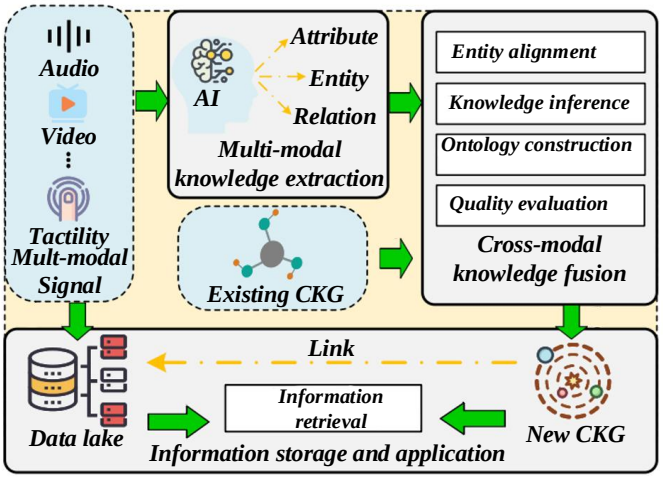


Fig. 1. Overview of the deep reinforcement learning framework.

TABLE I. THE VALUES OF INFORMATION CRITERIA

Information criteria	Price
Aic	1579.70254
Bic	1602.10028
Hqic	1588.73045

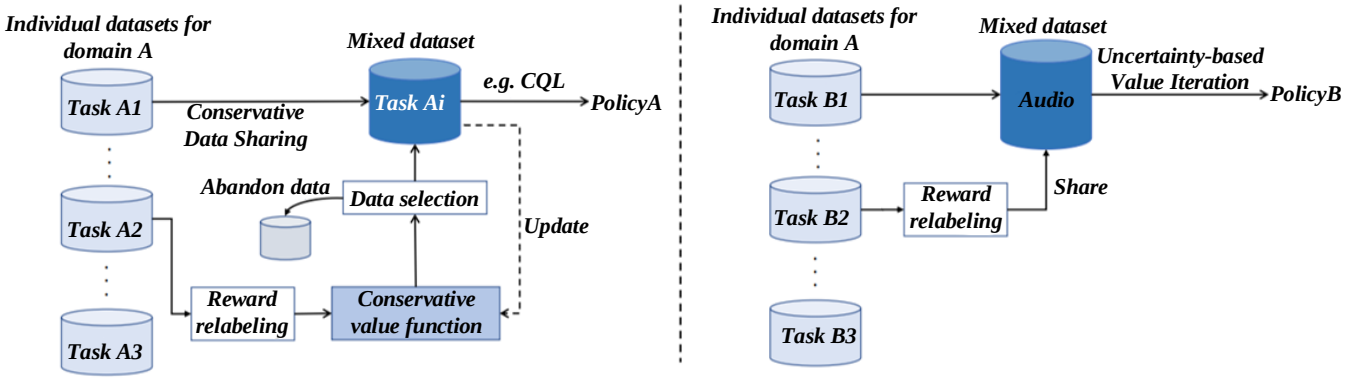


Fig. 2. State-update flow chart of the action value function.

B. Improve the DQN Customer Behavior Characteristic Analysis Model

After receiving the online advertising notification, advertisers use the known information to predict the click-through rate. After the advertiser gets the forecast result, the bid decision is made based on the value that the user's click brings to the brand, the advertising budget, the importance of the advertising space and other information. In most cases, the advertisers' bids are independent of each other. After receiving the quotation returned by each advertiser, DSP will also use known information to predict shot advertisements, calculate eCPC and rank online ads. In this calculation, advertisers with low click-through rate predictions cannot get advertising opportunities with high offers alone. To verify the proposed model, this paper introduces the famous advertising company, the published data training set for the global RTB algorithm competition. After pre-processing the data, it was found that the exposure of the advertisement changed periodically, and the exposure during the weekend period was significantly higher than the weekday's period. Fig. 3 is a graph of feature selection and network structure optimization algorithm. At the same time, the exposure at 12 and 22 points during the rest time is significantly higher than the working and sleep time. Before modeling the time series, the original sequence is required to verify the stationarity requirements.

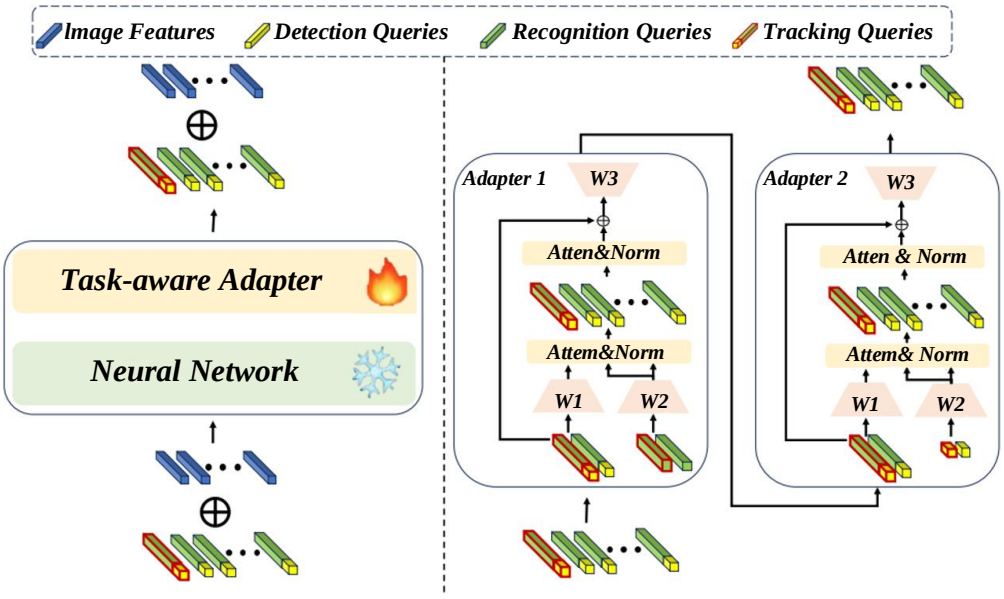


Fig. 3. Feature selection and network structure optimization algorithm.

The pattern of the characteristics of non-stationary sequences at various time points change randomly, which needs to be transformed by differential or logarithmic processing, and then modeling analysis. The randomness test is also known as a white noise test. In general, a white noise test of the stationary sequence is required before modeling to determine whether it has an analytical value. Table II shows the cross-entropy of different algorithms. In the time series model, the more unknown parameters contained, the more the independent variables are included, and the more flexible the corresponding model changes, the higher the accuracy of fitting the model and the greater the likelihood function value; however, when the number of unknown parameters in the model increases, the instability of the model becomes more difficult to fit the model. In general, a time series model is a good model when it considers the fitting accuracy and the number of unknown parameters.

TABLE II. CROSS-ENTROPY OF THE DIFFERENT ALGORITHMS

Algorithm Name	Logloss Cross Entropy
LR	0.423
XGBoost	0.414
ARMA-XGBoost	0.391

In the AD click rate prediction, the classification regression tree is generally chosen as the weak learners. The nature of the classification tree: First, make the basic assumptions, assume that the model is a basic binary tree, and then divide the nodes in the tree by constantly learning the features in the data set, and finally generate the classification tree that meets the expectations. In this article, By dividing it from the top down, The CART classification tree construction using a greedy strategy, Each child node in the tree is split according to the impurity of the subset elements, The urity of the set D that requires training using the Gini Expo measure, For the dataset D included in each node, XGBoost The model training goal is to computational solve the model to find the most appropriate division criterion and the final division value, Making the absolute value of the Gini index difference before and after the split, Fig. 4 shows a comparative evaluation chart of the effect of advertising channels, DA represents the size of the information gain: XGBoost has made great improvements in the algorithm and engineering application of GBDT. GBDT iterates on a set of weak learners, such as a decision tree, and outputs the final prediction results. Specific can be described as the error rate of the iterative decision tree to update the weight of the training set, and through the result of multiple decision tree accumulation as the prediction results output, through the

precise algorithm for all the information gain calculation value, select the maximum feature again using accurate algorithm for segmentation, get the corresponding training feature barrel. However, this order is not the optimal solution. According to Bayesian decision theory, the learning goal of the user response model should be determined by the final bid utility. In the paper, it is proposed that the accuracy of click rate required to predict and the computing resources are not the same within the range of all advertisements to be predicted. The prediction methods and the allocation of computing power should focus on cases with higher return on investment and learn how to predict more accurately. Therefore, the market price and competitive performance are included in the model, and if the investment return of an advertising space is relatively high, the confidence of whether the advertiser can take the advertisement will be predicted. When the confidence is low, the optimization method of click-through rate prediction and the allocated computing resources should be more concentrated than the advertising space with low investment return on ratio.

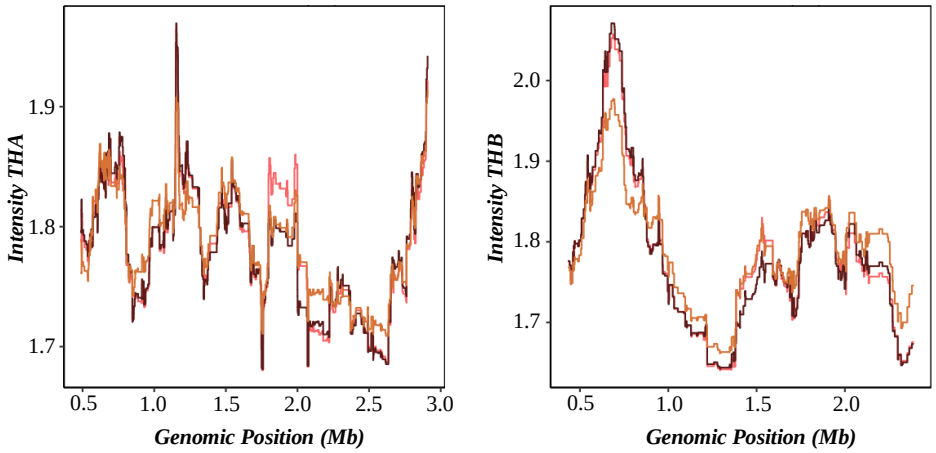


Fig. 4. Comparison and evaluation diagram of the effect of advertising channels.

IV. Research on Customer Behavior Characteristics Analysis and Online Advertising Optimization Based on Deep Reinforcement Learning

Grid search is a model hyperparameter optimization technique, which is essentially an exhaustive method. For each hyperparameter to be determined, a smaller finite set is chosen to explore. Then, we find the Cartesian products of the selected parameters to obtain several combinations of the hyperparameters. The grid search method trains the model using a combination of hyperparameters and selects the combination of hyperparameters that minimize the validation set error as the final parameter of the model. In the process of practical model training, the cross-validation method and the grid search method are usually combined, as the method of parameter evaluation, and this comprehensive method is recorded as the cross-validation grid search method. Fig. 5 shows the time evaluation chart of customer click behavior, and divides the labeled training data in the data set into n-folds for cross-validation. First in the super parameter grid search parameter calculation, and then each set of super parameters into the model for n fold training, select the score of the highest super parameter combination into model, using the model to train the training set data, while using the validation set data validation model training results, get the final results. In this paper, the area enclosed between the receiver working characteristic curve and the coordinate axis and the cross-entropy Logistic Loss is used to evaluate the prediction results of the click rate prediction model. AUC value evaluation index. In the online advertising auction, in order to ensure the maximum benefit and disturb users as little as possible, the accuracy index of the model is very important.

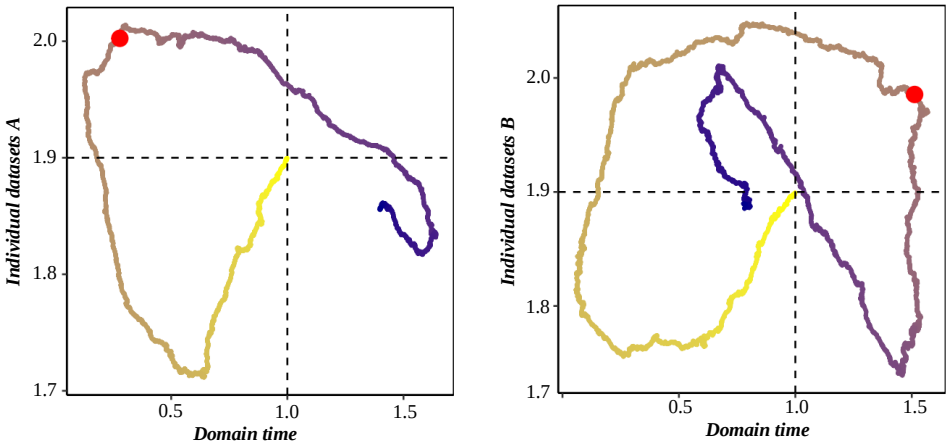


Fig. 5. Time assessment chart of customer click behavior.

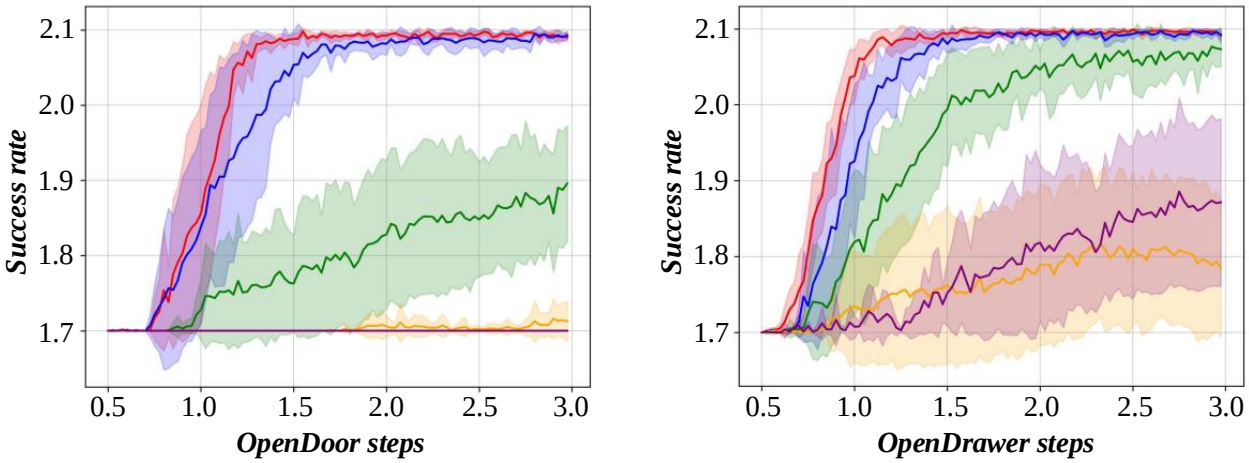


Fig. 6. Assessment chart of AD exposure and conversion rate.

The AUC value is the area obtained by calculus calculation of the ROC curve and the coordinate axis. It indicates the probability that the two samples are randomly selected and predicts the probability that the click rate. Where Rins, i represents the serial number of the sample, M represents the number of positive cases, and N represents the number of counterexamples. AUC value range between [0, 1], when the model AUC value of 0.5 represents a random classifier, the selection of threshold. The higher the AUC score indicates that the better the classifier predicts in the AD click rate prediction problem, the more meaningful the online advertisement. Game theory is based on rigorous mathematical models and introduces the limits of confrontational conflict in the real world. Fig. 6 shows the evaluation chart of advertising exposure and conversion rate, which defines the dominant strategy as follows: assuming that during the game process, the participants will adopt a certain strategy for any strategy choice of the other party. On this basis, if the strategy combination of all the players in the game is the dominant strategy for the other side, then the strategy combination of the two players is called the Nash equilibrium. The RTB online advertising problem discussed in this paper constitutes information asymmetry because the two parties do not know the other party's prediction of the click-through rate. But traditional game theory requires a complete set of situation information and strategies, such as in Texas Hold 'em, the opponent's cards must be a subset of the complete cards; in Go, the opponent's next strategy must be one of the feasible positions on the board. Therefore, the RTB online advertising problem discussed in this paper does not satisfy the incomplete information game; but it can make references for the incomplete information and it's solving methods.

In actual online advertising, the demand parties need to compete with each other in order to win the first auction round. In addition, in the process of continuous auction, both parties can push back the asymmetric information about the click rate by observing the auction results and testing the bid many times. Thus, it develops in the direction of an evolutionary game and achieves the equilibrium of the game through trial and error. In the first round of an auction, different demand parties may offer different bids, but they all face the DSP when predicting the click rate to calculate the eCPC ranking. Therefore, the scenario of the first round of auction is simplified, considering only the interaction between a single AD demand side and the DSP, and not the competition between the demand sides. In the process of continuous auction, the AD demand side and DSP may speculate positively or reverse the information about each other to achieve the dynamic equilibrium of evolution. However, in the early stage of the auction, designing the bid model for the unequal information between the parties is still beneficial to gain more benefits, and helps to seize the first advantage in evolution. In recent years, the ability to process large-scale data, discover the underlying features and extract the underlying features, so as to achieve specific goals more accurately. Fig. 7 shows the evaluation diagram of customer interest preferences at different ages. As a learning method with interactive ability, the online advertising decision model based on reinforcement learning is essentially a Markov decision process. Markov decision process has the characteristics of interacting with the environment, so this kind of model has the natural characteristics of modeling online advertising decision behavior, and the application of exploration mechanism can make the agent more fully explore the state and action space, and improve the accuracy and diversity of decision results to a certain extent. Among them, the online advertising decision system based on reinforcement learning consists of the current environment of the agent, the budget, the remaining advertising space and the information mastered by the DSP. In the process of the interaction between the agent and the environment, the decision agent constantly explores the decision scheme for the agent to obtain the maximum revenue according to the given budget, obtained traffic and other information, decide the bid at each auction time, and finally spends the budget to presents a decision scheme.

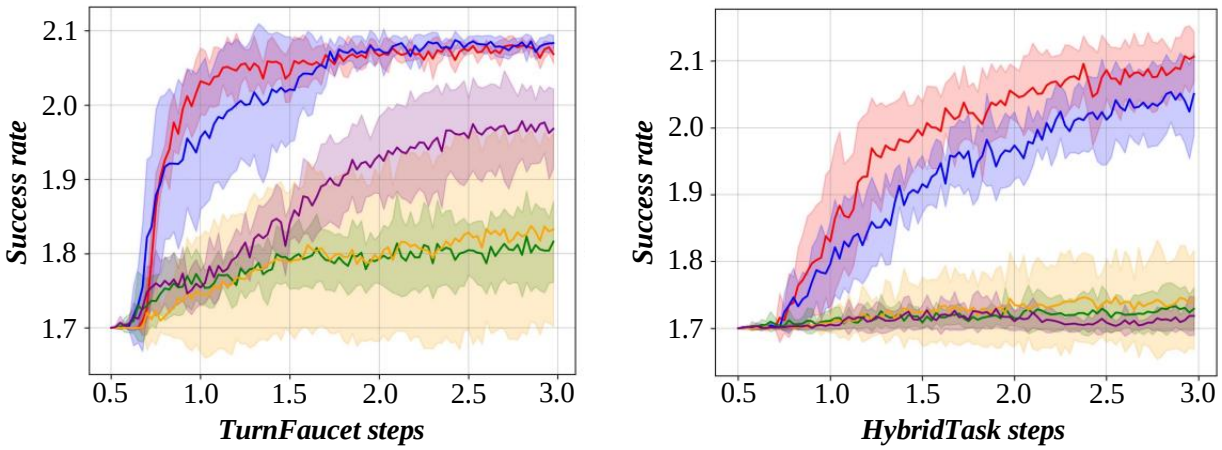


Fig. 7. Assessment chart of customer interest preferences at different ages.

Action space A: Action A represents the current moment, when the decision agent decides for the probability that the decision agent takes action a understate s, updates the budget and click rate, and moves to the next state S '. If the AD is won; if the AD is not won, one online advertising opportunity is lost. The decision agent can be converted to an immediate reward *R (s, a)* based on the feedback results. Discount factory: the factor that determines the value of long-term rewards at the current moment, the decision agent uses the greedy method, which is only considered for the immediate rewards; when y=1, the subsequent rewards have the same value at the current moment. Strategy π represents the basis for an individual taking an action, with its data described as a conditional probability distribution, namely the probability of taking an action *a* at state *s*. Action value function: the decision agent takes action A according to the strategy. The value function is an expectation function. Although Rt + 1, if only referring to the delay reward, it is easy to ignore the global situation and fall into the local optimal solution. Therefore, it is necessary to consider the delay reward of the current action and the potential delayed reward of the subsequent action. This section refers to the CFR algorithm for solving incomplete information games and improves the setting of the reward function in DQN. The regret value calculation method in CFR is introduced to set the reward function by taking the bid action and getting the environmental feedback regret value. The traditional Q-Learning reinforcement learning algorithm uses Q table to store data when making online advertising decisions. However, when the bidding environment is complex and the advertising space becomes more, Q table becomes huge and causes storage problems, and the search problem caused by the increase of data volume is easy to lead to the explosion of algorithm dimension. At present, in order to solve this problem, the industry generally uses other solutions to replace the Q value table, the most widely used is to use function approximation to replace. However, because the function approximation is calculated by calculating the value function, this alternative method is prone to the instability of the algorithm model and the failure to converge. The DQN algorithm combines the advantages of the traditional reinforcement learning algorithm Q-Learning and the deep neural network, which significantly improves the instability problem when the algorithm approximates the value function, which solves the problem of model instability and convergence to a certain extent.

V. Experimental Analysis

The core idea of the DQN algorithm is experience playback. During the training process, the reward results and model state update of each environment interaction are saved to the specified position, so that they can be used for the subsequent update operation of the target Q value. Fig. 8 for customer purchase path evaluation diagram, generally speaking, through the Q network calculated Q value and through experience playback target Q value will have some error, when the need to reduce the error can through the Q value gradient backpropagation to update the deep neural network parameter w, when the parameter w convergence, can get less error of approximate Q value.

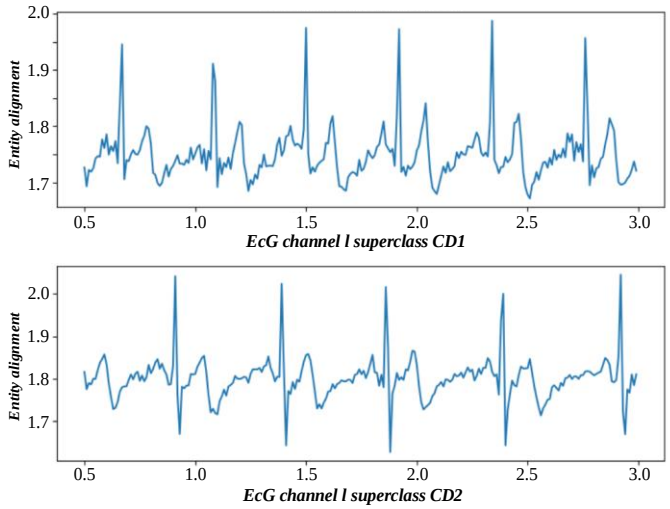


Fig. 8. Customer purchase path evaluation chart.

In reinforcement learning, the agent needs to explore in the current state to make better decisions, so as to obtain better returns, and choose the corresponding strategy to make the expected decision accordingly. The commonly used action selection strategies are the e-greedy algorithm. This paper selects the e-greedy algorithm. E-Greedy algorithm is a multi-arm gambling algorithm improved based on the greedy algorithm. Fig. 9 shows the evaluation diagram of the correlation of advertising creative type and click-through rate. In order to

avoid the agent always choosing the action with the current highest return to make the calculation probability of action execution. To obtain a better selection strategy, the strategy is randomly selected at probability E in the initial state.

In order to verify and analyze the prediction model and the subsequent online advertising model, the data set published by Tencent's 2019 advertising algorithm competition was selected. Fig. 10 is the evaluation chart of advertisement exposure during active time, which will hereinafter referred to as Tencent Data Set. In RTB transactions, Tencent can act as an advertiser to promote their games, or as an advertiser to publish ads for other brands on social platforms.

Therefore, Tencent's data set includes the whole process data and log of online advertising from online advertising to display and complete transformation: including historical exposure log, user characteristic and attribute data, advertising static data and advertising operation data. According to the assumptions of the model, the information inequality between the DSP and the demand side is discussed separately. Fig. 11 shows the cost-benefit analysis and evaluation chart of advertising, considering the following three experimental designs: the click rate predicted by the preset DSP is higher than the demand side, the click rate predicted by the preset DSP is lower than the forecast result of the demand side; and the real prediction results of the participants after dividing the data set. In the ideal stage of the initial transaction, by analyzing the winning results feedback by DSP, the estimated range of the DSP click rate is obtained.

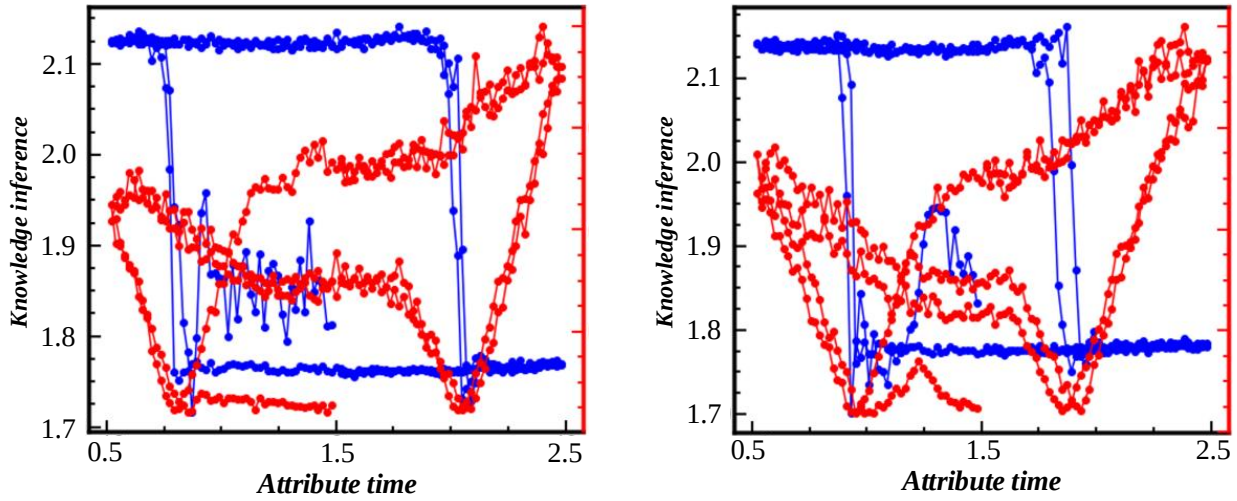


Fig. 9. Evaluation diagram of the correlation between advertising creative type and click-through rate.

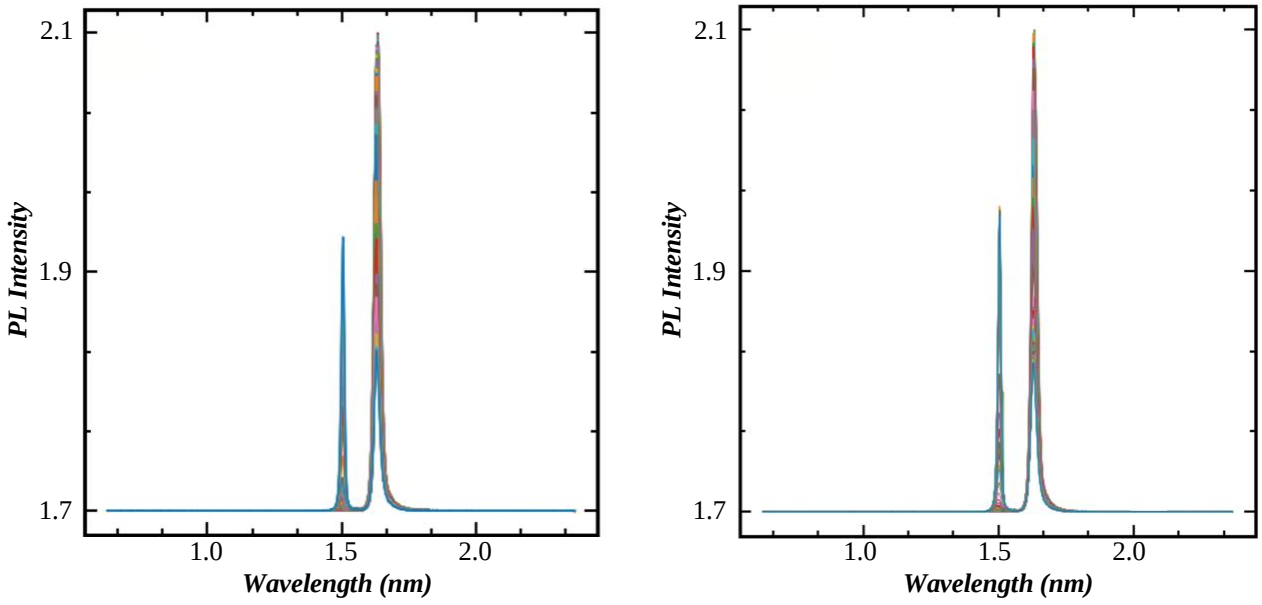


Fig. 10. Evaluation chart of advertising exposure during user active time.

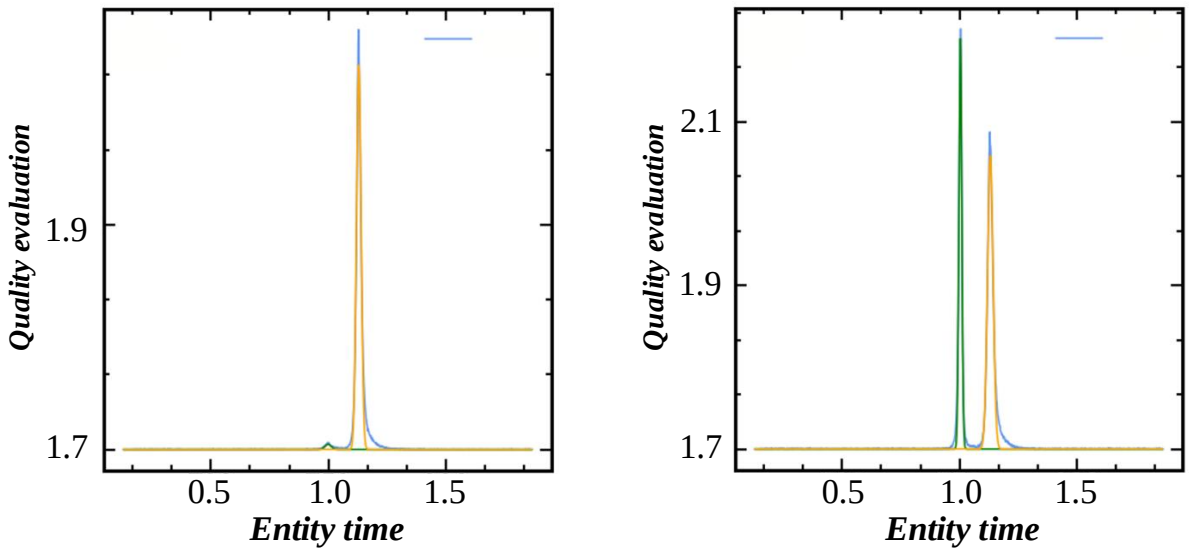


Fig. 11. Cost-benefit analysis and evaluation chart of advertising delivery.

VI. Conclusion

This paper introduces the common settlement method of online advertising, and introduces the eCPC formula of calculating online advertising ranking in the real production environment. Public information and private information in RTB transactions are defined and divided. This paper analyzes the problems caused by information asymmetry in the process of online advertising. Considering the difference in DSP and advertisers in the first auction, there is information asymmetry in online advertising. In this paper, the reinforcement learning model is designed, and the AD click rate is underestimated or overestimated by DSP, and the real prediction situation is designed respectively. The experiment shows that in the three scenarios, the click rate of improved DQN is higher than the traditional online advertising model. In the case where the click-through rate of advertising is overestimated by DSP, the high-frequency bid low-price advertising can maximize the interests of advertisers; in the situation where the click-through rate is underestimated by DSP, the low frequency for high-value advertising space is conducive to the exposure of the advertisement; in real scenes, using 10% of the budget to explore the environment, the price distribution of the final bid of the decision agent is the same as the distribution of click-through rate prediction results.

DSP and advertisers' click-through rate estimates follow a random distribution, and the prediction results are different in different ads. The prediction model of the final DSP has 23 inputs for training features and advertisers have 21 inputs for training features. According to the prediction results, the gap of DSP and advertisers within 0.02 was 15.7%, DSP higher than advertisers 41.7%, and DSP lower than advertisers 42.6%. LinBid Online advertising of Experiment 1 and 2, the LinBid winning rate always increases linearly with the budget. In contrast, the winning rate of the proposed improved DQN algorithm fluctuates greatly with the budget, and the improved DQN algorithm increases significantly when the budget increases. If the DSP predicted click-through rate is higher than the advertisers, the advertisers adopt the "positive" bidding strategy, high frequency bid low price ads, and achieve the click target; if the DSP predicted click-through rate is lower than the advertisers, the advertisers adopt the "cautious" bidding strategy, low frequency bid high price advertising, high price advertising, can significantly improve the return-on-investment ratio. When DSP has accumulated the user information of the brand and the transaction environment is gradually complex, DSP may overestimate or underestimate the click-through rate prediction results of advertisers: when the budget is low, advertisers can consider using the relatively stable LinBid online advertising model to bid based on the historical transaction price and the clicks predicted by advertisers. When the budget is high, and the pursuit of high profit, 10% of the budget to explore the proportion of the number of low-priced and high-priced ads in the final bid follows the proportion of the number of ads whose click rate is overestimated and undervalued by DSP.

References

[1] Abadi, Z. J. K., Mansouri, N., & Javidi, M. M. (2024). Deep reinforcement learning-based scheduling in distributed systems: a critical review. Knowledge and Information Systems, 74.

[2] Abdulazeez, D. H., & Askar, S. K. (2023). Offloading Mechanisms Based on Reinforcement Learning and Deep Learning Algorithms in the Fog Computing Environment. Ieee Access, 11, 12554-12585.

[3] Alipio, M., & Bures, M. (2023). Deep Reinforcement Learning Perspectives on Improving Reliable Transmissions in IoT Networks: Problem Formulation, Parameter Choices, Challenges, and Future Directions. Internet of Things, 23, 20.

[4] Allaoui, T., Gasmi, K., & Ezzedine, T. (2024). Reinforcement learning based task offloading of IoT applications in fog computing: algorithms and optimization techniques. Cluster Computing-the Journal of Networks Software Tools and Applications, 26.

[5] Almazrouei, K., Kamel, I., & Rabie, T. (2023). Dynamic Obstacle Avoidance and Path Planning through Reinforcement Learning. Applied Sciences-Basel, 13(14), 20.

[6] Chung, J. H., Fayyad, J., Al Younes, Y., & Najjaran, H. (2024). Learning team-based navigation: a review of deep reinforcement learning techniques for multi-agent pathfinding. Artificial Intelligence Review, 57(2), 36.

[7] Delgado, J. M. D., & Oyedele, L. (2022). Robotics in construction: A critical review of the reinforcement learning and imitation learning paradigms. Advanced Engineering Informatics, 54, 24.

[8] Dong, L., He, Z. C., Song, C. W., & Sun, C. Y. (2023). A review of mobile robot motion planning methods: from classical motion planning workflows to reinforcement learning-based architectures. Journal of Systems Engineering and Electronics, 34(2), 439-459.

[9] Esteso, A., Peidro, D., Mula, J., & Díaz-Madroñero, M. (2023). Reinforcement learning applied to production planning and control. International Journal of Production Research, 61(16), 5772-5789.

[10] Faria, R. D., Capron, B. D. O., Secchi, A. R., & de Souza, M. B., Jr. (2022). Where Reinforcement Learning Meets Process Control: Review and Guidelines. Processes, 10(11), 31.

[11] Frikha, M. S., Gammar, S. M., Lahmadi, A., & Andrey, L. (2021). Reinforcement and deep reinforcement learning for wireless Internet of Things: A survey. Computer Communications, 178, 98-113.

[12] Gao, Q. H., & Schweidtmann, A. M. (2024). Deep reinforcement learning for process design: Review and perspective. Current Opinion in Chemical Engineering, 44, 10.

[13] Ghotbi, M., & Zahedi, M. (2024). Predicting price trends combining kinetic energy and deep reinforcement learning. Expert Systems with Applications, 244, 12.

[14] Greguric, M., Vujic, M., Alexopoulos, C., & Miletic, M. (2020). Application of Deep Reinforcement Learning in Traffic Signal Control: An Overview and Impact of Open Traffic Data. Applied Sciences-Basel, 10(11), 25.

[15] Gupta, S., Singal, G., & Garg, D. (2021). Deep Reinforcement Learning Techniques in Diversified Domains: A Survey. Archives of Computational Methods in Engineering, 28(7), 4715-4754.

[16] Han, D., Mulyana, B., Stankovic, V., & Cheng, S. (2023). A Survey on Deep Reinforcement Learning Algorithms for Robotic Manipulation. Sensors, 23(7), 35.

[17] Hasan, Z., & Roy, N. (2021). Trending machine learning models in cyber-physical building environment: A survey. Wiley Interdisciplinary Reviews-Data Mining and Knowledge Discovery, 11(5), 13.

[18] Hickling, T., Zenati, A., Aouf, N., & Spencer, P. (2024). Explainability in Deep Reinforcement Learning: A Review into Current Methods and Applications. Acm Computing Surveys, 56(5), 35.

[19] Hou, H. H., Jawaddi, S. N. A., & Ismail, A. (2024). Energy efficient task scheduling based on deep reinforcement learning in cloud environment: A specialized review. Future Generation Computer Systems-the International Journal of Escience, 151, 214-231.

[20] Hua, J., Zeng, L. C., Li, G. F., & Ju, Z. J. (2021). Learning for a Robot: Deep Reinforcement Learning, Imitation Learning, Transfer Learning. Sensors, 21(4), 21.

[21] Ibrahim, A. M., Yau, K. L. A., Chong, Y. W., & Wu, C. (2021). Applications of Multi-Agent Deep Reinforcement Learning: Models and Algorithms. Applied Sciences-Basel, 11(22), 40.

[22] Jogunola, O., Adebisi, B., Ikpehai, A., Popoola, S. I., Gui, G., Gacanin, H., & Ci, S. (2021). Consensus Algorithms and Deep Reinforcement Learning in Energy Market: A Review. Ieee Internet of Things Journal, 8(6), 4211-4227.

[23] Ju, H., Juan, R. S., Gomez, R., Nakamura, K., & Li, G. L. (2022). Transferring policy of deep reinforcement learning from simulation to reality for robotics. Nature Machine Intelligence, 4(12), 1077-1087.

[24] Khoei, T. T., Slimane, H. O., & Kaabouch, N. (2023). Deep learning: systematic review, models, challenges, and research directions. Neural Computing & Applications, 35(31), 23103-23124.

[25] Li, C. X., Zheng, P., Yin, Y., Wang, B. C., & Wang, L. H. (2023). Deep reinforcement learning in smart manufacturing: A review and prospects. Cirp Journal of Manufacturing Science and Technology, 40, 75-101.

[26] Lin, B. H. (2024). Reinforcement learning and bandits for speech and language processing: Tutorial, review and outlook. Expert Systems with Applications, 238, 32.

[27] Massaoudi, M., Chihi, I., Abu-Rub, H., Refaat, S. S., & Oueslati, F. S. (2021). Convergence of Photovoltaic Power Forecasting and Deep Learning: State-of-Art Review. Ieee Access, 9, 136593-136615.

[28] Massaoudi, M. S., Abu-Rub, H., & Ghrayeb, A. (2023). Navigating the Landscape of Deep Reinforcement Learning for Power System Stability Control: A Review. Ieee Access, 11, 134298-134317.

[29] Mohammed, M. Q., Chung, K. L., & Chyi, C. S. (2020). Review of Deep Reinforcement Learning-Based Object Grasping: Techniques, Open Challenges, and Recommendations. Ieee Access, 8, 178450-178481.

[30] Munikoti, S., Agarwal, D., Das, L., Halappanavar, M., & Natarajan, B. (2023). Challenges and Opportunities in Deep Reinforcement Learning With Graph Neural Networks: A Comprehensive Review of Algorithms and Applications. Ieee Transactions on Neural Networks and Learning Systems, 21.

Logistics Transportation Vehicle Monitoring and Scheduling Based on the Internet of Things and Cloud Computing

Kang Wang*, Xin Wang
School of Economics and Management, Jiaozuo University, Jiaozuo, China

***Abstract*—This paper addresses challenges in the logistics industry, particularly information lag, inefficient resource allocation, and poor management, exacerbated by global economic integration and e-commerce growth. An advanced logistics and transportation vehicle monitoring and scheduling system is designed using IoT and cloud computing technologies. This system integrates Yolov5 for real-time vehicle location, DeepSort for continuous tracking, and a space-time convolutional network for vehicle status analysis, forming a comprehensive monitoring model. An improved multi-objective particle swarm optimization algorithm optimizes vehicle scheduling, balancing objectives like minimizing travel distance, time, and carbon emissions. Experimental results demonstrate superior performance in real-time monitoring accuracy, scheduling efficiency, arrival time prediction, road condition forecasting, and failure risk prediction. Notable achievements include 95% vehicle utilization, a 0.25 RMSE for predicted arrival times, and a 0.20 MAE for failure risk prediction. While the system significantly enhances operational efficiency and supports resource optimization, future work will focus on data security, system stability, and practical deployment challenges. This research contributes to transforming the logistics industry into a smarter, greener, and more efficient sector.**

***Keywords*—Internet of Things; cloud computing; logistics and transportation; vehicle monitoring; vehicle scheduling**

I. Introduction

In the context of accelerated global economic integration and the rise of e-commerce, logistics plays a critical role as the bridge between production and consumption. However, the traditional logistics model faces significant challenges, including information lags, inefficient resource allocation, and low management effectiveness, which hinder the industry's potential and efficiency [1].On one hand, the slow pace of information updates does not match the rapidly evolving business environment. In traditional systems, the lack of a real-time, transparent information flow mechanism leads to asymmetric information across the supply chain, impacting decision-making and causing issues like cargo delays and retention. On the other hand, inefficient resource utilization is another pressing issue [2]. Common problems include empty vehicles, idle warehouses, and redundant manpower, indicating significant room for optimization in capacity planning, warehouse layout, and human resource management. These inefficiencies increase logistics costs and undermine sustainability. Furthermore, limitations in management efficiency are evident, with traditional models and techniques making it difficult to achieve refined and intelligent management in areas such as order processing, distribution scheduling, and customer service [3].

To address these challenges, the logistics industry must embrace new technologies like IoT, cloud computing, big data, and AI to develop smart and efficient logistics and transportation vehicle monitoring and scheduling systems. This will enable the industry to innovate and upgrade, transitioning from information technology to intelligence, and ensuring competitiveness in the global logistics landscape. A technology share diagram for vehicle monitoring and scheduling is shown in Fig. 1.

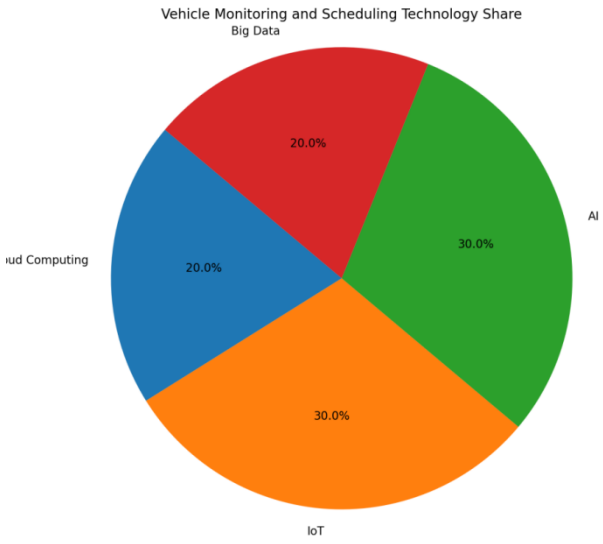


Fig. 1. Technology shares in-vehicle monitoring and scheduling (Source: UCI machine learning repository).

In the context of global economic integration and the rapid evolution of e-commerce, the logistics industry, as a core link connecting production and consumption, is becoming a key driver of global trade and economic growth. However, the traditional logistics model faces significant challenges, including information lags, unbalanced resource allocation, and low management efficiency, which hinder the industry's potential and efficiency [3].

*Corresponding Author.

Information transmission lags and the absence of real-time, transparent information sharing mechanisms lead to asymmetric information across the supply chain, affecting decision-making and causing issues like cargo delays and delivery holdups [4]. Unreasonable resource allocation, such as empty vehicles and wasted storage space, highlights significant optimization opportunities. These inefficiencies increase costs and undermine sustainability goals. Extensive management practices further limit service quality and customer satisfaction [5].

While there have been discussions about using new technologies to improve logistics efficiency, key issues remain unresolved, such as how to integrate IoT and cloud computing technologies to achieve intelligent upgrades, and how to address real-time data processing, resource optimization, security, and cost control [6].

To address these challenges, this study proposes an innovative logistics transportation vehicle monitoring and scheduling system based on IoT and cloud computing technologies. The system aims to promote the intelligent transformation of the logistics industry by focusing on (1) Building a comprehensive logistics vehicle monitoring network using IoT technology for transparent supply chain management. (2) Leveraging cloud computing for efficient data processing and storage, enabling deep analysis of logistics data to improve scheduling accuracy and efficiency. (3) Optimizing vehicle scheduling through intelligent algorithms to reduce idle time, optimize routes, and minimize energy consumption. Additionally, enhancing safety management during logistics transportation. (4) Optimizing logistics resource allocation through data analysis to improve the efficiency and service quality of the entire logistics chain.

The research innovations include: (1) Combining advanced computer vision technology (Yolov5 for vehicle positioning) with DeepSort tracking and spatio-temporal convolutional networks to create a comprehensive vehicle condition monitoring system. (2) Introducing and optimizing a particle swarm optimization algorithm, particularly focusing on multi-objective optimization strategies to resolve common conflicts in logistics transportation, such as balancing cost, time, and environmental impact. (3) Designing an end-to-end technical framework from data acquisition to intelligent decision-making, ensuring the solution's comprehensiveness and operability.

This paper addresses the challenges faced by the logistics industry, proposing an advanced logistics and transportation vehicle monitoring and scheduling system using IoT and cloud computing technologies. The system integrates Yolov5 for real-time vehicle location, DeepSort for continuous tracking, and a space-time convolutional network for vehicle status analysis. An improved multi-objective particle swarm optimization algorithm optimizes vehicle scheduling, balancing objectives like minimizing travel distance, time, and carbon emissions. Experimental results demonstrate superior performance in real-time monitoring accuracy, scheduling efficiency, arrival time prediction, road condition forecasting, and failure risk prediction. Notable achievements include 95% vehicle utilization, a 0.25 RMSE for predicted arrival times, and a 0.20 MAE for failure risk prediction. The paper is structured as follows: Section II provides a literature review and related work; Section III describes the methodology and technical framework; Particle Swarm Algorithm is given in Section IV; Section V presents the technical framework; Section VI presents the experimental setup and results; finally, Section VII concludes the paper and outlines future research directions.

II. Literature Review

A. Vehicle Monitoring Methods

In today's logistics and transportation industry, IoT technology has become an important tool for realizing efficient vehicle monitoring, which significantly enhances the transparency and controllability of the logistics and transportation process by integrating a variety of sensing devices to collect real-time and accurate information about the status of vehicles and their cargo.

1) Application of IoT technology in vehicle monitoring and control: The use of Internet of Things (IoT) technology in logistics vehicle monitoring has gained widespread scientific attention and practical application. Technologies such as GPS global positioning systems [8] are commonly embedded inside vehicles and can continuously provide real-time geographic location of vehicles, which not only supports precise geographic navigation, but also can be used to track the trajectory of logistics vehicles to ensure compliance and efficiency of transportation routes [9]. In addition, on-board rfid (radio-frequency identification) tags and reader systems [10] can monitor the identity and status changes of goods in real time, effectively preventing misallocation or loss of goods. Environmental monitoring equipment such as temperature and humidity sensors [11] can monitor the temperature and humidity conditions of the goods in real time to ensure the safe preservation of perishable or special goods during transportation.

2) Data transmission and processing: The massive amount of data generated by IoT devices must be efficiently transmitted and processed before it can be transformed into valuable information. Advances in wireless communication technologies, such as 4g/5g wide-area networks [12] and narrow-band IoT (nb-IoT) technology [13], provide high-speed, stable channels for data transmission from IoT devices, ensuring real-time transmission of vehicle monitoring data to cloud servers. The cloud computing platform plays a crucial role in this process. Through the infrastructure provided by cloud service providers such as alicloud, aws, azure, etc., the data collected by IoT can realize large-scale and highly concurrent data storage [14]. In addition, cloud computing platforms use their powerful data processing capabilities to clean, integrate, and analyze the collected data in real time, and even perform deep mining through machine learning algorithms [15] to report on operating conditions, predict failures, and optimize scheduling strategies. For example, a study [16] successfully realized remote monitoring and intelligent warning of logistics vehicles by building a cloud

computing-based IoT platform, improving the safety and efficiency of the transportation process. The study shows that by combining IoT data with cloud computing, it can not only effectively solve the problem of information silos in logistics and transportation, but also greatly improve the management effectiveness and customer service satisfaction of logistics enterprises.

Despite the advancements in IoT technology for vehicle monitoring, several limitations persist. One critical challenge lies in the interoperability and standardization of IoT devices across different manufacturers and platforms, which can result in data inconsistencies and compatibility issues. Moreover, the sheer volume of data generated often surpasses the analytical capabilities of some organizations, leading to a gap between data collection and actionable insights. There is also a need for more robust cybersecurity measures, as the increased connectivity exposes logistics systems to higher cyberattack risks [26]. Addressing these limitations requires the development of universal standards, enhancing data analytics capabilities, and implementing advanced security protocols.

B. Vehicle Scheduling Model

1) Traditional vehicle scheduling models: Traditional vehicle scheduling models have always been the core theoretical basis for logistics and transportation optimization, and the most representative ones include the capacitated vehicle routing problem (cvrp) and vehicle routing problem with time windows (vrptw). The cvrp model the cvrp model focuses on solving the problem of how to plan the shortest total distance traveled path to serve all customers while ensuring that each vehicle does not exceed its cargo capacity [17]. The vrptw model, on the other hand, adds complexity to this by not only considering vehicle capacity constraints but also ensuring that each customer is served within a preset time window. Although these models play an important role in rational allocation of resources and cost reduction, they mainly rely on pre-set static information and cannot adapt to changes in the external environment in real-time. For example, in real logistics scenarios, the uncertainty caused by traffic congestion, sudden demand changes, vehicle failures, etc., makes the traditional scheduling model show obvious limitations in solving the real-time scheduling problem.

2) Intelligent scheduling model Based on IoT and cloud computing: With the rapid development of the Internet of Things (IoT) technology and cloud computing technology, a new type of vehicle scheduling model with highly flexible and intelligent features has emerged. This model makes full use of the real-time sensing capability provided by the Internet of Things and the advantages of large-scale data processing and high-speed computing of the cloud computing platform and greatly improves the shortcomings of the traditional scheduling model in dealing with dynamic and complex environments. Various sensors, GPS positioning systems, and in-vehicle communication devices deployed by IoT technology in the logistics and transportation chain are able to collect and update multifaceted information such as vehicle location, status, and road conditions in real time [18]. After these real-time data are uploaded to the cloud computing platform, they are rapidly integrated and mined through big data analytics technology [19] to form a panoramic view reflecting the current overall operational situation. On this basis, advanced machine learning algorithms such as reinforcement learning (rl) [20] and deep learning (dl) [21] are applied to dynamic route planning and real-time scheduling optimization, enabling the system to respond optimally and quickly in the face of various uncertainties. Intelligent scheduling systems are able to adjust travel routes, reassign tasks, and predict potential delay risks in real time, thus effectively reducing the idle rate and waiting time, and greatly improving the efficiency of logistics and transportation and the quality of service [22].

Comprehensively speaking, the intelligent scheduling model based on IoT and cloud computing has realized the transformation from static to dynamic and from lagging to instantaneous compared with the traditional model, which can better adapt to the ever-changing market demand and operating conditions, and bring unprecedented level of refined management and efficient operation for the logistics and transportation industry.

Traditional vehicle scheduling models, despite their contributions, often struggle with real-time adaptability due to their reliance on predetermined parameters. These models may not effectively handle unexpected events such as sudden weather changes, traffic incidents, or urgent customer requests, which can lead to suboptimal route planning and inefficient resource allocation. Additionally, the computational complexity of these models escalates rapidly with the increase in the number of vehicles and delivery points, which can strain computational resources. To overcome these limitations, there is a pressing need for models that incorporate real-time data processing and predictive analytics to enhance decision-making flexibility and accuracy under dynamic conditions.

C. Application of Monitoring and Scheduling of Logistics and Transportation Vehicles Based on Internet of Things and Cloud Computing

1) Practical application cases: Nowadays, many leading domestic and international logistics companies have begun to adopt vehicle monitoring and dispatching systems based on IoT and cloud computing technologies, which have achieved significant practical benefits. For example, sf express has introduced IoT equipment and cloud computing platform in its logistics network, effectively realizing remote monitoring and intelligent dispatching of its huge fleet of vehicles through real-time monitoring of vehicle location and status information [23]. Through IoT technology, real-time transmission of vehicle gps data, driving status data, etc. To the cloud platform, combined with big data analysis technology, the system is able to accurately predict and plan the optimal driving routes, reduce unnecessary empty mileage, and improve the loading rate, thereby saving fuel costs and improving logistics efficiency [24]. In practice, cainiao

network has also created an intelligent logistics system using IoT and cloud computing to realize dynamic vehicle scheduling and real-time monitoring [25]. Through the sensors installed on the vehicle and mobile communication technology, the system can provide real-time feedback on the vehicle's operating status, cargo status and driver behavior, etc. The cloud computing platform carries out rapid processing and analysis of these data to optimize the vehicle scheduling program in real time, reduce operating costs, and improve service quality.

2) Application challenges and countermeasures: Although the logistics vehicle monitoring and dispatching system based on IoT and cloud computing has made remarkable achievements, it still faces a series of challenges in the process of practical application. First, data security and privacy protection is a major challenge. The large amount of data generated by IoT devices may be subject to malicious attacks or illegal theft during transmission and storage. In order to ensure information security, researchers are actively exploring and applying advanced encryption technologies, such as lightweight encryption algorithms and blockchain technology, to ensure the integrity and confidentiality of data during transmission and storage. Secondly, the stability and real-time responsiveness of the system is also an issue that should not be ignored. Large-scale IoT device access may lead to data flooding, affecting the processing efficiency and response speed of the cloud computing platform [28]. In order to solve this problem, researchers propose to adopt an edge computing strategy, i.e., offloading part of the data processing and analysis tasks to edge nodes close to the data source, reducing the pressure on the central cloud platform and improving the real-time response performance of the system. Optimizing communication protocols to ensure efficient and stable data transmission is also a research hotspot. By optimizing wireless communication technologies such as item and 5G, the network coverage and data transmission rate are improved, and the delay is reduced to ensure the real-time transmission of monitoring data.

While the integration of IoT and cloud computing in logistics has demonstrated substantial benefits, practical implementation faces several hurdles. Integration complexity, especially in legacy systems, poses a significant challenge as it requires seamless interfacing between various hardware components and software platforms. Additionally, the cost associated with the initial setup and ongoing maintenance of IoT infrastructure and cloud services can be prohibitive for smaller logistics companies. Ensuring continuous power supply for IoT devices in remote locations and managing the overwhelming amount of data generated without compromising data quality remains another challenge. To mitigate these issues, strategies such as phased implementation, leveraging cloud-based pay-as-you-go models, and investing in advanced data filtering and cleaning techniques are imperative. Furthermore, fostering collaboration among stakeholders to establish common standards and best practices can facilitate smoother integration and wider adoption of these advanced technologies in the logistics sector.

In preparation for this study, this paper conducted extensive literature research to fully understand the current state of the field of vehicle monitoring and scheduling in logistics transportation. This paper searched academic databases (e.g. Web of Science, Scopus, IEEE Xplore, SpringerLink, and Google Scholar) for relevant literature from the past decade, using keywords such as "IoT Logistics", "Cloud Computing Dispatch", "Vehicle Monitoring System", "Intelligent Logistics", etc. for precise searches. Through careful screening, this paper focusses on those representative and innovative research papers in technology implementation, algorithm optimization, system design and practical application effect evaluation. In particular, documents [1] to [5] provide us with a macro perspective of the challenges and opportunities facing the logistics industry, pointing out the key role of information technology, especially the Internet of Things and cloud computing technologies, in logistics modernization. Documents in study [6] and [7] discuss in depth the latest advances in vehicle monitoring technology and how they improve logistics management through real-time data transmission and intelligent analysis. The study [8] to [10] focuses on the development of vehicle scheduling models, from traditional optimization methods to dynamic scheduling strategies based on intelligent algorithms, which provide the theoretical basis for model design. In addition, this paper also refers to a number of case studies [23], [25] that demonstrate the successful application of IoT and cloud computing technologies in real-world logistics operations, providing valuable lessons learned and implementation strategies. Although there have been studies on the application of the Internet of Things and cloud computing in logistics, there is still a lack of in-depth and systematic research on how to deeply integrate these technologies, realize seamless connection from real-time vehicle monitoring to intelligent scheduling strategy, and how to effectively deal with the resulting data security and system stability problems. In addition, the application effect evaluation and parameter optimization method of multi-objective optimization scheduling model in actual logistics scenarios are also weak.

Based on the findings of the above literature review, this paper defines research orientation: to build an integrated logistics vehicle monitoring and scheduling system integrating advanced Internet of Things monitoring technology, cloud computing processing capabilities and multi-objective particle swarm optimization scheduling algorithm. The system was designed to address several challenges identified in the literature, including improving information transparency and decision efficiency, optimizing resource allocation, ensuring data security, and increasing scheduling flexibility. Through this innovative research, this paper not only deepens and expand the application of existing logistics technology but also provides a new theoretical basis and practical guidance for the intelligent transformation of the logistics industry.

III. VEHICLE MONITORING MODEL

A. Constructing a Vehicle Localization Module Based on Yolov5

Yolov5 is a real-time target detection model that outputs all target classes and their locations in an image in a single prediction. In a vehicle monitoring system, yolov5 is responsible for the initial localization of vehicles. Its network structure employs techniques such as cross-stage partial connectivity (csp) and cross-scale feature pyramid networks (fpn) to achieve fast and accurate vehicle detection. The output of the yolov5 model is a two-dimensional tensor containing the predicted values of multiple bounding boxes for each grid cell, including the confidence, class probability) and the center coordinates, width and height of the bounding box. The location loss function l (location loss) can be expressed as:

$$L_{loc}=\sum_{i=0}^{S^2}\sum_{j=0}^{B}x_{ij}^{obj}\left[\begin{array}{l}\lambda_{coord}\sum_{xywh}\mathrm{L}_1(pred_{ij}^{xywh},truth_{ij}^{xywh})\\+\mathrm{L}_1(pred_{ij}^{obj},truth_{ij}^{obj})\end{array}\right]$$

Where S^2 is the number of grids, B is the number of bounding boxes predicted for each grid, B is an indicator variable indicating the presence or absence of an object, $pred_{ij}^{xywh}$ and $truth_{ij}^{xywh}$ are the bounding box coordinates for the predicted and true values, respectively, $pred_{ij}^{obj}$ and $truth_{ij}^{obj}$ are the confidence level predicted and true values, λ_{coord} is a balancing coefficient, and L_1 is the mean squared error (MSE) or the huber loss function.

B. Deepsort-Based Vehicle Tracking Module

On the basis of vehicle localization, deepsort algorithm is used for continuous vehicle tracking. deepsort combines kalman filter for state prediction and deep learning methods (e.g., reid model) to extract vehicle features, and trajectory matching is performed by calculating similarity and iou values between detection frames.

The core of the correlation algorithm is to calculate the correlation score between the detection frame and the previous trajectory, which can be expressed as eq:

$$AssociationScore_{ij}=\exp\left(-\frac{\|f(Detection_i)-f(Track_j)\|_2^2}{2\sigma^2}\right)\times\mathrm{IOU}(Detection_i,Track_j)$$

C. Vehicle Condition Monitoring Module Combining Spatio-Temporal Convolutional Networks

Temporal convolutional networks (tcns) are used to analyze vehicle state time-series data, such as speed, acceleration and other dynamic features. Accurate monitoring of vehicle states is realized by capturing short-term and long-term dependencies of time series through deep convolutional layers [20].

In the tcn model, the convolution operation at each layer can be represented as:

$$H_t^n=f\left(\sum_{k=\max(0,t-K)}^{\min(T-1,t+K)}W_k^n*X_{t-k}^{n-1}+b^n\right)$$

Where, H_t^n is the output feature of the n th layer at time step t, X_{t-k}^{n-1} is the input feature of the previous layer at time step $t-k$, W_k^n is the weight of the temporal convolution kernel, b^n is the bias term, f is the nonlinear activation function, K is the time span of the convolution kernel, and T is the total length of the time series [26].

D. Integrated Monitoring Modeling Framework

The above three modules are organically combined to build a comprehensive vehicle monitoring system, as shown in Fig. 2. First, yolov5 performs real-time vehicle detection on the video stream to generate vehicle location information; next, deepsort receives this location information and combines it with historical data for vehicle tracking to maintain the tracking of the vehicle's continuous motion state; finally, the spatio-temporal convolutional network analyzes the vehicle's time-series state data in order to provide more comprehensive monitoring of the vehicle's state [21].

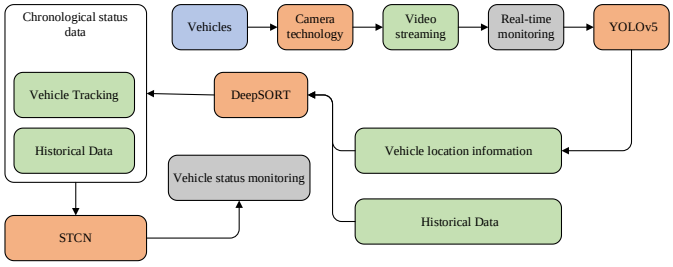


Fig. 2. Framework model.

The proposed integrated vehicle monitoring system (see Fig. 2) integrates the above three modules to form a highly coordinated and comprehensive solution. Starting with YOLOv5's real-time positioning, to DeepSORT's continuous, accurate tracking, to TCNs 'deep analysis of vehicle dynamics, this framework enables closed-loop monitoring from initial vehicle detection to detailed behavior analysis. This system not only optimizes the real-time monitoring efficiency, reduces the false detection rate and missed detection rate, but also significantly enhances the adaptability and robustness to complex environments through the application of deep learning technology. Its innovation lies in:

Efficient Integration: Seamless integration of cutting-edge computer vision and deep learning technologies to build an end-to-end solution from object detection to behavioral analysis.

Accurate tracking: Through the optimization of DeepSORT algorithm, continuous and accurate tracking of vehicles in dynamic and complex scenes is realized, and the stability of the overall system is improved.

In-depth analysis: The application of TCNs breaks through the limitations of traditional monitoring systems, realizes a deep understanding of vehicle dynamic states, and provides possibilities for advanced applications such as intelligent scheduling and safety warning.

Real-time response: The entire framework design focuses on real-time, ensuring that the monitoring system can respond quickly, process and feedback vehicle status information in a timely manner, and improve logistics transportation efficiency and safety.

To sum up, the proposed system not only has clear structure and strict logic, but also significantly improves the intelligent and refined level of vehicle monitoring through technological innovation, which brings innovation to the logistics industry and other fields related to large-scale vehicle management.

IV. Particle Swarm Algorithm-based Vehicle Scheduling Model in Logistics Transportation

A. Modeling of Logistics Transportation Vehicle Scheduling

In logistics transportation, vehicle scheduling problems are usually manifested in the form of capacitated vehicle routing problem (cvrp) or vehicle routing problem with time windows (vrptw). Factors considered include vehicle cargo capacity, customer delivery demand, traveling distance, service time window, driver working time constraints, and other dimensions. When modeling, the problem can be transformed into an optimization problem where the objective is to find one or more vehicle travel paths that satisfy all customer demands while minimizing the total travel distance, total travel time, or total cost. In logistics and transportation, the mathematical model of a vehicle scheduling problem usually involves the following key elements: x_{ij} Represents 1 if the vehicle travels directly from customer i to customer j, and 0 otherwise. Q represents the maximum cargo capacity of the vehicle. q_i Represents the demand of customer i. d_{ij} Represents the distance from customer i to customer j. T represents the maximum working time of the driver. s_i Represents the service time at customer i. e_i, l_i represent the service time windows of logistics task i with start and end times, respectively [34].

The objective function is to minimize the total distance traveled: $\min \sum_{i=1}^{n} \sum_{j=1, j\neq i}^{n} d_{ij} x_{ij}$. The constraints mainly include that each customer can only be visited once: $\sum_{i=1, i\neq j}^{n} x_{ij} = 1 \quad \forall j \in 1, \ldots, n$, the vehicle load does not exceed the maximum load q: $\sum_{i=1}^{n} q_i x_{ij} \leq Q \quad \forall j \in 1, \ldots, n$, the driver's working time does not exceed t: $\sum_{i=1}^{n} \sum_{j=1, j\neq i}^{n} (d_{ij} + s_i) x_{ij} \leq T$, and the service time window constraints: $e_i \leq t_i \leq l_i \quad \forall i \in 1, \ldots, n$ Where, t_i denotes the time when the vehicle arrives at the customer i [22].

B. Traditional Particle Swarm Algorithm for Scheduling of Logistics and Transportation Vehicles

Particle swarm algorithm in solving logistics transportation vehicle scheduling problem, each feasible vehicle scheduling program as a "Particle", its position vector indicates the driving route of each vehicle, the speed vector indicates the possibility of changing the driving route. The main process of particle swarm algorithm is as follows:

1) Initialization: Setting the swarm size, maximum number of iterations, initial velocity and position, as well as parameters inertia weights w and acceleration constants c_1, c_2.

2) Evaluate the fitness: Calculate the fitness value (e.g., total distance traveled, total cost) corresponding to each particle's position (i.e., vehicle scheduling scheme).

3) Update personal optimal solution (pbest): If the fitness value corresponding to the position of the current particle is better than its historical optimal solution, then update the personal optimal solution of this particle.

4) Update the globally optimal solution (gbest): Find the particle with the best fitness value in the whole particle swarm and take its position as the global optimal solution [23].

5) Update speed and location:

$$v_{i,d}(t+1) = w \cdot v_{i,d}(t) + c_1 \cdot r_1 \cdot (pbest_{i,d} - x_{i,d}(t)) + c_2 \cdot r_2 \cdot (gbest_d - x_{i,d}(t))$$

The value of particle i personal optimal solution and global optimal solution in the d_{th} dimension is given by $x_{i,d}(t+1) = x_{i,d}(t) + v_{i,d}(t+1)$ and. Where $(v_{i,d}(t)$ and $x_{i,d}(t)$ denote the velocity and position of particle i in the dth dimension, respectively, and $pbest_{i,d}$ and $gbest_d$ denote the values of the personal and globally optimal solutions of particle i in the dth dimension, respectively.

6) Judge the stop condition: Check whether the maximum number of iterations is reached, if not, then return to step 2 to continue iteration.

C. Improved Multi-Objective Particle Swarm Algorithm Applied to Logistics Transportation Vehicle Scheduling

In practical logistics and transportation, the vehicle scheduling problem often involves multiple conflicting objectives, such as minimizing the driving distance, minimizing the total transportation time, and reducing carbon emissions. At this time, multi-objective particle swarm algorithm (mopso) can be used. In mopso, each particle has multiple objective function values, forming a pareto front

solution set. The algorithm process is basically the same as the traditional pso, but the adaptation evaluation and the selection of the optimal solution need to take into account multiple objectives. The fitness function can adopt multi-objective optimization strategies such as hierarchical weighting method or objective space decomposition method.

For multiple objective values of a particular particle i $f_1(x_i), f_2(x_i),..., f_m(x_i)$, its position in the pareto front can be computed using the non-dominated ordering and the crowding distance. When updating the velocity and position, not only the optimal solution of a single objective is considered, but also the distribution of the whole pareto front is taken into account. In the improved multi-objective particle swarm algorithm, the velocity update formula becomes:

$$v_{i,d}(t+1) = w \cdot v_{i,d}(t) + \sum_{k=1}^{m} c_k \cdot r_k \cdot (dominant_solution_{i,k,d} - x_{i,d}(t))$$

Where $dominant_solution_{i,k,d}$ is the position of particle i in the objective k dimension that dominates its solution in the dth dimension. The improved multi-objective particle swarm algorithm is able to find the pareto optimal solution set for the logistics and transportation vehicle scheduling problem, thus providing the decision maker with multiple choices of optimal scheduling solutions that consider multiple objectives in a balanced manner [24].

V. Technical Framework

The overall architecture of the system is designed around building an advanced and efficient solution for monitoring and scheduling of logistics transportation vehicles, taking full advantage of Internet of Things (IoT) technology and cloud computing, aiming to comprehensively improve the overall intelligence level of the logistics industry chain. The following is a description of the overall framework of the system after refinement:

In the data collection layer, this paper deploys a complete set of on-board IoT devices and sensor components. The core role of this layer is to capture rich operational data in real-time, covering various status information of the vehicle itself, such as vehicle position, speed, running status, etc.; in addition, it also includes cargo status data, such as temperature, humidity and other environmental parameters, as well as identification and tracking of cargo with the help of rfid tags. Various types of vehicle terminals, such as high-precision GPS locators, temperature and humidity sensors, vehicle cameras (which can realize safety monitoring or driver behavior analysis) and load sensors and other multifaceted equipment work together to weave a tight and detailed data collection network [25].

As data is continuously generated from the first layer, the second network communication layer acts as a bridge to efficiently and reliably transmit this real-time data to the data center. This process relies on the power of modern communication technologies, including but not limited to 4g/5g mobile networks, satellite communications, Wi-Fi, and even lpwan technology for long-distance, low-power scenarios, to ensure seamless data transmission, whether it's on city streets or in remote areas [26].

After data transmission to the data center, this paper step into the third layer of the system - data storage and processing layer. This layer mainly relies on a powerful and stable cloud computing platform, through the server cluster to build a "Data warehouse". That can accommodate massive real-time monitoring data. Advanced cloud storage technology is utilized to ensure secure data storage and on-demand expansion, while significantly improving data access efficiency. In addition, a specialized big data processing module is set up to perform a series of cleaning, integration and pre-processing operations on the raw data received, and tools such as hadoop and spark are used to realize rapid analysis and mining of large-scale data using distributed computing frameworks [27].

After the data has been effectively processed, it comes to the fourth layer of the system - the intelligent analysis and decision-making layer. This layer includes a number of key modules, in which the real-time monitoring module, with accurate real-time data, can not only track the location of the vehicle in real time, but also reproduce the vehicle's driving trajectory, as well as timely identification of irregular or abnormal driving behavior. The intelligent prediction module utilizes the powerful prediction capability of machine learning algorithms to make precise and forward-looking assessments of vehicle arrival time, dynamic changes in road conditions, and even potential risks of mechanical failure. Intelligent scheduling module is the intelligent core of the whole system, which makes use of optimization methods such as particle swarm algorithm to make fine and flexible intelligent scheduling decisions based on real-time data and prediction results, taking into account the reasonable allocation of capacity resources, the degree of matching of order demand and the constraints of route planning, so as to realize the maximum efficiency of the capacity [28, 29].

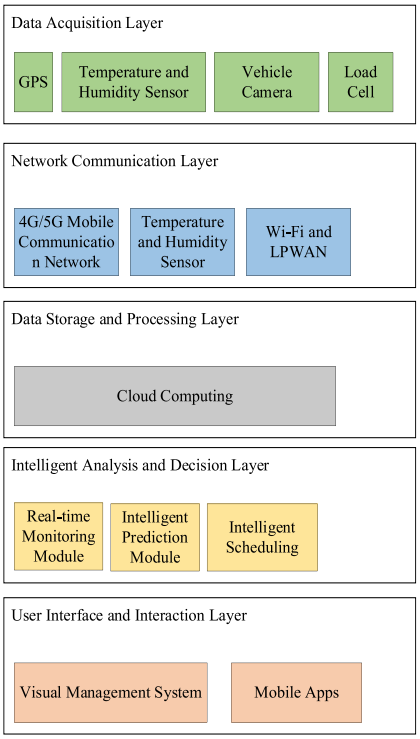


Fig. 3. Logistics transportation vehicle monitoring and scheduling solution.

The last layer, the user interface and interaction layer, is the key interface between the system and people. This layer contains two key parts: Visualization management system and mobile application. The visualization management system provides logistics managers with rich and vivid vehicle dynamic information display, dispatch result query and all kinds of business statistics report through the intuitive interface of web terminal or mobile terminal, and also supports remote control and real-time issuance of dispatching instructions [30].

In summary, the system builds a complete information link from bottom to top, from data acquisition, transmission, storage and processing, to intelligent analysis and decision-making, until the final human-computer interaction, forming a set of highly integrated and fully functional solutions for monitoring and scheduling of logistics and transportation vehicles, whose process is specifically shown in Fig. 3. This set of solutions effectively combines the Internet of things technology and cloud computing closely, and strongly promotes the intelligent process of the logistics industry.

VI. Experimental Research on Monitoring and Scheduling Systems for Logistics Transportation Vehicles Based on Internet of Things and Cloud Computing

A. Experimental Design and Baseline Modeling

In this chapter, this paper provides an in-depth exploration of the Internet of Things (IoT) and cloud computing based monitoring and scheduling system for logistics and transportation vehicles and provides an exhaustive comparative analysis of its performance with eight different baseline models, which include a fixed route model, a priority assignment model, a proximity principle scheduling model, a capacity matching model, a total trip minimization model, a fuel consumption minimization model, reactive scheduling model based on GPS real-time location monitoring and statistical analysis model. Among them, is the fixed route model**:** This model performs tasks according to a preset route without considering changes in real-time road conditions? Priority assignment model: The transportation tasks are assigned according to the urgency or importance of the goods. Proximity principle scheduling model: The nearest vehicle is selected to perform the task in order to reduce the waiting time. Capacity matching model: Matching based on cargo size and vehicle capacity to improve loading efficiency. Total trip minimization model: It aims to reduce the total distance traveled by vehicles. Fuel consumption minimization model: Optimizes routes to reduce fuel consumption and costs.

B. Experimental Environment and Assessment Indicators

Conducted in an environment equipped with cutting-edge IoT facilities and an efficient cloud computing platform, the experiment utilizes real-world logistics and transportation datasets to comprehensively evaluate the models against six key evaluation metrics - real-time monitoring accuracy, dispatch efficiency (including vehicle utilization, on-time rate, and empty rate), arrival time prediction accuracy (measured by RMSE and MAE), road condition change prediction accuracy (evaluated by RMSE), breakdown risk prediction accuracy (judged using MAE), and monitoring response latency. (Measured by RMSE and MAE), road condition change prediction accuracy (evaluated by RMSE), failure risk prediction accuracy (judged by MAE), and monitoring response latency - are comprehensively evaluated for each model. Among them, real-time monitoring accuracy: Assesses the accuracy and reliability of system monitoring data. Dispatch efficiency: Includes vehicle utilization rate, on-time rate, and empty rate to measure the overall efficiency of the dispatch system. Arrival time prediction accuracy: Assesses the accuracy of prediction using root mean square error (RMSE) and mean absolute error (MAE).

The real logistics and transportation dataset used in this paper is "Logistics Operation Insights Dataset", which is publicly published on Kaggle platform (kaggle.com/datasets/logisticsoperation/real-world-logistics-data) and covers multiple dimensions such as cargo information (such as name, quantity, weight, volume), transportation information (including transportation mode, freight, transportation time), delivery details (delivery address, consignee information), storage status and inventory changes. These data directly come from real logistics business operations, aiming to improve transportation efficiency, optimize cost control, intelligently plan transportation routes and refine inventory management through detailed analysis, providing powerful support for logistics enterprises to realize management optimization and intelligent decision-making with data insight.

During the experimental design and baseline modeling phase, this paper not only selected eight models covering a wide range of strategies for comparison, but also paid special attention to the closeness of the experimental setup to ensure the validity and universality of the results. To ensure the authenticity and comprehensiveness of the data, the logistics and transportation datasets used cover multiple dimensions, including but not limited to historical transportation routes, vehicle performance data, real-time traffic information, weather condition records and incident logs. Data preprocessing steps include data cleansing, outlier removal, missing value imputation, and data normalization to ensure that all models are evaluated based on consistent and high quality data.

When building the Internet of Things scheduling system, this paper make full use of the real-time sensing capability of Internet of Things devices and the massive data processing capability of cloud computing platforms. IoT devices installed on vehicles continuously collect vehicle status, cargo information and environmental data, and upload the data to the cloud through a stable wireless communication link. The cloud server integrates multi-source data using advanced data fusion algorithms and performs in-depth analysis through machine learning models to support vehicle scheduling decisions. The system designs a dynamic adjustment mechanism that can quickly re-plan the optimal path according to real-time road conditions, vehicle conditions and customer demand changes, thereby maximizing transportation efficiency and reducing costs while ensuring timeliness.

During the experiment, this paper pay special attention to the real-time response ability of the system. In the simulation

scheduling test, the IoT scheduling system demonstrated excellent performance, and its monitoring response latency was much lower than other models, ensuring that scheduling instructions could be quickly communicated to vehicle drivers to effectively respond to emergencies. In addition, the built-in fault prediction module of the system can warn potential faults in advance by analyzing abnormal patterns in vehicle operation data, which is verified by the mae index of fault risk prediction accuracy in Table V. The low score of 0.15 of Internet of Things dispatching system highlights its advantages in preventive maintenance.

It is worth noting that in the experiment, this paper also implemented a series of stress tests, simulating complex scenarios such as peak logistics demand surge, route change caused by extreme weather, and temporary emergency task insertion to verify the stability and flexibility of the IoT scheduling system. The results show that the system can maintain a high level of scheduling efficiency and accuracy, vehicle utilization and punctuality remain high, and empty rate remains low, even in a highly stressed logistics environment, which proves the effectiveness and robustness of the system design.

To sum up, through comprehensive experimental design and detailed performance evaluation, this study not only verifies the superiority of logistics transportation vehicle monitoring and scheduling system based on Internet of Things and cloud computing, but also reveals its specific performance in different application scenarios, providing powerful technical support and practical reference for intelligent upgrading of logistics industry. Future work will further explore how intelligent the system can be, such as by integrating more advanced prediction algorithms and optimization strategies, as well as enhancing the system's ability to adapt to uncertainties and external disturbances, to continuously improve the efficiency and reliability of logistics operations.

C. Experimental Results

The experimental process follows rigorous steps: First, this paper collects and preprocess a large amount of logistics transportation data to ensure data quality and consistency; second, this paper simulates the scheduling of the data by using the above baseline model and recording the performance of each evaluation index; then, this paper simulate the scheduling by using the self-developed IoT scheduling system and record the relevant indexes; then, this paper analyze the differences between the models in different evaluation indexes by comparing and analyzing their advantages and disadvantages. Assessment indicators to reveal their advantages and disadvantages.

Table I lists the performance of different models in terms of real-time monitoring accuracy in terms of percentage. The IoT scheduling system demonstrates excellent performance in terms of accuracy and reliability of real-time monitoring data, reaching 98%.

Table II shows the performance of different models in terms of dispatching efficiency in terms of vehicle utilization, on-time performance, and idle rate. The IoT dispatch system achieves optimal results in all three key metrics, with vehicle utilization as high as 95%, on-time performance at 98%, and idling rate reduced to a minimum level of 5%.

TABLE I. COMPARISON OF REAL-TIME MONITORING ACCURACY

Model name	Real-time monitoring accuracy
Fixed route model	85%
Prioritization model	90%
Proximity principle dispatch model	88%
Capacity matching model	87%
Total travel minimization model	89%
Fuel consumption minimization model	92%
GPS real-time location monitoring	95%
Statistical analysis model	90%
Model name	Real-time monitoring accuracy

TABLE II. COMPARISON OF SCHEDULING EFFICIENCY

Model name	Vehicle utilization rate
Fixed route model	70%
Prioritization model	75%
Proximity principle dispatch model	78%
Capacity matching model	80%
Total travel minimization model	83%
Fuel consumption minimization model	85%
GPS real-time location monitoring	88%
Statistical analysis model	82%
IoT dispatch system	95%

Tables III and IV evaluate the accuracy of the arrival time prediction by each model through two statistical metrics, RMSE (root mean square error) and MAE (mean absolute error), respectively. The IoT scheduling system again outperforms the other baseline models in terms of prediction accuracy, with significantly lower values for both RMSE and MAE, indicating that it is more accurate in predicting arrival times.

TABLE III. COMPARISON OF ARRIVAL TIME PREDICTION ACCURACY (RMSE)

Model name	Time of arrival forecasting (RMSE)
Fixed route model	0.50
Prioritization model	0.45
Proximity principle dispatch model	0.40
Capacity matching model	0.42
Total travel minimization model	0.38
Fuel consumption minimization model	0.35
Gps real-time location monitoring	0.30
Statistical analysis model	0.40
IoT dispatch system	0.25

TABLE IV. COMPARISON OF ARRIVAL TIME PREDICTION ACCURACY (MAE)

Model name	Time of arrival forecast (mae)
Fixed route model	0.40
Prioritization model	0.35
Proximity principle dispatch model	0.30
Capacity matching model	0.32
Total travel minimization model	0.28
Fuel consumption minimization model	0.25
Gps real-time location monitoring	0.20
Statistical analysis model	0.30
IoT dispatch system	0.15

Table V reflects the accuracy of the models in predicting changes in road conditions, again evaluated in terms of RMSE. The IoT dispatch system still maintains its lead in predicting changes in road conditions, with a RMSE value of 0.30, which is lower than the other baseline models, reflecting its strong ability to analyze and respond to real-time data.

Fig. 4 compares the accuracy of the models in predicting the risk of failure through the MAE metric. The MAE value of the IoT dispatching system is 0.20, which is much lower than the other baseline models, indicating that the system is able to more accurately predict and prevent the risk of possible failures, thus reducing the possibility of operational disruptions. Through a series of exhaustive data analyses and table comparisons, the IoT and cloud computing-based logistics and transportation vehicle monitoring and scheduling system achieves excellent results in a number of core metrics, such as real-time monitoring accuracy, scheduling efficiency, and prediction accuracy, which highlights its great advantage over traditional baseline models.

TABLE V. COMPARISON OF ROAD CONDITION CHANGE PREDICTION ACCURACY (RMSE)

Model name	Road condition change prediction (RMSE)
Fixed route model	0.60
Prioritization model	0.55
Proximity principle dispatch model	0.50
Capacity matching model	0.52
Total travel minimization model	0.48
Fuel consumption minimization model	0.45
GPS real-time location monitoring	0.40
Statistical analysis model	0.50
IoT dispatch system	0.30

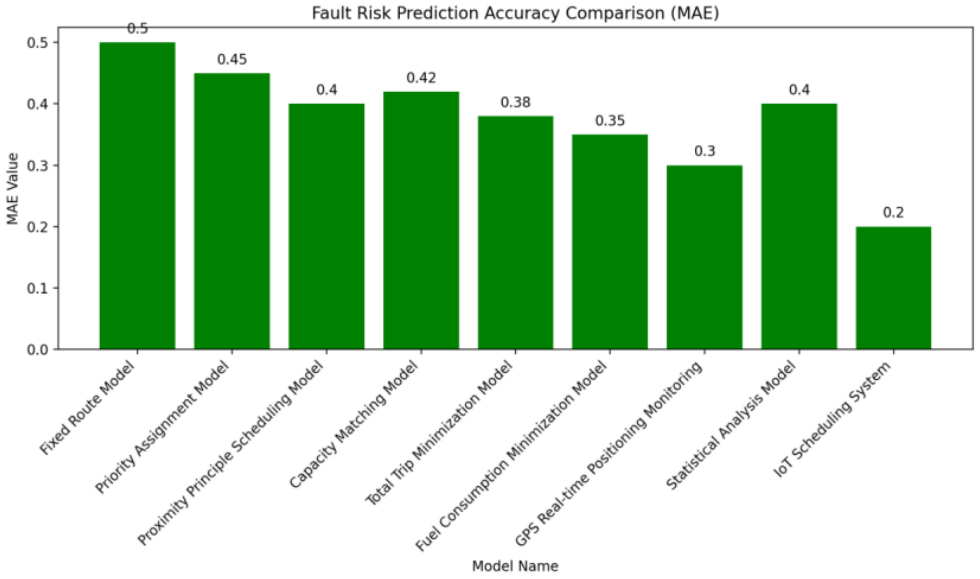


Fig. 4. Comparison of failure risk prediction accuracy (MAE).

Table VI reveals the performance differences between 10 different logistics models under their respective optimization scenarios by comparing and analyzing their performance on different data sets. In urban logistics scenarios, GPS real-time location monitoring systems can increase vehicle utilization to 85%, demonstrating their ability to efficiently utilize resources. In the field of rural logistics, the Internet of Things scheduling system stands out with 97% real-time monitoring accuracy, showing extremely high monitoring accuracy. For the arrival time prediction of long-distance freight, the RMSE of GPS system is only 0.34, indicating that its prediction accuracy is high. For seasonally varying failure risk predictions, IoT scheduling systems showed high accuracy of their predictions with an MAE of 0.23. Overall, IoT scheduling systems perform best across all datasets, with clear advantages in real-time monitoring and failure risk prediction. In contrast, the fixed-route model performs relatively poorly on each dataset, suggesting that we should choose the most appropriate model for a particular scenario in practice.

TABLE VI. COMPARATIVE ANALYSIS OF PERFORMANCE ON DIFFERENT DATA SETS

Model Name	Dataset A: Urban Logistics (Vehicle Utilization Rate)	Dataset B: Rural Logistics (Real-Time Monitoring Accuracy)	Dataset C: Long-Distance Freight (Arrival Time Prediction RMSE)	Dataset D: Seasonal Variability (Failure Risk Prediction MAE)
Fixed Route Model	68%	83%	0.52	0.35
Prioritization Model	72%	87%	0.48	0.33
Proximity Principle Dispatch Model	75%	86%	0.45	0.32
Capacity Matching Model	77%	85%	0.43	0.31
Total Travel Minimization Model	80%	88%	0.40	0.30
Fuel Consumption Minimization Model	82%	90%	0.37	0.28
GPS Real-Time Location Monitoring	85%	93%	0.34	0.26
Statistical Analysis Model	83%	91%	0.36	0.27
IoT Dispatch System	92%	97%	0.29	0.23

Through the above experimental process, this paper can comprehensively assess the advantages of the IoT and cloud computing-based scheduling system compared to the traditional model in terms of real-time monitoring, scheduling efficiency, prediction accuracy, and response speed. These advantages not only improve the efficiency and reliability of logistics transportation but also provide new ideas and solutions for future logistics transportation.

VII. CONCLUSION

Based on the urgent needs of modern intelligent traffic management and logistics and transportation industries, this study designs and implements an efficient vehicle monitoring and scheduling framework with the support of the Internet of Things (IoT) and cloud computing technologies. The research process covers the whole process from real-time vehicle detection to intelligent scheduling decision-making: Firstly, the advanced yolov5 model is used to implement real-time vehicle recognition on the video stream and generate high-precision position information in real-time; secondly, the deepsort algorithm is introduced to integrate real-time position data and historical trajectory information to ensure the accurate tracking of the vehicle's continuous motion state; finally, spatio-temporal convolutional network is applied to finally, the use of spatio-temporal convolutional network to deeply excavate the time series characteristics of the vehicle state greatly enhances the integrity of the state monitoring.

A. Innovation Points

1) A set of comprehensive vehicle monitoring models integrating yolov5, deepsort and spatio-temporal convolutional networks is constructed to realize the whole chain processing from position detection to state analysis.

2) An improved multi-objective particle swarm algorithm is proposed and successfully applied to the logistics transportation vehicle scheduling model, which improves the scientificity and effectiveness of scheduling decisions.

3) The scheduling system built using the Internet of Things and cloud computing technology has effectively improved real-time monitoring capability, scheduling efficiency and forecast accuracy, shortened response time and made significant progress compared with traditional models.

B. Deficiencies

1) The adaptability of the current system still needs to be further enhanced, especially the accuracy of vehicle detection and tracking in complex environments still needs to be improved.

2) Although spatio-temporal convolutional networks can better handle time-series data, when dealing with large-scale vehicle state data, the consumption of computational resources is large and the optimization space still exists.

3) Improved multi-objective particle swarm algorithms may require more diversified optimization strategies to cope with various uncertainties when facing extremely complex logistics scenarios.

In future work, this paper will focus on the following aspects: First, although this study successfully demonstrated the efficiency and accuracy of the logistics transportation vehicle monitoring and scheduling system based on IoT and cloud computing, as the technology continues to evolve, this paper will continue to explore and integrate the latest advances in emerging technologies such as 5G communication, edge computing and artificial intelligence algorithms to further optimize data transmission speed, improve system responsiveness and intelligent decision-making. Second, given the growing importance of environmental sustainability and green logistics, future research will aim to incorporate carbon

footprint calculations and environmental path optimization capabilities to ensure that systems not only improve logistics efficiency, but also support corporate sustainability goals and reduce the environmental impact of logistics activities. Finally, in order to promote and validate the system's broad applicability, this paper plan to conduct field pilots in logistics enterprises of different sizes and types, collect more diverse data, and conduct long-term follow-up studies to assess the long-term benefits and potential improvements of the system. Through interdisciplinary collaborations, bringing together operations management, information technology and social science perspectives, this paper will also delve into the socioeconomic impacts of technology implementation to ensure that technological advances benefit the entire logistics ecosystem.

VIII. FUNDING

(1) Fujian Provincial Department of Education Postgraduate Education Reform Project "Teaching Practice Exploration of 'Studio' System for Professional Master of Map Situation" (Project No. FBJG20190046); Undergraduate Education reform project of Fujian Normal University "Information Resource Construction" Curriculum Innovation and Practice Reform Exploration (Project No. I201812013), Closing item, main participant. (2) Fujian Normal University Postgraduate Education Reform Project: "Research on the Cultivation Mode of China's Professional Master in Mapping (MLIS)" (No. JZ160282) and Research topic on the Educational reform and Practice of Professional Master in Mapping (MLIS) under the "Studio" mode, Conclusion, main participants. (3) The 2022 National Social Science Fund Project "Research on the Development Strategy of Chinese Library Science Subject towards a Cultural Powerful Country" (Project No. 22BTQ035) is under research and is the main participant. (4) Key scientific research Project of colleges and universities in Henan Province: "Research on Rural E-commerce Live Streaming to Help Rural Revitalization and Development -- A Case study of Wenxian, Henan Province" (No. 22B790010), Closing item, main participants.

REFERENCES

[1] A. N. Assuncao, A. L. L. Aquino, R. Santos, R. L. M. Guimaraes, and R. A. R. Oliveira, "Vehicle driver monitoring through the statistical process control," Sensors, vol. 19, no. 14, 2019.

[2] V. Barth, R. de Oliveira, M. de Oliveira, and V. do Nascimento, "Vehicle speed monitoring using convolutional neural networks," IEEE Latin America Transactions, vol. 17, no. 6, pp. 1000-1008, 2019.

[3] M. M. Chen, H. T. Ding, M. M. Liu, Z. G. Zhu, D. D. Rui, Y. Chen, et al., "Vehicle operation status monitoring based on distributed acoustic sensor," Sensors, vol. 23, no. 21, 2023.

[4] G. D'urso, S. L. Smith, R. Mettu, T. Oksanen, and R. Fitch, "Multi-vehicle refill scheduling with queueing," Computers and Electronics in Agriculture, vol. 144, pp. 44-57, 2018.

[5] T. Filkin, N. Sliusar, M. Ritzkowski, and M. Huber-Humer, "Unmanned aerial vehicles for operational monitoring of landfills," Drones, vol. 5, no. 4, 2021.

[6] L. Gong, Y. Z. Li, and D. J. Xu, "Combinational scheduling model considering multiple vehicle sizes," Sustainability, vol. 11, no. 19, 2019.

[7] P. C. Guedes, D. Borenstein, M. S. Visentini, O. C. B. de Araújo, and A. F. K. Neto, "Vehicle scheduling problem with loss in bus ridership," Computers & Operations Research, vol. 111, pp. 230-242, 2019.

[8] J. Hartleb, M. Friedrich, and E. Richter, "Vehicle scheduling for on-demand vehicle fleets in macroscopic travel demand models," Transportation, vol. 49, no. 4, pp. 1133-1155, 2022.

[9] D. L. Iruthayaraj, R. R. P. Arockiam, and J. Subbaian, "Real-time indoor environment quality monitoring for vehicle cabin," Environmental Engineering And Management Journal, vol. 22, no. 11, pp. 1801-1811, 2023.

[10] N. A. Khan, N. Z. Jhanjhi, S. N. Brohi, R. S. A. Usmani, and A. Nayyar, "Smart traffic monitoring system using unmanned aerial vehicles (UAVs)," Computer Communications, vol. 157, pp. 434-443, 2020.

[11] S. U. Khan, N. Alam, S. U. Jan, and I. S. Koo, "IoT-enabled vehicle speed monitoring system," Electronics, vol. 11, no. 4, 2022.

[12] P. S. Klein and M. Schiffer, "Electric vehicle charge scheduling with flexible service operations," Transportation Science, vol. 57, no. 6, pp. 1605-1626, 2023.

[13] O. Kloster, C. Mannino, A. Riise, and P. Schittekat, "Scheduling vehicles with spatial conflicts," Transportation Science, vol. 56, no. 5, 2022.

[14] P. Kuwalek and G. Wiczynski, "Monitoring single-phase LV charging of electric vehicles," Sensors, vol. 23, no. 1, 2023.

[15] E. Lam, P. van Hentenryck, and P. Kilby, "Joint vehicle and crew routing and scheduling," Transportation Science, vol. 54, no. 2, pp. 488-511, 2020.

[16] H. Li, D. Son, and B. Jeong, "Electric vehicle charging scheduling with mobile charging stations," Journal of Cleaner Production, vol. 434, 2024.

[17] M. Li, L. Zhen, S. Wang, W. Y. Lv, and X. B. Qu, "Unmanned aerial vehicle scheduling problem for traffic monitoring," Computers & Industrial Engineering, vol. 122, pp. 15-23, 2018.

[18] T. Y. Li, H. Y. Liu, Z. W. Zhang, and D. L. Ding, "Shift scheduling strategy development for parallel hybrid construction vehicles," Journal of Central South University, vol. 26, no. 3, pp. 587-603, 2019.

[19] S. Limmer, J. Varga, and G. R. Raidl, "Large neighborhood search for electric vehicle fleet scheduling," Energies, vol. 16, no. 12, 2023.

[20] J. Liu, C. Bondiombouy, L. Mo, and P. Valduriez, "Two-phase scheduling for efficient vehicle sharing," IEEE Transactions on Intelligent Transportation Systems, vol. 23, no. 1, pp. 457-470, 2022.

[21] M. L. Liu, X. Z. Yao, J. Y. Huang, and C. Zhang, "Optimization of unmanned vehicle scheduling and order allocation," International Journal of Simulation Modelling, vol. 21, no. 3, pp. 477-488, 2022.

[22] S. Mancini and M. Gansterer, "Vehicle scheduling for rental-with-driver services," Transportation Research Part E-Logistics and Transportation Review, vol. 156, 2021.

[23] M. Milovancevic, J. S. Marinovic, J. Nikolic, A. Kitic, M. Shariati, N. T. Trung, et al., "UML diagrams for dynamical monitoring of rail vehicles," Physica A-Statistical Mechanics and its Applications, vol. 531, 2019.

[24] A. Mishra, S. H. Lee, D. Kim, and S. Kim, "In-cabin monitoring system for autonomous vehicles," Sensors, vol. 22, no. 12, 2022.

[25] R. Montemanni, M. dell'Amico, and A. Corsini, "Parallel drone scheduling vehicle routing problems with collective drones," Computers & Operations Research, vol. 163, 2024.

[26] M. A. Nguyen, G. T. H. Dang, M. H. Ha, and M. T. Pham, "The min-cost parallel drone scheduling vehicle routing problem," European Journal of Operational Research, vol. 299, no. 3, pp. 910-930, 2022.

[27] O. Pandithurai, M. Jawahar, S. Arockiaraj, and R. Bhavani, "IoT technology-based vehicle pollution monitoring and control," Global Nest Journal, vol. 25, no. 10, pp. 25-32, 2023.

[28] P. R. T. Peddinti, H. Puppala, and B. Kim, "Pavement monitoring using unmanned aerial vehicles: An overview," Journal of Transportation Engineering Part B-Pavements, vol. 149, no. 3, 2023.

[29] U. Shafi, A. Safi, A. R. Shahid, S. Ziauddin, and M. Q. Saleem, "Vehicle remote health monitoring and prognostic maintenance system," Journal of Advanced Transportation, 2018.

[30] H. Y. Shang, Y. P. Liu, W. X. Wu, and F. X. Zhao, "Multi-depot vehicle scheduling with multiple vehicle types on overlapped bus routes," Expert Systems with Applications, vol. 228, 2023.

Performance and Accuracy Research of the Large Language Models

Nicoleta Cristina GAITAN
Faculty of Electrical Engineering and Computer Science, Stefan Cel Mare University of Suceava, Suceava, Romania
Integrated Center for Research-Development and Innovation in Advanced Materials-Nanotechnologies and Distributed Systems for Fabrication and Control (MANSiD), Stefan cel Mare University, Suceava, Romania

***Abstract*—Starting with the end of 2022, there has been a massive global interest in Artificial Intelligence and, in particular, in the technology of large language models. These reduced the resolution of many problems dailies of varying degrees of complexity at a level accessible to every individual, whether it was an academic, business or social environment. A multitude of digital products have begun to use large language models to offer new functionalities such as intelligent messaging applications trained to respond efficiently depending on the specific parameters of a company, virtual assistants for programmers (GitHub Copilot), video call summarization functionality (Zoom), interpretation and extraction rapid drawing of conclusions from massive data (Big Data). These are just a few of the many uses of these technologies. Therefore, the general objective of this paper is the comparative analysis between three large language models such as ChatGPT, Gemini, and Llama3. Each model's strengths and constraints are analyzed, offering insights into their optimal use cases. This analysis provides a comprehensive understanding of the current state of large language models powered by deep learning, capable of executing various natural language processing (NLP) tasks, guiding future developments and applications in the field of artificial intelligence (AI).**

***Keywords*—Large language models; artificial intelligence; ChatGPT; natural language processing**

I. INTRODUCTION

Large language models (LLMs) are artificial intelligence (AI) systems capable of understanding and generating human language by processing vast amounts of text data [1] [2]. A large language model is a deep learning algorithm that can perform a variety of natural language processing (NLP) tasks.

Since the release of the first public version of ChatGPT in 2022 by the company OpenAI [3], there have been a lot of other versions of ChatGPT, but also other completely distinct models, as a result of a very close competition between the giants of the international technology industry (Big Tech): Meta (Facebook), Alphabet (Google), and Amazon, in an attempt to hold the key to winning and innovative technology.

A multitude of digital products have begun to use large language models to offer new functionalities such as intelligent messaging applications trained to respond efficiently according to the specific parameters of a company, virtual assistants for programmers (GitHub Copilot), summarization functionality of a video call (Zoom), interpreting and quickly drawing conclusions from massive data (Big Data). These are just a few of the many uses of this technology. Currently, most digital products offer at least some functionality based on artificial intelligence, which in reality is based on large language models.

If traditionally machines and computers were programmed by humans using different programming languages, in the context of this new technology, this can be done using natural language. Specialists named all these techniques "Prompt Engineering".

The general objective of this paper is the comparative analysis between three large language models: ChatGPT, Gemini and Llama 3. Initially, I set out to carry out this research on the accuracy and performance of large language models using the newest model released in April of this year: MetaAI. Unfortunately, this model is only available for use in the United States of America, plus a few countries in Asia and Africa. So, we replaced the newest MetaAI model with Llama 3, which is also part of the same company.

The specific objectives of the paper are the evaluations of large linguistic models according to certain selected criteria. Each criterion will be applied to each of the models separately, following their evaluation following the answers provided. In this paper, text responses representing the outputs of large language models will be evaluated. The image generation is not covered in this benchmark.

This paper presents a comparative analysis of three leading large language models: ChatGPT (GPT-4), Gemini, and Llama3. These models represent the forefront of natural language processing (NLP) advancements, each showcasing unique architectural designs, training paradigms, and application capabilities. ChatGPT, developed by OpenAI, utilizes an extensive Transformer-based architecture optimized for conversational tasks and general-purpose language understanding. Gemini, from Google DeepMind, integrates sophisticated contextual understanding with Google's vast data resources, excelling in multilingual and domain-specific applications. Llama3, created by Meta, prioritizes computational efficiency while maintaining high performance in text generation and real-time interaction tasks. ChatGPT is noted for its versatility in various NLP tasks and robust conversational abilities. Gemini leverages proprietary data for enhanced contextual reasoning and problem-solving. Llama3 stands out for its resource-efficient design, making it suitable for lightweight applications.

This paper presents a study about performance and accuracy research of the most used large language models. Section II

reviews state of the art of these models, while Section III describes the large language models, and evaluation criteria is given in Section IV. Finally, discussion and conclusion is given in Section V and Section VI respectively.

II. Related Works

In 1950, the first experiments with neural networks and neural information processing systems were carried out to allow computers to process natural language. Researchers at the Georgetown University and IBM have created a system that would be able to automatically translate phrases from Russian to English. Being a real demonstration of machine translation, it can be said that research in this field started from there.

The notion of a large language model was first introduced with the creation of Eliza in the 1960s (see Fig. 1), which was the world's first chatbot [2]. Designed by MIT researcher Joseph Weizenbaum, the Eliza chatbot marked the beginning of research into natural language processing (NLP), providing the basis for future more complex language models.

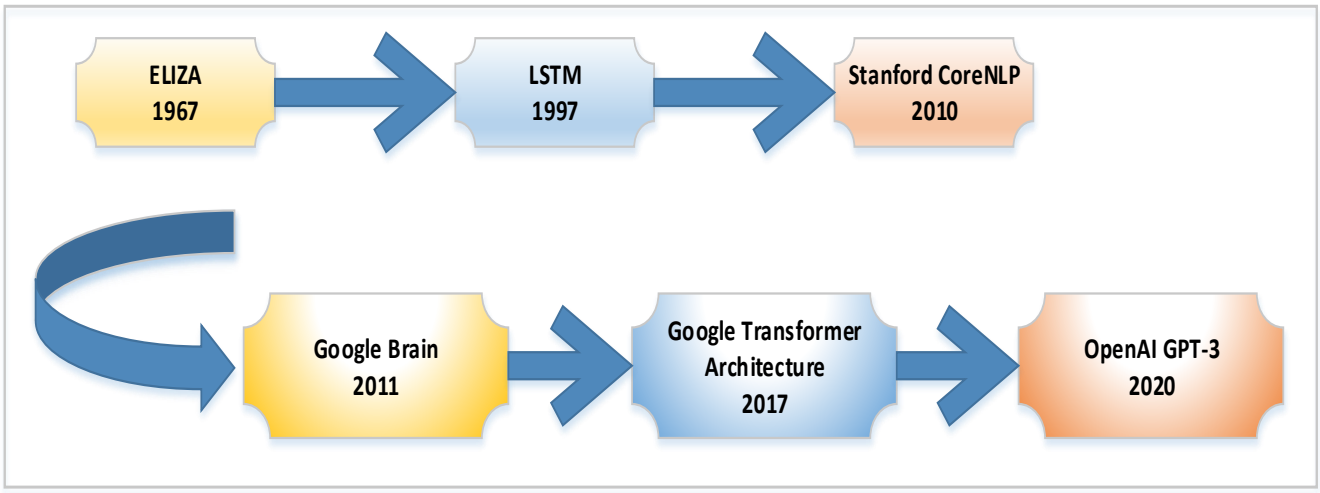


Fig. 1. The history of large language models.

In the year 1997, deeper and more complex neural networks appeared that managed larger amounts of data, based on Long Short-Term Memory (LSTM) networks.

Three years later, in 2010, Stanford's CoreNLP suite was introduced, allowing developers to perform sentiment analysis and named entity recognition.

Later, in 2011, a smaller version of Google Brain appeared with advanced features such as word embedding, which allowed NLP systems to gain a clearer sense of context.

Google researchers introduced a 340 million parameter bidirectional model, BERT (Bidirectional Encoder Representations from Transformers), in 2019, the third largest model of its kind. This model was able to understand the relationships between words by being pre-trained by self-supervised learning on a very large data set. Thus, BERT became the basic tool for natural language processing tasks, and more than that, it was behind every English query made through Google Search.

OpenAI brought to the scene the GPT-2 (Generative Pre-trained Transformer) transformer model with 1.5 billion parameters, and in 2020 release the GPT-3 with 175 billion parameters.

Fig. 1 shows those models that dictated the standard for large language models and formed the basis of ChatGPT that was released in November 2022 [3] [5]. Most recently, OpenAI introduced GPT-4, which is estimated to have one trillion parameters - of five times larger than GPT-3 and approximately 3,000 times larger than BERT when it first appeared.

III. Large Language Models

A. Selection of Large Language Models

1) Chat GPT4: What most captured the public's imagination was OpenAI's launch of ChatGPT, which reached 100 million users in just two months, making it the fastest-growing consumer app in history. This was the main reason why we chose this model for the comparative analysis.

The largest model in OpenAI's GPT series, Generative Pre-trained Transformer 4 was released in 2023 [3] [5]. Like other large language models, it is a transformer-based model. The main differentiating factor between the versions is that the number of parameters is more than 170 trillion. It can easily process and generate both language and images, and it can analyze data and produce graphs and charts. It features a system message that allows users to specify the tone of voice and task. It also powers Microsoft's Bing AI chatbot.

2) Gemini: Gemini [6] entered the generative artificial intelligence market in late 2023 after the rebranding of BARD, being the improved version of which initially faced several problems. Trained by Google, one of the world's largest technology companies and a giant in the field of artificial intelligence, Gemini wants to be positioned as a competitor to ChatGPT. This is the main motivation for choosing Gemini for comparison and evaluation. Is Gemini a competitor to ChatGPT? We will find out in what follows.

3) Llama 3: The reason why I chose the Llama 3 model is the fact that it was launched in April 2024 [7]. An important

leap compared to the Llama 2 is the fact that the Llama 3 comes with two variants of parameters, 8B and 70B.

One way that Llama 3 differs from other big language models such as Gemini and GPT is that Meta has released the model as open source - meaning that it is available for research as well as commercial purposes. However, the license is customized and requires users to follow specific regulations to avoid misuse. The Llama 3 models are available on AWS, Databricks, Google Cloud, Hugging Face, Kaggle, IBM Watson X, Microsoft Azure, NVIDIA NIM, and Snowflake.

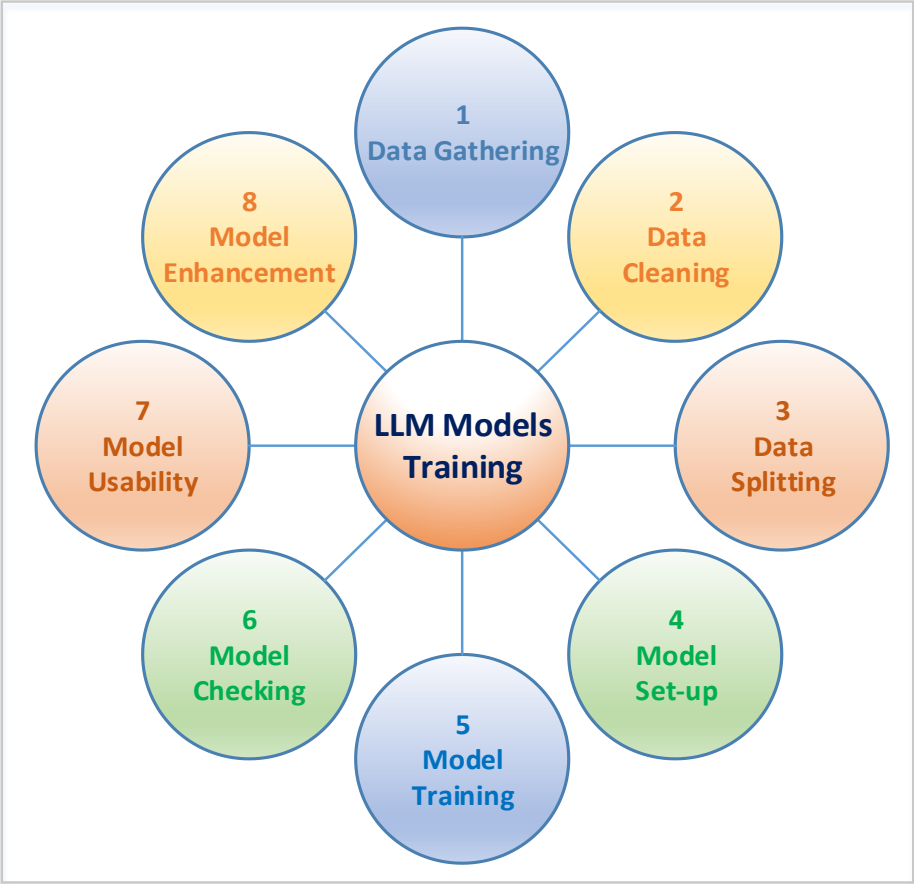


Fig. 2. Large language model training.

Llama 3 achieved compliance with major data protection standards, reducing data breaches in tested environments by over 40%. Understanding the critical importance of data security, Llama 3 incorporates enhanced privacy features that ensure the safe management of user data. Training Large Language Models.

An example of training large language models [9] is presented in Fig. 2.

B. How Large Language Models Work?

The large language models work by continuously predicting the next token, (about three-quarters of a word), starting from what was in the request.

Tokens are the basic elements of text in natural language processing. Each new token is selected according to the probability of appearing next, with a random element, controlled by the temperature parameter. The temperature parameter of large language models influences the output of the language model. Thus, it is determined if the result is more random and creative or more predictable. A higher temperature will result in a lower probability, i.e., more creative results. A lower temperature will output a higher probability, meaning more predictable results. This means that, through adjustment, the fine modelling of the model's performance will be obtained.

The large language models are based on a class of deep learning architectures called transform networks. A transformer model is a neural network that learns context and meaning by looking for relationships in sequential data.

The architecture of a transformer was introduced in a paper in the field of natural language processing (NLP). This work is called "Attention is All You Need" [4]. Because of their unique design and efficiency, transformers have become the foundation of natural language processing tasks. A transformer has an encoder-decoder type structure. The best performing models connect the encoder and decoder through an attention mechanism.

C. What is Normalization in the Context of Large Language Models?

The normalization [8] is a crucial step in the operation of large language models. Normalization helps ensure that the model will efficiently process and understand the data, being

used in both preprocessing and training and inference. In the context of large language models, inference refers to the process of obtaining an answer from the trained model by querying or asking the user.

In the preprocessing step, normalization is used to standardize and scale the input data. This helps reduce redundancy and ensure that the data is in a format that the model can easily understand. In the training step, normalization is used to ensure that the model is not biased towards any particular feature or data point. In the inference step, normalization is used to ensure that the output of the model is in a standard format.

D. What are Activation Functions?

Activation functions play a critical role in determining the complexity and capacity of neural networks [10]. The activation function in a neural network is a mathematical function that determines a neuron's output. The neuron takes as an argument the weighted sum of the inputs and the bias and produces an output that is used as input for the next layer in the network. The activation function is responsible for transforming the input signal into an output signal, and this output is what decides whether a particular neuron will be activated or not.

The training of large linguistic models involves going through the eight stages shown in Fig. 2, such as:

1) Data gathering: Training a large language model starts with collecting a huge amount of unstructured text data, data that comes from various sources such as books, web pages, articles or social networking platforms.

2) Data cleaning: This process is called preprocessing. The collected data must be cleaned and prepared for training. This involves removing unwanted characters, breaking the text into smaller parts called tokens, and putting it into a format that the model can work with.

3) Data splitting: The previously cleaned data is split into two sets. One set, the training data, will be used to train the model. The other set, the validation data, will be used later to test the performance of the model.

4) Model set-up: In this step, the structure of the large language model, known as the architecture, is defined. This involves selecting the type of neural network and deciding on various parameters such as the number of layers and hidden units in the network.

5) Model training: Now the actual training begins. The large language model learns by analyzing training data, making predictions based on what it has learned so far, and then adjusting internal parameters to reduce the gap between its predictions and the actual data.

6) Model checking: The learning of the large language model is verified using the validation data. This allows viewing the model's operation and modifying its settings for better performance.

7) Model usability: After training and evaluation, the large language model is ready for use and can be integrated into applications.

8) Model enhancement: By using updated data or adjusting settings based on real-world feedback and usage, the large language model can be further refined over time.

This training process requires massive computing resources such as powerful processing units and large storage as well as specialized machine learning knowledge. The amount of unstructured data that the model will learn through self-supervised learning starts at a size of at least 1000 GB, having billions of parameters. A parameter in the context of large linguistic models defines the behaviour of the artificial intelligence model, being used in predictions. An infrastructure with multiple GPUs is essential to be able to train such large models. Purchasing such a large number of GPUs is not feasible for most organizations. Even OpenAI, the creator of the ChatGPT model, did not train its models on its own infrastructure, but relied on Microsoft's Azure cloud platform. In 2019, Microsoft invested $1 billion in OpenAI, and it is estimated that much of the money was spent on training large language models on Azure cloud resources.

Although incredible advances have been made with the introduction of large language models, it is important to understand the limits of these models to avoid potential pitfalls and ensure responsible use. Misinformation, malware, discriminatory content, plagiarism and information that is untrue can lead to unwanted or dangerous results.

E. What is a Hallucination in the Context of Large Language Models?

Hallucinations are the occurrences where large linguistic patterns produce coherent and grammatically correct but incorrect or nonsensical results. In this context, the term "hallucination" means the generation of false or misleading information. The main factors causing hallucinations are training data limitations, model biases, inherent complexity of language, vague questions, and semantic gaps.

IV. Evaluation Criteria

Before evaluating the models and giving them instructions, an important aspect in the field of large language models should be mentioned, namely prompting. Prompt engineering is the process of discovering cues that reliably provide useful or desired results. A request is a natural language text that describes the task that an artificial intelligence model should perform. The prompt serves as a set of instructions that the model uses to predict the desired response.

These criteria come in the form of questions, prompts, or sets of instructions that I provide as input to large language models. There are many ways to ask an AI model to do a task, and even slight changes can make a big difference. The large language models work by continuously predicting the next token (about three-quarters of a word) from what was in the request provided as input. Each new token is selected based on the probability of appearing next, with a random element.

The large language models are trained on essentially the entire text of the Internet and then later tuned to provide useful answers. Average requests will return average responses, causing some people to be disappointed when their results don't live up to expectations. What is entered in the request changes

the probability of each generated word, which matters a lot for the results that will be obtained. These models have seen the best and worst of what humans have produced and are capable of emulating almost anything if you know how to ask.

A. *Criterion 1: Mathematics Problem, Topic Taken from the National Mathematics Olympiad*

Request 1**:** "Imagine you are a math teacher and you have the following 5th grade math problem to solve: A store sold 235 robots in the 12 months of a year. Every month either 16, 20 or 25 robots were sold. Determine the number of months in which exactly 20 robots were sold. Provide a detailed explanation of the solution and the correct solution to the problem."

ChatGPT Response, like in Fig. 3.

Rating from ChatGPT – ChatGPT chose a solution method by which it tested every possible variant for the value of x. It's not wrong, but if we have a higher number of attempts, how will it perform? However, it provided the correct solution. He also explained in detail how to solve the problem, so he complied with the requirements in the request.

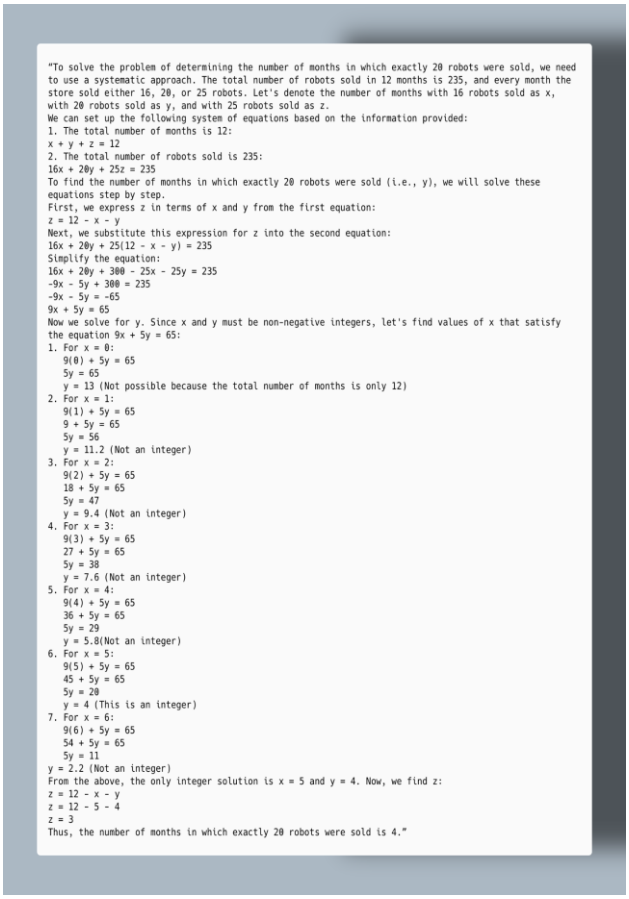

"To solve the problem of determining the number of months in which exactly 20 robots were sold, we need to use a systematic approach. The total number of robots sold in 12 months is 235, and every month the store sold either 16, 20, or 25 robots. Let's denote the number of months with 16 robots sold as x, with 20 robots sold as y, and with 25 robots sold as z.
We can set up the following system of equations based on the information provided:
1. The total number of months is 12:
x + y + z = 12
2. The total number of robots sold is 235:
16x + 20y + 25z = 235
To find the number of months in which exactly 20 robots were sold (i.e., y), we will solve these equations step by step.
First, we express z in terms of x and y from the first equation:
z = 12 - x - y
Next, we substitute this expression for z into the second equation:
16x + 20y + 25(12 - x - y) = 235
Simplify the equation:
16x + 20y + 300 - 25x - 25y = 235
-9x - 5y + 300 = 235
-9x - 5y = -65
9x + 5y = 65
Now we solve for y. Since x and y must be non-negative integers, let's find values of x that satisfy the equation 9x + 5y = 65:
1. For x = 0:
9(0) + 5y = 65
5y = 65
y = 13 (Not possible because the total number of months is only 12)
2. For x = 1:
9(1) + 5y = 65
9 + 5y = 65
5y = 56
y = 11.2 (Not an integer)
3. For x = 2:
9(2) + 5y = 65
18 + 5y = 65
5y = 47
y = 9.4 (Not an integer)
4. For x = 3:
9(3) + 5y = 65
27 + 5y = 65
5y = 38
y = 7.6 (Not an integer)
5. For x = 4:
9(4) + 5y = 65
36 + 5y = 65
5y = 29
y = 5.8(Not an integer)
6. For x = 5:
9(5) + 5y = 65
45 + 5y = 65
5y = 20
y = 4 (This is an integer)
7. For x = 6:
9(6) + 5y = 65
54 + 5y = 65
5y = 11
y = 2.2 (Not an integer)
From the above, the only integer solution is x = 5 and y = 4. Now, we find z:
z = 12 - x - y
z = 12 - 5 - 4
z = 3
Thus, the number of months in which exactly 20 robots were sold is 4."

Fig. 3. ChatGPT response.

Gemini Response, like in Fig. 4.

Rating from Gemini – Gemini has a particular style of detailing and structuring problem-solving methods, offering a very explicit and logical description of the solution steps. It seems that he is not able to solve equations, the solution offered by Gemini not being correct.

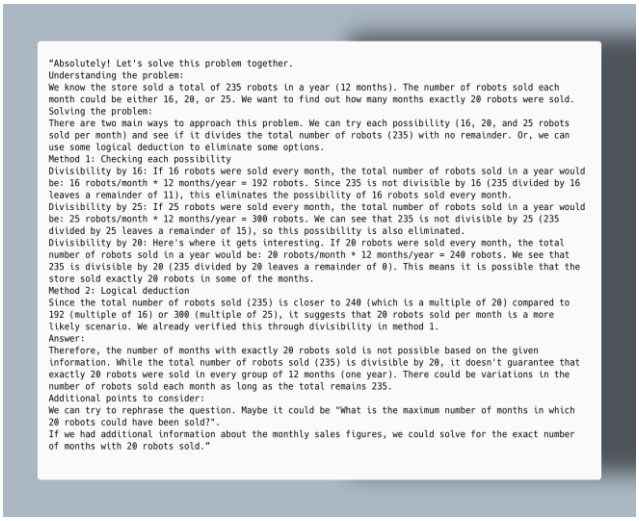

"Absolutely! Let's solve this problem together.
Understanding the problem:
We know the store sold a total of 235 robots in a year (12 months). The number of robots sold each month could be either 16, 20, or 25. We want to find out how many months exactly 20 robots were sold.
Solving the problem:
There are two main ways to approach this problem. We can try each possibility (16, 20, and 25 robots sold per month) and see if it divides the total number of robots (235) with no remainder. Or, we can use some logical deduction to eliminate some options.
Method 1: Checking each possibility
Divisibility by 16: If 16 robots were sold every month, the total number of robots sold in a year would be: 16 robots/month * 12 months/year = 192 robots. Since 235 is not divisible by 16 (235 divided by 16 leaves a remainder of 11), this eliminates the possibility of 16 robots sold every month.
Divisibility by 25: If 25 robots were sold every month, the total number of robots sold in a year would be: 25 robots/month * 12 months/year = 300 robots. We can see that 235 is not divisible by 25 (235 divided by 25 leaves a remainder of 15), so this possibility is also eliminated.
Divisibility by 20: Here's where it gets interesting. If 20 robots were sold every month, the total number of robots sold in a year would be: 20 robots/month * 12 months/year = 240 robots. We see that 235 is divisible by 20 (235 divided by 20 leaves a remainder of 0). This means it is possible that the store sold exactly 20 robots in some of the months.
Method 2: Logical deduction
Since the total number of robots sold (235) is closer to 240 (which is a multiple of 20) compared to 192 (multiple of 16) or 300 (multiple of 25), it suggests that 20 robots sold per month is a more likely scenario. We already verified this through divisibility in method 1.
Answer:
Therefore, the number of months with exactly 20 robots sold is not possible based on the given information. While the total number of robots sold (235) is divisible by 20, it doesn't guarantee that exactly 20 robots were sold in every group of 12 months (one year). There could be variations in the number of robots sold each month as long as the total remains 235.
Additional points to consider:
We can try to rephrase the question. Maybe it could be "What is the maximum number of months in which 20 robots could have been sold?".
If we had additional information about the monthly sales figures, we could solve for the exact number of months with 20 robots sold."

Fig. 4. Gemini response.

Llama3 Response, like in Fig. 5.

Rating from Llama3 – It seems that Llama 3 offered detailed explanations, logically structured the solution to the problem, however, the solution is not correct. The bottom line is that Llama 3 needs precise directions, but even so, it's sometimes unstable. The less positive part in the case of Llama 3 is that you cannot provide feedback, it does not have this feature implemented. So far, Llama3 has provided correct answers, even if in a different language than the received request. This response is evidence of the model's first hallucination.


"What a great problem! Let's dive into it step by step.
Step 1: Understand the problem We know that a store sold 235 robots in 12 months, and every month, they sold either 16, 20, or 25 robots. Our goal is to find the number of months in which exactly 20 robots were sold.
Step 2: Identify the possibilities Let's think about the possible combinations of robots sold each month:
16 robots/month: 12 months x 16 robots/month = 192 robots (short of 235)
20 robots/month: 12 months x 20 robots/month = 240 robots (more than 235)
25 robots/month: 12 months x 25 robots/month = 300 robots (way more than 235)
Since we need to reach 235 robots, we know that we can't have all months with the same number of robots sold. We need to mix and match the possibilities.
Step 3: Create a system of equations Let's introduce variables to represent the number of months for each type of robot sale:
Let x be the number of months with 16 robots sold.
Let y be the number of months with 20 robots sold.
Let z be the number of months with 25 robots sold.
We know that the total number of months is 12, so we can write an equation:
x + y + z = 12 ... (Equation 1)
We also know that the total number of robots sold is 235. We can set up another equation:
16x + 20y + 25z = 235 ... (Equation 2)
Step 4: Solve the system of equations We can solve Equation 1 for x: x = 12 - y - z
Substitute this expression for x into Equation 2:
16(12 - y - z) + 20y + 25z = 235
Expand and simplify:
192 - 16y - 16z + 20y + 25z = 235
Combine like terms:
4y + 9z = 43
Now, we need to find integer solutions for y and z. Since y and z represent months, they must be whole numbers.
Step 5: Find the solution After some trial and error, we find a solution:
y = 3 (months with 20 robots sold) z = 3 (months with 25 robots sold)
Substitute these values into Equation 1 to find x:
x + 3 + 3 = 12 => x = 6 (months with 16 robots sold)
Step 6: Verify the solution Let's calculate the total number of robots sold:
6 months x 16 robots/month = 96 robots
3 months x 20 robots/month = 60 robots
3 months x 25 robots/month = 75 robots
Total robots = 96 + 60 + 75 = 235
The solution checks out!
Answer: The store sold exactly 20 robots in 3 months."

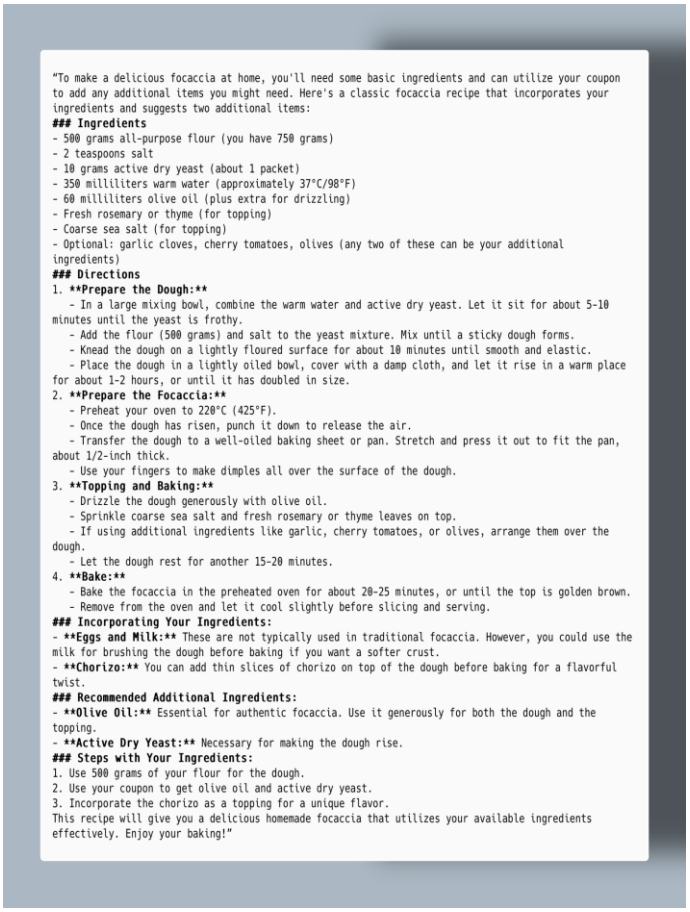
Fig. 5. Llama3 response.

B. *Creating a Culinary Recipe, Having a List of Ingredients Available*

Request 2: "I have the following ingredients: 500 milliliters of milk, six eggs, 750 grams of flour, 250 grams of chorizo and a coupon for two additional ingredients (it can be whatever I decide in whatever quantity). I want to make a focaccia but I

don't know which are the exact ingredients and the quantities for that. Please tell me a good recipe of focaccia to make in house."

ChatGPT Response**,** like in Fig. 6.



"To make a delicious focaccia at home, you'll need some basic ingredients and can utilize your coupon to add any additional items you might need. Here's a classic focaccia recipe that incorporates your ingredients and suggests two additional items:

Ingredients

- 500 grams all-purpose flour (you have 750 grams)
- 2 teaspoons salt
- 10 grams active dry yeast (about 1 packet)
- 350 milliliters warm water (approximately 37°C/98°F)
- 60 milliliters olive oil (plus extra for drizzling)
- Fresh rosemary or thyme (for topping)
- Coarse sea salt (for topping)
- Optional: garlic cloves, cherry tomatoes, olives (any two of these can be your additional ingredients)

Directions

1. ****Prepare the Dough:****
   - In a large mixing bowl, combine the warm water and active dry yeast. Let it sit for about 5-10 minutes until the yeast is frothy.
   - Add the flour (500 grams) and salt to the yeast mixture. Mix until a sticky dough forms.
   - Knead the dough on a lightly floured surface for about 10 minutes until smooth and elastic.
   - Place the dough in a lightly oiled bowl, cover with a damp cloth, and let it rise in a warm place for about 1-2 hours, or until it has doubled in size.
2. ****Prepare the Focaccia:****
   - Preheat your oven to 220°C (425°F).
   - Once the dough has risen, punch it down to release the air.
   - Transfer the dough to a well-oiled baking sheet or pan. Stretch and press it out to fit the pan, about 1/2-inch thick.
   - Use your fingers to make dimples all over the surface of the dough.
3. ****Topping and Baking:****
   - Drizzle the dough generously with olive oil.
   - Sprinkle coarse sea salt and fresh rosemary or thyme leaves on top.
   - If using additional ingredients like garlic, cherry tomatoes, or olives, arrange them over the dough.
   - Let the dough rest for another 15-20 minutes.
4. ****Bake:****
   - Bake the focaccia in the preheated oven for about 20-25 minutes, or until the top is golden brown.
   - Remove from the oven and let it cool slightly before slicing and serving.

Incorporating Your Ingredients:

- ****Eggs and Milk:**** These are not typically used in traditional focaccia. However, you could use the milk for brushing the dough before baking if you want a softer crust.
- ****Chorizo:**** You can add thin slices of chorizo on top of the dough before baking for a flavorful twist.

Recommended Additional Ingredients:

- ****Olive Oil:**** Essential for authentic focaccia. Use it generously for both the dough and the topping.
- ****Active Dry Yeast:**** Necessary for making the dough rise.

Steps with Your Ingredients:

1. Use 500 grams of your flour for the dough.
2. Use your coupon to get olive oil and active dry yeast.
3. Incorporate the chorizo as a topping for a unique flavor.

This recipe will give you a delicious homemade focaccia that utilizes your available ingredients effectively. Enjoy your baking!"

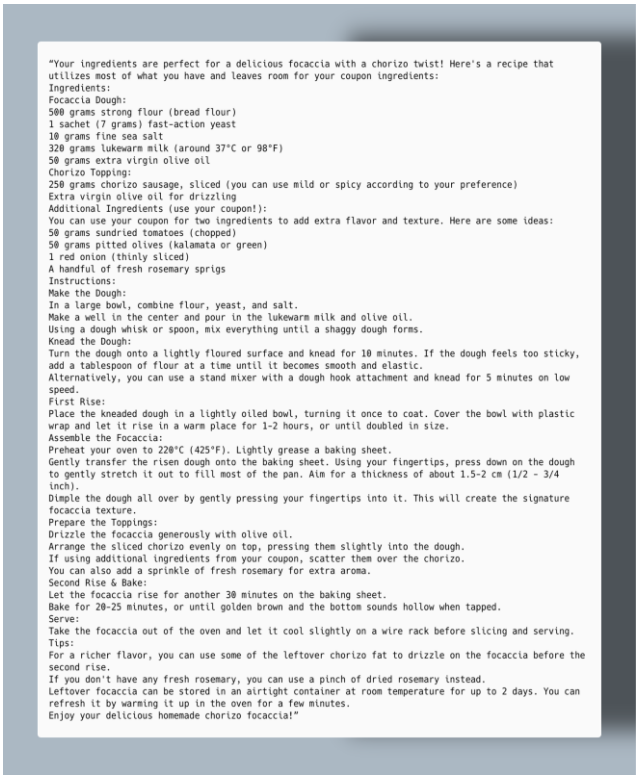
Fig. 6. ChatGPT response to request2.

Rating from ChatGPT – This is by far the best, most comprehensive and detailed answer ChatGPT has provided yet. Not only did he describe in detail every step for preparing and cooking the product, but he also made a number of correct suggestions regarding the coupon mentioned in the request: how I can use it and what missing ingredients from the list I can purchase with it.

Then, he mentioned that only 500g of the 750g of flour should be used for the proposed recipe, and he also realized that from the list of ingredients provided, two of them are not used for the focaccia recipe. Awesome, ChatGPT!

Gemini Response, like in Fig. 7.

Rating from Gemini – Gemini, this time had a good performance. The answer provided is correct, well structured, explained in detail. The ingredients used are the correct ones for a focaccia, as well as the method of preparation and cooking of the product. He also gave some tips at the end on how to improve the recipe, not bad at all, but not up to ChatGPT's level. Also, he didn't exactly understand the use of the coupon, the idea of the request was to purchase ingredients missing from the list with its help. However, on the whole it performed well the Gemini model, which until now has accustomed us to not very accurate answers.



"Your ingredients are perfect for a delicious focaccia with a chorizo twist! Here's a recipe that utilizes most of what you have and leaves room for your coupon ingredients:

Ingredients:

Focaccia Dough:

500 grams strong flour (bread flour)
1 sachet (7 grams) fast-action yeast
10 grams fine sea salt
320 grams lukewarm milk (around 37°C or 98°F)
50 grams extra virgin olive oil

Chorizo Topping:

250 grams chorizo sausage, sliced (you can use mild or spicy according to your preference)
Extra virgin olive oil for drizzling

Additional Ingredients (use your coupon!):

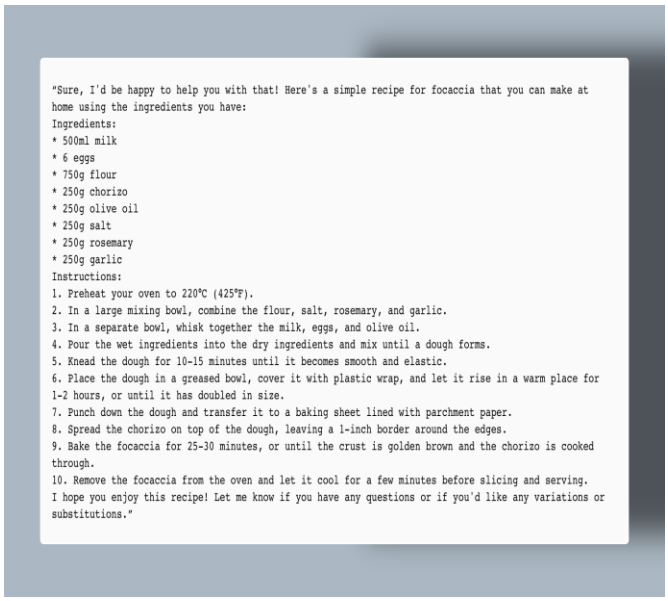
You can use your coupon for two ingredients to add extra flavor and texture. Here are some ideas:
50 grams sundried tomatoes (chopped)
50 grams pitted olives (kalamata or green)
1 red onion (thinly sliced)
A handful of fresh rosemary sprigs

Instructions:

Make the Dough:
In a large bowl, combine flour, yeast, and salt.
Make a well in the center and pour in the lukewarm milk and olive oil.
Using a dough whisk or spoon, mix everything until a shaggy dough forms.
Knead the Dough:
Turn the dough onto a lightly floured surface and knead for 10 minutes. If the dough feels too sticky, add a tablespoon of flour at a time until it becomes smooth and elastic.
Alternatively, you can use a stand mixer with a dough hook attachment and knead for 5 minutes on low speed.
First Rise:
Place the kneaded dough in a lightly oiled bowl, turning it once to coat. Cover the bowl with plastic wrap and let it rise in a warm place for 1-2 hours, or until doubled in size.
Assemble the Focaccia:
Preheat your oven to 220°C (425°F). Lightly grease a baking sheet.
Gently transfer the risen dough onto the baking sheet. Using your fingertips, press down on the dough to gently stretch it out to fill most of the pan. Aim for a thickness of about 1.5-2 cm (1/2 - 3/4 inch).
Dimple the dough all over by gently pressing your fingertips into it. This will create the signature focaccia texture.
Prepare the Toppings:
Drizzle the focaccia generously with olive oil.
Arrange the sliced chorizo evenly on top, pressing them slightly into the dough.
If using additional ingredients from your coupon, scatter them over the chorizo.
You can also add a sprinkle of fresh rosemary for extra aroma.
Second Rise & Bake:
Let the focaccia rise for another 30 minutes on the baking sheet.
Bake for 20-25 minutes, or until golden brown and the bottom sounds hollow when tapped.
Serve:
Take the focaccia out of the oven and let it cool slightly on a wire rack before slicing and serving.
Tips:
For a richer flavor, you can use some of the leftover chorizo fat to drizzle on the focaccia before the second rise.
If you don't have any fresh rosemary, you can use a pinch of dried rosemary instead.
Leftover focaccia can be stored in an airtight container at room temperature for up to 2 days. You can refresh it by warming it up in the oven for a few minutes.
Enjoy your delicious homemade chorizo focaccia!"

Fig. 7. Gemini response to request2.

Llama3 Response**,** like in Fig. 8.



"Sure, I'd be happy to help you with that! Here's a simple recipe for focaccia that you can make at home using the ingredients you have:

Ingredients:

* 500ml milk
* 6 eggs
* 750g flour
* 250g chorizo
* 250g olive oil
* 250g salt
* 250g rosemary
* 250g garlic

Instructions:

1. Preheat your oven to 220°C (425°F).
2. In a large mixing bowl, combine the flour, salt, rosemary, and garlic.
3. In a separate bowl, whisk together the milk, eggs, and olive oil.
4. Pour the wet ingredients into the dry ingredients and mix until a dough forms.
5. Knead the dough for 10-15 minutes until it becomes smooth and elastic.
6. Place the dough in a greased bowl, cover it with plastic wrap, and let it rise in a warm place for 1-2 hours, or until it has doubled in size.
7. Punch down the dough and transfer it to a baking sheet lined with parchment paper.
8. Spread the chorizo on top of the dough, leaving a 1-inch border around the edges.
9. Bake the focaccia for 25-30 minutes, or until the crust is golden brown and the chorizo is cooked through.
10. Remove the focaccia from the oven and let it cool for a few minutes before slicing and serving.

I hope you enjoy this recipe! Let me know if you have any questions or if you'd like any variations or substitutions."

Fig. 8. Llama3 response to request2.

Rating from Llama3 – The Llama3 model used all the ingredients on the list plus some added, and invented a recipe. We could say that he was creative, but comparing the answer with that of the ChatGPT model or even Gemini, I think more that he did not understand exactly the requirement. He did not

mention anything about the coupon, and he also used two ingredients that are not part of the focaccia recipe, ingredients that, at least ChatGPT specified, should not be used. It cannot be compared on this criterion with the first two models, once again remaining in last place. From what I have analyzed so far, it seems that the Llama 3 is not a stable model.

V. DISCUSSIONS

Following the comparative analysis carried out in this paper, we have learned some important aspects about these large linguistic patterns. First, these models are not stable. Although it is true that they were trained on massive data sets, these models have difficulties in solving more complex problems, logic, etc. Also, another argument to support this conclusion is the fact that their answers differ from one request to another, so much so that the final result can be completely different, and here I refer in particular to the Llama3 model. Of the 3 models chosen for comparative analysis, Gemini and Llama 3 are the most unstable.

An important aspect that deserves to be mentioned and put into practice is the observance of the principles of prompt engineering. An essential principle is to assess the quality of models when they do not provide a correct or true answer. Thus, these large language models can improve their performance and accuracy in future responses. This is confirmed by the analysis where ChatGPT's response improved after giving feedback that it did not respect the text size range specified in the request, ultimately providing the correct and complete response. The same can be said about the Gemini model, which improved its response after receiving feedback and respected the text size range specified in the request. Another principle of agile engineering that is very important to follow is providing direction. When interacting with a large language model, it is crucial to provide as much context as possible and to be as specific as possible about what we want to get from that model.

Regarding the performance and accuracy of the three models, observing the table. We can immediately conclude that ChatGPT is the language model that performed best in most of the criteria if we take into account the fact that after evaluating the quality of the answer and giving the feedback, it gave the correct final answer in case of the first evaluation criterion. Also, the ChatGPT language model provided accuracy and precision in most responses.

The Gemini model performed poorly overall, hallucinating several times and not giving accurate answers. These things are supported by several examples, namely: failure to respect the specified size range although we evaluated the quality and provided feedback on this aspect, it is not able to solve a math problem that also involves logic, as well as it is not even able to generate a code in the Python language that does not return an error and works correctly, and in the case of the interview simulator it performed very poorly, compared to the other two big language models, ChatGPT and Llama 3. A positive aspect in the Gemini model is the fact that it does not have an extremely important limitation, namely harmful or toxic results. Gemini, led by Llama 3, demonstrated that they are not limited from this point of view. We can't say the same about ChatGPT, unfortunately.

In terms of performance and accuracy of the answers provided, Llama 3 model is in the 2nd place, so it is located in the middle of the ranking. The worst hallucination from my point of view of this model was in the case of criterion 1 given by the wrong answer to the math problem and the incorrect focaccia recipe.

In conclusion, these large language models are far from stable, far from hallucinating, and even far from having a human understanding of a task. Although a lot is invested in this field, and many articles describe these models as being very capable, in essence, there is still a lot of work to be done before creating a machine with "human" characteristics. It is possible that in 10 years these models will have excellent performance and accuracy to match, but at the moment, although such a model can give you a very elaborate answer, most of the time, it is not useful in everyday reality, as we saw in the interview simulator. It is true, however, that such a model is really helpful, it can give you answers to guide you in your tasks if you know how to ask for these answers.

VI. CONCLUSIONS

Each model has its strengths and is designed to excel in different areas. ChatGPT (GPT-4) is versatile and strong in general-purpose conversational AI. Gemini leverages Google's extensive data and infrastructure for advanced contextual understanding and multilingual capabilities. Llama3 focuses on efficiency and performance, making it suitable for applications requiring lightweight and resource-efficient solutions. The choice between them would depend on the specific requirements of the application, such as accuracy, computational resources, and the need for real-time data integration.

Nowadays, these large language models are massively used in big companies and corporations. By incorporating these models in various fields, companies have witnessed improved customer interactions, improved content generation and efficient data analysis. The large language models have emerged as a transformative force in natural language processing, reshaping the way industries interact with and use text data.

As future directions, we can affirm that the insights gained from this comparative analysis highlight the importance of selecting the appropriate language model based on specific application needs and resource constraints. Future developments in large language models should focus first on improving accuracy and reducing bias where efforts should be made to enhance the accuracy of responses and minimize biases, particularly for models trained on diverse datasets; second on enhancing resource efficiency, continued innovation in model architecture improvement to optimize performance while reducing computational requirements that will be crucial, and third expanding accessibility and use cases that can increasing the accessibility of these models and broadening their applicability across different domains and industries that will drive further advancements in artificial intelligence.

In conclusion, ChatGPT, Gemini, and Llama3 each offer unique benefits tailored to different applications, and their continued evolution will play a significant role in the advancement of natural language processing technologies.

REFERENCES

[1] J. Phoenix and M Taylor, “Prompt engineering of generative AI, future-proof inputs of reliable AI outputs, ”in O'Reilly Media, vol. 1, 422 pag, 2024.

[2] Toloka. (n.d.), "History of LLMs". 09 May2024. [Online]. Available: https://toloka.ai/blog/history-of-llms/ [Accessed 18 June 2024].

[3] Wagh, A. “Open AI: Understand foundational concepts of ChatGPT and cool stuff you can explore“. 2 April 2023. Medium. [Online]. Available: https://medium.com/@amol-wagh/open-ai-understand-foundational-concepts-of-chatgpt-and-cool-stuff-you-can-explore-a7a77baf0ee3 [Accessed 18 June 2024].

[4] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., & Polosukhin, I, “All attention you need“, In Advances in Neural Information Processing Systems, pp. 5998-6008, 2017.

[5] ChatGPT. (2024). [Online]. Available: https://chatgpt.com/ [Accessed 18 March 2024].

[6] Google Gemini. (2024). [Online]. Available: https://gemini.google.com/app [Accessed 20 March 2024].

[7] Hugging Face. (2024). Llama 3 chatbot. [Online]. Available: https://huggingface.co/spaces/Be-Bo/llama-3-chatbot_70b [Accessed 28 March 2024].

[8] ChatGPT Guide. (2024, May 2). What is normalization? LLMs explained. [Online]. Available: https://www.chatgptguide.ai/2024/03/02/what-is-normalization-llms-explained/#:~:text=of%20the%20dataset.-,Normalization%20in%20LLMs,process%20and%20understand%20the%20data [Accessed 28 May 2024].

[9] Run:AI. (2024). A guide to large language model (LLM) training. [Online]. Available: https://www.run.ai/guides/machine-learning-engineering/llm-training [Accessed 30 May 2024].

[10] Shaip. (2024). A guide to large language model (LLM). [Online]. Available: https://ro.shaip.com/blog/a-guide-large-language-model-llm/ [Accessed 01 June 2024].

Image Generation Using StyleVGG19-NST Generative Adversarial Networks

Dorcas Oladayo Esan, Pius Adewale Owolawi, Chunling Tu
Department of Computer Systems Engineering, Tshwane University of Technology, Pretoria, South Africa

***Abstract*—Creating new image styles from the content of existing images is challenging to conventional Generative Adversarial Networks (GANs), due to their inability to generate high-quality image resolutions. The study aims to create top-notch images that seamlessly blend the style of one image with another without losing its style to artefacts. This research integrates Style Generative Adversarial Networks with Visual Geometry Group 19 (VGG19) and Neural Style Transfer (NST) to address this challenging issue. The styleGAN is employed to generate high-quality images, the VGG19 model is used to extract features from the image and NST is used for style transfer. Experiments were conducted on curated COCO masks and publicly available CelebFace art image datasets. The outcomes of the proposed approach when contrasted with alternative simulation techniques, indicated that the CelebFace dataset results produced an Inception Score (IS) of 16.57, Frecher Inception Distance (FID) of 18.33, Peak Signal-to-Noise Ratio (PSNR) of 28.33, Structural Similarity Index Measure (SSIM) of 0.93. While the curated dataset yields high IS scores of 11.67, low FID scores of 21.49, PSNR of 29.98, and SSIM of 0.98. This result indicates that artists can generate a variety of artistic styles with less effort without losing the key features of artefacts with the proposed method.**

***Keywords*—Artworks; VGG19; Neural Style Transfer; Generative Adversarial Network; inception score; StyleGAN**

I. INTRODUCTION

Drawings, paintings, and carvings with pencils, brushes, and cardboard were traditional tools used by artists to express their unique creativity and ideas [1]. This infers that the production of artwork requires the craftsman who expects to integrate unique and special imaginative artistic styles to invest much time and energy to show their innovative skills, which can be tiring and overwhelming.

The introduction of Artificial Intelligence (AI) into computer technology applications in recent years has made it possible for artists to enhance their original and creative artwork styles incrementally and effortlessly [2]. GANs have received impressive consideration in artistic image generation due to their ability to learn deep representations without extensive training data. GANs utilize their generator engine to reconstruct the input image and discriminator engine to differentiate between the generated images and input images [3].

Traditional GANs are facing challenges of model inability to generate high-resolution images, poor choice of parameter optimisation methods, and difficulty in the generation of another image style from the content of existing images [4].

The lack of in-depth capability to capture intricate image features makes effective transferring of complex artistic styles to images, resulting in unappealing visual image-generated outputs [5]. Also, most of the existing techniques lack a robust starting point for training, which significantly affects the efficiency and stability of the GANs during the training process [6]. Furthermore, the inability of conventional image generation techniques to separate content and style representations effectively is challenging, thereby affecting the content structure of the original image and the desired style [7]. These issues have significantly affected many artists in the generation of closely related artworks from existing works, thereby creating hindrances for the artist to mitigate their innovative and creative styles in their artworks.

Several studies have made efforts to tackle these issues by using GAN models for artistic image style generation and manipulation [8, 9]. These models include Convolutional Neural Networks (CNNs) [10], Cycle GAN [11], Conditional GAN [12], Genetic Algorithm (GA) [13], etc., which have performed to varying degrees of success, but none have given conclusive solutions to address the challenging gap of creating perfect and realistic artwork due to difficulty in many of the methods to adjust the content structure in images which consequently results in the missing of some important features, distortion, and ambiguity creating local features. Hence, it is important to address the issue of style loss and the generation of imperfect and unrealistic images that most existing GAN models exhibit to assist artists in the improvement and enhancement of their artwork creativity.

To better generate perfect and realistic styled artwork based on existing artwork, this study introduces an innovative method to enhance the creation of artistic images without losing the art image contents. Leveraging StyleGAN, VGG19Net and Neural Style Transfer, StyleGAN generates top-notch images, while the VGG19Net model extracts features and NST is utilized to retain artistic image characteristics for the generation of artistic artefacts having high perceptual and realistic art images [14]. Utilizing this approach can serve as a mechanism for artists to manipulate and generate different artistic styles from existing artworks with minimal effort.

The following are the primary contributions of this study:

1) Fusion of models: The development of a new architecture that integrates multiple inputs and outputs, employing a StyleGAN with the application of VGG19Net for feature extraction, and NST for the preservation of image features, and generation of artistic artefacts having high perceptual and realistic art images. This offers an innovative approach to enhance the performance of art image generation.

2) Parameter optimization: Tweaking different parameters to enhance the aesthetic images generated for visual quantitative visual evaluation.

3) Evaluation metrics: Utilization of different performance evaluation metrics to determine the quantitative enhancement and generation of the newly generated images on the proposed model and other state-of-the-art models.

4) Computational time: Evaluation of the proposed model and other selected recent GAN models on the datasets used to determine their execution time on both CPU and GPU systems.

5) Comparative analysis: Benchmarking and comparing the proposed model performance on curated Coco Mask African and publicly available CelebFace datasets against other Art GAN models.

The remainder of this article follows this structure: Section II outlines related works and the theories of the proposed method. Section III provides an in-depth explanation of the proposed method. Section IV deliberates on the experimental outcomes and assesses the proposed model. Concluding remarks are presented in Section V.

II. Related Works

An extension to GANs, called ARTGAN, is proposed in a study in [15] to artificially generate more difficult and complex images, like abstract art. The suggested model can create artwork that looks natural because it learns more quickly and produces high-quality, realistic images based on the CIFAR-10 dataset, as demonstrated by the results the authors obtained. The authors measure the log-likelihood of the generated artwork using the trained GAN models. One of the limitations of this approach is that the generator works with a limited number of image samples and with hyper-parameter choices and generates imperfect images.

The instability of GAN training was addressed with the proposal of StackGAN [16]. By stacking several generators that can produce images with varying resolutions, the authors used a hierarchical structure. The outcomes showed that 256 by 256 resolution images produced by StackGAN can be visually appealing. In contrast to the more advanced technologies used for statistical data conception, visualizing textual data, particularly for creative text is still in its early stages of development. The limitation of this method is that it cannot handle more extreme and varied image transformations.

In study [17], the concept of image super-resolution using Super Resolved Generative Adversarial Networks (SRGAN) was introduced. To produce photo-realistic natural images, the authors incorporated a perceptual loss function that combines both adversarial and content losses. The adversarial loss incentivizes the solution to conform to natural image attributes by employing a discriminator network that has been trained to differentiate between super-resolved images and authentic photo-realistic images. The model was tested on the BSD100 dataset, where the deep residual network successfully reconstructed photo-realistic textures from significantly down-sampled images within the public benchmarks' dataset. Extensive Mean Opinion Score (MOS) testing highlighted substantial enhancements in perceptual quality with SRGAN. The MOS scores obtained from SRGAN were notably closer to those of the original high-resolution images compared to other prominent methods.

The recovering and restoring of artwork that has been damaged over time due to several factors was introduced in study [18]. The authors utilized a conditional Generative Adversarial Network that involves the generator combining adversarial loss and a discriminator that uses binary cross-entropy loss for optimization. The result of the experiment conducted shows that the method completely removes damage in most of the image and perfectly estimates the damage region. Although the method might be unstable causing the generator to output only a certain type of image and the discriminator to be unable to distinguish between the input image and generated image.

One can see that the existing literature regarding artistic image generation and style techniques has become apparent from the limitations. Previous research has contributed immensely but often lacked comprehensive and advanced approaches to deal with the intricacies of creating perfect and realistic artwork, style loss representation and in-depth capability in capturing intricate image features make effective transferring of complex artistic styles to images. Also, a robust starting point for training optimization of the GANs during the training process is hampered by the computing efficiency of real-time artistic image generation systems. Through a variety of noteworthy contributions like those listed above, this research greatly strengthens solutions to these weaknesses. It develops a novel approach integration of the GAN model which integrates principles of Style Transfer to generate high-quality image styles without losing any art image features.

III. Proposed Method

This section discusses the suggested approach. There are four steps in the suggested method: (a) image acquisition (b) the image preprocessing stage (c) the image feature extraction stage (d) the image generation stage. A detailed explanation is given in the following subsections.

A. Image Acquisition Stage

The experiments in this study used COCO African Mask art images [19]. The dataset is images of African masks that will help readers experience the pinnacle of African art. This dataset consists of 9,300 images of African art as in Fig. 1.



Fig. 1. Samples of the Coco African mask dataset [19].

The detailed proposed framework is shown below in Fig. 2.

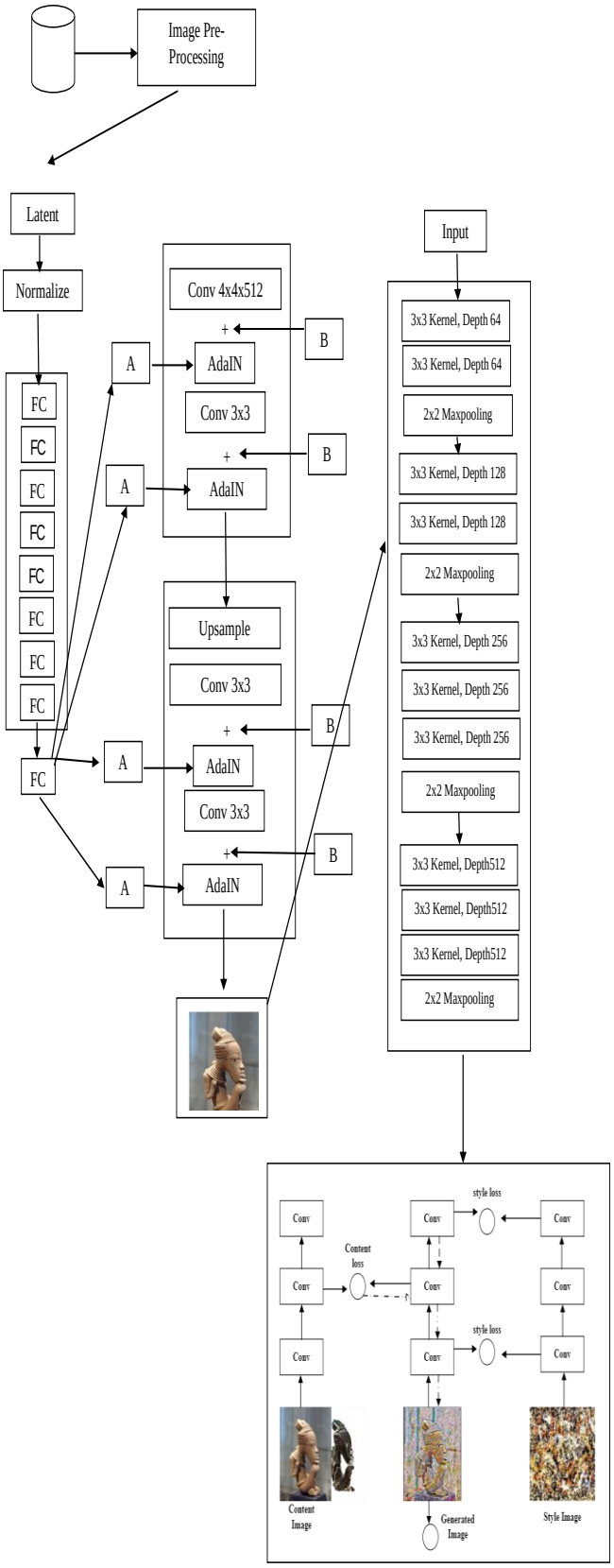


Fig. 2. Proposed styleVGG19-NST framework for artistic image generation.

B. Image Pre-Processing Stage

The input data is pre-processed to improve feature transfer. The pre-processing includes image noise removal and image segmentation. The image pre-processing is done to remove unnecessary and unwanted artefacts that can affect the performance of the GAN models used in this research. Fig. 3 illustrates all the steps in the pre-processing stage used for the implementation of this research.

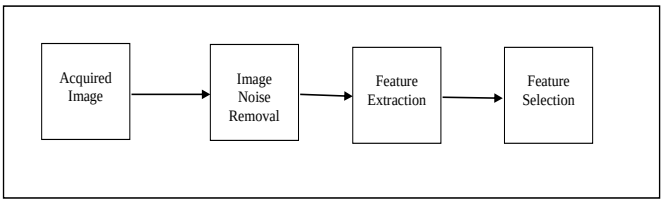


Fig. 3. Flowchart of pre-processing stage.

1) Image segmentation (background removal): In this study, an Otsu segmentation technique is used to model the image from a series of image frames to perform image segmentation. The foreground image's pixels are separated from the background using this technique. To accomplish this, subtract one image at a time (t-1) from the image generated at the time(t). Then the background subtraction is calculated as in Eq. (1).

$$B(x,y,t) = \left(I_{(t-1)}(x_{t-1}, y_{t-1}) - I_t(x_t, y_t)\right) > Thr \quad (1)$$

The selected threshold denoted by *Thr* is dynamically determined to adjust to the changes in frame surroundings. The background image is updated as in Eq. (2).

$$I_{t+1} = \begin{cases} I_t(x_t, y_t) > B(x,y,t), & foreground \\ otherwise & \\ I_t(x_t, y_t) < B(x,y,t), & background \end{cases} \quad (2)$$

2) Image noise removal: After extracting the image foreground, some environmental noises are still present in the foreground image, such as illumination, shadow, light intensity etc. This research adopted the median filtering technique where the noisy image pixels are replaced by the average value of their neighbouring pixels (mask) as in Eq. (3).

$$I'(x', y') = median\{g'(x'+i), (y'+j), i, j \in w'\} \quad (3)$$

Where, $I'(x', y')$ is the image median $g'(x', y')$ is the input image, and $j \in w'$denotes a 2-D image mask. The output of the enhanced image is passed to the feature extraction stage for further processing.

3) Feature extraction stage: At this stage, the objects are recognized objects based on certain characteristics they possess and the HSV colour feature extraction is utilized as in Eq. (4).

$$\mu_{HSV} = \frac{1}{N}\sum_{i,j}^{N} = 1\, P_{HSV}(i,j) \quad (4)$$

Where μ_{HSV} is the image HSV colour mean value, N is the pixel number, and $P_{HSV}(i,j)$ *the* colour component in image shape. The output of colour extraction is fed to feature selection to remove any redundant features.

4) Feature selection: Feature selection is the selection of a subset of relevant features with short dimensionality, short training time, and low overfitting. The extracted features are then spatially related to each other, but there are some semantic inconsistencies between them which can lead to overfitting. In this study, feature selection is performed using the correlation-based features as in Eq. (5).

$$correlation = \frac{\sum(F_{1i}-\bar{F}_1)\sum(F_{2i}-\bar{F}_2)}{\sqrt{\sum(F_{1i}-\bar{F}_1)^2\sum(F_{2i}-\bar{F}_2)^2}} \quad (5)$$

Where, (F_1, F_2) represents the cross-correlation between space F_1 and F_2. The correlations(F_1, F_2) of the two features are in the range of -1 to 1. If two features F_1 and F_2 are independent of each other, then the correlation is $(F_1, F_2) = 0$. The image generation stage receives the feature extraction output after it has been processed.

C. Image Generation Stage

Here, the output of the image extracted is fed to the image generation stage, where the generator (G) is used to generate a 512-dimensional latent vector Z, which is fed into 8 convolutional layers of the Mapping Network (MP). The latent vector Z is converted to a space w that defines the style of the resulting image. The latent code of the input image is continuously optimized for the parameters to achieve the differences between the input image and the generated. The latent code Z, often referred to as the latent spatial mapping of the image, is applied to reduce the painterly style. The vector Z is sampled from a predefined distribution (uniform Gaussian distribution) in the latent space Z which is mapped to the latent space N to produce w which is passed to the AdaIN module. After the model has been trained, the generator is applied to gradually increase the resolution of generated images with 8 convolutional layers from 512x512 to 1024x1024, and AdaIN to add noise to each layer. AdaIN (Adaptive Instance Normalization) converts the latent vector into two scalars (scaling and bias) to control the style of the image generated at each resolution level. In this module, the encoded information w obtained from the mapping network to the generated image. The latent code 'w' generated by the mapping network is passed to the affine transform and AdaIN layer for training. Affine transformations are implemented using two linear planes to create a style using the AdaIN in Eq. (6).

$$AdaIN(x, y) = y_{s,i}\frac{x_i-\mu(x_i)}{\sigma(x_i)} + y_{b,i} \quad (6)$$

Where x is the output feature map to the previous level. AdaIN first normalizes the "zero mean" and "uniform variance" of each channel x_i and then applies the scales y, s andy, b. This means that style y controls the stats of the next convolutional layer's feature map. Where y, s is the standard deviation and y, b denotes the mean.

The discriminator (D) receives the generated image afterwards. A backpropagation algorithm is used to modify the weights of the three networks to enhance the quality of the final image. Furthermore, the generated image from the StyleGAN model was fed into the VGG-19 model alongside the content image. StyleGAN uses a combination of Progressive Generative Adversarial Networks (PGGAN) and neurotransmission techniques [20]. StyleGAN has gained prominence due to its ability to transform low-resolution images into enhanced images [21]. The mean and variance of the feature map x_i generated by the layers in the synthetic network are altered by StyleGAN using reference style bias $y_{b,i}$ and scale $y_{s,i}$in equation (6). As shown in Fig. 4, the generator grows incrementally, adding new constants, scaling the image, and applying style and noise to each block as illustrated in Fig. 4.

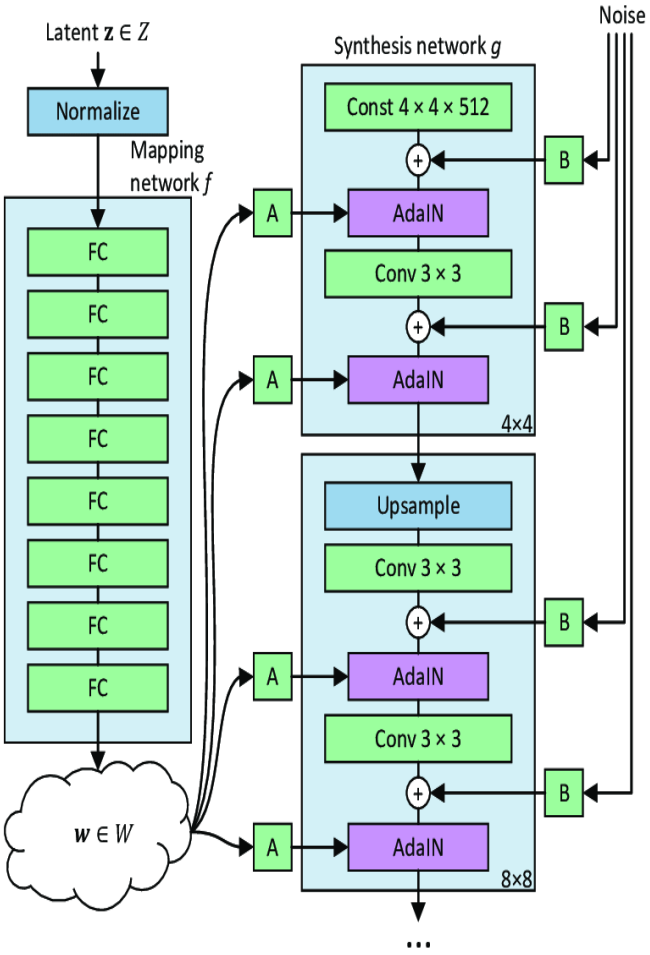


Fig. 4. Architecture of StyleGAN adopted from study [22].

D. Stochastic Variation with Noise Injection

Stochastic variations in an image are small details that do not change the context of the image. There are generators embedded in StyleGAN that try to learn how to generate image styles and content. Noise injection into the StyleGAN created before the AdaIN layer helps create such variations. The noise added to the feature map has zero mean and low variance compared to the feature map. Therefore, the overall context of the image is preserved as the feature map statistics remain the same. From the network, the latent W vector is employed to control the image style of the generated images.

E. VGG19-Network

The VGG-19 is a fully connected model with nineteen deep trainable convolutional layers that include dropout and max pooling layers. The convolutional layer is trained to extract features produced by the StyleGAN in this paper. model output

with a regularized dropout layer and a densely connected classifier [23]. To extend the depth, VGG-19 employs a 3 × 3 convNet configuration. To reduce dimensionality, max-pooling layers are utilized as handlers. The two FCN layers contain 4096 neurons each. To minimize the false positives, while testing, all lesions are considered, as VGG is trained on individual lesions. Convolution layers execute convolution operations across pixel by pixel, enabling the output to progress through the next layers. Filters within the convolution layer are typically 3 × 3 in size and are trained to extract features. After every series of convolutional layers, a Rectified Linear Unit (ReLU) layer and a max-pooling layer are included. ReLU is recognized as an effective non-linear activation function, permitting only the positive values from the input. The architecture of VGG-19 is illustrated in Fig. 5.

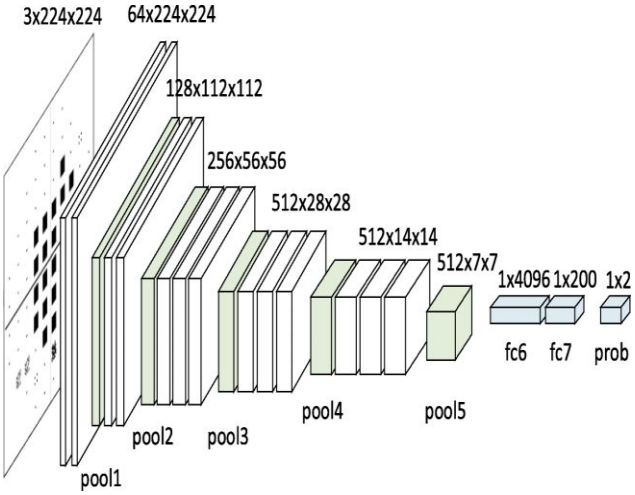


Fig. 5. VGG-19Net adopted from study [23].

F. Neural Style Transfer Model

Blending two images, one with content and the other with style can result in new artwork through Neural Style Transfer. The process of transferring an image's style while preserving its content is known as Neural Style Transfer. To add an artistic touch to your image, all that needs to be changed are the style configurations. The two sets of images that Neural Style Transfer works with are the content image and the Style image.

The content image can be replicated using this technique in the reference image's style. The creative style is applied from one image to another using Neural Networks. To synthesize features and transfer style from one image to another, NST uses a pre-trained Convolutional Neural Network with additional loss functions. NST specifies the following inputs:

- an input (generated) image (g) that contains the final result.
- a content image (c) that is the image to which a style is to be transferred.
- an input style image (s) that is the image from which the style is to be transferred.

The Neural Style Transfer architecture is depicted in Fig. 6.

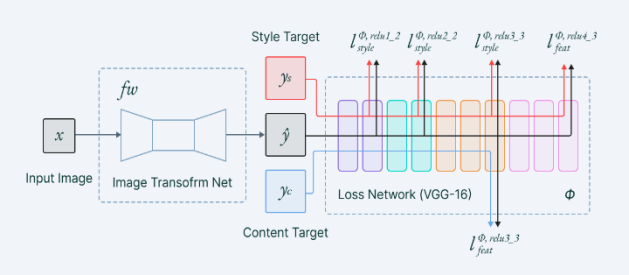


Fig. 6. Neural Style Transfer architecture adopted from study [24].

1) Content loss: Comparisons between the content image and the generated image are made easier by the content loss. The model's upper layers, intuitively, concentrate more on the characteristics seen in the image (the picture's general content). The equation for content loss computes the Euclidean distance between the input image (x) and the content image (p) at layer l, which correspond to the respective intermediate higher-level feature representations. The content loss is shown in Eq. (7).

$$L_{content}(\vec{p}, \vec{x}, l) = \frac{1}{2}\sum_{i,j}(F_{ij}^{l} - C_{ij}^{l})^2 \quad (7)$$

In this context, with the content image denoted as "C," the target image as *"x,"* and the processed layer as *"l," "F"* and *"P"* represent the feature representations of the content image and the target image, respectively, on layer "*l.*"

2) Style loss: Style loss, similar to content loss, uses the squared loss function to measure the difference in style between the synthesized image and the style image. The style loss involves calculating the Maximum Mean Discrepancy between two images, and it is determined using Eq. (8).

$$\mathcal{L}_{style}(\vec{p}, \vec{x}) = \sum_{i=0}^{L} w_l E_l \quad (8)$$

Where, w_l is a weight given to each layer during loss computation, the content image is *p*, and *x* is the target image.

Three key components are essential for generating a style transfer image: content image, style image and generated image. The content image and the style image are modified together to generate new artistic images. The style is the variation added to the content image that produces an entirely new image. The NST model produced stylized images resembling a blend of the content and style images.

Maintaining the generated image's proximity to the local textures of the style reference image was achieved by utilizing the style loss function. However, the generated image's high-level representation is maintained close to the base image by the content loss function. To ensure that the generated locally coherent, the total loss function is used.

G. Evaluation Mechanism

Quantitative and qualitative evaluation metrics are used in this research to analyse the performance of a proposed model. For the quantitative evaluation, the FID, PSNR, SSIM, and IS. For the qualitative evaluation, the enhancement of the image is visually inspected to show the performance of the models used in terms of image clarity. The quantitative metrics are further explained in the following sub-section.

1) Frecher Inception Distance (FID): This metric is used to quantitatively evaluate the quality of created images using the proposed model [25] as in Eq. (9).

$$FID(r,g) = \|\mu_r - \mu_g\|_2^2 + Tr(\sum_r + \sum_g + 2(\sum_r \sum_g)^{1/2} \quad (9)$$

Where the mean and covariance of the real and generated data are represented as $(\mu_g, \sum g)$ and$(\mu_r, \sum r)$.

2) Inception Score (IS): The quality of images generated by GANs is measured by the inception score [25], as in Eq. (10).

$$\exp(E_x[KL(p(m|n)\,||p(m))] = \exp(H_y - E_x[H(m|n)] \quad (10)$$

Where, $p(m|n)$ is the probability of marginal image distribution.

3) Peak Signal-to-Noise Ratio (PSNR): This compares the peak signal-to-noise ratio of two monochrome images, I and k, to determine how good a generating image is compared to a set of real images. As the PSNR (measured in dB) rises, the generated image's quality increases. It is computed as in Eq. (11).

$$PSNR(I;K) = 10log_{10}\left(\frac{max_i^2}{MSE}\right) = 20log_{10}(max^2 I) - 20log_{10}(MSE_{I,K}) \quad (11)$$

$MSE_{I,K} = \frac{1}{m}\sum_{i=0}^{m-1}\sum_{i=0}^{n-1}(I(m,n) - K(m,n))^2$ and Max_i is the minimum possible pixel value.

4) Structural Similarity Index Measure (SSIM): This is an indicator of how similar two images are to one another. The SSIM is expressed as in Eq. (12).

$$SSIM_{(x,y)} = [l(x,y)]^{\alpha}.[c(x,y)^{\beta}.[s(x,y)^{\gamma} \quad (12)$$

Where, α, β, and γ are the positive constants, and *l*, *s*, and *c* are the luminance, brightness, and contrast ranges, respectively, that are used to compare two images. On the other hand, the structure *s* is utilized to analyze the local luminance pattern of two images to determine their level of similarity or dissimilarity.

IV. Result and Discussion

This section presents different experiments carried out to achieve the generation of historical artistic images using the proposed method. The configuration and parameter settings and the experimental simulation are discussed in the subsequent section.

A. Configuration Experimentation and Parameter Setting

The experiment was done on a Central Processing Unit (CPU) and Graphic Processing Unit (GPU) computer using the Google Collaboratory with the Tensor Flow library installed independently. Experiments were quantitatively and qualitatively conducted on both the COCO Mask African dataset and publicly available CelebFace datasets which were validated on selected techniques in sections B - D. A total of 1,500 frames were selected during the simulation. This test data can assist in effective observations of the test performance of the pre-trained proposed model on the trained model. The image resolution is 512*512 and the training has been iterated 1500 times with 0.0001 learning rates and 250 batch numbers. The values chosen for learning rate and batch iterations improve stability and speed during training. Detailed experiments are described in the next section.

B. Hyperparameter Selection for the Proposed Model

During the implementation, the proposed SyleGANVGG19-NST model parameter values were set to the following as shown in Table I.

TABLE I. Hyperparameters Used in the Implementation of Training Stylegan Models

Parameters	Values
Learning rate	0.0001
Batch size	250
Epochs	1500
Beta	0.5
Adversarial loss mode	lsgan
loss weight	10
Identity loss weight	0
Pool size to store fake samples	60

C. Experiment 1.1: A Qualitative Evaluation of the Proposed Model and the Existing Recent Used GANS on the COCO African Mask Dataset

The objective of this study is to address the issue of image style transfer and also generate images with high resolution using the style transfer. Different image pre-processing methods were used, such as image enhancement and feature extraction.

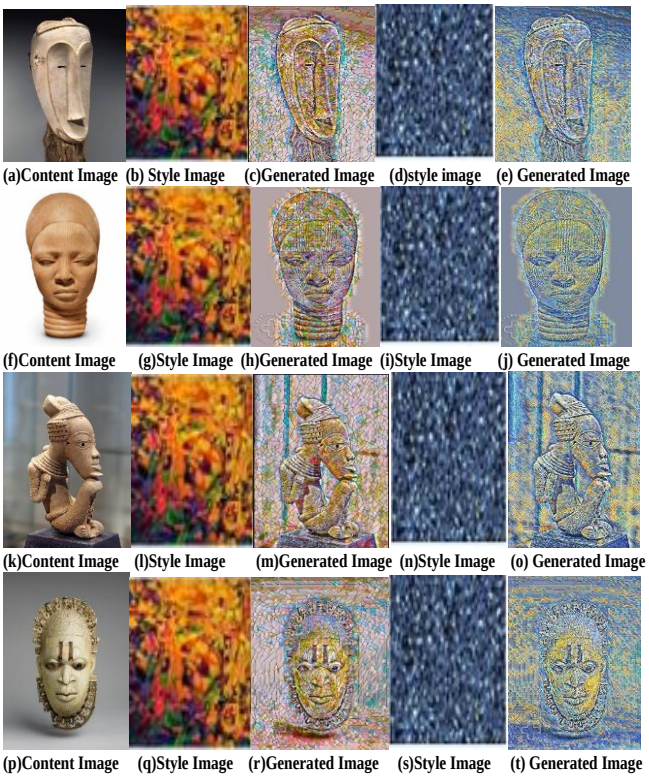


Fig. 7. Qualitative evaluation of SyleGANVGG19-NST model on COCO African mask dataset.

Fig. 7(a), 6(f), 7(k), and 7(p) consist of the original/content image, Fig. 7(b), 7(g), 7(l), and 7(q) is the style image to be matched with the content image, Fig. 7(c),7(h), 7(m), and 7(r) are the generated image and Fig. 7(d), 7(j), 7(n), and 7(s) is the second style image to be matched with the content image, Fig. 7(e), 7(j), 7(o), and 7(t) are the second generated images. This result shows that the proposed model was able to generate an artistic image that is different from the content image. Furthermore, the proposed method is compared with other recently used baseline methods as shown in Fig. 8.

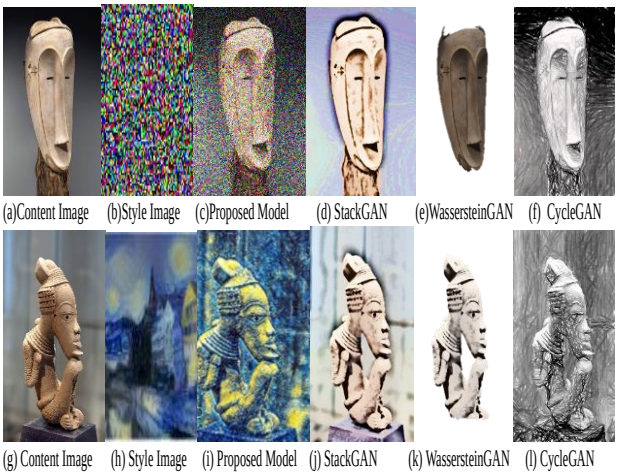


Fig. 8. Qualitative evaluation of the proposed model with other recent baseline art GAN models.

Fig. 8(a) and 8(g) show the content of the original image, Fig. 8(b) and 6(h) shows the style, Fig. 8(c) and (i) are the images generated by the proposed StyleVGG19-NST, Fig. 8(d) and (j) represent the image generated by the Stack GAN model, Fig. 8(e) and (k) and Fig. 8(f) and (l) are the images generated by the Wasserstein GAN and CycleGAN models respectively. One can see that the images generated by the Style method do not transfer the style image into the generated images, the Wasserstein GAN-generated images in Fig. 8(e) have a style transfer issue as some parts of the image generated have already been cut off. The image obtained in Fig. 8(c) and 8(i) a clearer image generated by the proposed StyleVG19-NST method which contains the style image compared with other methods. Loss values of the corresponding training of the selected models with the proposed model as in Table II.

Table III shows the iteration of training loss for all the models, one can observe that the proposed model has lower content loss values compared to other models, and this shows the consistency of the model in terms of generated image content with the original image. The findings from this experiment show that the optimization used in this experiment was able to generate a perfect image with epoch 1500 which is better in comparison to the image generated with optimization parameters in study [15].

D. Experiment 1.2: Quantitative Evaluation of the Proposed Model and the Existing Recent Used GANS on Curated Dataset

This section aims to quantitatively compute the generated images because of the difficulty in evaluating the model objectively using only subjective visual assessment of the synthetic image. The summary of the result generated in terms of FID, IS, PSNR and SSIM score is used in Table III.

TABLE II. GENERATOR AND DISCRIMINATOR LOSS VALUES WITH DIFFERENT EPOCHS FOR COCO AFRICAN MASK DATASET

Epoch	StackGAN		Wasserstein GAN		CycleGAN		Proposed model	
	Discriminator loss	Generator loss	Discriminator loss	Generator loss	Discriminator loss	Generator loss	Discriminator loss	Generator loss
200	0.3095	0.8521	0.3721	0.4176	0.3216	0.3987	0.2112	0.2021
500	0.3728	0.5711	0.3364	0.3792	0.3411	0.3769	0.3462	0.3291
800	0.3111	0.4277	0.3516	0.3618	0.3423	0.3591	0.3063	0.2537
1000	0.3693	0.3631	0.3433	0.3536	0.3482	0.3419	0.3146	0.3093
1500	0.4331	0.2911	0.4036	0.3240	0.3855	0.3892	0.3021	0.2187

TABLE III. METRICS RESULT OF THE PROPOSED MODEL WITH OTHER RECENTLY USED ART GAN MODELS

Models	FID	IS	PSNR	SSIM
StackGAN	24.83	6.13	21.85	0.61
Wasserstein GAN	25.56	7.76	24.19	0.68
CycleGAN	29.32	5.72	25.59	0.78
Proposed Model	21.49	11.67	29.98	0.98

From Table III, the art image generated using the proposed StyleVGG19-NST model has improvements in FID, IS, PSNR, and SSIM compared with the Stack GAN, Wasserstein GAN, and CycleGAN models. The proposed model has a higher IS of 11.67, a lower FID of 21.49, an SSIM of 0.98 and a higher PSNR of 29.98 The higher IS, SSIM, and PSNR signifies that the better image quality produced by the proposed StyleVGG19-NST model, and the lower FID indicates that the proposed model generated images that have more structural features than the original image. The findings here show that the proposed method has a better image generation in terms of FID in comparison to the method used in study [16].

E. Experiment 2: Benchmarking the Proposed Methodology on Publicly Available Celebface Dataset

Since there is no common set of image data of similar artistic for the different existing techniques, the validation performance of the proposed technique is tested on a publicly available CelebFace dataset. The face image is randomly selected. The result obtained from the proposed model is compared with some existing recent state-of-the-art Art GAN techniques in terms of image style generation and the use of qualitative evaluation

metrics such as FID, IS, PSNR, and SSIM. The details of the experiments are presented in the following sections.

F. Experiment 2.1: A Qualitative Assessment of the Proposed Model and the Existing Recent Art GANs on the Celebface Dataset

The objective is to assess the dependability and strength of the proposed model using the CelebFace dataset, which is openly accessible. The performance of the image pre-processing stages in image enhancement and feature extraction on curated datasets is qualitatively evaluated as shown in Fig. 9(a)-(o).

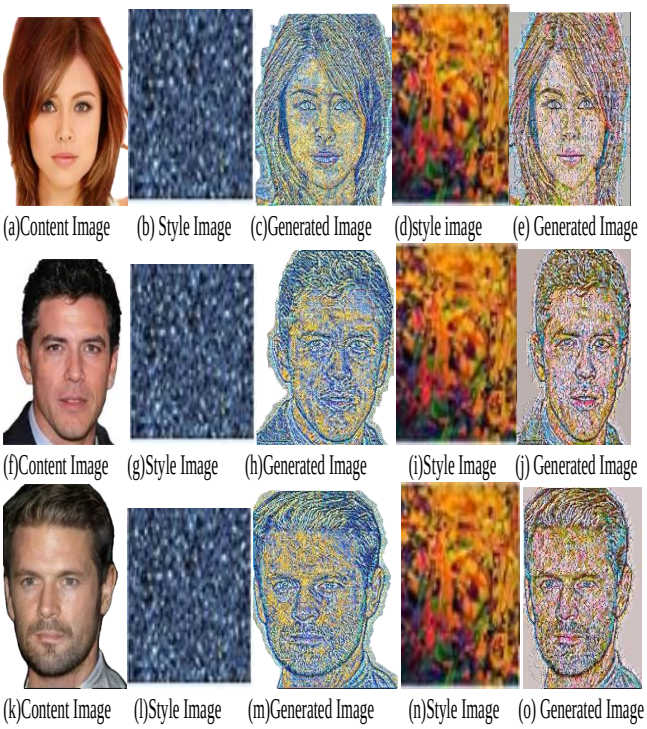


Fig. 9. Qualitative evaluation of the proposed model on the CelebFace dataset.

Fig. 9(a), 9(f), and 9(k) consists of the original/content image, Fig. 9(b), 9(g), 9(l) is the style image to be matched with the content image, Fig. 9(c), 9(h), and 9(m) are the generated image and Fig. 9(d), 9(i), and 9(n) are second style images to be matched with the content image, Fig. 9(e), 9(j), and 9(o) are the second generated images. This result shows that the proposed model was able to generate a stylistic artistic image that is different from the content image. Furthermore, the proposed model and the other existing recent Art GAN models as shown in Fig. 10(a) - (l).

Fig. 10(a) and 10(g) show the images Content image, Fig. 10(b) and 10(h) shows the style, Fig. 10(c) and 10(i) are the images generated by the proposed model, Fig. 10(d) and 10(j) represent the image generated by the Stack GAN model, Fig. 10(e) and 10(k) and Fig. 10(f) and 10(l) are the images generated by the Wasserstein GAN and CycleGAN models respectively. The image obtained image in Fig. 10(c) and 10(i) a clearer image generated by the proposed image which contains the style image compared with other methods.



Fig. 10. Qualitative evaluation of the proposed model with other recent baseline art GAN models.

The generator and discriminator loss values of the corresponding pre-trained selected models with the proposed model are shown in Table IV.

TABLE IV. GENERATOR AND DISCRIMINATOR LOSS VALUES WITH DIFFERENT EPOCHS

Epoch	StackGAN		Wasserstein GAN		CycleGAN		Proposed model	
	Discriminator loss	Generator loss	Discriminator loss	Generator loss	Discriminator loss	Generator loss	Discriminator loss	Generator loss
200	0.3811	0.4473	0.3931	0.5211	0.2652	0.2827	0.3667	0.3117
500	0.3693	0.3652	0.3672	0.3633	0.2977	0.3081	0.3406	0.3513
800	0.3511	0.3630	0.4021	0.4213	0.3211	0.2953	0.3542	0.3911
1000	0.3271	0.4019	0.3163	0.3177	0.3123	0.3078	0.3183	0.3485
1500	0.3522	0.3586	0.3011	0.3033	0.3433	0.3988	0.3496	0.3698

Table IV shows the iteration of training loss for all the art GAN models, it is observed that the proposed model has lesser content loss values compared to other existing recent Art GAN models, and this shows the consistency of the model in terms of generated image content with the original image.

G. Experiment 2.2: Quantitative Evaluation of the Proposed Technique and the Existing Recent Art GANs on the Celebface Dataset

This section aims to quantitatively evaluate the generated images because of the difficulty in evaluating the model objectively using only subjective visual assessment of the synthetic image. The summary of the result generated in terms of FID, IS, PSNR and SSIM score is used as shown in Table V.

TABLE V. METRIC RESULT OF THE PROPOSED TECHNIQUE WITH OTHER EXISTING RECENT ART GAN MODELS

Models	FID	IS	PSNR	SSIM
StackGAN	45.32	9.31	21..32	0.78
Wasserstein GAN	31.45	10.29	24.25	0.83
CycleGAN	35.92	8.11	22.73	0.89
Proposed StyleVGG19-NST method	18.33	16.54	28.33	0.93

From Table VI, one can observe that the proposed StyleVGG19-NST method used in this research on the input image has improvements in terms of FID, IS, PSNR and SSIM compared with the Cycle GAN, DC GAN, and C-GAN models. The proposed model has a higher IS of 16.54, a lower FID of 18.33, an SSIM of 0.93 and a higher PSNR of 28.33 The higher IS, SSIM, and PSNR indicate that the proposed model generates better image quality, and the low FID score signifies that the proposed StyleVGG19-NST model produced images that have more structural features than the original image. Furthermore, the FID of the proposed method is better than the author in study [18].

H. Computational Time

Since most of the image-generated models evaluate the computational time of the model, the computational time of the selected recent Art GAN models is also measured in this research using the same image resolution, the epoch of 1500, and GPU processor on both curated Coco African Mask and publicly available CelebFace datasets. The computational time for each model is shown in Fig. 10(a) and (b) respectively.

From Fig. 11(a), the proposed model has a lower computational time on the same dataset with the same GPU when compared to other techniques. Also, Fig. 11(b) exhibits that the proposed model has a reduced computational time of 45 hours in comparison to other models.

I. Benchmarking Proposed Model with Other Existing Recent Art GAN Techniques in Literature

Validating the performance of the proposed model on CelebFace datasets by comparing it to widely used GAN techniques for artistic image generation is one of the study's goals. The comparison of the proposed StyleVGG19-NST model with recent Art GAN approaches in terms of the dataset, IS score, FID, PSNR, and SSIM score is shown in Table VI.

From Table VI, the proposed model shows significantly better performance on the celeb datasets with FID of 21.49, IS of 11.69, PSNR of 29.98 and SSIM of 0.98 compared to other models. It's worth noting that the higher the IS, PSNR and SSIM of the model the better image quality generated, and the lower the FID the better the structural features that the proposed methodology exhibits when compared with selected baseline recent Art GAN techniques.

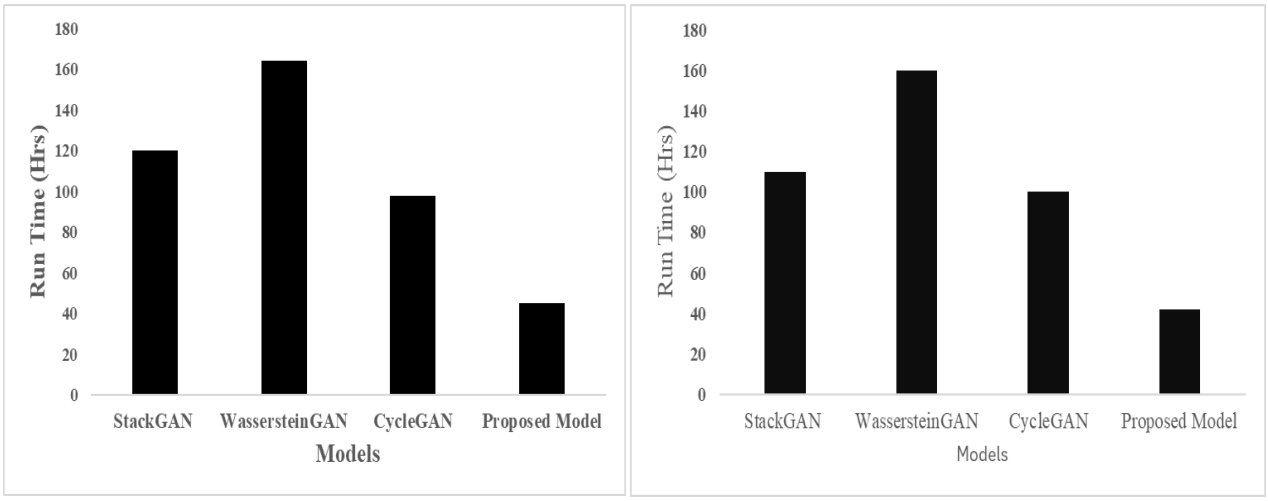


Fig. 11. Computational time of the proposed model with other recent art GAN models; (a) Coco african mask dataset and (b) publicly available CelebFace dataset.

TABLE VI. COMPARATIVE OF PROPOSED MODEL WITH OTHER EXISTING RECENT ART GAN MODELS

Ref	Model	Dataset	FID	IS	PSNR	SSIM
[26]	Boundedness and Continuity GAN (BC-GAN).	Celeb Face	22.8	8.40	19.76	0.76
[27]	Layered Recursive GAN (LR-GAN)	Celeb Face	-	7.17	-	0.89
[28]	Orthonormal	Celeb Face	27.40	2.9	13.56	-
[29]	Denoising Feature	Celeb Face	37.72	-	-	-
[30]	MSGAN	Celeb Face	28.44	-	17.78	-
Proposed Model	StyleVGG19-NST	Celeb Face	21.49	11.67	29.98	0.98

V. CONCLUSION

In this study, the proposed StyleVGG19-NST model was applied to the Coco African Mask artistic dataset and the publicly available CelebFace dataset to generate realistic artistic. The trained model network generated convincing artistic images compared to other baseline model images due to the ability of the proposed model to learn the rich and varied distribution of images. The use of a generator and discriminator network allows the proposed method to capture the spatial structure of an image, which is essential for many artistic image generation tasks. The application of this proposed model on curated artistic dataset images can transform works of art into different styles. Also, both generative loss and adversarial loss values are presented to apply constraints on brightness, colour contrast, and structure of the generated image. This allows the network to converge faster and retain more image detail as a result.

Qualitative and quantitative simulations were performed on the publicly available CelebFace dataset and the curated artistic dataset using the proposed method and other selected baseline methods. The qualitative comparison results show that the proposed model produces better and higher image quality in terms of structural and texture features compared with the baseline models. From the quantitative analysis perspective, the results of the proposed technique on the curated dataset have a high IS score of 11.67, a low FID score of 21.49, PSNR of 29.98 and SSIM of 0.98 while on the CelebFace dataset, the IS of 16.57, FID of 18.33, PSNR of 28.33 and SSIM of 0.93 which is superior compared to other methods used in the simulation. Furthermore, the computational period of the proposed method and baseline models on both curated and publicly available CelebFace datasets with the same training iterations processes show that the proposed technique has a lower computational time than to other models used in the simulation with 48 hours. The overall results of this research exhibit the potential of the proposed methodology for artistic image generation and suggest that the proposed model can be used for extensive image-generation tasks.

Further research is needed to explore how the proposed model performs. Also, future work can be investigated on how inherent biases in the training data of the proposed model can translate to the generated images. Nevertheless, the findings presented in this study can help artists in the generation of different artistic styles with less effort.

ACKNOWLEDGMENT

The authors would like to thank the Department of Computer Systems Engineering for their financial support.

REFERENCES

[1] S. Chakrabarty, R. F. Johnson, M. Rashmi, and R. Raha, "Generating Abstract Art from Hand-Drawn Sketches Using GAN Models," Proceedings of International Joint Conference on Advances in Computational Intelligence pp. 539–552, 2023, doi: https://doi.org/10.1007/978-981-99-1435-7_45

[2] H. Taherdoost and M. Madanchian, "AI Advancements: Comparison of Innovative Techniques," AI, vol. 5, no. 1, pp. 38-54, 2024. [Online]. Available: https://www.mdpi.com/2504-4990/5/1/3

[3] G. Iglesias, E. Talavera, and A. Díaz-Álvarez, "A survey on GANs for computer vision: Recent research, analysis and taxonomy," Computer Science Review vol. 48, p. 100553, 2023, doi: https://doi.org/10.1016/

[4] O. N. Oyelade, A. E. Ezugwu, M. S. Almutairi, A. K. Saha, L. Abualigah, and H. Chiroma, "A generative adversarial network for synthetization of regions of interest based on digital mammograms," Scientific Reports, vol. 12, no. 6166, 2022.

[5] Y. Deng et al., "StyTr2: Image Style Transfer with Transformers," CVF, pp. 11326-11336, 2022.

[6] N. Singh and T. Sandhan, "Learnable GAN Regularization for Improving Training Stability in Limited Data Paradigm," In Kaur, H., Jakhetiya, V., Goyal, P., Khanna, P., Raman, B., Kumar, S. (eds) Computer Vision and Image Processing. CVIP 2023. Communications in Computer and Information Science, vol. 2010, 2024, doi https://doi.org/10.1007/978-3-031-58174-8_45.

[7] S. He et al., "Context-aware layout to image generation with enhanced object appearance," In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp. 15049–15058, 2021.

[8] H. Guo, Z. Ma, X. Chen, X. Wang, J. Xu, and Y. Zheng, "Generating Artistic Portraits from Face Photos with Feature Disentanglement and Reconstruction," Electronics vol. 13, no. 5, p. 955, 2024, doi: https://doi.org/10.3390/electronics13050955.

[9] D. O. Esan, P. A. Owolawi, and C. Tu, Generative Adversarial Networks: Applications, Challenges, and Open Issues (intechopen). 2023.

[10] J. Z. Laith Alzubaidi, Amjad J. Humaidi, Ayad Al-Dujaili, Ye Duan, Omran Al-Shamma, J. Santamaría, Mohammed A. Fadhel, Muthana Al-Amidie, Laith Farhan "Review of deep learning: concepts, CNN architectures, challenges, applications, future directions," Journal of Big Data vol. 8, no. 53, pp. 1-73, 2021, doi https://doi.org/10.1186/s40537-021-00444-8.

[11] C. Dewi, R.-C. Chen, Y.-T. Liu, and H. Yu, "Various Generative Adversarial Networks Model for Synthetic Prohibitory Sign Image Generation," Applied Sciences, vol. 11, no. 7, p. 2913, 2021. [Online]. Available: https://www.mdpi.com/2076-3417/11/7/2913.

[12] R. T. A. Guna, R. Benitez, and O. K. Sikha, "Interpreting CNN Predictions using Conditional Generative Adversarial Networks," arXiv preprint arXiv:2301.08067, 2023, doi: https://arxiv.org/abs/2301.08067.

[13] R. Carbonne, S. Gauthier, and J. Leclerc, "Generative Model based on Genetic Algorithm for Artistic Image Generation," arXiv preprint arXiv:2301.08067, 2023. [Online]. Available: Retrieved from https://arxiv.org/abs/2301.08067.

[14] T. Zhu, J. Chen, R. Zhu, and G. Gupta, "StyleGAN3: Generative Networks for Improving the Equivariance of Translation and Rotation," arXiv preprint arXiv:2307.03898, 2023. [Online]. Available: Retrieved from https://arxiv.org/abs/2307.03898.

[15] W. R. Tan, C. S. Chan, H. a. E. Aguirre, and K. Tanaka, "ArtGAN: Artwork Synthesis With Conditional Categorical GANs," 2020.

[16] T. X. H. Zhang, and H. Li, "StackGAN: Text to Photo-Realistic Image Synthesis With Stacked Generative Adversarial Networks," IEEE International Conference on Computer Vision, pp. 5908–5916, 2020.

[17] Y. Jiang and J. Li, "Generative Adversarial Network for Image Super-Resolution Combining Texture Loss," Applied Sciences, vol. 10, no. 5, p. 1729, 2020, doi: https://doi.org/10.3390/app10051729.

[18] B. J. Sowmya, Meeradevi, and S. Shedole, "Generative adversarial networks with attentional multimodal for human face synthesis," Indonesian Journal of Electrical Engineering and Computer Science, vol. 33, no. 2, pp. 1205-1215, 2024, doi: 10.11591/ijeecs.v33.i2.pp1205-1215.

[19] D. Victor, "COCO-AFRICA: A Curation Tool and Dataset of Common Objects in the Context of Africa," Conference on Neural Information Processing, 2nd Black in AI Workshop, 2018.

[20] A. Bhattad, D. McKee, D. Hoiem, and D. Forsyth, "Examining Pathological Bias in a Generative Adversarial Network Discriminator: A Case Study on a StyleGAN3 Model," arXiv. 2023., 2023.

[21] S.-W. Park, J.-S. Ko, J.-H. Huh, and J.-C. Kim, "Review on Generative Adversarial Networks: Focusing on Computer Vision and its Applications," Electronics, vol. 10, 2021.

[22] T. Kramberger, "LSUN-Stanford Car Dataset: Enhancing Large-Scale Car Image Datasets Using Deep Learning for Usage in GAN Training," Applied Sciences, 2020, doi: 10. 10.3390/app10144913.

[23] X. Jia, S. Liu, and Y. Chen, " Enhanced Feature Extraction with VGG-19 for StyleGAN-based Image Synthesis," International Journal of Computer Vision and Machine Learning, vol. 12, no. 4, pp. 45-60, 2023, doi: https://doi.org/10.1016/j.ijcvml.2023.03.004.

[24] A. A. Justin Johnson, and Li Fei-Fei, "Perceptual Losses for Real-Time Style Transfer and Super-Resolution " 2016.

[25] I. Vaccari, V. Orani, A. Paglialonga, E. Cambiaso, and M. Mongelli, "A Generative Adversarial Network (GAN) Technique for Internet of Medical Things Data," Sensors, vol. 3726, no. 21, pp. 1-14, 2021, doi: https://doi.org/10.3390/s21113726.

[26] K. Liu and G. Qiu, "Lipschitz constrained GANs via boundedness and continuity," Neural Computing and Applications, vol. 32, pp. 18271-18283, 2020.

[27] Y. J, K. A, and B. D, "Lr-Gan: Layered Recursive Generative Adversarial Networks for Image Generation," 2017.

[28] T. Miyato, T. Kataoka, and M. Koyama, "Spectral Normalization for Generative Adversarial Networks," 2018.

[29] D. Warde-Farley and Y. Bengio, "Improving Generative Adversarial Networks with Denoising Feature Matching," presented at the ICLR 2017, 2017.

[30] A. Karnewar and O. Wang, "MSG-GAN: Multi-Scale Gradients for Generative Adversarial Networks," arXiv:1903.06048v4 [cs.CV] 12 Jun 2020, pp. 1-18, 2020.

Applied to Art and Design Scene Visual Comprehension and Recognition Algorithm Research

Yuxin Shi
College of Arts and Media Science and Technology, College of Hubei University of Arts and Science, Xiangyang 441025, Hubei, China

***Abstract*—Combining advanced intelligent algorithms to improve the scene visual understanding and recognition method for art design can not only provide more inspirations and creative materials for artists, but also improve the efficiency and quality of art creation, and provide scientific and accurate references of artworks. Focusing on the art design scene visual understanding and recognition problem, a scene visual understanding and recognition method based on the intelligent optimisation algorithm to optimise the structural parameters of the multilayer perception machine is proposed. Firstly, the scene visual recognition method is outlined and analyzed, and the application scheme of multilayer perceptron in the understanding and recognition problem is designed; then, for the problems of the multilayer perceptron model, such as the training does not generalize, combined with the Pond's optimization algorithm, the training parameters of the multilayer perceptron model are optimized, and the visual understanding and recognition scheme of the art design scene is designed; finally, the proposed model is verified with the image dataset, and the scene visual understanding and recognition accuracy reaches 0.98, compared with other models, the proposed method has higher recognition accuracy. This research solves the problem of scene visual understanding and recognition, and applies it to the field of art design to improve the efficiency of art design assistance.**

Keywords*—Art design; scene visual understanding and recognition; multilayer perceptron; pond goose algorithm; image dataset

I. INTRODUCTION

Traditional art design methods are carried out using software or hand-drawing and so on [1]. With the rapid development of multimedia technology, the art design method based on computer technology has gradually been highly valued by experts and scholars in the field of art design [2]. At present, art design software presents diversified types and functions, which enriches the creativity of visual effects, improves the working method of art design, and becomes one of the future trends of art design research [3]. Scene visual understanding and recognition, as one of the key technologies most widely used in the field of art, uses computer vision and image processing technology to deeply understand and analyse the scenes in images and videos [4]. Art design that combines scene visual understanding and recognition algorithms can provide artists with more inspiration and creative materials, and can also improve the efficiency and quality of art creation and provide scientific and accurate references for artwork [5]. Therefore, the study of art design methods combining scene visual understanding and recognition algorithms is an important theoretical research significance for art design to assist decision-making and creation. According to the principle of design process, the research on scene visual understanding and recognition for art design generally includes the research contents of scene image segmentation, extraction of target features, and understanding and recognition model construction [6]. Scene image segmentation processes the image from colour segmentation, morphological processing, etc.; extracting target features captures the key information of the image, i.e., shape, colour, texture, edges, etc.; as a key part of the scene visual understanding and recognition problem [7], deep learning or machine learning algorithms are trained and constructed based on the annotated feature sample set. According to the principle of the core approach, scene visual understanding and recognition methods can be classified into search-based scene visual understanding and recognition methods [8], template matching-based scene visual understanding and recognition methods [9], and language model-based scene visual understanding and recognition methods [10].Sandrine et al [11] obtained image descriptions by constructing the mapping relationship between images and texts using similarity to obtain the compliant utterances; Mustafa et al [12] proposed a visual scene description scene based on coordinate position by analysing a large number of images and annotation information; Daniel et al [13] improved convolutional neural network using encoding-decoding network to further improve the accuracy of the scene visual understanding recognition model; Chen et al [14] combined the current state of the art of domestic and international research and used encoding-decoding network and attention mechanism model to construct and analyse the visual scene understanding model. Although the current scene visual understanding and recognition algorithms for art design have achieved a lot of results and application progress, there are still some challenges, such as the generalisation ability needs to be improved, more sensitive to the smile change of the data, the real-time needs to be further improved, and the consumption of computational resources is more [15].

This paper focuses on the scene visual understanding and recognition problem for art design, combines the intelligent optimization algorithm to optimize the network structure parameter framework paradigm [16], and puts forward a scene visual understanding and recognition method based on the optimization of the Pond's Goose algorithm to improve the multilayer perceptual machine model. Aiming at the characteristics of the scene visual understanding and recognition problem for art design, the scene visual understanding and recognition model algorithm and its

optimisation strategy are analysed and introduced, and at the same time, the method of Pond's Goose algorithm optimisation recognition is applied to the scene visual understanding and recognition problem for art design, and the art-related data is used to compare the other five recognition models, which verifies the high efficiency of the proposed method and the high accuracy and real-time performance.

II. Overview of Identification Methods

A. Analysis of the Problem

1) Introduction to the issue: In the field of art design, scene visual understanding recognition algorithms are mainly used to automatically identify and classify images in art works, or automate the processing of art works to improve the efficiency and quality of art creation [17]. In the art market, scene visual understanding algorithms can be used for art appraisal and valuation to provide a more scientific and accurate reference for art transactions. The main applications of scene visual understanding recognition algorithms in the field of art design are shown in Fig. 1.

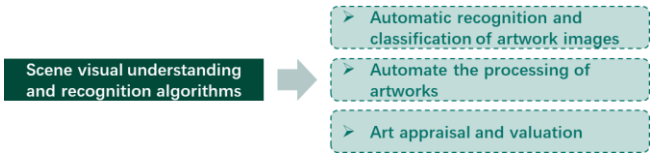


Fig. 1. Scene visual understanding recognition algorithm application.

In this paper, the scene visual understanding recognition algorithm is used to extract the target features in the complex scene image, complete the combination design of the extracted target, and realise the art assistance, so the research in this paper is mainly used to automatically identify and classify the images in the art work. For the problem of target recognition and classification in complex images, this paper adopts multi-layer perceptron to construct the mapping relationship between target features and target information, as shown in Fig. 2.

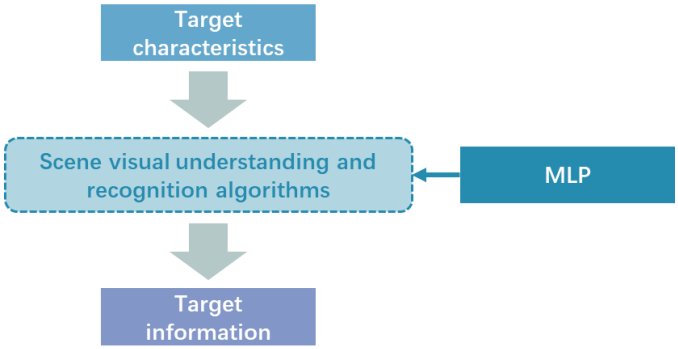


Fig. 2. Scene visual understanding and recognition solution architecture.

2) Multimodal feature extraction: In order to obtain as much information about the image target features and improve the accuracy of the understanding recognition algorithm, this paper firstly rubs the coarse segmentation technique [18] for colour segmentation of the scene image; then, the effective description region is processed morphologically; finally, the scale invariant feature transform technique is used to extract the target features in the candidate region. The multimodal feature extraction is shown in Fig. 3, and the principle of Scale Invariant Feature Transform (SIFT) [19] technique is shown in Fig. 4.

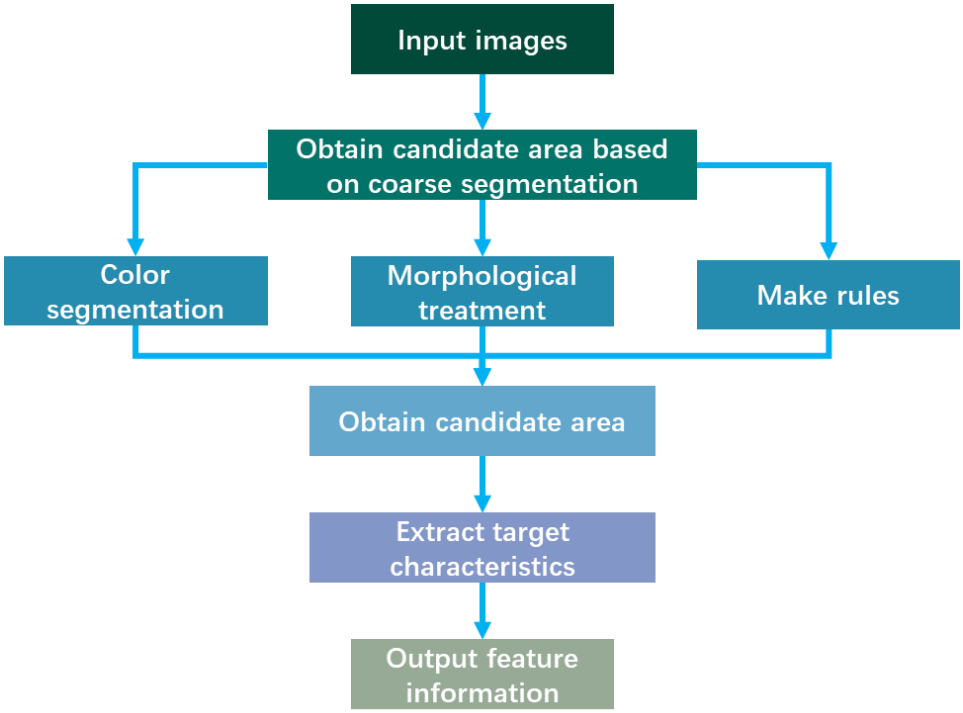


Fig. 3. Multimodal feature extraction.



Fig. 4. Principle of SIFT technology.

B. Overview of Multilayer Perceptron

Multilayer perceptrons (MLP) [20] is the introduction of multiple hidden layers on top of a single-layer neural network. MLP networks can acquire complex relationships between inputs through hidden layer nodes, the inputs undergo an activation function, which produces the final prediction through the output layer. MLP networks require at least three layers of artificial neurons. The MLP input layer will be make each node connected to the hidden layer according to the input sample set vector dimensions. The hidden layer generally uses a sigmoid function to make the feature data from the hidden layer to the output layer very smooth.

$$f(x)=1\Big/1+\exp\left(-\sum_j w_j x_j - b\right) \tag{1}$$

Where, $f(x)$ denotes the hidden layer output, w_j denotes the node weights and b denotes the node bias.

The parameters of MLP network include the number of hidden layer nodes, hidden layer node excitation function, and connection weight. For the selection of the number of hidden layer nodes, this paper adopts an experimental approach, taking different numbers of nodes respectively, observing the recognition accuracy, and taking the number of nodes with the largest accuracy. For the hidden layer node excitation function, this paper adopts the sigmoid function as the activation function [21]; for the connection weight, MLP generally adopts the back propagation algorithm [22]. Backpropagation can be used to train feed-forward artificial neural networks with any layers and any number of hidden units, but the practical limitations of computational power will constrain the ability of backpropagation, therefore, this paper adopts the Pond's algorithm [23] as the optimisation algorithm for the selection of MLP network structure parameters.

$$C(w,b)=\frac{1}{2n}\sum_x \left\|y(x)-a\right\|^2 \tag{2}$$

$$w_k = w_k - \frac{\eta}{m}\sum_j \frac{\partial C_{X_j}}{\partial w_k} \tag{3}$$

$$b_k = b_k - \frac{\eta}{m}\sum_j \frac{\partial C_{X_j}}{\partial b_k} \tag{4}$$

where w and b denote the structural weights and biases of the MLP network, respectively.

MLP is applied to many fields, including image recognition, speech processing, language analysis, etc., and is mainly applied to complex nonlinear mapping relationship processing.

C. Application of Multilayer Perceptron Machine in Comprehension Recognition

The application of MLP to the problem of English recognition for visual understanding of scenes oriented to the field of art and design is mainly shown in Fig. 5. The input vector of MLP is the target features extracted from the image, and the output vector is the category of the target.

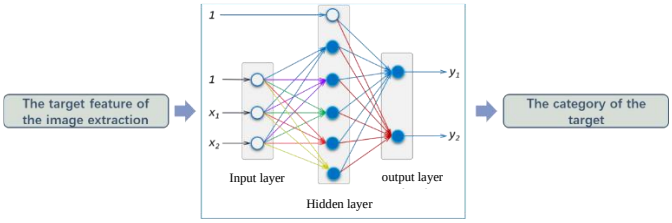


Fig. 5. MLP applications

III. Recognition Method Optimisation Algorithm

In order to improve the recognition accuracy of the multilayer perceptron network, this paper chooses the pond goose algorithm as the recognition method optimisation algorithm, the specific optimisation process is as follows:

A. Pond's Goose Algorithm

The Gannet Optimization Algorithm (GOA) [24] is a natural heuristic optimization algorithm that mimics the behaviour of the pond goose. The algorithm is used to explore optimal solutions in the search space by mathematically modelling the U- and V-diving behaviours of the Pond Goose during foraging as well as sudden rotations and random wandering. The GOA is designed to balance the capabilities of global exploration and local exploitation in order to improve the efficiency of solving engineering optimisation problems.

1) Initialisation phase: The GOA algorithm uses a random initialisation strategy for population generation, which generates uniformly distributed GOA population locations from each dimension using upper and lower bounds of the search space.

2) Exploration phase: The exploration phase of the GOA algorithm simulates the U- and V-diving behaviour of the pond goose after finding prey in the air (shown in Fig. 6) for global search.

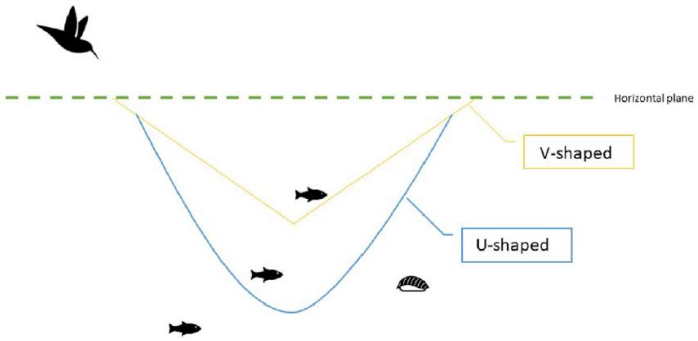


Fig. 6. The GOA algorithm U- and V-dive behaviour.

Pond geese dive in a U-shaped dive when prey is at greater depths:

$$a = 2 \times \cos\left(2 \times \pi \times r_2\right) \times t_1 \tag{5}$$

Pond geese dive in a V-dive when prey is in a relatively shallow position:

$$a = 2 \times V\left(2 \times \pi \times r_3\right) \times t_1 \tag{6}$$

$$V(x) = \begin{cases} -\frac{1}{\pi} \times x + 1 & x \in (0, \pi) \\ \frac{1}{\pi} \times x11 & x \in (\pi, 2\pi) \end{cases} \tag{7}$$

$$t_1 = 1 - \frac{t}{T_{\max}} \tag{8}$$

Where t denotes the number of contemporary iterations, $T_{\max}$ denotes the maximum number of iterations, and r_2 and r_3 denote random numbers, respectively.

The position of the pond goose is updated by introducing a random variable q and randomly selecting either a U-dive approach or a V-dive approach:

$$MX_i(t+1) = \begin{cases} X_i(t) + u_1 + u_2 & q \geq 0.5 \\ X_i(t) + v_1 + v_2 & q < 0.5 \end{cases} \tag{9}$$

$$u_2 = A \times \left(X_i(t) - X_r(t)\right) \tag{10}$$

$$v_2 = B \times \left(X_i(t) - X_m(t)\right) \tag{11}$$

$$A = \left(2 \times r_4 - 1\right) \times a \tag{12}$$

$$B = \left(2 \times r_5 - 1\right) \times b \tag{13}$$

Where r_4 and r_5 denote random numbers, u_1 is a random number between -a and a, v_1 is a random number between -b and b, $X_i(t)$ denotes the position information of the i-th song pond goose individual in the t-th iteration, $X_r(t)$ denotes the position of the randomly selected pond goose individual, $X_m(t)$ denotes the average of the position of all the pond geese individuals, and the calculation is as follows:

$$X_m(t) = \frac{1}{N} \sum_{i=1}^{N} X_i(t) \tag{14}$$

3) Development phase: The GOA algorithm development phase simulates the capture behaviour of the pond goose in the water, and based on the capture ability of the pond goose decides whether to perform a random swim or a sudden rotation to capture the prey (as shown in Fig. 7).

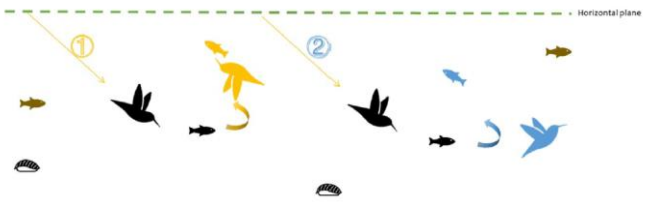


Fig. 7. GOA algorithm development phase.

When the pond goose enters the water, its catching ability C is greater than or equal to c, it will suddenly rotate to catch fish; its catching ability C is less than c, it will give up catching fish and march randomly, the Levy flight model is used to simulate the marching of the pond goose, and the specific model of the position update is as follows:

$$MX_i(t+1) = \begin{cases} t_1 \times \delta \times \left(X_i(t) - X_{best}(t)\right) + X_i(t) & C \geq c \\ X_{best}(t) - \left(X_i(t) - X_{best}(t)\right) \times P \times t_2 & C < c \end{cases} \tag{15}$$

$$C = \frac{1}{R \times t_2} \tag{16}$$

$$t_2 = 1 + \frac{t}{T_{\max}} \tag{17}$$

$$R = \frac{M \times vel^2}{L} \tag{18}$$

$$L = 0.2 + \left(2 - 0.2\right) \times r_6 \tag{19}$$

$$\delta = C \times \left|X_i(t) - X_{best}(t)\right| \tag{20}$$

$$P = Levy(D) \tag{21}$$

$$Levy(D) = 0.01 \times \frac{\mu \times \sigma}{|v|^{\frac{1}{\beta}}} \tag{22}$$

$$\sigma = \left(\frac{\Gamma(1+\beta) \times \sin\left(\frac{\pi\beta}{2}\right)}{\Gamma\left(\frac{1+\beta}{2}\right) \times \beta \times 2^{\frac{\beta-1}{2}}} \right)^{\frac{1}{\beta}} \tag{23}$$

Where, r_6 is a random number, M denotes the mass of the pond goose, vel denotes the velocity of the pond goose with the value of 1.5 m/s, c is a constant, which generally takes the values of c=0.2 and β=1.5, $X_{best}(t)$ denotes a constant, and μ and σ are random numbers, respectively.

The pseudo-code and flowchart of the GOA algorithm are shown in Fig. 8 and Fig. 9, respectively.

```
Algorithm1: Gannet Optimization Algorithm (GOA)
Set GOA parameters;
Initialize GOA population based on random distribution;
Obtain memory matrix MX;
Calculate X fitness;
While t<=Max_t
    if rand>0.5
        for MXi do
            if q>=0.5, update position using U-shaped;
            else, update position using V-shaped;
            end if
        end for
    else
        for Mxi do
            if c>=0.2, update position with a sudden tuning;
            else, update position with Levy movement;
        end
    end
    Calculate MXi fitness, and update better position;
End while
```

Fig. 8. Pseudo-code of the GOA algorithm.

B. *Optimisation of Recognition Methods by the Pond's Goose Algorithm*

In order to increase the accuracy of the multilayer perceptron recognition method, this paper uses the Pond's Goose algorithm to optimise the multilayer perceptron network. The paradigm of the Pond's Goose algorithm to optimise the multilayer perceptual machine network is shown in Fig. 10, and the GOA algorithm takes w and b as the optimisation variables, and $C(w,b)$ as the fitness function. The flow of the MLP scene visual understanding recognition method based on the GOA algorithm is shown in Fig. 11, and the specific steps are as follows:

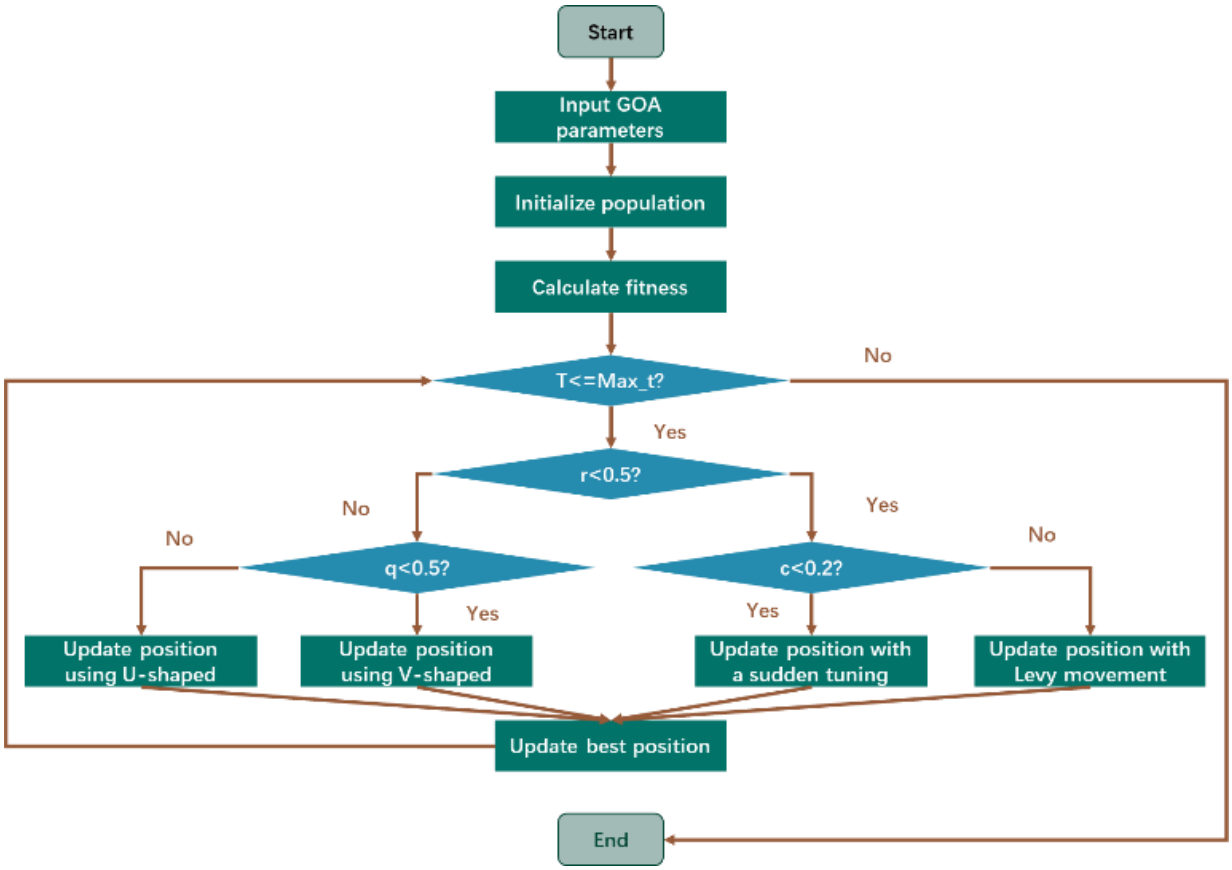


Fig. 9. Flowchart of GOA algorithm.

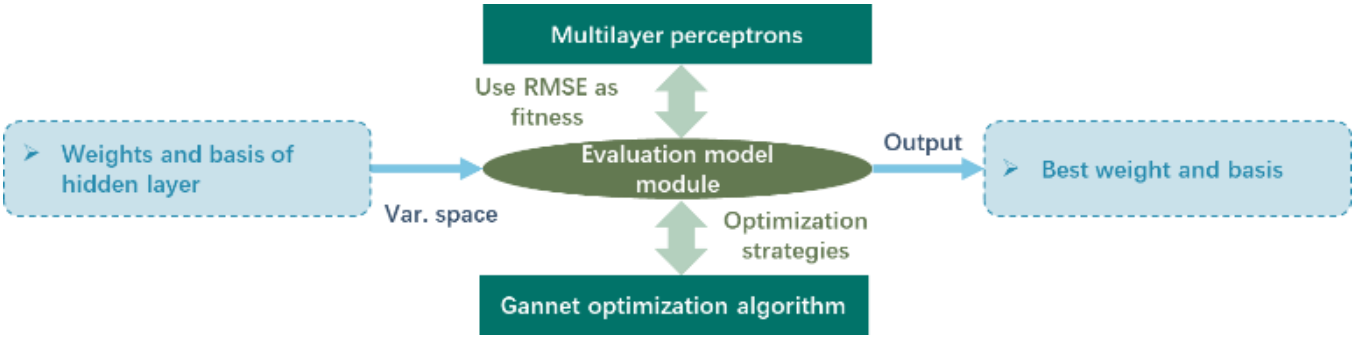


Fig. 10. GOA algorithm to optimise the MLP network paradigm.

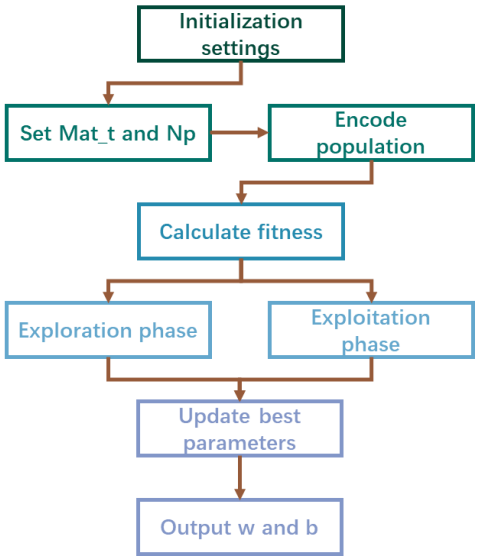


Fig. 11. GOA-MLP step-by-step diagram.

- Step 1: Initialisation setup. 1) Set the maximum number of iterations as well as the number of populations and algorithmic control parameters for the GOA algorithm optimised MLP network; 2) Initialise the GOA algorithm populations using real number coding;
- Step 2: Calculate the $C(w,b)$ value and determine the optimal structural parameters for the current number of iterations based on the error value;
- Step 3: Behavioural simulation models such as U- and V-diving behaviours during foraging, as well as sudden rotations and random wandering, were used to update information on the location of individuals in the population;
- Step 4: Calculate the fitness value, update the structural parameters, and determine whether the maximum number of iterations or the optimal solution of the GOA-MLP algorithm is no longer changing;
- Step 5: Output the structural parameters of the optimal MLP model.

IV. Understanding the Application of Recognition Methods in Art and Design

A. Application Programmes

The visual understanding and recognition method of art design scene based on GOA-MLP model performs colour segmentation and morphological processing of the scene through coarse segmentation technology and formulates rules to extract candidate scene regions, uses SIFT technology to extract internal features from the feature regions, annotates the sample data, and trains the scene visual understanding and recognition algorithm based on the GOA-MLP model, to complete the visual understanding and recognition of the scene based on the art design [25]. Focusing on the principle of visual understanding and recognition method of art design scene based on GOA-MLP model, this section designs the visual understanding and recognition scheme of art design scene based on GOA-MLP model (shown in Fig. 12), and gives the application analysis.

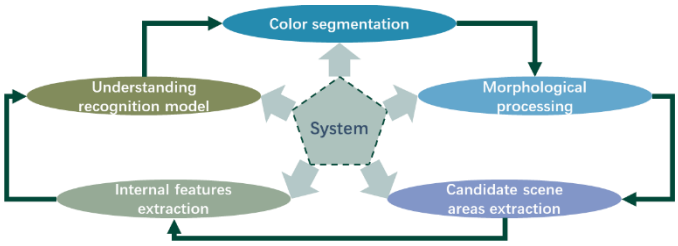


Fig. 12. Artistic design scene visual understanding and recognition programme.

In Fig. 12, the visual understanding and recognition system for art and design scenes based on GOA-MLP model consists of modules such as colour segmentation, morphological processing, extraction of candidate scene regions, extraction of internal features, and construction of scene visual understanding and recognition model.

B. Application Steps

Based on the analysis of the Art and Design Scene Visual Understanding and Recognition Programme, the modules are described below:

1) Colour segmentation module: For the colour segmentation problem, HSV colour space is used, where H is converted by the following formula:

$$H=\begin{cases}0 & \max=\min\\ 60\times\dfrac{G-B}{\max-\min} & \max=R\,\&\,G\geq B\\ 60\times\dfrac{G-B}{\max-\min}+360 & \max=R\,\&\,G<B\\ 60\times\dfrac{G-B}{\max-\min}+120 & \max=G\\ 60\times\dfrac{G-B}{\max-\min}+240 & \max=B\end{cases}\tag{24}$$

Where $\max$ and $\min$ denote the maximum and minimum values of the pixel for each channel in the RGB colour space, respectively.

2) Morphological processing module: To address the problem of noise and breakage in colour segmented images, this subsection uses morphological processing methods to reduce the effect of noise and to obtain connected regions. In morphological processing technique, the broken portion of the candidate region is re-cracked using an expansion operation to fill the broken contour lines [26].

3) Candidate scene area analysis module: Aiming at the problem that there are still uneliminated interference regions in the image after morphological processing, this paper uses constraints to extract candidate scene regions. The specific satisfaction conditions are as follows:

$$\begin{cases} S_i \geq S_{\min} \cap S_i \leq S_{\max} \\ \dfrac{L_i}{W_i} \geq \left(\dfrac{L}{W}\right)_{\min} \cap \dfrac{L_i}{W_i} \leq \left(\dfrac{L}{W}\right)_{\max} \\ \dfrac{S_i}{L_i \times W_i} \geq \left(\dfrac{S}{L \times W}\right)_{\min} \end{cases} \quad (25)$$

Among them, L , W and S denote the width, height and area of the connected area, respectively. $S_{\min}$, $S_{\max}$ and $\left(S/\left(L\times W\right)\right)_{\min}$ denote the minimum and maximum values of the area of the connected area, respectively; $\left(L/W\right)_{\min}$ and $\left(L/W\right)_{\max}$ denote the maximum value of the aspect ratio of the connected area, respectively; and denotes the minimum value of the duty cycle.

4) Internal feature extraction module: In order to reduce the computational complexity and retain the key information, this subsection adopts the SIFT technique to extract the target features inside the candidate region. The idea of SIFT-based target feature extraction method is to take the SIFT feature point as the centre point, calculate the gradient and direction of the pixels within the range of 16×16 , and at the same time use dense sampling to obtain more information.

5) Scene visual understanding recognition model building module: In order to complete the art design method based on the scene visual understanding recognition algorithm, this paper uses GOA-MLP algorithm training to construct the scene visual understanding recognition model, and the specific training process is shown in Fig. 13. Firstly, the data is divided, then the MLP model is improved by using GOA optimisation, secondly, the MLP structure parameters are reconstructed by using the training set, and finally the understanding recognition model is used to complete the art design.

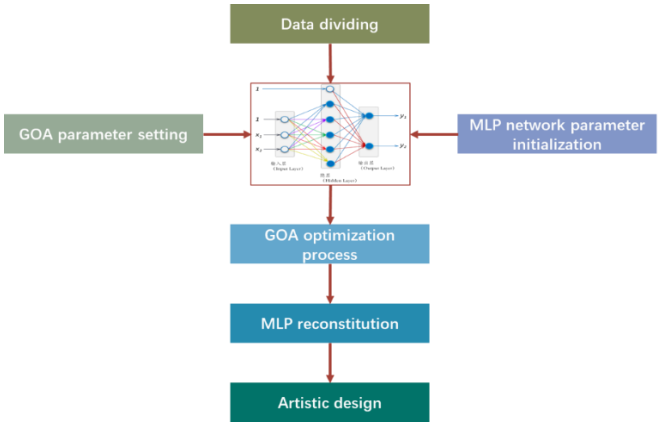


Fig. 13. Scene visual understanding recognition model building module.

V. Simulation Experiment

A. Experimental Set-up

In order to verify that the GAO algorithm improves the accuracy of the MLP model, this paper uses the sine-cosine optimisation algorithm (SCA) [27], the Harris hawk optimisation algorithm (HHO) [28], the butterfly optimisation algorithm (BOA) [29], the crow's tern optimisation algorithm (STOA) [29], and the gannet optimisation algorithm (GOA) to make comparisons with the specific parameter settings as shown in Table I. The structure of the MLP model to be optimised includes one input layer, three hidden layers and one output layer, the activation function is a Sigmoid function, the number of nodes in the hidden layer is divided into, the number of populations of SCA, HHO, BOA, STOA and GOA algorithms is 100, the maximum number of iterations is 500, and the condition for satisfying the optimal result output is to reach the maximum number of iterations.

TABLE I. Experimental Parameter Settings

Arithmetic	Parameterisation
MLP	The activation function is Sigmoid function and the classifier is Softmax, including input layer, hidden layer (3 layers), output layer, hidden layer nodes are [30]
SCA-MLP [27]	The MLP structure parameters are set as above, with a set to 2
HHO-MLP [28]	The MLP structure parameters are set as above, E0 is a random number and E1 decreases linearly from 2 to 0
BOA-MLP [29]	The MLP structural parameters are set as above, with a transition probability of 0.6, a force index of 0.1 and a perceptual mode value of 0.01
STOA-MLP [30]	Sa decreases linearly from 2 to 0.
GOA-MLP	MOP_max is 1, MOP_min is 0.2, Alpha is 5, Mu is 0.499

This paper uses the scene visual understanding feature extraction method parameter settings are shown in Table II.

TABLE II. Parameter Settings for Visual Feature Extraction Methods

Project	Parameter values	Note
Optimal thresholds	90	RGB colour space
Optimal lower and upper thresholds	200, 280	HSV colour space
S_{max}	400	Maximum area of connected areas
S_{min}	(L x W)/2	Minimum area of connected areas
$(L/W)_{max}$	0.2	Maximum Aspect Ratio for Connected Areas
$(L/W)_{min}$	4.6	Minimum aspect ratio of connected areas
$(S/(L\times W))_{min}$	0.7	Duty Cycle Min.

The hardware platform for the experimental development environment is Pentium IV 3.7 GHz CPU and 4 RAM with a memory size of 3 GB. The software environment includes VisualStudio, Win10 operating system, Multigen Creator, Visual C++ 4.0 programming language, Matlab2021a programming language.

The dataset was adopted from the image dataset of the Massachusetts Institute of Technology (MIT), including 2688 scene images, including scenes of mountains, cities, coasts, flowers, etc. as shown in Fig. 14.



Fig. 14. Sources of data sets.

B. Analysis of Performance Test Results

In order to verify the effectiveness as well as the efficiency of the method proposed in this paper, this paper firstly analyses the visual processing methods of scenes oriented to artistic design, and secondly, tests and analyses the recognition methods for comparison.

1) Scene visual processing analysis: According to the above simulation environment, parameter settings and data set for target information extraction, the extraction results are obtained as shown in Fig. 15.

Obviously, as obtained from Fig. 15, the scene visual understanding algorithm can effectively extract the floral information of the known image, solve the problem of poor clarity of the output image, reduce the influence of noise, and improve the overall design of the image.

2) *Analysis of identification test results:* In order to enhance the effect of validation experiments, this paper randomly extracts five sets of data sets from the database to carry out experimental analysis, each time the extracted data sets include 2000 images, of which the training set reaches 1500, the test set is 350, and the rest is the validation set, and the specific test results are shown in Table III, Fig. 16, and Fig. 17.

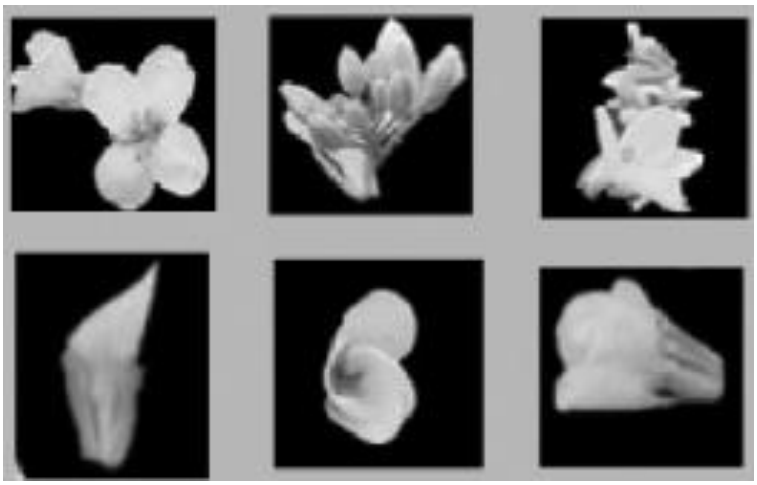


Fig. 15. Target information extraction results.

From Table III and Fig. 16, it is easy to find that from the dataset point of view, the accuracy of GOA-MLP comprehension and recognition method is better than that of MLP, SCA-MLP, HHO-MLP, BOA-MLP, and STOA-MLP.It can be seen that the accuracy of MLP comprehension and recognition method based on optimisation algorithms is better than that of MLP method, and the accuracy of GOA optimised MLP model is better than other optimisation algorithms.

Fig. 17 demonstrates the optimisation convergence curves for each algorithm of SCA-MLP, HHO-MLP, BOA-MLP, STOA-MLP, and GOA-MLP. From Fig. 17, it can be obtained that: with the increase in the number of iterations, the value of the adaptation function of the comprehension recognition accuracy of the GOA optimised MLP model increases, and the accuracy of the convergence is higher than that of the algorithms such as SCA, HHO, BOA, and STOA; and the optimised convergence of the GOA-MLP model is stable up to the vicinity of 0.98.

TABLE III. DATA SET TEST IDENTIFICATION ACCURACY RESULTS (%)

Data set	MLP	SCA-MLP	HHO-MLP	BOA-MLP	STOA-MLP	GOA-MLP
1	67.34	86.98	87.32	84.22	93.90	97.98
2	66.21	86.10	87.35	83.78	92.53	97.46
3	69.88	88.38	89.25	87.90	95.16	98.85
4	65.20	86.12	86.95	83.15	90.74	97.00
5	67.58	87.24	88.65	85.37	92.88	98.05

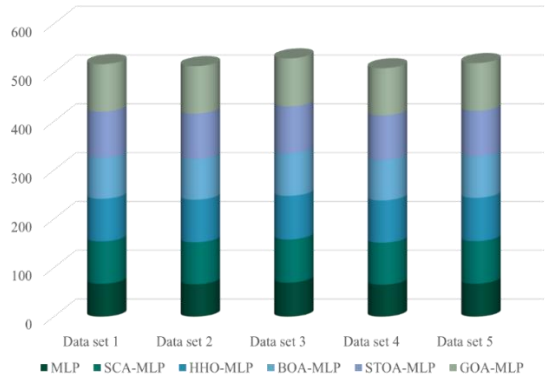


Fig. 16. Recognition results of each algorithm.

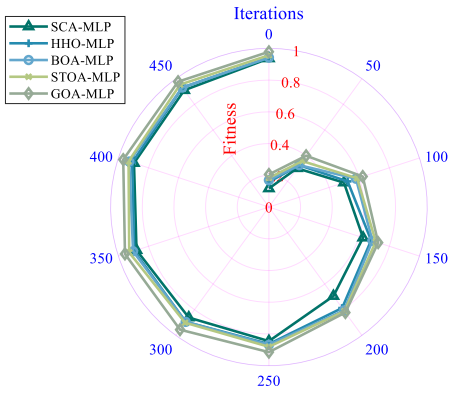


Fig. 17. Convergence process of each optimisation algorithm.

VI. Conclusion

Aiming at the problem of visual understanding and recognition effect of diverse scenes of art design, this paper proposes a visual understanding and recognition method of scenes oriented to art design based on GOA-MLP, which makes the GOA algorithm optimise the MLP structural parameters and improve the understanding and recognition accuracy. Experiments with MIT image datasets lead to the following conclusions: using the GOA algorithm to optimise the MLP structural parameters and using it in the art design scene visual understanding and recognition problem improves the art scene understanding and recognition accuracy and increases the art design effect. By analysing five types of image datasets, the GOA-MLP based scene visual understanding and recognition method has higher recognition accuracy. This study only focuses on the accuracy of the understanding and recognition method, and the subsequent re-recognition time method still requires in-depth research.

References

[1] Zhang F S , Ge D Y , Song J , Xiang W J. Outdoor scene understanding of mobile robot via multi-sensor information fusion[J]. Information Integration, 2022.

[2] Glavan A ,Talavera, Estefanía.InstaIndoor and multi-modal deep learning for indoor scene recognition[J].Neural Computing and Applications,. 2022, 34(9):6861-6877.

[3] Zhang W , Wang Y , Ni B , Yang X. Fully context-aware image inpainting with a learned semantic pyramid[J]. Pattern Recognition Society, 2023:143.

[4] Rafique A A , Ghadi Y Y , Alsuhibany S A , Chelloug S A, Jalal A, Park J. CNN Based Multi-Object Segmentation and Feature Fusion for Scene Recognition[J]. Computers, Materials and Continuum (English), 2022.

[5] Nourali K , Dolkhani E .Scene text visual question answering by using YOLO and STN[J].International Journal of Speech Technology, 2024, 27(1):69-76 .

[6] Su T , Shi Y , Xie C , Luo W, Ye H, Xu L. A hybrid loss balancing algorithm based on gradient equilibrium and sample loss for understanding of road scenes at basic-level[J].Pattern Analysis and Applications, 2022, 25(4):1041-1053.

[7] Narazaki Y , Pang W , Wang G Chai W. Unsupervised Domain Adaptation Approach for Vision-Based Semantic Understanding of Bridge Inspection Scenes without Manual Annotations[J].Journal of Bridge Engineering, 2024, 29(2):4023118.1-4023118.16.

[8] Bendall R C A , Eachus P , Thompson C .The influence of stimuli valence, extraversion, and emotion regulation on visual search within real-world scenes [J].Scientific reports, 2022, 12(1):948.

[9] Arif A , Ghadi Y Y , Alarfaj M , Jalal A, Kamal S, Kim D S. Human Pose Estimation and Object Interaction for Sports Behaviour[J]. Computers, Materials and Continuum (English), 2022(7):18.

[10] Pan Y .On visual understanding[J].Frontiers of Information Technology & Electronic Engineering, 2022, 23(9):1287-1289.

[11] Sandrine B , Cohen A A , Paul F .How iconic news images travel: republishing and reframing historic photographs in Israeli newspapers[J]. Communication, 2022(1):1.

[12] Mustafa A , Russell C , Hilton A .4D Temporally Coherent Multi-Person Semantic Reconstruction and Segmentation[J]. Computer Vision, 2022, 130(6):1583-1606.

[13] Daniel S P .Recognition of Family Life by Children Living in Kinship Care Arrangements in England[J].

[14] Chen S , Demachi K , Dong F .Graph-based linguistic and visual information integration for on-site occupational hazards identification[J]. Automation in construction, 2022(5):137.

[15] Pan Y .On visual understanding[J]. Frontiers in Information and Electronic Engineering: English Edition, 2022, 23(9):1287-1289.

[16] Ogiela U , Snasel V .Predictive intelligence in evaluation of visual perception thresholds for visual pattern recognition and understanding[J]. Information Processing & Management: Libraries and Information Retrieval Systems and Communication Networks: an International Journal, 2022(2):59.

[17] Tu J , Wu G , Wang L .Dual Graph Networks for Pose Estimation in Crowded Scenes[J].International Journal of Computer Vision, 2024(3):132.

[18] Fang Y .Instantaneous visual imaging of latent fingerprints in water[J]. Science in China:Chemistry, 2022.

[19] Kishinami H , Itoyama K , Nishida K , Nakadai K. Visual Scene Reconstruction based on Echolocation with a Generative Adversarial Network[J]. of the Robotics Society of Japan, 2022, 40(4):351-354.

[20] ZHANG Xiaoyan, XIANG Mian, ZHU Li, ZHOU Bintao, LIU Hongxiao, DUAN Yaqin. Ultra-short-term wind power prediction based on MLP-BiLSTM-TCN combination[J]. Journal of Hubei University for Nationalities (Natural Science Edition),2023,41(04):513-519+529.

[21] Yanni Lu. Atmospheric temperature prediction based on MLP and Transformer model[J]. Journal of Yuncheng College,2024,42(03):43-47.

[22] HOU Kepeng,BAO Guangtuo,SUN Huafen.Application of SSA-MLP model in rocky slope stability prediction[J]. Journal of Safety and Environment,2024,24(05):1795-1803.

[23] QI Huiling, HU Hongping, BAI Yanping, HOU Qiang. Prediction of novel coronavirus outbreak based on optimised BP neural network with improved pond goose algorithm[J]. Journal of Shanxi University (Natural Science Edition),2023,46(06):1283-1292.

[24] Shieh C S .A Parallel Compact Gannet Optimization Algorithm for Solving Engineering Optimization Problems[J].Mathematics, 2023, 11.

[25] Lu K, Miao Tenghui. Research on scene visual understanding algorithm applied to art assisted design[J]. Modern Electronic Technology,2020,43(09):37-40.

[26] Z. M. Wang, X. Wang, G. Li, F. T. Zhang. A review of visual scene understanding[J]. Journal of Xi'an University of Posts and Telecommunications,2019,24(01):1-15.

[27] Liu LQ,Chen F. PCNN parameter-optimised image fusion with sine-cosine dynamic interference Harris Hawk algorithm[J]. Software Guide,2024,23(03):62-70.

[28] Murugesan S , Suganyadevi M V .Performance Analysis of Simplified Seven-Level Inverter using Hybrid HHO-PSO Algorithm for Renewable Energy Applications[J].Iranian Journal of Science and Technology, Transactions of Electrical Engineering, 2024, 48(2):781-801.

[29] Badi M , Mahapatra S .Optimal reactive power management through a hybrid BOA-GWO-PSO algorithm for alleviating congestion[J].International Journal of System Assurance Engineering and Management, 2023, 14:1437 - 1456.

[30] He Huan. Fault diagnosis of rolling bearing of wind turbine based on STOA-XGBoost[J]. Heilongjiang Science,2024,15(12):30-33.

Quantitative Measurement and Preference Research of Urban Landscape Environmental Image Based on Computer Vision

Yan Wang
School of Art and Design, HuangHuai University, Zhumadian 463000, China

***Abstract*—At present, research on landscape preferences mostly uses traditional questionnaire surveys to obtain public aesthetic attitudes, and the analysis method still relies on manual coding with small sample sizes. However, the research on landscape preference of applying network big data and computer vision technology is rare, and the research content and algorithm application are limited. In order to improve the research effect of quantitative measurement and preference of urban landscape environment image, the algorithm proposed in this paper combines two-dimensional analysis modules, two-dimensional visual domain analysis and three-dimensional visual analysis, and makes full use of the advantages of the two analysis modules, and analyzes the scale from large scale to medium and micro scale based on different accuracy urban digital models. Through image classification and content recognition, image semantic segmentation and image color quantification, the landscape feature information in pictures is mined, and the dimension of landscape image is put forward based on this. In addition, this paper combines experimental analysis to verify that the method proposed in this paper has certain results. It is not only suitable for visual analysis of landmark buildings and landmark structures in cities, but also can analyze the visual characteristics of natural landscapes as urban images in cities. Therefore, the quantitative method of urban visual landscape analysis proposed in this paper can provide reliable data support for the follow-up urban design work.**

***Keywords*—Computer vision; urban landscape; environmental image; quantification; measure**

I. Introduction

With the development of the times, people's living standards are constantly improving, and the public's aesthetic concept is also constantly improving. More and more people are beginning to pay attention to the spiritual ascension. Moreover, some once neglected art forms, such as abstraction and performance art, are increasingly recognized by people through deformation and exaggeration. Although more information can be obtained through the domestic literature retrieval system at present, it is still found that many researches on image modeling mainly focus on the field of art, such as traditional Chinese painting, oil painting and sculpture. At the same time, although the development of graphic design, product design and other design fields has made some achievements, most design works still focus on specific artistic design, and there is little research on the specific environment in urban public space.

Image is the bridge of two-way communication between people and landscape, and any experience in landscape places is related to images. Landscape image experience focuses on the spiritual communication relationship between landscape quality and people, including memory, imagination, thinking and accompanying positive emotions. Kevin Lynch first put forward the concept of urban image. After analysis, he concluded that environmental image should consist of three parts: personality, structure and implication. Although his concept of environmental image is put forward at the macro level of the city, it is undeniable that these three characteristics are also the basis of forming landscape image [1].

The personality of landscape image refers to the distinguishability and recognizability as a landscape place. Nowadays, driven by the rapid and urgent production environment of globalization and economic interests, the newly built squares are the same, the landscape avenues are the same, and even the street lamps and signs are in a unified mode. Similar pedestrian commercial streets, similar architectural forms, similar residential landscapes and similar pocket parks make cities and landscapes similar. Because landscape designers don't fully tap the physical and humanistic characteristics of the site, the potential energy of the site can't be revealed and its uniqueness can't be created. The personality of the urban landscape disappears, and the cultural threads of the urban landscape also break. In this case, the landscape loses its narrative (information component) and its poetic experience (rich expression) [2].

In order to improve the research effect of quantitative measurement and preference of urban landscape environmental image, the algorithm proposed in this paper combines two-dimensional analysis modules: two-dimensional visual domain analysis and three-dimensional visual analysis. In addition, this paper combines experimental analysis to verify that the method proposed in this paper has certain results. It is not only suitable for visual analysis of landmark buildings and landmark structures in cities, but also can analyze the visual characteristics of natural landscapes as urban images in cities. Therefore, the quantitative method of urban visual landscape analysis proposed in this paper can provide reliable data support for the follow-up urban design work.

II. Related Works

Landscape imagery is a landscape imagery pattern gradually formed by people in the process of landscape cognition, which refers to the interactive relationship and

perceptual experience between subjects and objects. It includes both the presentation based on the "image" of the landscape object and the perception based on the "meaning" of the viewing subject. The coastal landscape imagery is a mapping of the public's perception of coastal landscapes and the cognitive evaluation of different sea areas, including the landscape projection imagery created by relevant propaganda agencies through abstracting urban landscape elements and the landscape perception imagery formed by the public based on their understanding and evaluation of destination urban landscape elements [3]. Previous studies have shown that landscape perception imagery is an important theoretical support for landscape style renewal and creation, and plays an important role in enhancing landscape attractiveness. It has the characteristics of personalization, locality, and sociality [4]. At present, domestic and foreign scholars mainly focus on exploring the constituent elements, spatial distribution, perceptual characteristics, and formation mechanisms of landscape perception imagery, as well as analyzing the perceptual differences of landscape imagery from different perspectives. In addition, some scholars have conducted quantitative evaluations by constructing a landscape image related evaluation system [5].

The public's perception of landscapes often stems from the material carriers and social interaction activities in the natural environment. Traditional research on landscape perception imagery is mostly based on cognitive maps and combined with methods such as questionnaire surveys, participatory surveys, and interviews. There are significant limitations on research time and sample size. With the continuous development of Internet technology and the popularization of artificial intelligence applications, network data based on mass media has become one of the media of landscape image cognition, and it has the advantages of high accuracy, wide source, fast update, large volume and convenient data acquisition [6]. In this context, different scholars have gradually shifted their research on landscape perception imagery to the mining and content analysis of online data. Network data includes travel commentary text data and landscape photography photo data, among which text data has been widely used in the study of landscape perception imagery. For example, some scholars have explored tourists' cognitive preferences for destination landscapes by filtering Weibo texts, or quantitatively evaluated landscape perception images based on the emotions expressed in the texts. At the same time, with the popularization of social media software and the iterative updating of photography equipment, people often spontaneously upload landscape images taken through smartphones and cameras during tourism or sightseeing to social media, photo sharing platforms, and tourism websites. As the public's condensation of local culture and landscape features, these network images contain a large amount of image information and geographical coordinates, as well as other social metadata, which provide extensive and real data for the research of landscape perception images [7]. However, these photo data, as intuitive analysis materials for visual content, are more suitable for application in research. Existing research often relies on manual encoding and other processing methods when processing image data, which has limitations such as strong subjectivity and small research scale. However, the introduction of computer vision algorithms has made up for this deficiency. Currently, research is mostly based on image semantic segmentation technology to identify objects such as people, animals, plants, buildings, and signs in images. The research will be applied to fields such as landscape ecosystem cultural services, landscape aesthetics, urban street scenes, and campus green spaces [8].

In terms of multi-objective adaptive landscape, multi-objective optimization differs from single objective optimization mainly in two aspects. The first point is the objective layer, where each conflicting objective is composed of multiple fitness functions, rather than traditional single fitness functions. The second point is the solution layer. Compared to a single optimal solution in a single objective scenario, the Pareto optimal solution set in a multi-objective scenario is the optimal solution balanced by multiple fitness functions [9]. The author in [10] applied standard fitness landscape analysis techniques to erroneous landscapes. The research content includes the impact of search space boundaries on landscape analysis, the impact of regularization on error surfaces, the impact of architectural settings on landscape morphology, and the impact of different loss functions on attractive basins. With the development of adaptive landscape models, many models for specific scenarios have been further developed, such as those for numerical optimization or discrete optimization, as well as for co-evolution, constraint optimization, multi-objective optimization, and other directions. The Local Optimal Networks (LONs) model has been further developed and has become one of the most widely used landscape analysis techniques today. Meanwhile, the local optimal network is regarded as a composite landscape model that specifically captures the number and distribution of local optimal values in the fitness landscape, and these landscape features have a significant impact on the search heuristic performance [11]. In addition, LONs are mainly applied in discrete optimization and are designed for discrete search spaces. Given the global structure of the landscape that affects search behavior, LONs can be generated for a series of problems to compare algorithm success and failure or the global structure of the search space [12].

Nowadays, social media text data mainly comes from sources such as Weibo, Twitter, and online travelogues. Compared to traditional text data (such as interviews and surveys), it has advantages such as strong real-time performance, rich content, large production volume, and clear semantics. Moreover, it can reflect human behavioral characteristics at a smaller granularity unit. At the same time, social media text data has been applied in tourism geography, such as perceiving tourist destination imagery, analyzing tourist behavior, studying tourist preference characteristics, analyzing tourist attraction popularity, urban social perception, etc. [13]. The consistency between the projected and perceived images of the target image is determined by principal component analysis using online commentary text data [14]. Text data has strong subjectivity, relatively low spatial coupling, high dependence of text semantics on context, and insufficient depiction of common spatial details and backgrounds. However, network photo data has high accuracy, wide geographical coverage, and a large amount of

information, which can not only reflect people's subjective feelings, but also comprehensively depict the overall image of tourist destinations from multiple dimensions [15]. Compared to text, images can better record what tourists see in real time, providing visually impactful and realistic information for other potential tourists, and making intangible tourism experiences tangible. Image data is intuitive and easy to understand, with rich visual semantic information, which can objectively depict what tourists see [16]. In addition, tourism managers also actively create the image of tourist destinations by taking beautiful photos. The people, objects, and scenery in the photos are cognitive elements of tourists, and they are all carriers of tourist destination imagery. At present, the main sources of image data include street view data, social media image data, and media promotional image data [17]. Images are a data source with high application value in the future and an important supplement to existing research data. Tourism photos are an important component of tourism activities. Tourists share their travel process in the form of photos, which is a common means of disseminating tourism experiences and destination images. Tourism photos have the function of image dissemination, which enables domestic and foreign scholars to use them as a data source to study the role of tourism photos in the dissemination of destination images [18]. The tourism gaze theory suggests that potential tourists can have psychological imagination and travel motivation when browsing travel photos [19]. At the same time, photos can attract potential tourists to generate imagination and desire to participate in travel activities, and through the analysis of photo content, tourists can understand their interests, preferences, travel behavior, and spatial distribution characteristics in different destinations [20].

With the development of remote sensing technology, some scholars have applied it to landscape pattern analysis. Common landscape classification methods include unsupervised classification and supervised classification. Meftaul et al. [21] have shown through experiments that using mixed classification (i.e. a combination of supervised and unsupervised classification) to classify images can achieve higher classification accuracy, and with the rise of neural networks, some scholars have already used backpropagation (BP) neural networks (BP) for classification.

RAL network classifies images and obtains better results. Based on this discovery, Liu et al. [22] proposed a landscape measurement calculation method, which is widely used in urban landscape pattern analysis, such as vegetation pattern, urban growth and other disciplines. Landscape measurement can quantitatively describe and monitor the changes in landscape spatial structure over time, but cannot specifically reflect the direction and rationality of landscape changes.

Deng et al. [23] proposed an "inverse S-shaped function" to represent the density of urban land, which can identify the size of urban built-up area, measure the compactness and expansion intensity of urban land. However, it only discusses the spatial distribution of urban land use and does not analyze the changes in land use structure during urban development.

At present, there are two problems in the research of landscape pattern. The unclear interpretation characteristics of images of different landscape types have led to the prevalence of the phenomenon of "same spectrum foreign objects and same objects but different universality". These analyses are mostly based on landscape metrics, which are relatively simple research methods and cannot reflect the complexity of landscape pattern changes during the rapid development of cities.

III. Quantitative Measurement Method of Urban Landscape Environment Image

A. Research Methods

The urban visual landscape analysis framework proposed in this paper includes two analysis routes: the traditional model and the point cloud-solid hybrid model, includes two-dimensional and three-dimensional analysis processes, and mainly expounds the visual analysis based on traditional model. Traditional models mainly include urban digital surface model (UDSM) and urban 3D solid model (urban 3D solid model).

The sight urban visual landscape analysis framework proposed in this paper mainly consists of three parts: urban digital model construction, visibility analysis and visual feature extraction. The analysis framework is shown in Fig. 1. The method proposed in this paper combines two-dimensional analysis modules of two-dimensional visual domain analysis and three-dimensional visual analysis, makes full use of the advantages of the two analysis modules, and analyzes the scale from large scale to medium and micro scale on the basis of urban digital models with different accuracy. Models with different accuracy can support visibility analysis results with different degrees of fineness, and the calculation time and model construction cost also increases with the improvement of accuracy. In the analysis model of this paper, the viewing points and the viewed landscape objects will be discretized in the digital model, and the spatial points with elevation attributes will be taken as representatives to participate in the calculation.

For an urban space without strict design control, in addition to the information accumulated in daily life, there may also be some missing public spaces where the target urban landscape can be observed, and visual domain analysis can help research and judge these areas that may be ignored by existing planning and design but have ornamental potential (as shown in Fig. 2).

Potential viewing points are extracted from the visual public space and carried out forward operation in the visual domain analysis, that is, these viewing points V_V are analyzed as viewpoints in the visual domain analysis, and the viewshed of viewpoints in the city can be obtained. The visible area VA_{TO} of target objects from open spaces in the city can be obtained by superimposing the obtained urban visible range and the overall range of the target object. It is worth noting that there is the following relationship of VA_{TO} , A_T and V_V :

$$VA_{TO} = A_T \cap V_V \tag{1}$$

If and only if that entire extent of the target object is visible to the public space,

$$VA_{TO} = A_T \in V_V \quad (2)$$

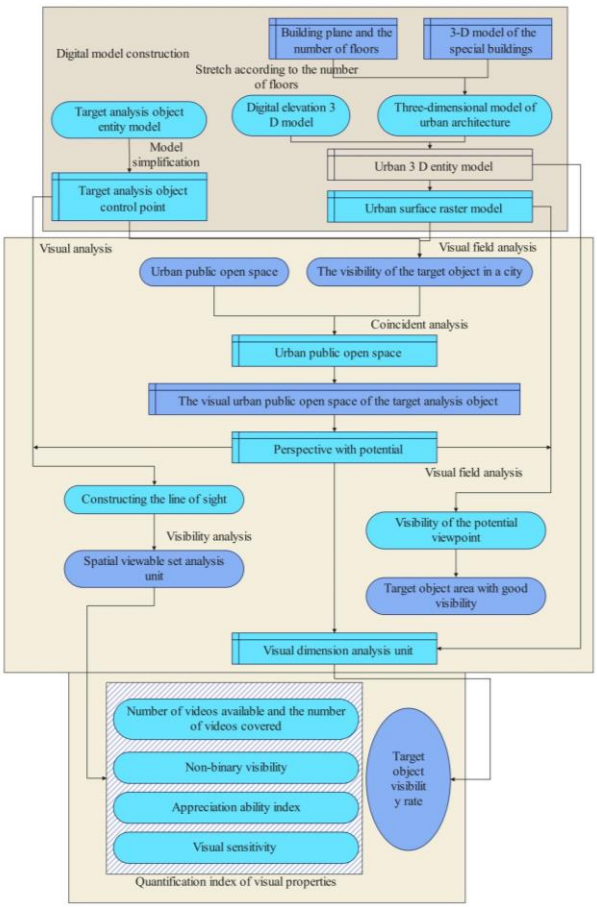
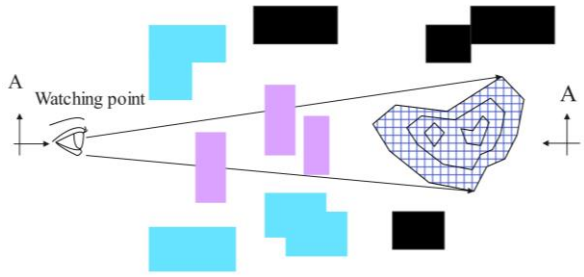
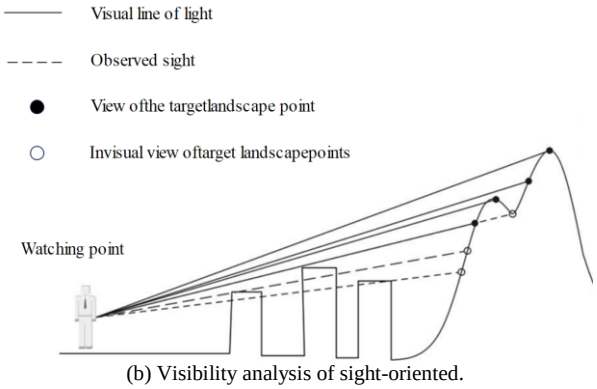


Fig. 1. Flowchart of sight-oriented urban visual landscape analysis.



(a) Visual analysis plane view.



(b) Visibility analysis of sight-oriented.

Fig. 2. Visual domain analysis.

The analysis results of this step can preliminarily test the visibility of the viewing target object, and make more in-depth exploration of areas with relatively high cumulative visibility, that is, areas with high viewing probability in urban space.

The principle of human eye imaging is similar to that of convex lens imaging. If the size of an object remains the same, the closer the object is to the human eye, the larger the image it forms on the retina will gradually become, and the better the visibility of the object will be, and vice versa. That is, the well-known phenomenon of "near big and far small", as shown in Fig. 3.

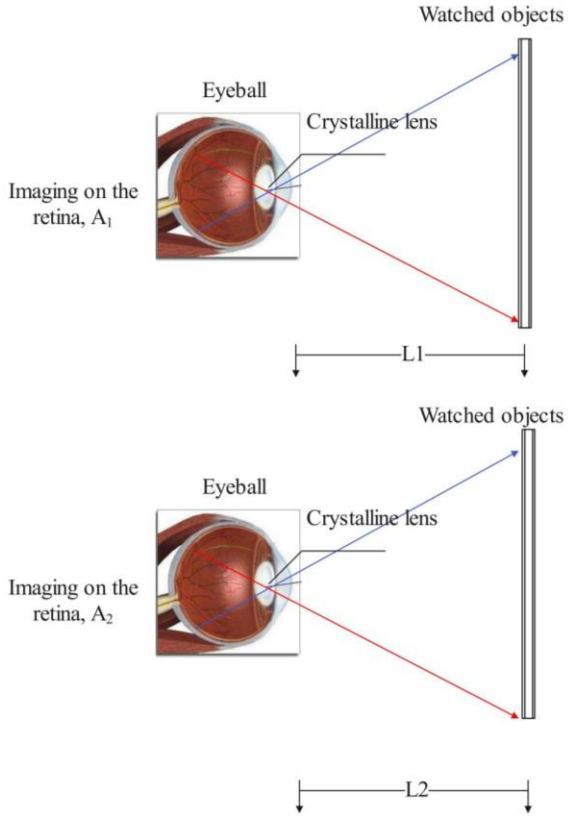


Fig. 3. Imaging size of human eye under different visual distances.

The visibility V_{ij} between the i-th viewpoint and the j-th target point is as follows:

$$V_{ij} = \begin{cases} 0, Line\ of\ sight\ obstruction\ or\ L_{ij} > L_{max} \\ 1, L_{ij} \le D_0 \\ k_{ij} = \dfrac{L_{max} - L_{ij} + D_0}{L_{max}}, D_0 < L_{ij} < L_{max} \end{cases} \quad (3)$$

Among them, D_0 is the optimal line of sight, k_{ij} is the distance attenuation coefficient, L_{max} is the theoretical maximum visual range, that is, the farthest distance that the human eye can see the object clearly, and L_{ij} is the distance between the i-th viewpoint and the j-th target point.

The value range of the optimal sight distance D_0 should be selected according to the height and volume of the target object. The field of view where a clear landscape image and a relatively complete composition effect can be obtained without turning the head is 45 ° ~ 60 ° horizontal angle α and 26 ° ~ 30 ° vertical viewing angle β. Beyond this range, the head must be observed up and down, and the overall composition impression of the scene is not complete enough. It is assumed that the horizontal length of the target scene is W, the vertical height is H, and the human eye height is h. According to the calculation, the sight distance D_{OW} that can obtain the best horizontal field of view is about (0.9 ~ 1.2), as shown in Fig. 4. For the convenience of calculation, $D_{OW} = W$ and $D_{OH} = 4(H - h)$ can be obtained by taking the median of the two.

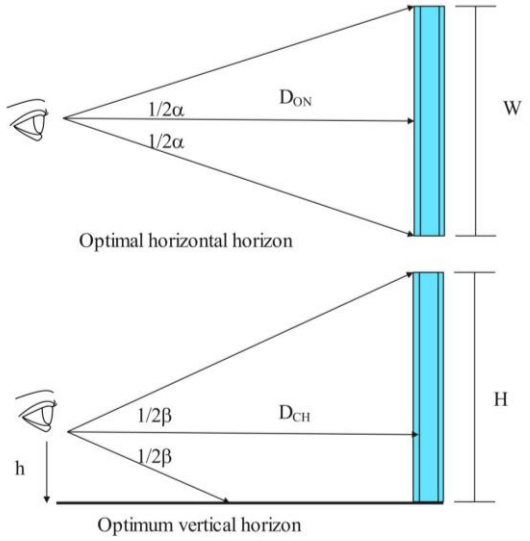


Fig. 4. Schematic diagram of optimal sight distance in horizontal and vertical field of view.

The correction formulas for curvature and atmospheric refraction are as follows:

$$Z = \frac{\left[Z_0 + D^2(R-1)\right]}{d} \tag{4}$$

Among them, Z is the corrected elevation, Z_0 is the surface elevation, D is the horizontal distance, d is the diameter of the earth (12740 km), and R is the refractive index of light (0.13 in standard atmospheric cases). ArcGIS's through-vision sight analysis is based on the analysis and calculation of three-dimensional line segments (rays between the viewpoint and the target point), which retains spatial visual information to the greatest extent, such as the spatial position of sight obstacle points (Fig. 5).

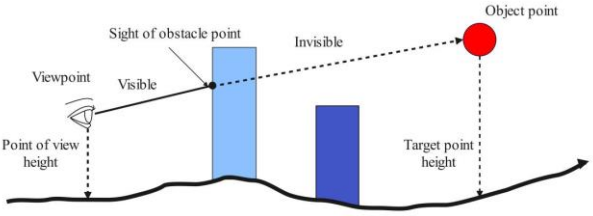


Fig. 5. Analysis of visual sight.

The spatial relationship of sight accessibility is shown in Fig. 6. Taking the target object as a natural mountain as an example, the angle θ_{ij} between the line of sight and the normal vector of the mountain plane where the target point is located determines the proportion of the target mountain displayed in the eyes of the observer. On the other hand, the distance influence can be expressed in non-binary visibility to consider the visual influence of "near is large and far is small".

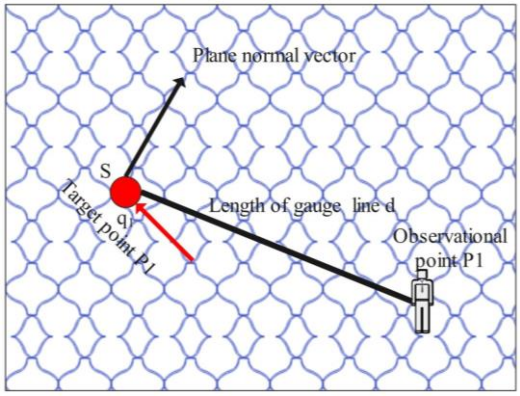


Fig. 6. Schematic diagram of sight accessibility.

It is assumed that there are the i-th viewpoint $P_i(x_i, y_i, z_i)$, the j-th target point $P_j(x_j, y_j, z_j)$ and the plane S_j where P_j are visible to each other, and the cosine of the angle θ_{ij} between the through sight line $\vec{P_J P_l}$ and the plane S_j is:

$$sin\,\theta_{ij} = \frac{\vec{n_J} \cdot \vec{P_J P_l}}{\left|\vec{P_J P_l}\right|} \tag{5}$$

Among them, $\vec{n_J}$ is the unit normal vector of plane S_j. The sight accessibility between the plane S_j represented by the target point and the viewpoint P_i is:

$$ASL_{ij} = V_{ij} \cdot sin\,\theta_{ij} \tag{6}$$

In the formula, V_{ij} is the non-binary visibility between the observation point and the target point.

If it is assumed that there are n visible target object points for the i-th viewing point P_i in the analysis, the calculation formula of the viewing ability index VAI of P_i is as follows:

$$VAI = \sum_{j=1}^{n} ASL_{ij} \tag{7}$$

Among them, V_{ij} is the sight accessibility between the i-th viewpoint and the j-th target object point. Since the viewpoint and target point of the through-sight line are selected, $0 < V_{ij} \le 1$.

The average viewing ability index measures the average viewing ability of a certain viewpoint to the visible target point. The larger the value, the closer the distance between the viewpoint and the target viewing object, and the better viewing effect can be obtained, and the distance has less influence on it. The smaller the value, the farther the viewpoint is from the target viewing object, and the viewing effect is worse because of the greater influence of the distance on it.

$$VAI_{avg} = \frac{1}{n} \times VAI \tag{8}$$

Visual exposure of target objects (VE) describes the degree of attention of a certain target point in the urban visual landscape environment. Small changes in areas with high visual exposure will be keenly perceived by the observer, while the observer will be slow to change in areas with low visual exposure.

If there are m viewpoints $(P_1, P_2, ..., P_m)$ visible to the target point P_T, that is, the number of videos of the target point is m, the visual exposure of P_T is defined as:

$$VE = \sum_{j=1}^{m} LOSA_{ij} \tag{9}$$

Among them, $LOSA_{ij}$ is the sight accessibility between the i-th viewpoint and the j-th target object point that are visible to each other. Visual exposure reveals the overall gaze degree of mountain points on a selected series of viewing spots. The higher the value, it means that the area represented by the target point is easier to appreciate the whole picture for the selected viewing spots. Otherwise, it means that the area represented by the target point is less exposed in the viewing spots and cannot fully display its whole picture.

The average visual exposure of line of sight measures the average gaze degree of a visible target mountain point in urban space. A low average visual exposure indicates that the average projected area of the point in the viewing spot is small, and the average gaze degree is low. The average visual exposure of the mountain point is:

$$VE_{avg} = \frac{1}{m} \times VE \tag{10}$$

The visible percentage of target objects (VP) refers to the area extraction of the components of the obtained digital image under the condition of simulating the visual picture and calculating the proportion of the target analysis object in the picture. Then, the visibility of the target object observed for the viewpoint i is:

$$VP = \frac{A_T}{A_i} \times 100\% \tag{11}$$

Among them, A_i is the total area of the picture obtained by the viewpoint i, and A_T is the area of the target object in the picture.

B. Experimental Study

This paper takes the landscape of urban lakes and parks as an example. This study uses image data from social media to analyze the public's landscape perception preference characteristics. In order to reduce the influence of a single social media user's preference on the results, this study selects the comment pictures on the social media platform of mainstream travel websites with high audience, wide coverage and large amount of evaluation data as the data source. The web crawler is used to retrieve and crawl all tourist comment pictures in 20 lakes and parks on the Internet. Because the sources of the obtained pictures are complex, it is necessary to clean the picture data, check the obtained social media pictures, and eliminate blurred pictures.

Through image classification and content recognition, image semantic segmentation and image color quantification of image data, the landscape feature information in pictures is mined. Based on this, three dimensions of landscape image are proposed: landscape composition, landscape proportion and landscape color, in which landscape composition can be decomposed into landscape type, spatial scale and landscape elements, landscape proportion includes green vision, sky visibility and architectural visibility, and landscape color is described by HSV color features, and a multi-dimensional landscape image quantitative measurement framework based on public perception is constructed (Fig. 7).

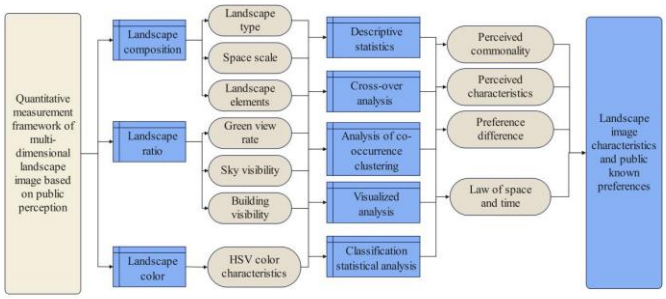


Fig. 7. Quantitative measurement framework of multi-dimensional landscape image based on public perception.

1) Image label analysis data set: The screened landscape pictures are stratified according to different parks, and simple random sampling is carried out in each layer through stratified ratio. A total of 500 pictures are sampled for image label analysis.

2) Model training data set: The screened landscape pictures are stratified according to different parks. Firstly, simple random sampling is carried out in each layer through stratified ratio, and a total of 1,200 pictures are sampled for preliminary model training and adjustment. Secondly, the training data set is adjusted according to the validity of model evaluation, and image data is added to labels with poor recognition effect. Finally, it is adjusted to 1,500 pictures for building Auto ML models.

C. Results

The number of pictures in the training set, validation set and test set of the landscape feature classification AutoML (Automated Machine Learning) model is 1197, 151 and 152, respectively. Through model evaluation, it can be obtained that the average accuracy AP value of the AutoML model for landscape feature classification is 0.933, and when the confidence threshold is 0.5, the accuracy P is 88.36%, the recall R is 84.87%, and the F1 value is 0.866. The P-R curve is shown in Fig. 8. Generally speaking, the accuracy of the model is high, and all indexes are greater than 0.8. It is considered that the model can be used to classify image landscape elements.

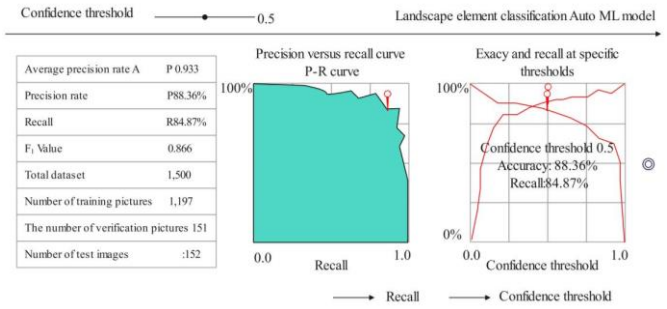


Fig. 8. P-R curve of Auto ML model for landscape feature classification.

The AP values, precision, recall, and F1 values of each label are shown in Table I, and the P-R curve of each label is plotted according to precision and recall, as shown in Fig. 9.

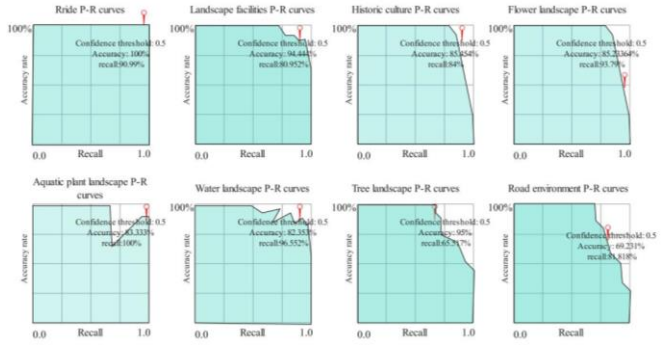


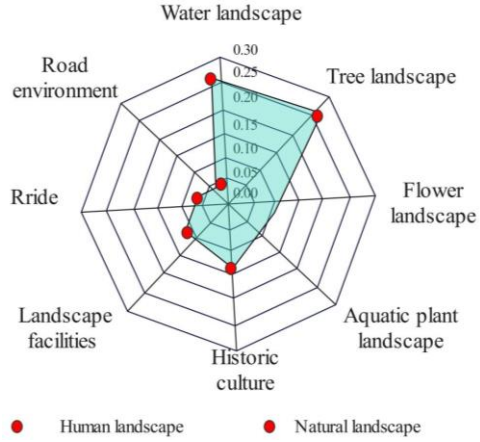
Fig. 9. P-R curves of various landscape types.

TABLE I. EVALUATION INDICATORS OF AUTO ML MODEL FOR LANDSCAPE ELEMENT CLASSIFICATION

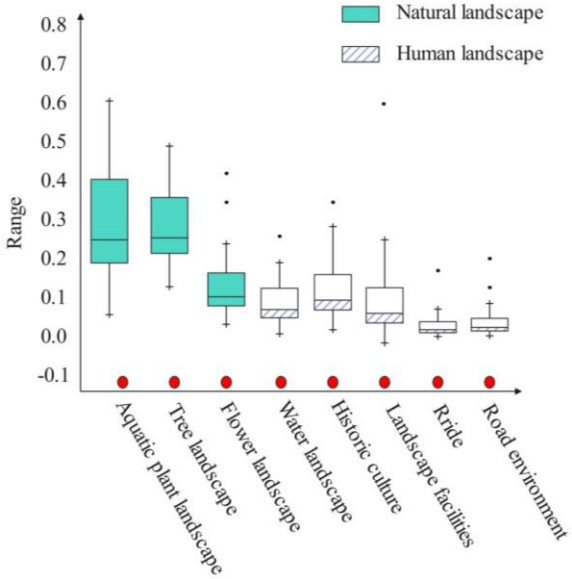
Landscape Type	AP	Precision	Recall	F1-score
Amusement facilities	1	99.00%	90.00%	0.942
Landscape facilities	0.971	93.50%	80.14%	0.863
History and culture	0.959	94.51%	83.16%	0.885
Flower landscape	0.944	87.36%	92.81%	0.900
Aquatic plant landscape	0.936	82.50%	99.00%	0.900
Water landscape	0.933	81.53%	95.58%	0.880
Forest landscape	0.901	94.05%	64.86%	0.768
Road environment	0.895	68.54%	81.00%	0.743

Note: The values of Precision, Recall, and F1-score are the results when the confidence threshold is 0.5.

After sorting out, the statistical results of perception frequency of each landscape type in lake park are obtained (Table II), and the perception results of landscape types are shown in Fig. 10.



(a) Landscape type of perceived intensity.



(b) Boxplots of landscape type perception frequency.

Fig. 10. Landscape type perception results.

TABLE II. STATISTICAL RESULTS OF PERCEPTION FREQUENCY OF LANDSCAPE TYPES IN LAKE PARKS

Landscape Type	Natural landscape				
	Water landscape	Forest landscape	Flower landscape	Aquatic plant landscape	Summary
Number/N	8180	8075	3324	2605	22185
Perceived frequency/%	25.28%	24.97%	10.28%	8.05%	68.58%
Perceived frequency Min	5.73%	11.13%	0.82%	0.34%	35.70%
Perceived frequency Max	57.68%	45.95%	41.11%	21.26%	86.78%
Standard deviation	0.1303	0.0895	0.1036	0.0609	0.1347
Landscape Type	Human landscape				
	History culture	Landscape facilities	amusement facilities	road environment	Summary
Numbe/N	3819	3130	1722	1168	9840
Perceived frequency/%	11.81%	9.67%	5.33%	3.61%	30.42%
Perceived frequency Min	0.00%	3.83%	0.00%	0.82%	12.22%
Perceived frequency Max	31.67%	60.67%	17.32%	15.72%	63.30%
Standard deviation	0.1055	0.1238	0.0383	0.0384	0.1347

D. Analysis and Discussion

By examining the false and negative cases predicted in road environment, water landscape and forest landscape, it is found that the reason for the confusion is that the pictures of the two groups of landscape types contain overlapping or similar landscape contents. For example, a picture taken with a water plank road as the main scene contains a large area of water at the same time, which may cause the model to encounter difficulties in classifying road environment and water landscape, or a picture taken with street trees as the main scene will also cause the model to deviate when classifying forest landscape and road environment. Considering the complexity of the landscape and the manual recognition will also have a certain degree of error, this part of the error can be considered as a normal result. Generally speaking, the trained AutoML model of landscape element classification has high validity, and can be applied to batch prediction of landscape images and identify the types of landscape elements.

The landscape composition of lake park can be divided into three aspects: landscape type, spatial scale and landscape element. This paper focuses on the analysis of the perception commonality and perception characteristics of landscape composition, discusses the public preference orientation and its reasons, obtains the various landscape composition and combination patterns that people prefer, and reveals the seasonal preference differences of landscape composition.

According to the perception intensity of each landscape type [Fig. 10(a)], the perception of natural landscape in lake park is significantly higher than that of human landscape, the perception frequency of water landscape and forest landscape in natural landscape is higher, and the perception frequency of historical culture and landscape facilities in human landscape is higher. On the whole, water landscape, forest landscape and historical culture are the most perceived landscape types in lake parks, accounting for more than 60%. Therefore, these three types of landscapes can be considered as the representative and core landscape types of lake parks. It can be seen from Fig. 10(b) that the perceived changes of natural landscapes are more diverse than those of human landscapes, indicating that the differences in natural landscape characteristics in different parks are more significant.

Therefore, the waterscape in lake park, as the most important landscape element, is also the most popular landscape element, which also reflects that people's hydrophilic psychology and yearning for natural scenery greatly affect their perception preference. Forest landscape is the basic component of urban green space, and its landscape quality directly affects the overall landscape quality of garden green space. It covers all aspects of the landscape environment, making it easier for tourists to be highly perceived. In addition, the reason why history and culture are highly perceived is that it represents the important landscape features of the city, and the cultural landscapes of many tourist destinations in the city show strong tradition and historical customs, which arouse the strong perception of tourists.

The research shows that natural landscape is the core landscape image of urban landscape, and people's hydrophilic psychology and yearning for natural landscape greatly affect their perception preference. Therefore, in the construction and optimization of urban landscape, the naturalness of the landscape should be improved and the authenticity of the natural landscape should be maintained. Besides preserving and protecting the original landscape, it is also very important to restore and reproduce the natural features of the surface by artificial means. According to the natural landscape characteristics and public preference orientation of the study site, the optimization of the natural landscape quality of Wuhan's urban landscape can be considered from the following points. Firstly, the transformation of ecological natural revetment can take the form of lawns and aquatic plants, or natural stone and wood bottom protection. Moreover, the landscape design integrates sponge city ecological technology, reflects the characteristics of lakes and wetlands in Wuhan, and meets the dual needs of ecology and beauty. Secondly, it is necessary to build an open waterfront space, improve the

hydrophilicity of waterfront space, and create a recreational landscape place that can be close to nature. Third, it is necessary to plant characteristic plant communities to form a rich plant landscape, pay attention to the construction and development of flower gardens and various flower shows, and attract people to watch and play.

It is necessary to pay attention to the balanced development of landscape construction at macro, meso and micro scales, and create landscapes with reasonable sense of scale, diverse changes in spatial scale, distinct spatial sequence levels and significant differences in visual perception, so as to enrich tourists' emotional experience. Specifically, it can be optimized from the following points. First, it is necessary to coordinate the sense of scale of landscape elements. When creating different landscape spaces, it is necessary to meet the functionality of different landscape elements and their own specific scale attributes, such as the close-range viewing function of sculpture or art installation landscape, which limits its spatial scale from being too large, and requires attention to the design of landscape details. Moreover, it is necessary to highlight the eye-catching image and control function of physical scenery elements in landscape space. Second, it is necessary to create diversified landscape spatial scales. The spatial scale of the landscape within a certain range should not be static, but should be different from the adjacent space, so as to ensure the spatial integrity and richness of the landscape environment and create a variety of environmental atmospheres. Thirdly, it is necessary to construct a distinct hierarchical landscape spatial sequence. Moreover, the macro, medium and micro-scale landscape nodes should be reasonably arranged on the tour route, so as to provide tourists with a variety of spaces for distribution, communication and privacy, enhance the sense of viewing hierarchy of the landscape along the route, and meet people's needs for diversified spatial scales, so as to optimize the utilization of landscape space.

IV. Conclusion

This paper presents a sight-oriented quantitative analysis method of urban visual landscape, which is designed on the basis of the whole three-dimensional model of the city and combines the advantages of two-dimensional and three-dimensional spatial visual analysis. It is not only suitable for the visual analysis of landmark buildings and structures in the city, but also can analyze the visual characteristics of natural landscape as the image of the city in the city. Therefore, the quantitative method of urban visual landscape analysis proposed in this paper can provide reliable data support for the follow-up urban design work.

Because the traditional model analysis lacks consideration of vegetation, on the basis of it, the point cloud model can be further used to consider the visual analysis of vegetation. In the quantitative analysis framework of urban visual landscape proposed in this paper, the analysis process among two-dimensional model, three-dimensional traditional solid model and point cloud mixed model is progressive.

This paper crawls the comment pictures on social media platforms, uses a variety of computer vision algorithms to parameterize the pictures, extracts and quantifies the landscape features in the pictures, and based on this, proposes three dimensions of landscape images: landscape composition, landscape proportion and landscape color, constructs a multi-dimensional landscape image quantitative measurement framework based on public perception, quantitatively analyzes the commonalities and differences of landscape images, and explores public perception preferences.

The data query efficiency of the model in this paper is low, so in future research, methods such as voxels will be used to segment point cloud data or seek more efficient big data query strategies to improve computational efficiency and accuracy.

Acknowledgment

This project was supported by 2024 Henan Province Soft Science Project. (Project Name: Rural Aesthetic Education Service Rural Revitalization Research, Project No.: 242400411048).

References

[1] L. Feng, and J. Zhao, “Research on the construction of intelligent management platform of garden landscape environment system based on remote sensing images,” Arab. J. Geosci., vol. 14, no. 2, pp. 1-19, 2021.

[2] H. Gu, and Y. Wei, “Environmental monitoring and landscape design of green city based on remote sensing image and improved neural network,” Environ. Technol. Inno., vol. 23, no. 2, pp. 101718-101730, 2021.

[3] N. He, and G. Li, “Urban neighbourhood environment assessment based on street view image processing: A review of research trends,” Environmental Challenges, vol. 4, no. 1, pp. 100090-100098, 2021.

[4] Y. Kang, F. Zhang, S. Gao, H. Lin, and Y. Liu, “A review of urban physical environment sensing using street view imagery in public health studies,” Ann. GIS, vol. 26, no. 3, pp. 261-275, 2020.

[5] T. Hu, and W. Gong, “Urban landscape information atlas and model system based on remote sensing images,” Mob. Inf. Syst., vol. 2021, no. 1, pp. 9613102-9613112, 2021.

[6] H. I. Jo, and J. Y. Jeon, “Overall environmental assessment in urban parks: Modelling audio-visual interaction with a structural equation model based on soundscape and landscape indices,” Build. Environ., vol. 204, no. 2, pp. 108166-108177, 2021.

[7] A. Jahani, and M. Saffariha, “Aesthetic preference and mental restoration prediction in urban parks: An application of environmental modeling approach,” Urban For. Urban Gree., vol. 54, no. 2, pp. 126775-126485, 2020.

[8] L. Mirza, and H. Byrd, “Measuring view preferences in cities: A window onto urban landscapes,” Cities & Health, no. 2, pp. 250-259, 2023.

[9] X. Huang, H. Wang, L. Shan, and F. **ao, “Constructing and optimizing urban ecological network in the context of rapid urbanization for improving landscape connectivity,” Ecol. Indic., vol. 132, no. 2, pp. 108319-108330, 2021.

[10] X. Ma, C. Ma, C. Wu, Y. **, R. Yang, N. Peng,... and F. Ren, “Measuring human perceptions of streetscapes to better inform urban renewal: A perspective of scene semantic parsing,” Cities, vol. 110, no. 1, pp. 103086-103095, 2021.

[11] S. Huai, and T. Van de Voorde, “Which environmental features contribute to positive and negative perceptions of urban parks? A cross-cultural comparison using online reviews and Natural Language Processing methods,” Landscape Urban Plan., vol. 218, no. 3, pp. 104307-104317, 2022

[12] D. Shen, “Application of GIS and Multisensor Technology in Green Urban Garden Landscape Design,” Journal of Sensors, vol. 2023, no. 1, pp. 9730980-9730991, 2023.

[13] L. Kong, Z. Liu, X. Pan, Y. Wang, X. Guo, and J. Wu, “How do different types and landscape attributes of urban parks affect visitors' positive emotions?” Landscape Urban Plann., vol. 226, no. 3, pp. 104482-104490, 2022.

[14] S. Y. Cao, and X. J. Hu, "Dynamic prediction of urban landscape pattern based on remote sensing image fusion," Int. J. Environ. Techno., vol. 24, no. (1-2), pp. 18-32, 2021.

[15] L. Deng, H. Luo, J. Ma, Z. Huang, L. X. Sun, M. Y. Jiang, ... and X. Li, "Effects of integration between visual stimuli and auditory stimuli on restorative potential and aesthetic preference in urban green spaces," Urban For. Urban Gree., vol. 53, no. 2, pp. 126702-126712, 2020.

[16] X. Huang, Y. Wang, J. Li, X. Chang, and Y. Cao, "Landscape analysis of the 42 major cities in China using ZY-3 satellite images," Sci. Bull., vol. 65, no. 12, pp. 1039-1048, 2021.

[17] Z. Li, X. Han, L.Wang, T. Zhu, and F. Yuan, "Feature Extraction and Image Retrieval of Landscape Images Based on Image Processing," Trait. Signal, vol. 37, no. 6, pp. 55-641, 2020.

[18] M. Masoudi, D. R. Richards, and P. Y. Tan, "Assessment of the Influence of Spatial Scale and Type of Land Cover on Urban Landscape Pattern Analysis Using Landscape Metrics," J. Geovis. Spat. Anal., vol. 8, no. 1, pp. 8-18, 2024.

[19] D. M. Zhou, C. Y. Chen, M. J. Wang, Z. W. Luo, L. T. Kang, and S. Wu, "Gradient and directional differentiation in landscape Pattern characteristics of urban ecological space based on optimal spatial scale: A case study in Changsha City, China," Journal of Ecology and Rural Environment, vol. 38, no. 5, pp. 566-577, 2022.

[20] B. Czarnecki, and M. P. Chodorowski, "Urban environment during post-war reconstruction: Architectural dominants and nodal points as measures of changes in an urban landscape," Land, vol. 10, no. 10, pp. 1083-1094, 2021.

[21] I. M. Meftaul, K. Venkateswarlu, P. Annamalai, A. Parven, and M .Megharaj, "Glyphosate use in urban landscape soils: Fate, distribution, and potential human and environmental health risks," J. Environ. Manage., vol. 292, no. 2, pp. 112786- 112798, 2021.

[22] K. Liu, X. Li, S. Wang, and X. Gao, "Assessing the effects of urban green landscape on urban thermal environment dynamic in a semiarid city by integrated use of airborne data, satellite imagery and land surface model," Int. J. Appl. Earth Obs., vol. 107, no. 2, pp. 102674-102684, 2022.

[23] L. Deng, X. Li, H. Luo, E. K. Fu, J. Ma, L. X. Sun,... and Y. Jia, "Empirical study of landscape types, landscape elements and landscape components of the urban park promoting physiological and psychological restoration," Urban Forestry & Urban Greening, vol. 48, no. 3, pp. 126488-126498, 2020.

Clustering Algorithms to Analyse Smart City Traffic Data

Praveena Kumari M K[1], Manjaiah D H[2], Ashwini K M[3]*
MCA Dept. NMAM Institute of Technology, Nitte (Deemed to be University), Nitte, India[1, 3]
Department of Computer Science, Mangalore University, Mangalore, India[2]

***Abstract*—Urban transportation systems encounter significant challenges in extracting meaningful traffic patterns from extensive historical datasets, a critical aspect of smart city initiatives. This paper addresses the challenge of analyzing and understanding these patterns by employing various clustering techniques on hourly urban traffic flow data. The principal aim is to develop a model that can effectively analyze temporal patterns in urban traffic, uncovering underlying trends and factors influencing traffic flow, which are essential for optimizing smart city infrastructure. To achieve this, we applied DBSCAN, K-Means, Affinity Propagation, Mean Shift, and Gaussian Mixture clustering techniques to the traffic dataset of Aarhus, Denmark's second-largest city. The performance of these clustering methods was evaluated using the Silhouette Score and Dunn Index, with DBSCAN emerging as the most effective algorithm in terms of cluster quality and computational efficiency. The study also compares the training times of the algorithms, revealing that DBSCAN, K-Means, and Gaussian Mixture offer faster training times, while Affinity Propagation and Mean Shift are more computationally intensive. The results demonstrate that DBSCAN not only provides superior clustering performance but also operates efficiently, making it an ideal choice for analyzing urban traffic patterns in large datasets. This research emphasizes the importance of selecting appropriate clustering techniques for effective traffic analysis and management within smart city frameworks, thereby contributing to more efficient urban planning and infrastructure development.**

***Keywords*—Clustering; smart city; traffic; analyze**

I. Introduction

In the current era of rapid urbanization and technological advancement, cities are increasingly adopting "smart city" initiatives aimed at improving the quality of urban life through data-driven decision-making. A smart city leverages technology, particularly networked sensors and data analytics, to optimize urban services, including traffic management. The massive influx of data generated from various sources—such as government records, online platforms, and IoT devices—is a vital resource for transforming urban environments into innovation hubs. However, this wealth of data presents challenges in extraction, analysis, and application, particularly in traffic management, where traditional methods struggle to keep pace with the growing complexity.

Transportation systems are essential to economic stability and social development, yet they also contribute to urban challenges such as congestion and air pollution. Effective traffic management, a cornerstone of smart city initiatives, requires a deep understanding of traffic patterns and behaviors. Data mining techniques, particularly clustering algorithms, offer powerful tools for uncovering these patterns, facilitating traffic forecasting, and enabling informed decision-making.

This paper focuses on applying clustering algorithms to analyze urban traffic data, specifically within the context of smart city development. By employing techniques such as DBSCAN, K-Means, Affinity Propagation, Mean Shift, and Gaussian Mixture, we aim to uncover meaningful insights into traffic flow patterns and identify factors influencing congestion. The chosen methods are evaluated on their clustering performance and computational efficiency, as these factors are crucial in real-time traffic management systems.

Previous research, such as the work by Pattanaik, Singh, Gupta, and Singh (2016) [1] proposed a real-time traffic congestion estimation system for urban roads, incorporating K-Means clustering and other clustering techniques. Their approach involves collecting and analyzing real-time traffic data to classify and estimate congestion levels accurately. The system dynamically categorizes traffic patterns to provide timely insights and support effective traffic management. This research highlights the practical application of clustering methods in enhancing urban traffic control and decision-making.

The remainder of this paper is structured as follows: Section II provides a literature review, identifying the state-of-the-art in traffic analysis using clustering techniques. Section III outlines the methodology, detailing the data collection process and the specific algorithms used. Section IV provides the performance measures used to analyze the clustering algorithms Section V presents the results and discussion, comparing the performance of the clustering methods. Finally, Section VI concludes with a summary of findings, implications for smart city initiatives, and suggestions for future research.

II. Literature Review

First, confirm that you have the correct template for your paper size. This template has been tailored for output on the US-letter paper size. If you are using A4-sized paper, please close this file and download the file “MSW_A4_format”. The management of urban traffic is a critical component of smart city initiatives, where data-driven strategies are employed to enhance the efficiency and sustainability of urban transportation systems. Clustering algorithms have gained significant attention for their ability to analyze and interpret large-scale traffic data. This section reviews key studies that have applied clustering techniques to traffic data, highlighting the methods used, the

contexts in which they were applied, and the findings that contribute to the understanding of urban traffic management.

A. Clustering Techniques in Traffic Analysis

Clustering, a form of unsupervised learning, has been widely utilized to uncover patterns in traffic data. Various algorithms, such as K-Means, DBSCAN, and Gaussian Mixture Models, have been applied in different contexts to group similar traffic patterns and identify congestion hotspots.

Rouky et al. (2024) [2] investigated traffic congestion in Casablanca using K-Means and DBSCAN clustering methods. Their study focused on identifying congestion patterns and hotspots within the city, contributing to improved traffic management strategies. The findings indicated that clustering could effectively segment traffic data into meaningful clusters, providing insights into daily and seasonal traffic variations. Similarly, Wang et al. (2016) [3] discovered that vehicular traffic levels on metropolitan highways in Shanghai differ dramatically during times of day and night. They discovered that at peak hours, vehicle numbers increase and speeds decrease owing to bottlenecks. Off-peak hours, on the other hand, see a reduction in the amount of traffic and a rise in speed, resulting in a smoother flow of traffic. Their research emphasizes the significance of comprehending these differences for successful traffic control and urban planning in dynamic metropolis contexts such as Shanghai.

Shi et al. (2021) [4] used a density-based moving object clustering technique to determine the spatiotemporal extends of traffic jams. Their approach combines density-based clustering techniques with moving object data to precisely identify crowded zones and their lengths of stay. This technique improves the exactness of bottleneck identification throughout time frames, providing vital knowledge for improving metropolitan traffic control and planning initiatives.

Yang et al. (2017) [5] investigated transportation state fluctuation trends in urban roadways using spectral clustering. Their work used spectral clustering methods to categorize and analyse various traffic situations using spatiotemporal data. The results demonstrated the usefulness of spectral clustering various traffic patterns, such as bottleneck patterns and flow variances across urban regions. This technique offers important conclusions for enhancing metropolitan traffic management techniques and infrastructure design by effectively recording and analyzing complicated traffic behaviors.

Angmo et al. (2021) [6] proposed an enhanced clustering approach that integrates density-based and hierarchical clustering algorithms. Their methodology, which combines spatiotemporal data with advanced clustering techniques, demonstrated increased accuracy and efficiency in identifying relevant locations for traffic management. This study emphasizes the importance of adapting traditional clustering methods to the specific requirements of urban traffic analysis.

Asadi and Regan (2019) [7] presented a method for spatiotemporal clustering of data on traffic through deep-embedded clustering. Their method uses deep learning techniques to embed traffic data in a latent space, allowing for the finding of significant clusters based on geographical and temporal trends. The study sought to increase traffic data analysis and forecast accuracy by utilizing deep-embedded clustering algorithms. This methodological development offers the potential for improving our knowledge about city transportation habits and optimizing roadway safety tactics.

Sfyridis and Agnolucci (2020) [8] created a technique for predicting Annual Average Daily Traffic (AADT) in England and Wales. They used clustering and regression modeling to increase the correctness of AADT predictions. Clustering methods were employed to aggregate road segments that have similar traffic patterns, models and regression were utilised inside these clusters to calculate traffic numbers. The results showed that this comprehensive strategy improved the accuracy of traffic estimates, providing helpful information for planning transportation and construction of infrastructure.

Wang et al. (2020) [9] devised a technique for identifying hotspots in travel security and safety with an emphasis on the regular fluctuation of the flow of traffic and crash data. Their method combines statistical evaluation and geographical information systems (GIS) to detect high-risk regions for vehicular crashes. By taking into account daily traffic changes and collision data, the study improves hotspot identification accuracy, offering useful information for adopting targeted safety measures and enhancing transportation infrastructure.

Taamneh et al. (2017) [10] published research on categorizing road crashes using clustering-based algorithms. They suggested a hybrid technique that combined hierarchical clustering with artificial neural networks (ANNs). Hierarchical clustering was utilised to divide accident data into clusters with comparable features and ANNs were used to categorise and forecast the severity of accidents within each cluster. The findings showed that this combination strategy increased the accuracy of accident categorization and gave useful information for improving road safety measures and tactics.

Acun and Gol (2021) [11] discovered that data levels on large-scale traffic networks follow different patterns, which may be efficiently analyzed by employing ARIMA and K-means clustering. By using K-means to combine comparable traffic flow patterns and ARIMA models inside these clusters for prediction, they discovered that this hybrid strategy considerably increases traffic volume estimates. Their findings imply that taking into account both temporal trends and geographical clustering results in more accurate traffic management and prediction on large road systems.

Zou, X., & Chung, E. (2024) [12] proposed a novel traffic prediction approach that combines clustering and deep transfer learning to address the challenge of limited data availability. The study utilizes clustering techniques to group similar traffic patterns and then applies deep transfer learning to enhance prediction accuracy across different datasets. This method significantly improves traffic prediction performance, particularly in scenarios where data is sparse, demonstrating its potential in urban traffic management.

Nguyen, T. T., et al. (2019) [13] proposed a method for feature extraction and clustering analysis to study highway congestion. By utilizing clustering techniques, the study identifies and classifies different congestion patterns based on key traffic features extracted from highway data. The approach

focuses on understanding the underlying causes of congestion and its temporal variations. The findings provide valuable insights for designing more effective traffic management strategies on highways, helping to mitigate congestion and improve traffic flow efficiency. The research highlights the importance of data-driven approaches in addressing complex transportation challenges.

B. Smart Cities and Traffic Management

The concept of smart cities involves the use of information and communication technologies (ICT) to optimize urban services, including transportation. Traffic management in smart cities relies heavily on the ability to analyze vast amounts of data generated by sensors, GPS devices, and social media platforms.

Xu et al. (2020) [12] suggested an approach for anticipating traffic jams in Shanghai based on multiperiod hotspot clustering. To forecast bottleneck trends, they use hotspot clustering methods applied across different periods. The work improves overcrowding forecasting accuracy by analyzing past traffic data and using clustering algorithms. This methodological innovation promotes proactive traffic management tactics, hence improving public transportation and the construction of infrastructure in Shanghai.

C. Comparative Analysis of Clustering Algorithms

Several studies have conducted comparative analyses of clustering algorithms to determine their effectiveness in various traffic management scenarios.

Chen et al. (2022) [14] created an automobile flow forecasting system employing a Graph Attention Network (GAT) and spatial-temporal clustering. Their technique makes use of GATs to capture geographical and temporal connections in traffic data, which improves traffic flow prediction precision. The investigation also uses spatial-temporal clustering to organize traffic patterns, which helps manage complicated transportation patterns. The findings show that this combination of approaches increases the accuracy of traffic flow estimates, resulting in improved administration and planning in smart transportation networks.

D. Gaps and Future Directions

While the literature demonstrates the effectiveness of clustering techniques in traffic analysis, several gaps remain. Most studies have focused on traditional algorithms, with limited exploration of novel methods that could offer improved performance or new insights. Moreover, there is a need for more comprehensive evaluations of these techniques within the smart city framework, considering factors such as scalability, adaptability to dynamic data, and integration with other smart city systems.

Additionally, the literature often lacks a critical comparison of how these clustering methods perform in different urban environments, which can vary significantly in terms of traffic patterns and data availability. Future research should focus on developing and testing new algorithms that can address these challenges and contribute to more robust and adaptive traffic management solutions in smart cities.

III. Study Area and Methodology

A. Dataset Description

This research used traffic information from the city of Aarhus to analyze traffic habits in an urban region. Aarhus is Denmark's second-biggest city and main cultural center. Due to its northern latitude, Aarhus experiences significant variations in daylight hours between summer and winter. Various clustering approaches are used to examine data from February 2014 to June 2014 to identify transportation trends. Sensors were set at two nearby places to collect data, tallying the amount of automobiles moving by during a certain period. Every route's report contains a traffic count taken by the sensor on a certain day and time. Table I describes the attributes in the dataset.

TABLE I. Attributes in the Dataset

Attributes	Description
avgMeasuredTime	Shows the duration in seconds of the sensor collecting data.
avgSpeed	This signifies the mean pace of the cars moving within the reported period.
extID	Signifies the distinct identification allocated for every route.
medianMeasuredTime	Like avgMeasuredTime
TIMESTAMP	Represents the beginning time measurement of transportation on a route.
vehicleCount	Quantity of cars traveling across each pair of observations.
_id	A distinctive number, issued for every traffic estimate.
REPORT_ID	Denotes the distinct number of every observation point.

B. Data Preprocessing

The traffic data is processed to enable traffic pattern analytics. Routes with inadequate data are populated using filling missing values techniques. Preprocessing guarantees that the study's final data include comprehensive traffic counts for each road on a given day. Following numerous evaluations of quality, all processed records are grouped into 1-hour intervals. For transportation statistics, we just require an automobile count at periods throughout the day. Therefore, data such as the extID, avgSpeed, avgMeasuredTime and medianMeasuredTime, and other particulars are unneeded. We have trimmed our data collection to contain just the fields important for extracting patterns. Each entry for a certain roadway now has only two fields: date and automobile frequency.

A few rows in the dataset lack accurate timestamps or interval bounds. A defective timestamp indicates that what was observed has changed in time. To resolve this, automobile counts are proportionally adjusted for inadequate timestamps or intervals. The adjusted values are subsequently used for a variety of performance assessments. To eliminate rows with wrong values, the following types of verification procedures are used:

1) Poor-quality detectors may record excessively elevated transport volumes, affecting approximation. Rows with traffic counts exceeding 120 are eliminated from data collection.

2) If nothing is achieved for three hours consecutively in a day, the day is declared worthless. If over thirty percent of the available days for an area are unreliable, it is removed from further investigation.

C. Models

Clustering methods are used when there are not any classes to forecast but instead, instances need to be classified into natural groupings. For the investigation of streets, we employed the DBSCAN, K-Means, Affinity Propagation, Mean Shift, and Gaussian Mixtures method to cluster days of the week. Each method takes an alternate strategy to the problem of identifying natural groupings in data. We selected 372 streets and clustered them over all days in the dataset. Fig. 1 shows the data source collection location. Fig. 2 is the proposed methodology of the suggested approach to analyze the traffic pattern.



Fig. 1. Data source location.

DBSCAN (Density-Based Clustering of Applications with Noise) is a clustering technique that detects clusters and outliers in a dataset using density. The algorithm has two parameters: epsilon(ε), which determines the radius for neighborhood searches, and MinPts, which specifies the minimal of points needed to build dense regions. Core points have at least MinPts neighbours within ε, whereas border points are close to a core point but do not have enough neighbors to be core points themselves. The noise points do not belong to any cluster.

DBSCAN performs well for arbitrary clusters and can withstand noise, but its efficiency is significantly dependent on its selection of ε and MinPts.

K-Means is a common clustering technique that divides a dataset into K separate, not interconnected groups. The technique initiates K centroids and allocates every data point to the closest centroid, resulting in K clusters. It then continuously updates the centroids by taking the average of every point in each cluster and reassigning them depending on the new centroids. This procedure repeats until the centroids stop changing appreciably. K-Means is effective and easy to construct, but the number of clusters (K) must be identified in advance. It works with spherical collections but struggles with clusters of different shapes and densities.

Affinity Propagation is a clustering approach that detects exemplars (representative points) inside data by delivering messages between data points. Unlike, K-Means, the total number of groups does not have to be preset. The algorithm operates by repeatedly transferring two sorts of messages: "responsibility", which shows how well-suited the data point is to serve as an exemplar for another point, and "availability" which represents a point's appropriateness for assignment to an exemplar. These notifications will be maintained until convergence. Affinity propagation may detect clusters of variable sizes and does not presume an ideal structure for the clusters, making it a useful tool for a variety of grouping jobs. However, it can be extremely costly for huge datasets.

Mean Shift is a nonparametric clustering approach that seeks to identify groups by repeatedly moving data points to the densest region of the feature space. The technique begins by constructing a window (or kernel) around each data point and then computes the average of the points within that window. Each data point is then relocated to the mean, and the procedure is continued until convergence, which causes points to cluster around the density function's modes (peaks). Mean Shift, unlike K-Means, doesn't need a pre-specified amount of clusters and may discover clusters of any form. However, its performance is significantly reliant on the selection of the bandwidth option, which controls the window size.

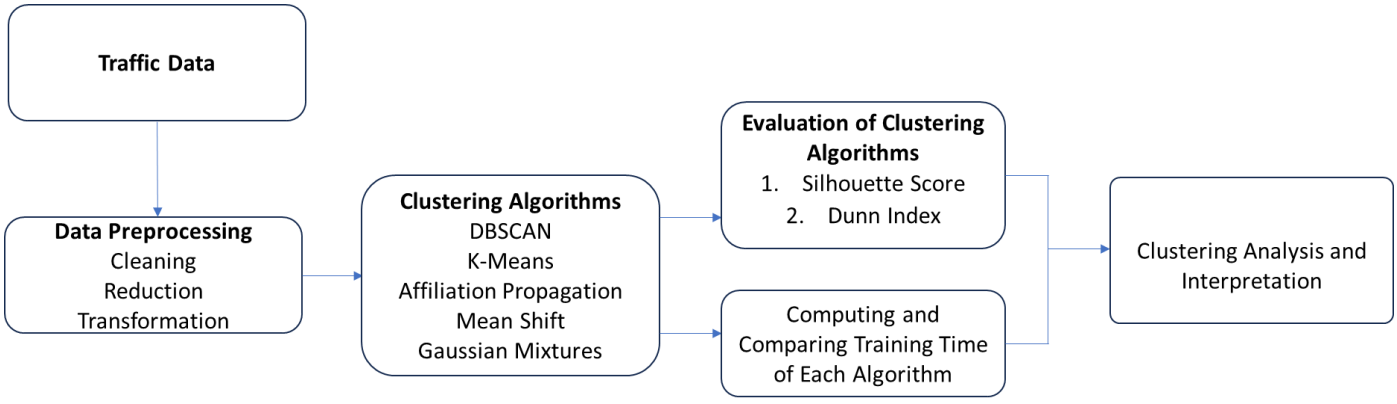


Fig. 2. Methodology of the suggested approach.

Gaussian Mixture Models (GMM) are probabilistic clustering methods that presume data is created by a combination of multiple Gaussian distributions with unidentified variables. Each Gaussian component has a mean and a covariance, and the model is described by the components' mixture weights, means, and covariances. GMMs employ the

Expectation-Maximization (EM) method to continuously modify these variables to maximize the probability of the data being observed. Unlike K-Means, GMMs may detect the elliptical form of clusters and give soft categorization, which means that each data point has a chance to be assigned to each group. This makes GMMs more adaptable and stronger when modelling real-world data with different cluster shapes and densities. However, they can be sensitive to startup and may require careful adjustment.

IV. Study Area and Methodology

To determine which clustering approach is optimal for our model, we apply the following validity indices.

1) *Silhouette score*
2) *Dunn index*

A. The Silhouette Score

The Silhouette Score is a way to evaluate the quality of clusters generated by a clustering algorithm. It determines how comparable a data point is to its group relative to different groups. The score goes from -1 to 1, with a large number indicating that the data points are compatible within the same cluster but not sufficiently matched with neighboring groups. The score for each point is determined and averaged to produce an overall evaluation. A higher average Silhouette Score indicates a more clearly defined and suitable grouping structure. This method is useful for determining the ideal number of clusters and evaluating the effectiveness of various clustering algorithms. The Silhouette Score for a single data point is given by the formula:

$$s(i) = \frac{z(j)-y(j)}{max(y(j),z(j)} \quad (1)$$

where:

- y(j) is the average distance between j and all other points in the same cluster (also known as intra-cluster distance).
- z(j) is the minimum average distance from j to all points in the nearest cluster that j is not a part of (also known as the nearest cluster distance).

The Silhouette Score for the entire dataset is the mean Silhouette Score of all individual data points. The formula is:

$$S = \frac{1}{n}\sum_{i=1}^{n} s(i) \quad (2)$$

where n is the total number of data points. The overall score indicates the clustering quality, with values closer to 1 indicating better clustering.

B. Dunn Index

The Dunn Index is a tool for evaluating the effectiveness of algorithms for clustering that measures group density and separation. It can be defined as the ratio of the lowest intercluster distance to the maximal intra-cluster distance. A higher Dunn Index implies better grouping since it implies that groups are clearly distinguished and densely packed. Here is the formula:

$$D = \frac{(max_{1\le i<j<k} d(C_i C_j)}{(max_{1<l<k} \delta(C_l)} \quad (3)$$

where $d(C_i, C_j)$ is the distance between the two clusters C_i and C_j, and $\delta(C_l)$ is the diameter of the cluster C_l. The Dunn Index is particularly useful for comparing different clustering results on the same dataset. However, it can be computationally expensive for large datasets due to the need to compute all pairwise distances between cluster.

V. Experiments and Results

Everyday travel patterns from various days are clustered using the DBSCAN, K-Means, Affiliation Propagation, Mean Shift, and Gaussian Mixture algorithms. Records for working and nonworking days were combined for all routes from February to June.

A. Clustering Analysis

Clustering was performed on 372 streets, each having enough information to analyze. The number of clusters and related traffic patterns varied according to location. Affinity Propagation clustering divided days into two categories: weekdays and weekends. In DBSCAN, K-Means, Mean Shift, and Gaussian Mixture clustering days were divided into four main groups. A thorough investigation of several roadways was conducted, providing reliable outcomes. The results of DBSCAN, K-Means, Affinity Propagation, Mean Shift and Gaussian Mixture clustering are presented in Fig. 3, Fig. 4, Fig. 5, Fig. 6, and Fig. 7 respectively. Except Affinity Propagation in other clustering techniques each street's days were separated into four groups.

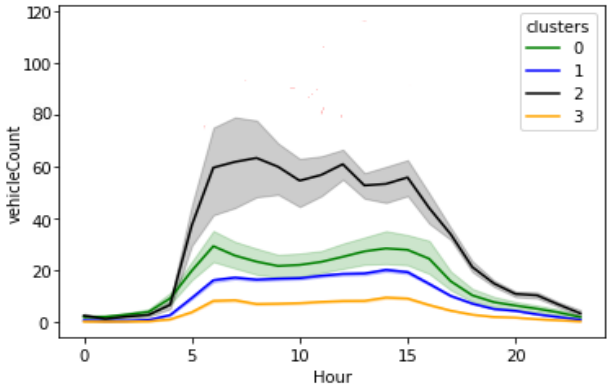


Fig. 3. Traffic pattern representing four clusters using DBSCAN.

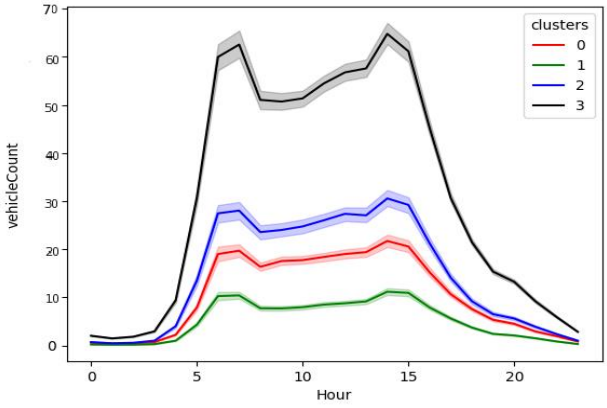


Fig. 4. Traffic pattern representing four clusters using K-Means.

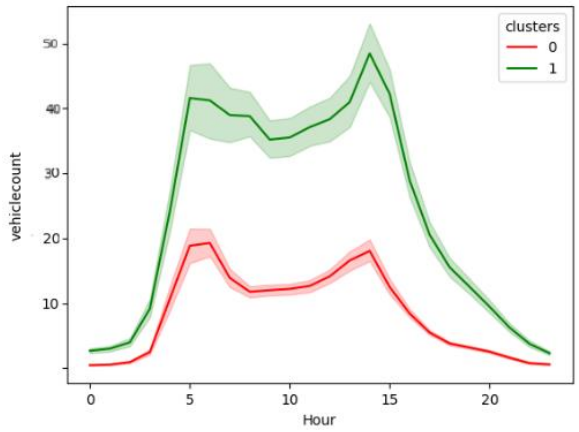


Fig. 5. Traffic pattern representing four clusters using affinity propagation.

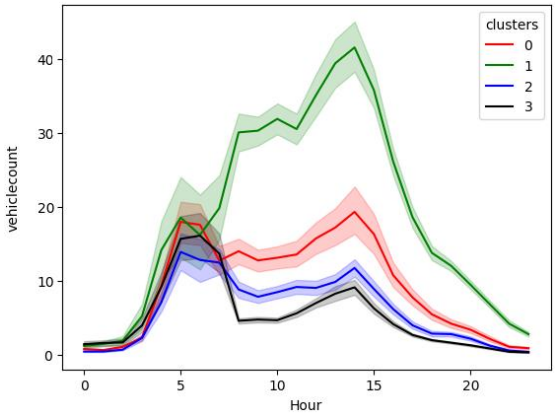


Fig. 6. Traffic pattern representing four clusters using mean shift.

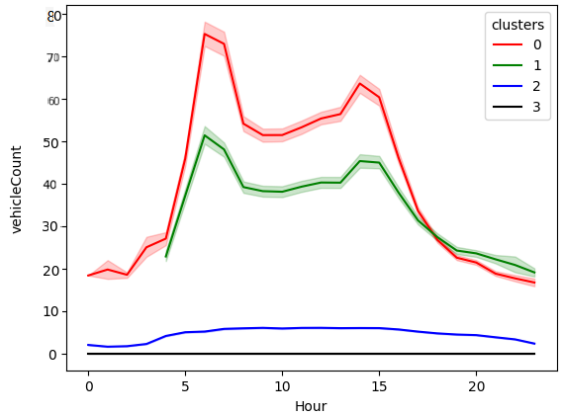


Fig. 7. Traffic pattern representing four clusters using Gaussian Mixture.

The lower level cluster in Fig. 3 to Fig. 7 includes Saturdays, Sundays, and certain weekdays. Weekdays with vacations or wet weather have significantly lower transportation than regular days. This cluster consists of routes with low to medium traffic throughout the day. The other groups include a variety of weekdays.

NearestNeighbors is used to calculate the estimated closeness between points of data and it is utilized as eps in DBSCAN. The _n_neighbors_ in NearestNeighbors is used as the minimum sample point in DBSCAN. The elbow approach determines the ideal number of groups in K-Means clustering. To assess which clustering strategy is best for our model Silhouette Score and Dunn Index are used. The DBSCAN gives better values for the Silhouette coefficient and Dunn Index as seen in the results shown above in Table II.

TABLE II. SILHOUETTE SCORE AND DUNN INDEX OF DIFFERENT CLUSTERING TECHNIQUES

Technique	Silhouette Score	Dunn Index
DBSCAN	0.3664	0.03752
K-Means	0.3119	0.00395
Affinity Propagation	0.0017	0.00123
Mean Shift	0.0042	0.00350
Gaussian Mixture	0.2032	0.00390

We have also compared the training times taken by each clustering algorithm. Fig. 8 depicts the training time for each investigated clustering algorithm. Fig. 9 depicts the training time for each clustering algorithm excluding Affinity Propagation. On the scale of the training timeframes for the Mean Shift and Affinity Propagation algorithms, the training durations of the other methods completely evaporate. Furthermore, Affinity Propagation is eliminated to compare the training durations of the remaining algorithms on a more reasonable scale.



Fig. 8. Training time for each clustering algorithm.

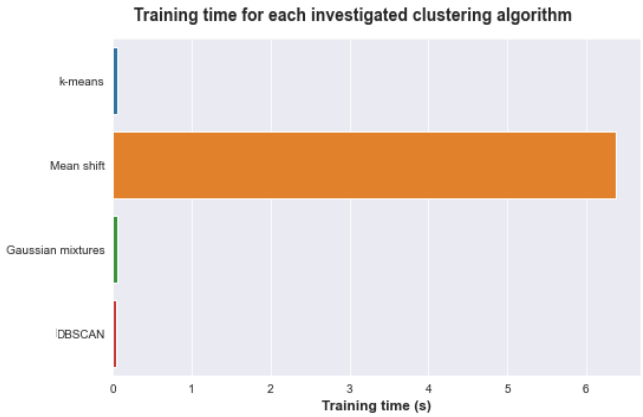


Fig. 9. Training time for each clustering algorithm excluding affinity propagation.

Compared to Affiliation Propagation and Mean Shift, training time is less for DBSCAN, K-Means, and Gaussian Mixture.

VI. Conclusion and Future Work

Various clustering methods have been employed to extract patterns from urban traffic data, including DBSCAN, K-Means, Affinity Propagation, Mean Shift, and Gaussian Mixture, notably DBSCAN, which produced enlightening results. DBSCAN proved to be very successful in its capacity to manage noise and detect groups of arbitrary shape, making it ideal for the complex and variable traffic patterns seen across multiple streets, the clustering procedure showed various traffic patterns, in some categorizing days a weekdays or weekends and many cases into four primary clusters. The rigorous clustering method enabled a thorough understanding of the traffic behavior, allowing for more focused traffic control tactics. DBSCANs exceptional efficiency when handling and analyzing traffic data demonstrates its potential to improve urban traffic systems.

For future studies, it is recommended to use big data technology to handle and analyze large amounts of traffic data collected from diverse sources such as GPS, detectors, and online platforms. This strategy will give more specific information on traffic trends. Furthermore, utilizing clustering algorithms to undertake comparative evaluations of travel habits across multiple cities can aid in identifying common difficulties and best practices, as well as revealing trends and solutions that may be broadly implemented.

References

[1] Pattanaik, V., Singh, M., Gupta, P. K., & Singh, S. K. (2016, November). Smart real-time traffic congestion estimation and clustering technique for urban vehicular roads. In 2016 IEEE region 10 conference (TENCON) (pp. 3420-3423). IEEE.

[2] Rouky, N., Bousouf, A., Benmoussa, O., & Fri, M. (2024). A spatiotemporal analysis of traffic congestion patterns using clustering algorithms: A case study of Casablanca. Decision Analytics Journal, 10, 100404.

[3] Wang, X., Qu, X., & Jin, S. (2020). Hotspot identification considering daily variability of traffic flow and crash record: A case study. Journal of Transportation Safety & Security, 12(2), 275-291.

[4] Shi, Y., Wang, D., Tang, J., Deng, M., Liu, H., & Liu, B. (2021). Detecting spatiotemporal extents of traffic congestion: A density-based moving object clustering approach. International Journal of Geographical Information Science, 35(7), 1449-1473.

[5] Yang, S., Wu, J., Qi, G., & Tian, K. (2017). Analysis of traffic state variation patterns for urban road network based on spectral clustering. Advances in Mechanical Engineering, 9(9), 1687814017723790.Stathopoulos, A., & Karlaftis, M. (2001). Temporal and spatial variations of real-time traffic data in urban areas. Transportation Research Record, 1768(1), 135-140.

[6] Angmo, R., Aggarwal, N., Mangat, V., Lal, A., & Kaur, S. (2021). An improved clustering approach for identifying significant locations from spatio-temporal data. Wireless Personal Communications, 121(1), 985-1009.

[7] Asadi, R., & Regan, A. (2019, November). Spatio-temporal clustering of traffic data with deep embedded clustering. In Proceedings of the 3rd ACM SIGSPATIAL International Workshop on Prediction of Human Mobility (pp. 45-52).

[8] Sfyridis, A., & Agnolucci, P. (2020). Annual average daily traffic estimation in England and Wales: An application of clustering and regression modelling. Journal of Transport Geography, 83, 102658.

[9] Wang, W., Shao, F., Du, B., & Lu, G. (2016). Analysis of vehicular traffic flow characteristics on urban expressways based on traffic flow data in Shanghai. Transportation Research Part C: Emerging Technologies, 69, 30-49.

[10] Taamneh, M., Taamneh, S., & Alkheder, S. (2017). Clustering-based classification of road traffic accidents using hierarchical clustering and artificial neural networks. International journal of injury control and safety promotion, 24(3), 388-395.

[11] Acun, G., & Gol, E. (2021). Traffic prediction and management using ARIMA and K-means clustering on large-scale traffic networks. Journal of Intelligent Transportation Systems, 25(5), 385-399.

[12] Zou, X., & Chung, E. (2024). Traffic prediction via clustering and deep transfer learning with limited data. Computer-Aided Civil and Infrastructure Engineering.

[13] Nguyen, T. T., Krishnakumari, P., Calvert, S. C., Vu, H. L., & Van Lint, H. (2019). Feature extraction and clustering analysis of highway congestion. Transportation Research Part C: Emerging Technologies, 100, 238-258.

[14] Chen, Y., Li, Z., & Zhang, J. (2022). Automobile flow forecasting with Graph Attention Network and spatial-temporal clustering. Computer-Aided Civil and Infrastructure Engineering, 37(3), 230-245.

Prediction of Outpatient No-Show Appointments Using Machine Learning Algorithms for Pediatric Patients in Saudi Arabia

Abdulwahhab Alshammari[1], Fahad Alotaibi[2], Sana Alnafrani[3]
Health Informatics Department, College of Public Health and Health Informatics[1]
Information Technology Department[2], Pediatric Department[3]
King Saud bin Abdulaziz University for Health Sciences, Riyadh, Saudi Arabia[1, 2, 3]
King Abdullah International Medical Research Center, Riyadh, Saudi Arabia[1, 2, 3]
Ministry of the National Guard-Health Affairs, Riyadh, Saudi Arabia[1, 2, 3]

Abstract*—Patient no-shows are prevalent in pediatric outpatient visits, leading to underutilized medical resources, increased healthcare costs, reduced clinic efficiency, and decreased access to care. The use of machine learning techniques provides insights to mitigate this problem. This study aimed to develop a predictive model for patient no-shows at the Ministry of National Guard Health-Affairs, Saudi Arabia, and evaluate the results of various machine learning algorithms in predicting these events. Four machine learning algorithms - Gradient Boosting, AdaBoost, Random Forest, and Naive Bayes - were used to create predictive models for patient no-shows. Each model underwent extensive parameter tuning and reliability assessment to ensure robust performance, including sensitivity analysis and cross-validation. Gradient Boosting achieved the highest area under the receiver operating curve (AUC) of 0.902 and Classification Accuracy (CA) of 0.944, while the AdaBoost model achieved an AUC of 0.812 and CA of 0.927. The Naive Bayes and Random Forest models achieved AUCs of 0.677 and 0.889 and CAs of 0.915 and 0.937, respectively. The confusion matrix demonstrated high true-positive rates for no-shows for the Gradient Boosting and Random Forest models, while Naive Bayes had the lowest values. The Gradient Boosting and Random Forest models were most effective in predicting patient no-shows. These models could enhance outpatient clinic efficiency by predicting no-shows. Future research can further refine these models and investigate practical strategies for their implementation.

Keywords*—No-show; pediatric; machine learning; algorithms; prediction; outpatients

I. Introduction

Patient no-shows are one of the main challenges in the healthcare sector, disturbing the workflow or affecting cost load, reflecting on the quality and performance [1]. Reducing the number of no-shows significantly impacts healthcare institutions' services, reducing financial costs and effectively utilizing resources to improve patient service [2]. The issue of no-shows is a recurring problem that hinders the efficient utilization of human resources [2,3]. In addition, it increases patient waiting time and negatively impacts the workflow by wasting time for healthcare providers [3].

No-shows are a significant concern for healthcare institutions and may be expensive and inconvenient [3]. Capacity is underutilized, and costly assets are underused [4]. Researchers have found that eliminating non-cancelled no-show appointments may considerably influence productivity, profitability, and clinical outcomes [4]. Machine learning techniques would provide a solution [2]. Thus, finding a way to predict no-shows would facilitate the effective utilization of hospital resources and enhance the satisfaction of both providers and patients, ultimately improving healthcare quality [3–5]. There are some practices used to overcome no-shows, such as overbooking and walk-ins, but these methods are still not the ideal solutions to address the issue of no-shows; there is no effective tool in the electric healthcare system to detect patients at a higher risk of not showing up [6].

Prediction is the most challenging part of human behavior; presuming and predicting this pattern or behavior of a no-show [7]. Finding associations with variables and attributes would facilitate the prediction of no-show appointments [8]. Enabling a prediction model to predict no-shows would help to effectively utilize human resources, reduce financial losses, and increase patient satisfaction [9]. It would also help to improve appointment scheduling, reduce waiting time, and increase the number of patients seen daily [10]. Therefore, a no-show prediction tool is an added value for any organization [11].

A study by Alshammari aims to predict no-shows through machine learning [12]. The dataset includes more than thirty-three million outpatient appointments [12]. The dataset was extracted for a period of nearly three years (January 2016 - July 2019) from the Health Information System (HIS) at all facilities in the central region of the Ministry of National Guard Health Affairs (MNGHA), Saudi Arabia. The authors state that nearly 77,000 outpatient appointments were scheduled monthly at the MNGHA in the Riyadh Region. The patients' ages ranged from 5 to 69 years old. The highest no-show rate was observed among patients over 45 years old. Almost 85% of the no-shows were a national citizen. The study utilized three machine-learning algorithms: Deep Neural Network, AdaBoost, and Naive Bayes. The results of this study were promising, showing that it achieved a 98% precision rate using the deep learning model [12].

The authors of Alshammari's other research paper attempt to develop a prediction model based on a machine learning

algorithm for cases where patients do not attend their scheduled appointments [13]. The dataset was obtained from the Kaggle database of hospital appointments booked between April 29, 2016, and June 8, 2016. The dataset included (110,528) medical appointments. Recursive Feature Elimination (RFE) was implemented using Python to exclude unrelated items from the dataset. After running the RFE to remove the unrelated attributes, as including all variables can lead to highly complex modeling, nearly 83,000 appointments were included. The no-show rate was approximately 20%. The dataset has been divided into 70% for the training dataset and 30% for the test dataset. The machine learning algorithms used in this study are Decision Trees and AdaBoost. Multiple variables were used to determine the optimal model for predicting no-shows. The results showed high precision and recall, indicating that the Decision Tree outperformed the AdaBoost results [13].

A study by AlMuhaideb used machine learning to create a model for predicting no-shows in outpatients [14]. The research data were extracted from the health information system, which captures records of patient visits for outpatients. The dataset contains almost more than 1 million outpatient records. The period of this dataset was between January and December 2014, and the no-show rate was 11.3%. The machine prediction models used are JRip and Hoeffding tree algorithms. The machine learning software used was Weka. The dataset was cleansed and preprocessed to conduct the modeling analysis. Both the JRip and Hoeffding algorithms provided rational degrees of accuracy levels of almost 77%. The study showed that the no-shows could be predicated using a machine-learning model [14].

Hamdan, A., and Abu Bakar, A. published a study in 2023 on outpatient no-show appointments in a Malaysian tertiary hospital [15]. The study aimed to develop a model for predicting patient no-show appointments using machine learning. The data were collected through 2019 and included 246,943 appointment records with 14 attributes, including demographics and appointment data. The result shows that 69,173 patients did not attend their appointment, which accounts for about 28% of the dataset. The machine learning model used seven algorithms: logistic regression (LR), decision tree (DT), k-nearest neighbors (k-NN), Naïve Bayes (NB), random forest (RF), gradient boosting (GB), and multilayer perceptron (MLP). Three different train and testing splits were applied at 60:40, 70:30, and 80:20, and ten folding validations were performed on each split using Python. The evaluation metrics included accuracy, AUC value, and F1 score. The GB scored the highest accuracy of 78%.

Therefore, finding a way to predict the no-show or high no-show candidates will help healthcare organizations overcome this issue [16]. Developing a prediction model will help stakeholders mitigate the anticipated effects of no-shows and enhance healthcare efficiency by optimizing resource utilization [17]. The prediction model can help to identify patients at high risk of no-shows based on factors such as age, gender, appointment type, past behavior, and geographic location [18]. A machine learning model that predicts patient no-shows can enhance clinical efficiency by optimizing resource allocation, reducing wasted time through overbooking appointments without compromising patient care, and allocating efficient slot allocation by understanding no-show patient patterns.

The main objective of this study is to develop a machine-learning model capable of accurately predicting the likelihood of a pediatric patient missing a scheduled appointment. Other objectives include identifying key factors influencing pediatric no-show rates to inform targeted interventions and optimize appointment scheduling and resource allocation based on no-show predictions. This model can improve patient satisfaction by reducing wait times and increasing appointment availability. From a broader perspective, this model can help better understand pediatric patient behavior and healthcare utilization.

In this study, we aim to develop a predictive model and evaluate its performance using machine learning algorithms to predict pediatric patient no-shows in pediatric outpatient visits at the Ministry of National Guard - Health Affairs (MNGHA) using machine learning techniques. This study differs from the mentioned studies [12–15]. This study targets a more specific population of pediatric patients. It uses primary data extracted from a tertiary hospital and applies multiple or different machine learning algorithms in a single study.

II. METHOD

This study is a retrospective exploratory/predictive study. It aims to predict the no-show based on machine learning techniques and Machine Learning (ML) algorithms such as Gradient Boosting (GB), AdaBoost, Naive Bayes (NB), and Random Forest, which are all supervised Machine Learning algorithms.

This study was conducted ethically, following established guidelines and protocols to ensure patient privacy and data confidentiality. This study has been approved by an Institutional Review Board (IRB) committee from the King Abdullah International Medical Research Center (KAIMRC).

A. Study Area, Settings, and Subjects

The study was conducted on pediatric patients at the Ministry of National Guard Health Affairs (MNGHA) in Saudi Arabia. The data were extracted from the BESTCare health information system used in the MNGHA. They encompass appointments scheduled throughout the day from January 1, 2021, to May 5, 2022.

The patient records eligible for the study must fulfill the inclusion and exclusion criteria. All patients under the age of 14 who had a scheduled visit to a pediatric outpatient clinic (pediatric patients) were included. A patient no-show is a visit in which the patient fails to attend a scheduled appointment without providing prior notice. Canceled appointments before the clinic were not counted as no-shows to ensure all missed appointments were not intervenable. Emergency visits, unscheduled visits, such as walk-ins, and patients older than 14 years were excluded.

B. Data Collection, Management, and Analysis Plan

The dataset used in this study comprises 358,759 outpatient appointment visits, with a mix of nominal, ordinal, and numeric attributes related to patient demographics, appointment details, and medical history. The dataset includes data on patients' age groups, gender, nationality, appointment types, region, and appointment times to ensure a representative sample of the study

population across different age groups, genders, and appointment types.

This study utilizes historical data on patient visits to predict the likelihood of no-shows. The analysis plan encompasses the standard stages of the data mining process, including data collection and understanding, data preparation, model selection, model building, and model evaluation.

The study collected various attributes from the dataset to analyze and predict patient no-show appointments. Table I summarizes these attributes, including their descriptions and types. The attributes capture information such as visit ID, region, facility, department, clinic, patient demographics (gender, nationality, age), appointment details (date, time), diagnosis information, appointment message status, patient's address, sponsor eligibility, and more. One notable attribute is the lead time, which represents the difference between the booking and appointment dates. This table references the attributes used in analyzing and predicting no-show appointments.

The data was cleaned and preprocessed to ensure the quality of the dataset. This included handling missing values, removing irrelevant attributes, and transforming variables required for model development. For example, the lead time variable was derived by calculating the difference between the booking and appointment dates. In addition, the appointment time was categorized into AM and PM.

Relevant features were selected based on their potential impact on predicting no-show appointments. Factors such as age group, gender, nationality, appointment type, region, and appointment time were included in the analysis, as these were expected to contribute to the prediction of no-show appointments.

TABLE I. DESCRIPTION OF THE COLLECTED AND DERIVED ATTRIBUTES

Data Attributes			
No.	**Attribute Name**	**Description**	**Type**
1	Visit_ID	Visit ID	Numeric
2	Region	Region Name	Nominal
3	Facility	Facility Name	Nominal
4	HSP_TP_CD	Hospital or facility type code	Numeric
5	HSPI_TP_CD	Internal hospital or facility code	Numeric
6	Department	Medical Department	Nominal
7	Department_CD	Medical Department Code	Numeric
8	Clinic	Clinic name	Nominal
9	Clinic_CD	Clinic Code	Numeric
10	MRN	Patient's Medical record number	Numeric
11	Appointment_DT	Appointment Date	Ordinal
12	Appointment_TIME	Appointment Time	Ordinal
13	Appointment_DTM	Appointment Date & Time	Ordinal
14	Visit_Type	Patient Visit type	Nominal
15	Appointment_Booking_DTM	Appointment Booking Date & Time	Ordinal
16	Appointment_Booking_TIME	Appointment Booking Time	Ordinal
17	ICD10_CD	ICD10 code for the diagnosis	Nominal
18	Diagnosis	Diagnosis	Nominal
19	Flag	Show/no-show	Nominal
20	MSG_SENT_YN	Appointment message sends the status	Nominal
21	MSG_Status	Appointment message status	Nominal
22	Gender	Patient Gender	Nominal
23	Nationality	Patient Nationality	Nominal
24	Age	Patient Age	Ordinal
25	Address1	Patient Region or an area name or code of the patient's residence	Nominal
26	Address2	Patient district name or code of patient's residence	Nominal
27	Sponsor_Eligibility	Patient's Sponsor_Eligibility	Nominal
28	ETPR_PT_NO	Patient Enterprise record number	Numeric
29	Cachement_Area_CD	Area name of the patient's residence	Numeric
30	Cachement_Area_NAME	Area Code of the patient's residence	Nominal
31	Cachement_FCLT_NO	Facility Code of the patient's residence	Numeric
32	Cachement_FCLT_NAME	Facility name of the patient's residence	Nominal
33	Lead time	Difference between Appointment Booking and Appointment dates	Numeric
34	Appointment time AM/PM	AM/PM	Ordinal

C. Model Selection, Building, and Evaluation

We applied four machine learning algorithms to the preprocessed data: Gradient Boosting, AdaBoost, Naive Bayes, and Random Forest. Gradient Boosting, AdaBoost, Naive Bayes, and Random Forest are flexible, well-suited algorithms for handling complex relationships in large datasets. These algorithms are also well-suited for handling categorical features and numerical data [19,20]. We used 10-fold cross-validation to assess model performance and avoid overfitting. Each model evaluation was based on various metrics, including Area Under the Receiver Operating Characteristic Curve (AUC), Classification Accuracy (CA), F1 score, Precision, and Recall. Furthermore, the preprocessing steps and hyperparameters for each model were recorded, ensuring the integrity and consistency of the input data.

Gradient Boosting (XGBoost): The model was trained and constructed by combining 100 individual decision trees and a learning rate of 0.3, which determines the weight given to each tree's prediction when they are combined. In this case, a learning rate of 0.3 balances responsiveness and stability in the model's predictions. The maximum depth of individual trees was set at 20, providing a good balance between the model's complexity and its ability to learn the underlying patterns in the data.

Regularization was applied with a lambda value of 7 to prevent overfitting. Regularization helps prevent overfitting when a model becomes too complex and starts to memorize the training data instead of learning the underlying patterns.

The lambda value of 7 represents the strength of the regularization. A higher lambda value increases the penalty for complex models, encouraging the model to simplify its predictions and avoid overfitting. By applying regularization with a lambda value of 7, the model aims to balance capturing essential patterns in the data and avoiding excessive complexity.

We also experimented with different fractions of training instances and features for each tree, level, and split. The fraction was set to 1.0 in all cases to use all available data and features. We fixed the random seed for replicable training to ensure that specific conditions did not affect our results.

The preprocessing steps for Gradient Boosting were removing instances with unknown target values, customizing categorical variables using one-hot-encoding, removing empty columns, and imputing missing values with mean values.

AdaBoost: The model was built using a base estimator (a decision tree) and 100 additional estimators. A high learning rate of 0.999 was set to give more weight to the most recent data. The classification algorithm SAMME was used to boost the model. As in the case of XGBoost, we ensured the replicability of results by fixing the random seed.

AdaBoost's preprocessing steps included removing instances with unknown target values, customizing categorical variables using one-hot encoding, removing empty columns, and imputing missing values with mean values.

Naive Bayes: The Naive Bayes algorithm does not have specific hyperparameters like other algorithms, but preparing the data well for this model is essential. For Naive Bayes, the preprocessing step was removing empty columns.

Random Forest: The model was trained with 100 trees, and the number of attributes considered at each split was set to 5. This allowed the model to consider a balanced number of attributes at each node to achieve a good compromise between bias and variance. Growth control measures were applied to avoid creating complex models that could lead to overfitting. We ensured that subsets smaller than a certain threshold were not split.

Preprocessing of the Random Forest included removing instances with unknown target values, customizing categorical variables using one-hot encoding, removing empty columns, and imputing missing values with mean values.

III. Results and Analysis

This section presents the research project's findings based on the statistical analysis and the development of the machine learning model described in the previous section. The results are presented as descriptive statistics, model performance, and critical findings from the analysis.

A. Description Analysis

This section provides an overview of the dataset, showing patterns and trends in no-show appointments among different patient demographics and appointment attributes.

1) Age group: In Table II, the data show different age groups, including infants (0-12 months), toddlers (1-3 years), preschoolers (3-6 years), school-age children (6-12 years), and adolescents (12-14 years). The table displays the number of patients who attended their appointments and those who did not (no-shows) for each age group. The "Show" column represents the number of patients who attended their appointments, while the "No-Show" column represents the number of patients who did not show up. The "Total" column indicates the total number of patients in each age group. The school-age children (6-12 years) accounted for the highest proportion of no-show appointments, followed by infants (0-12 months) and preschoolers, see Table II.

TABLE II. Description and Distribution of Age Groups with No-Show Percentages

Age Groups Categories			
Age Group	**Show (%)**	**No-show (%)**	**Total**
Infant (0-12 Months)	79,948 (92.7%)	6,281 (7.3%)	86,229
Toddler (1-3 Years)	53,150 (92.1%)	4,460 (7.9%)	57,610
Preschool (3-6 Years	59,248 (91.1%)	5,740 (8.9%)	64,988
School-age (6-12 Years)	101,212 (91.1%)	9,831 (8.9%)	111,043
Adolescent (12-14 Years)	35,264 (90.6%)	3,624 (9.4%)	38,888

2) Gender: Table III presents the data for gender distribution. The table displays the number of patients who attended their appointments and those who did not (no-shows)

for each gender. The dataset contained a higher proportion of shows with male patients than with female patients.

TABLE III. DISTRIBUTION OF AND PERCENTAGES OF GENDER

Gender and Distribution			
Gender	**Show (%)**	**No-show (%)**	**Total**
Male	171,469 (91.9%)	15,382 (8.2%)	186,851
Female	157,353 (91.5%)	14,554 (8.5%)	171,907

3) Nationality: Table IV presents the data for individuals of Saudi and non-Saudi nationality. The majority of patients were of Saudi Nationality, with a small percentage of non-Saudi patients. The rate of no-shows is higher for non-Saudi patients.

TABLE IV. DISTRIBUTION OF AND PERCENTAGES OF NATIONALITY

Nationality and Distribution			
Nationality	**Show (%)**	**No-show (%)**	**Total**
Saudi	325,944 (91.7%)	29,409 (8.3%)	355,353
Non-Saudi	2,878 (84.6%)	527 (15.4%)	3,405

4) Appointment types: Table V presents data on different appointment types, including New Patient (NP), First Visit (FV), and Follow-up (FU). Follow-up appointments had the highest no-show rates, followed by first visits and new patient appointments.

TABLE V. DISTRIBUTION AND PERCENTAGES OF APPOINTMENT TYPES

Appointment Types and Distribution			
Appointment Type	**Show (%)**	**No-show (%)**	**Total**
New Patient (NP)	10,182 (95.9%)	446 (4.1%)	10,628
First visit (FV)	171,415 (92.6%)	13,681 (7.4%)	185,096
Follow-up (FU)	147,225 (90.3%)	15,809 (9.7%)	163,034

Table VI shows the data for the Central, Eastern, and Western regions, showing attendance and no-show numbers. Geographically, the proportion of no-shows was highest in the Central region, followed by the Western and Eastern regions.

TABLE VI. DISTRIBUTION OF AND PERCENTAGES OF INCLUDED REGIONS

Regions and Distribution			
Region Name	**Show (%)**	**No-show (%)**	**Total**
Central	218,151 (93.9%)	13,994 (6.1%)	232,145
Eastern	32,206 (93.1%)	2,388 (6.9%)	34,594
Western	78,465 (85.2%)	13,554 (14.8%)	92,019

5) Appointment time and hours: Table VII provides information on the appointment times of the day, namely AM and PM, indicating the number of patients who showed up and those who did not. The data showed that appointments in the morning (AM) had slightly higher no-show rates than afternoon (PM) appointments.

TABLE VII. DISTRIBUTION OF AND PERCENTAGES OF APPOINTMENT TIME OF THE DAY

Time of the Day and Distribution			
Gender	**Show (%)**	**No-show (%)**	**Total**
AM	166,642 (90.8%)	16,737 (9.2%)	183,379
PM	162,180 (92.5%)	13,199 (7.5%)	175,379

Table VIII provides information on appointment hours, including the number of no-shows and appointments attended for each period. The table displays data for time slots 7-9, 9-12, 12-15, 15-17, and beyond working hours. Table IX presents data on appointment days of the week, showing the number of no-shows and shows for each day.

TABLE VIII. DISTRIBUTION OF APPOINTMENT HOURS CATEGORIES

Appointment Hours Categories of the Day			
Appointment Hour	**Show (%)**	**No-show (%)**	**Total**
7-9	3,855 (8.7%)	40,597 (91.3%)	44,452
9-12	12,867 (9.6%)	121,583 (90.4%)	134,450
12-15	9,221 (7.9%)	106,923 (92.1%)	116,144
15-17	3,659 (9.8%)	33,723 (90.2%)	37,382
After working hours	334 (1.3%)	25,996 (98.7%)	26,330

TABLE IX. DISTRIBUTION OF APPOINTMENT DAY

Appointment Day			
Appointment Day	**Show (%)**	**No-show (%)**	**Total**
Sunday	6,040 (8.4%)	66,002 (91.6%)	72,042
Monday	6,495 (8.2%)	72,645 (91.8%)	79,140
Tuesday	7,423 (9.9%)	67,213 (90.1%)	74,636
Wednesday	5,662 (7.8%)	66,647 (92.2%)	72,309
Thursday	4,168 (7.7%)	49,965 (92.3%)	54,133
Friday	85 (2.8%)	2,958 (97.2%)	3,043
Saturday	63 (1.8%)	3,392 (98.2%)	3,455

B. Model Performance

The results suggest that all four models (Gradient Boosting (GB), AdaBoost, Naive Bayes (NB), and Random Forest) have performed reasonably well, but Gradient Boosting stood out as the most robust model in our study. Our sensitivity analysis, which involved varying the training and testing data splits and the machine learning algorithms' hyperparameters enhanced the reliability of our results. Comprehensive evaluation ensures that the models are reliable and not overly sensitive to the specific selection of parameters, confirming the robustness of the findings. The consistency in preprocessing across models further strengthens the credibility of the outcomes. The performance of the four machine learning models was evaluated using cross-validation with ten subsets and a separate testing set.

Table X summarizes the model performance, with Gradient Boosting demonstrating the best performance by achieving the highest values for AUC, CA, F1 score, precision, and recall. This indicates that Gradient Boosting outperformed the other models in accurately predicting the likelihood of no-show appointments.

TABLE X. MODEL'S ALGORITHMS AND EVALUATION METRICS PERFORMANCE

Algorithms and Evaluation Metrics					
Algorithm	**AUC**	**CA**	**FI**	**Precision**	**Recall**
Gradient Boosting	0.902	0.944	0.937	0.939	0.944
AdaBoost	0.812	0.927	0.926	0.924	0.927
Naive Bayes	0.677	0.915	0.877	0.861	0.915
Random Forest	0.889	0.937	0.925	0.931	0.937

The Receiver Operating Characteristic (ROC) analysis is a graphical representation that illustrates the performance of a binary classification model. In this study, the ROC plot (Fig. 1) demonstrates the performance of the machine learning models in predicting no-show appointments.

The x-axis represents the False Positive Rate (FPR), which measures the proportion of false positives (show appointments incorrectly classified as no-shows) to all actual negatives (show appointments). The y-axis represents the True Positive Rate (TPR), which measures the proportion of true positives (no-show appointments correctly classified as a no-show) to all actual positives (no-show appointments).

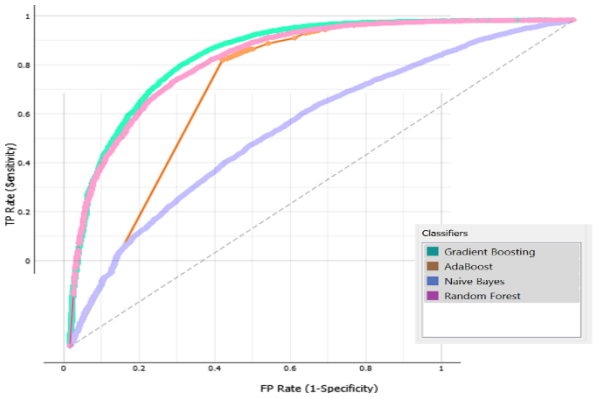


Fig. 1. ROC analysis diagram with classifier.

A curve on the ROC plot represents each model. The closer the curve is to the top-left corner, the better the model's performance. The ideal scenario is a model with a curve that reaches the top-left corner, indicating a high TPR and a low FPR.

By examining the ROC plot, we can observe that the Gradient Boosting model exhibits the highest performance among the four models. It shows the highest TPR for a given FPR threshold, indicating its ability to identify and classify no-show appointments accurately. The other models, including AdaBoost, Naive Bayes, and Random Forest, also demonstrate varying performance levels, with their respective curves positioned below that of Gradient Boosting.

The ROC Analysis Diagram visually represents the models' performance, distinguishing between show and no-show appointments. It helps evaluate and compare the predictive capabilities of models and select the most suitable one for accurately predicting no-shows in future scenarios.

Table XI shows the confusion matrix for the four represented models and the predicted versus actual outcomes for no-show and show appointments. The matrices provide insight into each model's performance by showing the True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN) rates.

TABLE XI. DISTRIBUTION OF APPOINTMENT DAY (CONFUSION MATRIX FOR THE FOUR REPRESENTED MODELS)

Confusion Matrix			
Gradient Boosting (GB) Algorithm			
		Actual Values	
		Positive	**Negative**
Predicted Values	Positive	TP (80.4 %)	FP (4.9 %)
	Negative	FN (19.6 %)	TN (95.1 %)
AdaBoost Algorithm			
		Actual Values	
		Positive	Negative
Predicted Values	Positive	TP (56.9 %)	FP (4.3 %)
	Negative	FN (43.1 %)	TN (95.7 %)
Naïve Bayes			
		Actual Values	
		Positive	Negative
Predicted Values	Positive	TP (24.8 %)	FP (8.3 %)
	Negative	FN (75.2 %	TN (91.7 %)
Random Forest			
		Actual Values	
		Positive	**Negative**
Predicted Values	Positive	TP (81.3 %)	FP (5.8 %)
	Negative	FN (18.7 %)	TN (94.2 %)

1) Gradient Boosting (GB):

- The Gradient Boosting model delivered the highest performance among all the evaluated models with an AUC of 0.902, CA of 0.944, F1 score of 0.937, precision of 0.939, and recall of 0.944.
- The model accurately predicted no-show appointments at 80.6% and showed appointments at 95.1%.
- The model misclassified 4.9% of actual no-show appointments and 19.6% of existing show appointments as no-show appointments.
- Gradient Boosting (GB) demonstrated high accuracy in predicting no-show appointments, with 80.4% True Positives (correctly predicted no-shows) and 95.1% True Negatives (correctly predicted shows). However, the model also exhibited a 19.6% False Positive rate (incorrectly predicted no-shows) and a 4.9% False Negative rate (incorrectly predicted shows).

2) AdaBoost:

- This model achieved an AUC of 0.812, CA of 0.927, F1 score of 0.926, precision of 0.924, and recall of 0.927.

- The model accurately predicted no-show appointments at 56.8% and showed appointments at 95.7%.
- The model misclassified 4.3% of actual no-show appointments and 43.1% of existing show appointments as no-show appointments.
- The AdaBoost model yielded 56.9% True Positives and 95.7% True Negatives, exhibiting 43.1% False Positive and 4.3% False Negative rates. This performance suggests a moderate ability to predict no-show appointments correctly.

3) Naive Bayes (NB):

- This model had an AUC of 0.677, CA of 0.915, F1 score of 0.877, precision of 0.861, and recall of 0.915.
- The model accurately predicted no-show appointments at 24.8% and showed appointments at 91.7%.
- The model misclassified 8.3% of actual no-show appointments and 75.2% of existing show appointments as no-show appointments.
- The Naive Bayes (NB) model achieved a lower accuracy in predicting no-show appointments, with a 24.8% True Positive rate and a 91.7% True Negative rate. The model had a high False Positive rate of 75.2% and a False Negative rate of 8.3%.

4) Random Forest (RF):

- This model reported an AUC of 0.889, CA of 0.937, F1 score of 0.925, precision of 0.931, and recall of 0.937.
- The model accurately predicted no-show appointments at 81.3% and showed appointments at 94.2%.
- The model misclassified 5.8% of actual no-show appointments and 18.4% of existing show appointments as no-show appointments.
- The RF model strongly predicted no-show appointments, with an 81.3% True Positive rate and a 94.2% True Negative rate. The model had an 18.4% False Positive rate and a 5.8% False Negative rate.

The Gradient Boosting model exhibits effective performance with a high true positive rate for no-show and show appointments and relatively low misclassification rates. Specifically, it correctly identified 80.4% of no-show appointments and 95.1% of show appointments. This level of performance indicates a strong ability of this model to distinguish between the two classes accurately.

On the contrary, the Naive Bayes model demonstrates the poorest performance, with the lowest true positive rate for no-show appointments (24.8%) and a relatively lower success rate for show appointments (91.7%). It had the highest misclassification rate for show appointments, signaling potential weaknesses in the model's ability to identify true positives in a balanced manner correctly.

The Random Forest model showed strong performance with a high true positive rate for no-show appointments (81.3%) and a high success rate for show appointments (94.2%). This reflects a balanced performance for both classes, making it a reliable model for this prediction task.

Finally, while not as proficient as Gradient Boosting or Random Forest, the AdaBoost model still showed a reasonable true positive rate for no-show appointments (56.9%) and a high success rate for show appointments (95.7%).

In conclusion, based on these results, the Gradient Boosting and Random Forest models demonstrate superior performance in predicting no-show appointments compared to the AdaBoost and Naive Bayes models. This comprehensive evaluation gives insights into each model's strengths and weaknesses. It provides valuable information for selecting the most suitable model for predicting no-show appointments in future studies.

IV. Discussion

The primary aim of this study was to evaluate and predict no-show appointments at MNGHA pediatric outpatient visits using machine learning models. Our research builds upon the existing literature by developing a predictive model specifically for pediatric outpatient settings, focusing on a large dataset that includes demographic, appointment-related, and geographic factors.

Our findings contribute to the existing body of knowledge by identifying patterns and trends in no-show appointments across various patient demographics and appointment attributes. This information can help healthcare providers better understand the factors contributing to no-show appointments and develop targeted strategies for reducing no-show rates [21,22].

The results of this study indicate that the GB and RF models outperformed other models in predicting no-show appointments. This superior performance can be attributed to the model's ability to capture complex relationships between various features in the dataset, making it particularly suitable for our research objective.

The strengths of this study include the large and diverse dataset, which allowed us to develop a robust and reliable predictive model. Moreover, the use of multiple machine learning models and the implementation of cross-validation for model evaluation ensure the validity of our findings [23].

A. Limitation

First, the data used in this study were limited to a single healthcare organization, "MNGHA," which may not be representative of other pediatric outpatient settings [24]. Future research could consider including data from multiple healthcare systems to further validate the predictive model's generalizability. Second, the dataset should have included certain factors such as socioeconomic status, transportation availability, weather conditions, and patient preference; including these factors might enhance the model's predictive capabilities [25].

Future research directions could involve the following:

1) Expanding the dataset to include additional pediatric outpatient settings to validate the predictive model's performance across different healthcare organizations [24].

2) Incorporating other relevant factors, such as socioeconomic status, transportation availability, and weather conditions, further enhances the model's predictive capabilities [25,26].

3) Investigating the potential impact of targeted interventions, such as appointment reminders or personalized follow-up, on reducing no-show rates based on the predictive model's output [25–27].

In conclusion, this study's results provide valuable insights into the factors associated with no-show appointments in pediatric outpatient settings. The machine learning model developed can aid healthcare providers in predicting no-show appointments, optimizing resource management, and improving patient care.

V. CONCLUSION

The primary contribution of this research project is developing a machine learning model to predict no-show appointments in pediatric outpatient settings. Our study has identified patterns and trends in no-show appointments by analyzing a large and diverse dataset. This analysis can assist healthcare providers in optimizing resource management and enhancing patient care. The GB and RF models emerged as the best performers in predicting no-show appointments, demonstrating their potential utility in pediatric outpatient settings.

Our findings build upon existing literature, highlighting the importance of understanding factors contributing to no-show appointments in pediatric populations. These insights can guide healthcare providers in developing targeted strategies for reducing no-show rates and enhancing overall healthcare delivery.

While our study has some limitations, such as focusing on a single healthcare system and excluding certain factors, it provides a solid foundation for future research. Expanding the dataset to include additional pediatric outpatient settings, incorporating other relevant factors, and investigating the impact of targeted interventions based on the predictive model's results may further deepen the understanding of no-show appointments and help improve healthcare management.

In conclusion, this study provides valuable insight into the factors associated with no-show appointments in outpatient pediatrics. It gives healthcare providers a powerful tool for effectively predicting and managing missed appointments. Through continued research and model improvement, we can further enhance our understanding of no-show appointments and optimize resource allocation in outpatient pediatric care.

ACKNOWLEDGMENT

The authors would like to acknowledge Reem Alamr, editor at the Office of Research, King Saud Bin Abdulaziz University for Health Sciences, for her contribution to editing and proofreading the manuscript.

REFERENCES

[1] Huang Y, Hanauer DA. Patient no-show predictive model development using multiple data sources for an effective overbooking approach. Appl Clin Inform. 2014;5(3):836–60.

[2] Denney J, Coyne S, Rafiqi S. Machine Learning Predictions of No-Show Appointments in a Primary Care Setting. SMU Data Sci Rev. 2019;2(1):1–32.

[3] Norris JB, Kumar C, Chand S, Moskowitz H, Shade SA, Willis DR. An empirical investigation into factors affecting patient cancellations and no-shows at outpatient clinics. Decis Support Syst. 2014 Jan;57:428–43.

[4] Huang Z, Ashraf M, Gordish-Dressman H, Mudd P. The financial impact of clinic no-show rates in an academic pediatric otolaryngology practice. Am J Otolaryngol. 2017;38(2):127–9.

[5] Mochón F, Elvira C, Ochoa A, Gonzalvez JC. Machine-Learning-Based No Show Prediction in Outpatient Visits. Int J Interact Multimed Artif Intell. 2018;4(Special Issue on Big Data and e-Health):29–34.

[6] Devasahay SR, Karpagam S, Ma NL. Predicting appointment misses in hospitals using data analytics. mHealth. 2017 Apr;3(12):1–9.

[7] Goffman RM, Harris SL, May JH, Milicevic AS, Monte RJ, Myaskovsky L, et al. Modeling Patient No-Show History and Predicting Future Outpatient Appointment Behavior in the Veterans Health Administration. Mil Med. 2017 May;182(5):e1708–14.

[8] Dantas LF, Hamacher S, Cyrino Oliveira FL, Barbosa SDJ, Viegas F. Predicting Patient No-show Behavior: a Study in a Bariatric Clinic. Obes Surg. 2019 Jan;29(1):40–7.

[9] Nelson A, Herron D, Rees G, Nachev P. Predicting scheduled hospital attendance with artificial intelligence. NPJ Digit Med. 2019 Apr;2(26):1–7.

[10] Roy S, Gupta A, Datta S, Das A, Pradhan S, Das T, et al. Prediction of heart diseases using machine learning. AIP Conf Proc. 2023 Nov;2851(1):020001.

[11] Alhamad Z. Reasons for missing appointments in general clinics of primary health care center in Riyadh Military Hospital, Saudi Arabia. Int J Med Sci Public Health. 2013;2(2):258–67.

[12] Alshammari R, Daghistani T, Alshammari A. The Prediction of Outpatient No-Show Visits by Using Deep Neural Network from Large Data. Int J Adv Comput Sci Appl. 2020;11(10):533–9.

[13] Alshammari A, Almalki R, Alshammari R. Developing a Predictive Model of Predicting Appointment No-Show by Using Machine Learning Algorithms. J Adv Inf Technol. 2021;12(3):234–9.

[14] AlMuhaideb S, Alswailem O, Alsubaie N, Ferwana I, Alnajem A. Prediction of hospital no-show appointments through artificial intelligence algorithms. Ann Saudi Med. 2019 Dec;39(6):373–81.

[15] Ahmad Hamdan AF, Abu Bakar A. Machine Learning Predictions on Outpatient No-Show Appointments in a Malaysia Major Tertiary Hospital. Malays J Med Sci MJMS. 2023 Oct;30(5):169–80.

[16] Dashtban M, Li W. Deep Learning for Predicting Non-attendance in Hospital Outpatient Appointments. In: Proceedings of the 52nd Hawaii International Conference on System Sciences. 2019. p. 3731–40.

[17] Harvey HB, Liu C, Ai J, Jaworsky C, Guerrier CE, Flores E, et al. Predicting No-Shows in Radiology Using Regression Modeling of Data Available in the Electronic Medical Record. J Am Coll Radiol JACR. 2017 Oct;14(10):1303–9.

[18] Carreras-García D, Delgado-Gómez D, Llorente-Fernández F, Arribas-Gil A. Patient No-Show Prediction: A Systematic Literature Review. Entropy. 2020 Jun 17;22(6):675.

[19] Tanha J, Abdi Y, Samadi N, Razzaghi N, Asadpour M. Boosting methods for multi-class imbalanced data classification: an experimental review. J Big Data. 2020 Sep 1;7(1):70.

[20] Wickramasinghe I, Kalutarage H. Naive Bayes: applications, variations and vulnerabilities: a review of literature with code snippets for implementation. Soft Comput. 2021 Feb 1;25(3):2277–93.

[21] Mohammadi I, Wu H, Turkcan A, Toscos T, Doebbeling BN. Data Analytics and Modeling for Appointment No-show in Community Health Centers. J Prim Care Community Health. 2018 Nov 17;9:2150132718811692.

[22] Daggy J, Lawley M, Willis D, Thayer D, Suelzer C, DeLaurentis PC, et al. Using no-show modeling to improve clinic performance. Health Informatics J. 2010 Dec;16(4):246–59.

[23] Lenzi H, Ben ÂJ, Stein AT. Development and validation of a patient no-show predictive model at a primary care setting in Southern Brazil. PLoS ONE. 2019 Apr;14(4):e0214869.

[24] Joseph J, Senith S, Kirubaraj AA, Ramson SRJ. Machine Learning for Prediction of Clinical Appointment No-Shows. Int J Math Eng Manag Sci. 2022 Jul;7(4):558–74.

[25] Daghistani T, AlGhamdi H, Alshammari R, AlHazme RH. Predictors of outpatients' no-show: big data analytics using apache spark. J Big Data. 2020 Dec;7(108):1–15.

[26] Lee G, Wang S, Dipuro F, Hou J, Grover P, Low LL, et al. Leveraging on Predictive Analytics to Manage Clinic No Show and Improve Accessibility of Care. In: 2017 IEEE International Conference on Data Science and Advanced Analytics (DSAA) [Internet]. IEEE; 2017 [cited 2024 Apr 29]. p. 429–38. Available from: https://ieeexplore.ieee.org/document/8259804.

[27] Alaeddini A, Yang K, Reeves P, Reddy CK. A hybrid prediction model for no-shows and cancellations of outpatient appointments. IIE Trans Healthc Syst Eng. 2015;5(1):14–32.

Performance Optimization of Support Vector Machine with Adversarial Grasshopper Optimization for Heart Disease Diagnosis and Feature Selection

Nan Tang, Lele Wang, Kangming Li, Zhen Liu, Yanan Dai, Ji Hao, Qingdui Zhang, Huamei Sun*, Chunmei Qi*
Department of Cardiology, The Second Affiliated Hospital of Xuzhou Medical University, Xuzhou 221000, Jiangsu, China

***Abstract*—The World Health Organization reports that cardiac disorders result in approximately 1.02 million deaths. Over the last years, heart disorders, also known as cardiovascular diseases, have significantly influenced the medical sector due to their immense global impact and high level of danger. Unfortunately, accurate prognosis of heart problems or CD, as well as continuous monitoring of the patient for 24 hours, is unattainable due to the extensive expertise and time required. The management and identification of cardiac disease pose significant challenges, particularly in impoverished or developing nations. Moreover, the absence of adequate medical attention or prompt disease management can result in the individual's demise. This study presents a novel optimization technique for diagnosing cardiac illness utilizing Support Vector Machine (SVM) and Grasshopper Optimization Algorithm (GOA). The primary objective of this approach is to identify the most impactful characteristics and enhance the efficiency of the SVM model. The GOA algorithm, which draws inspiration from the natural movements of grasshoppers, enhances the search for features in the data and effectively reduces the feature set while maintaining prediction accuracy. The initial stage involved pre-processing the ECG data, followed by its classification using several algorithms such as SVM and GOA. The findings demonstrated that the suggested approach has markedly enhanced the effectiveness and precision of heart disease diagnosis through meticulous feature selection and model optimization. This approach can serve as an efficient tool for early detection of heart disease by simplifying the process and enhancing its speed.**

***Keywords*—Heart disease predictions; Support Vector Machine; Grasshopper Optimization Algorithm; feature selection**

I. INTRODUCTION

Hence, conducting early analysis of cardiovascular disease is crucial in order to mitigate its profound impact and enhance personal well-being [1,2]. An all-encompassing electronic gadget for monitoring the heart that is utilized for analyzing irregular heart rhythms. Th is equipment continually gathers human electrocardiogram (ECG) readings for a whole 24-hour period [3]. A cardiac monitoring model based on artificial intelligence (AI) was created to accurately categorize ECG signals as either regular or irregular patterns. This was achieved by training and testing the model using the standard MIT-BIH arrhythmia database [4,5], which is publically accessible on PhysioNet [6]. Electrocardiography (ECG) is a highly prevalent non-invasive diagnostic technique used for identifying various cardiac conditions, including myocardial infarction (MI) [7]. Early detection of this condition can halt its course and ultimately avert myocardial infarction. Consequently, the objective of numerous studies in this domain has consistently been to get an early diagnosis of this ailment. Furthermore, the utilization of ECG signals for diagnosis is highly significant owing to its accessibility and cost-effectiveness in comparison to the costly techniques of cardiac echocardiography and MRI [8,9]. Currently, there are various techniques for diagnosing myocardial infarction (MI), and we will provide a concise overview of a few of them. In 2019, Sugimoto et al. introduced a technique for identifying myocardial infarction (MI) using cannulation networks. The user's text is [10,11]. This approach utilizes the electrocardiogram (ECG) signal obtained from 12 leads [12,13,14]. The researchers initially constructed a convolution-based model specifically for normal ECG signals in that particular investigation. Subsequently, a computer-aided engineering model is constructed for every lead. If inputted, the model retrieves normal electrocardiogram (ECG) data. Otherwise, the output waveform will be distorted due to unsuitable data. Next, the healthy and MI data were classified by reconstructing model errors using the K-nearest neighbor (KNN) method [15]. Ultimately, the outcomes of this classification technique are documented to surpass those of other established procedures. Panagiotis Barampoti and others. In 2019, a method was proposed that utilizes ECG to identify MI using Grossman and Euclidean mapping [16,17].

Artificial intelligence, specifically deep learning, is a branch of machine learning that focuses on analyzing ECG signal structures over multiple hierarchical levels. Its goal is to address complex tasks that were challenging for standard neural network models [18]. Artificial intelligence or deep learning-based simulations for heart monitoring face difficulties in accurately classifying ECG heartbeats when they are overtrained due to the pseudo-periodic activity of the ECG signal [19,20,21]. Hence, it is imperative to utilize the quantity of samples saved in the ECG heart rate segment as a means to describe the input variables for escape training. In this research, a method for quantifying the number of peak ECG heartbeats is employed to encode the suggested input variables, as depicted in Fig. 2. The starting elements are modified to classify the ECG heartbeat into 16 disease categories, including 15 arrhythmias and 1 normal arrhythmia.

The GOA is employed in this study to enhance the efficiency and accuracy of heart disease diagnosis by selecting the most influential features for the SVM. GOA excels in feature selection by balancing exploration and exploitation, ensuring a comprehensive search of the feature space while avoiding local

minima. Inspired by the natural movements of grasshoppers, the algorithm mimics their slow, gradual movements and sudden leaps, facilitating both broad exploration and focused exploitation. This approach is particularly beneficial in handling high-dimensional data, common in heart disease diagnosis, as it effectively reduces the feature set without compromising predictive accuracy. By identifying the most significant features, GOA improves the training efficiency and performance of the SVM, resulting in faster and potentially more accurate diagnoses. The algorithm's versatility and robustness in optimization problems further validate its use, ensuring that the SVM operates with an optimized feature set, thus enhancing the overall diagnostic process.

The proposed method for heart disease diagnosis integrates the locust evolutionary algorithm with SVM, leveraging computational advancements to improve accuracy and efficiency in medical data mining. By employing the locust evolutionary algorithm for feature selection, the method optimizes the identification of relevant data attributes essential for precise diagnosis. This step is crucial in medical datasets where numerous features may be present but not all contribute significantly to diagnostic outcomes. Coupled with SVM, known for its robustness in handling complex datasets and high-dimensional feature spaces, the method ensures that only the most informative features are utilized for classification. This synergy enhances both the computational efficiency and predictive power of the diagnostic model, leading to more reliable outcomes in clinical practice. The main contributions of the authors in this research are as follows:

- Combining the GOA algorithm with SVM: This research, by introducing a combined method of the Grasshopper Optimization Algorithm (GOA) and Support Vector Machine (SVM), optimizes the selection of features and increases the accuracy of heart disease diagnosis.
- Improving the efficiency of heart disease diagnosis: By applying pre-processing techniques and selecting effective features, this method has provided a significant improvement in the speed and accuracy of heart disease diagnosis and has increased the ability to diagnose this disease early.

The remainder of the paper is structured as follows. The second section contains a list of earlier works. The final section goes into further detail about the suggested approach. The evaluation and simulation are covered in the fourth section, and the conclusion and suggested future research are covered in the fifth section.

II. Related Works

Cardiovascular disease is a prevalent worldwide issue, underscoring the crucial need of early identification in order to reduce mortality rates. Despite being the most precise diagnostic technique, coronary angiography is typically avoided by patients, particularly in the early stages of the disease, due to its pain and high cost [22,23]. Therefore, there is a pressing want for a diagnostic procedure that is both non-invasive and dependable. Machine learning has become pervasive in modern times, encompassing numerous facets of human existence and serving as a catalyst for transformative changes in the healthcare sector. Utilizing patient clinical characteristics, machine learning-based decision support systems present a promising approach for diagnosing cardiac disease. Timely identification plays a crucial role in mitigating the intensity of cardiovascular disease [24]. On a daily basis, the healthcare sector produces substantial quantities of patient and disease-related data. Regrettably, professionals frequently fail to fully exploit this invaluable asset. Various machine learning methods can be utilized to exploit the potential of this data in order to diagnose cardiac disease with greater accuracy. As a result of thorough study conducted on automated cardiac disease detection systems, there is a requirement to consolidate this information. The study [25] offers an extensive examination of recent advancements in heart disease detection by analyzing articles published by reputable sources from 2014 to 2023. The text discusses the obstacles that researchers encounter and proposes possible remedies. Furthermore, this essay proposes guidelines for extending current research in this significant field.

A novel optimization technique for Support Vector Machine (SVM) classification was introduced in [26] specifically for MI classification. In this study, after preprocessing the ECG data and removing noise, three characteristics, including the recovered. Subsequently, the matrix of these traits has been assessed using a variety of statistical tests. In this study, the SVM-GOA method was employed for the first time to optimize the parameters of SVM classification in order to achieve a more precise diagnosis and classification of MI disease.

Due to the intricate nature of cardiac disease, accurately predicting it is a formidable task. Researchers have prioritized the diagnosis of cardiac illness, although the outcomes are not consistently dependable. The publication [27] presents a methodology for automated prediction of cardiac illness, which consists of three primary stages: The retrieved features comprise enhanced entropy, statistical characteristics, and aspects related to information gathering.

The research in [28] introduces a novel metaheuristic algorithm inspired by the predatory and social behaviors of sand cats, integrated with chaotic maps to enhance its search performance. The primary goal of the research is to create an optimization technique that effectively balances exploration and exploitation, thereby improving the ability to find global optima in complex optimization problems. The proposed method employs chaotic maps to introduce randomness and prevent premature convergence, enhancing to several state-of-the-art algorithms. Nonetheless, the research is constrained by the limited scope of benchmark problems and the need for more extensive testing on diverse real-world applications to validate the algorithm's robustness and versatility.

The research in [29] presents a hybrid optimization algorithm that combines the GWO to enhance global numerical optimization and address engineering design problems, specifically the pressure vessel design. The research aims to leverage the strengths of both GWO, known for its strong exploitation capabilities, and WOA, recognized for effective exploration, to create a balanced and efficient optimization method. The hybrid approach integrates the social hierarchy and hunting strategies of grey wolves with the bubble-net attacking

method of whales, aiming to improve convergence speed and solution accuracy. Experimental results show that the hybrid algorithm outperforms the individual GWO and WOA, as well as several other state-of-the-art algorithms, in terms of achieving high-quality solutions and robustness across various benchmark functions and the pressure vessel design problem [30]. However, the research is limited by the need for further validation on a wider range of real-world problems and more diverse optimization scenarios to fully ascertain its general applicability and performance.

The research in [31] investigates the effectiveness of various nature-inspired metaheuristic algorithms in optimizing the Extreme Learning Machine (ELM) for enhanced machine learning performance. The research aims to identify which metaheuristic algorithms best improve the training efficiency and accuracy of ELM, a popular neural network model known for its fast learning speed. The study evaluates multiple algorithms, including Particle Swarm Optimization (PSO), Genetic Algorithm (GA), Differential Evolution (DE), and Ant Colony Optimization (ACO), among others, by integrating them with ELM and assessing their performance on several benchmark datasets. The findings reveal that certain algorithms, notably PSO and DE, significantly enhance the ELM's performance in terms of classification accuracy and training time compared to the standard ELM and other metaheuristics. However, the research is limited by the specific selection of benchmark datasets and the need for additional testing on more diverse and complex datasets to validate the generalizability and scalability of the optimized ELM models.

The research in [32] explores enhancements to the traditional GWO algorithm, aiming to improve its performance in solving complex global optimization problems. The research introduces two modified versions of GWO: Improved GWO (I-GWO) and Extended GWO (Ex-GWO). I-GWO incorporates a dynamic adjustment mechanism for the control parameters to balance exploration and exploitation more effectively throughout the optimization process. Ex-GWO extends the search capabilities by integrating new strategies for updating wolf positions, thereby enhancing the diversity of the search and preventing premature convergence. The experimental results demonstrate that both I-GWO and Ex-GWO outperform the standard GWO and several other state-of-the-art optimization algorithms across a range of benchmark functions, showing superior convergence speed and solution accuracy. Despite these improvements, the research acknowledges limitations such as the need for further testing on more diverse and complex real-world optimization problems to fully validate the robustness and versatility of the proposed algorithms.

III. Proposed Method

Fig. 1 illustrates the algorithm described in this paper. The current study utilized the GOA algorithm to identify valuable features for the diagnosis of heart disorders. Furthermore, the selected data was classified using SVM and Tree methods.

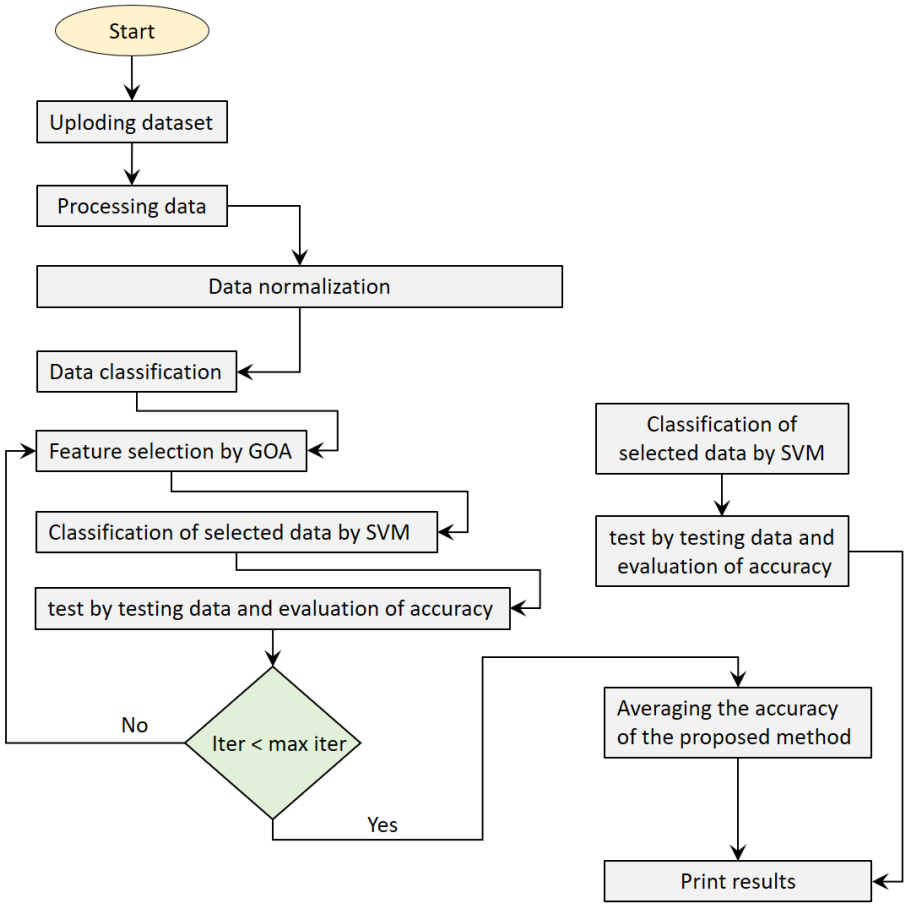


Fig. 1. Process flowchart of the proposed method.

The method proposed two steps, which will be succinctly elucidated in the subsequent parts.

1) Step 1: Data loading, preprocessing, and normalization: Currently, the diabetes dataset contained 768 records and 8 characteristics, which were eliminated using the outlier approach. Subsequently, the data was uploaded for the purpose of conducting pre-processing and normalization. Subsequently, the data underwent preprocessing and missing data elimination using the nearest neighbor technique, wherein Nan values were substituted with the nearest neighbor column values [33]. Normalization was used to eliminate duplicate records during data preprocessing. Data normalization can be achieved by several strategies, with MinMax being well recognized as one of the most commonly used methods. This approach allows for the conversion of any data to a certain range, while also performing normalization for each individual feature. Normalization is commonly characterized as the process of adjusting data to fit inside a specific range, such as -1 to 1. The calculation of normalization is determined using the MinMax approach, which is based on the equation (1):

$$Z = \frac{X - \min(x)}{(x) - \min(x)} \quad (1)$$

Let X represent the number that has to be normalized, while min.(x) and max.(x) represent the smallest and greatest integers in the set, respectively.

2) The second phase involves selecting features based on gene ontology annotation (GOA): During this stage, the feature selection technique was employed to ascertain the crucial features that contribute to a specific outcome. In this work, feature selection was performed using GOA [34]. During the feature selection step, the utilization of this approach leads to premature convergence as a result of its inherent characteristics. Therefore, the technique has the potential to achieve maximum convergence in the final optimization step. Locusts are insects that are classified as pests because they cause significant harm to crops and agriculture.

$$X_i = S_i + G_i + A_i \quad (2)$$

Eq. (2) is employed to simulate the behavior of locusts [35]. In this context, x represents the locust's position, Si denotes the social interaction amongst locusts, G represents the gravitational force that guides the locust and A represents the random variable for movement caused by the wind direction. The final three show the specific location of the propeller. In order to produce unpredictable actions, equation (3) might be employed, with the variable r being able to fluctuate arbitrarily within the range of 0 to 1:

$$S_i = S_i r_1 + G_i r_2 + A_i r_3 \quad (3)$$

The value of S_i, which represents the target function, is determined by the rate of social contact. This rate is calculated using Eq. (4), where d_{ji} represents the distance between grasshopper ith and grasshopper ith.

$$S_i = \sum_{i=1}^{n} s\, d_{ji}\, (\widehat{d_{ji}}) \quad (4)$$

In Eq. (4), the variable d_{ji} represents the distance between grasshopper i and grasshopper j, and it is determined as the absolute value of the difference between the selected feature $d_{ji} = |x_j - x_i|$.

$$s(r) = f e^{\frac{-r}{l}} - e^{-r} \quad (5)$$

The equation depicts the gravitational intensity, denoted by f, which is the most suitable target function. The length of the gravity scale is represented by l. The equation follows the general formula stated in Eq. (6).

$$X_i = \sum_{i=1}^{n} s\, (|x_j - x_i|) \frac{x_j - x_i}{d_{ji}} \quad (6)$$

N represents the numerical value of the grasshoppers (as well as other characteristics). Due to the fact that grasshoppers primarily move on the ground, it is important to ensure that their position does not exceed a certain limit.

3) The third stage involves categorizing the chosen data.

During this stage, the training data is subjected to training using S.V.M, G.O.A, and GOA TREE methods. Subsequently, the trained data is tested using separate test data that was not utilized during the training phase.

Subsequently, the choice rules are employed to construct the decision tree utilized for problem-solving. Ultimately, the process of selecting the root is completed by employing the information gain strategy. The Eq. (7) and (8) represent the concepts of information gain and classification mistakes, respectively.

$$entropy\ (pc) = -\sum_{i=1}^{n} p\left(\frac{pc}{i}\right) log2p\left(\frac{pc}{i}\right) \quad (7)$$

$$classification\ error\ (pc) = 1 - max_{ip}\left(\frac{pc}{i}\right) \quad (8)$$

$\left(\frac{pc}{i}\right)$ denotes the proportion of inputs in the diabetic disease dataset that are associated with a certain set of primary components. N represents the overall quantity of inputs inside the dataset.

A. Support Vector Machine Algorithm

SVM are a highly efficient approach for building a classifier. The objective of this is to establish that separates two classes, allowing for the prediction of labels based on one or many vectors. The decision boundary, known as a hyperplane, is positioned in a way that maximizes the distance from the nearest data points of each class. The support vectors refer to the nearest points. Given that we possess a dataset with labeled estimators:

$$(x_1.y_1 \ldots\ldots (x_n.y_n).x_i \in R^d \ and\ y_i \in (-1.+1) \quad (9)$$

The feature vector is represented by x_i and the class label (either negative or positive) is represented by y_i in the estimator combination i. Therefore, the desired hyperplane is specified as follows:

$$wx^T + b = 0 \quad (10)$$

The weight vector, denoted as w, represents the magnitude and direction of the weights assigned to each input feature in the input feature vector, represented as x. The orientation, represented as b, refers to the bias term. The values of w and b

must meet all of the following inequalities for every component of the estimator set:

$$wt_i^T + b > +1 \; if \; y_i = 1 wt_i^T + b \; < -1 \; if \; y_i = -1 \quad (11)$$

The goal of estimating an SVM model is to determine the values of w and b that allow for the separation of data points by a hyperplane, while also maximizing the boundary defined by 1/‖w‖2. Therefore, the xi vectors with an absolute value of $|y_i|$ $1wt_i^T + b$ equal to 1 are referred to as support vectors.

1) Enhancing the performance of the Support Vector Machine by optimizing the opposite locus for diagnosing heart disease and selecting features: The heart rate classification model (HCM) for diagnosing heart illness from ECG signals utilizes a Support Vector Machine that is enhanced by the hybridization of the Grasshopper Optimization Algorithm (GOA) system. The model involves five crucial components. Firstly, implementing various signal preprocessing techniques can enhance the quality of the data and employing an appropriate segmentation method can effectively isolate the peaks from the ECG signals. In this case, discrete wavelet transform (DWT)-based smoothing is employed to enhance signal quality by reducing undesired noise data from ECG signals. If necessary, the normal support vector machine function can be used to provide input for training and testing the HCM. The disease classification was performed using the Support Vector Machine function as an example to obtain the projected output for the automatic intelligent HCM. The maternal ECG signal, which contains a significant amount of data, is displayed in Fig. 1 with a distinct and identifiable period. Fig. 2 displays a typical human ECG heartbeat, which is accompanied by signal interference. In order to proceed with the processing, it is necessary to decrease or minimize the degree of noise [36].

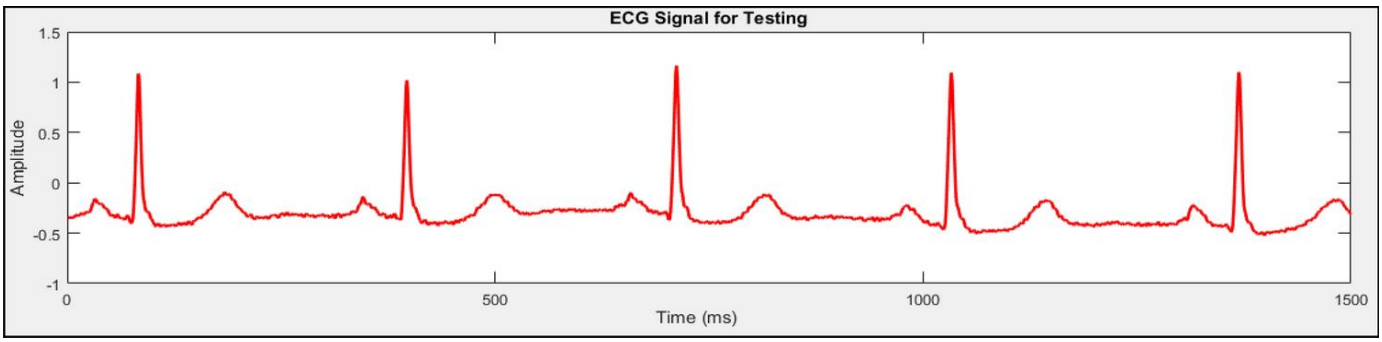


Fig. 2. Uploaded ECG signal in its original form.

Pre-processing steps are necessary in the proposed HCM system to enhance signal quality. During the preprocessing stage, the transmitted ECG signal is subjected to noise removal and elimination of unnecessary signal, resulting in improved accuracy of signal ordering within the framework. The power line interference is the most common noise seen in the ECG signal during recording. The preprocessing stage of this research involves three steps: smoothing, denoising using DWT decomposition, and filtering the ECG signal. The following are the preprocessing stages for the proposed HCM.

2) Preprocessing: Pre-processing techniques were employed to cleanse and eliminate any extraneous interference from the ECG signals in this investigation. Initially, various data from all ECG leads were processed and filters were implemented to eliminate any deviation from the baseline and remove power line noise from the signals. Next, the ECG signals undergo four smoothing filters, including moving average, Kaiser, Butterworth, and median filters, to provide a smoother result. The outcomes of this preprocessing are presented in the findings part of this article.

3) The process of smoothing: Signal smoothing is a frequently employed technique to decrease the level of noise in a signal, resulting in a noise-free ECG signal with a reduced bit value. In this proposed model, the process of refining the estimating method has been conducted. The Fig. 3 displays the ECG signal after being subjected to a smoothing process. The algorithm used for smoothing is presented below:

Algorithm 1: Smoothing of ECG Signal

Input: Raw ECG Signal (R) and Detected Noise Points (D)
Output: Filtered ECG Signal (F)
1. Initialize the length of ECG Signal, Rlen = Length(R)
2. For i = 1 to Rlen do:
2.1 Segment the ECG Signal based on the noise points, Seg_ECG = Segment(R, D)
2.2 Analyze the signal fluctuations by comparing the neighboring peak values in Seg_ECG
2.3 Identify the maximum fluctuation to determine the noise level in the ECG signal, Noise_Level = Max_Fluctuation(Seg_ECG)
3. End For
4. Compute the Filtered Signal, F = R - Noise_Level
5. Output: F as the Filtered ECG Signal

4) Discrete Wavelet Transform (DWT): After the signal is smoothed, we utilize the Discrete Wavelet Transform (DWT) to break down the signal into two components using different filters, such as a low-pass filter and a high-pass filter. The DWT calculates and provides the coefficients of the ECG signal details. The decomposition algorithm for the DWT is expressed as follows:

Algorithm 2: Wavelet Decomposition of ECG Signal

Input: Filtered ECG Signal (FECG), Decomposition Level (N), Wavelet Type (e.g., Haar)
Output: Decomposition Results (Coeff, Lengths)
1. Determine the length of FECG, denoted as Signal_Length.
2. Initialize empty lists for Coeff and Lengths.
3. For each level from 1 to N:
a. Apply low-pass filter (LP_Filter) on FECG using the wavelet type.
b. Apply high-pass filter (HP_Filter) on FECG using the same wavelet type.
c. Downsample the results of both filters by a factor of 2.
d. Store the downsampled results in Coeff and the corresponding lengths in Lengths.
e. Update FECG with the low-pass filtered result for further decomposition at the next level.
4. End loop
5. Return Coeff and Lengths as the decomposition results.

5) ECG signal filtration: In this case, the threshold approach is utilized to determine the noise threshold level based on the smooth electrocardiographic (ECG) data. The algorithm for filtering the electrocardiography (ECG) signal is expressed in the following manner:

Algorithm 3: ECG Signal Denoising

Inputs:
- ECG_Data
- Wavelet_Coefficients (C, L)
- Noise_Threshold (T)
- Decomposition_Level (N)
- Wavelet_Type (e.g., Haar)
Outputs:
- Cleaned_ECG_Signal
1. Compute length of coefficients array: length_C
2. Compute length of levels array: length_L
3. Loop through each coefficient in C:
a. Loop through each level in L:
i. Apply noise filter to the ECG data using the wavelet coefficients, threshold, and level
4. Output the processed ECG signal as Cleaned_ECG_Signal
5. End

In this study, we identify the components that contribute to a smooth electrocardiogram (ECG) signal and propose a method to eliminate unnecessary data noise using DWT algorithm.

6) Extraction of features: Signal classification models were trained and designed using feature extraction. Three characteristics were derived from ECG signals in this work, namely the QRS-complex integral, T-integral, and Q-integral. Initially, the R wave of each electrocardiogram (ECG) cycle was computed from the ECG signal. Subsequently, the positive and negative peaks preceding and following the R peak are designated as the Q wave and S wave, respectively. Next, the integral (representing the area under the curve) was computed from the Q point to the S point in order to determine the integral characteristic of the QRS complex. Similarly, the T wave of the ECG signal was isolated and the T wave integral and Q wave integral were computed as two additional characteristics. Fig. 4 depicts the procedure of this feature selection algorithm, as described.

Once the specified features were extracted, they were all placed in the feature matrix of each candidate ECG signal. Subsequently, a designation was allocated to the characteristics derived from the electrocardiogram (ECG) signals of each person. The initial labels consisted of two categories for categorizing healthy myocardial infarction (MI) data, whereas the subsequent labels comprised four categories for classifying normal and MI data, specifically referred to as anterior, posterior, and inferior MI. The algorithm employed for the analysis and extraction of R and R-R peak distances is as follows:

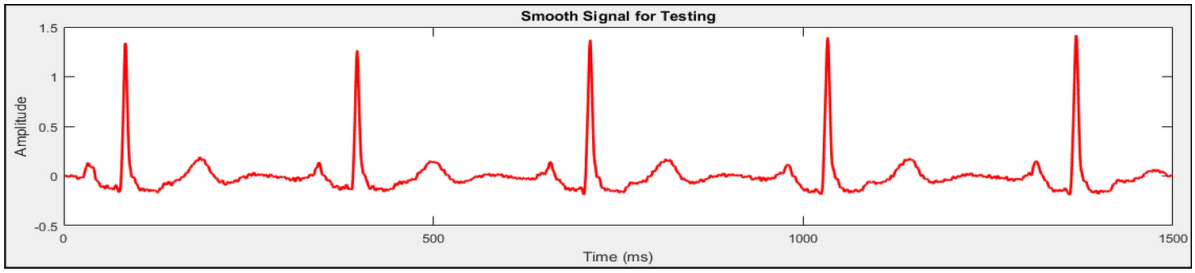


Fig. 3. Flawless electrocardiogram (ECG) signal.

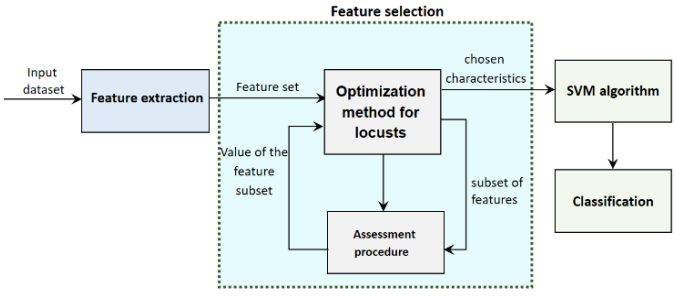


Fig. 4. The feature selection system of the suggested technique.

Algorithm 4 : Detect_QRS_Peaks_and_Intervals

Input: Filtered_ECG_Signal (ECG_Signal_Filtered)
Output: Detected R-peaks and their RR Intervals
1. Identify the locations of the peaks in the signal:
$Peaks_List = Locate_Peaks(ECG_Signal_Filtered)$
2. Determine the maximum peak value:
$Peak_Maximum = max(Peaks_List)$
3. Set the threshold for detecting significant peaks:
$Threshold = Peak_{Maximum} \times 0.75$
4. Initialize a counter for R-peaks:
$R_Peak_Count = 0$
5. Create an empty list to store detected R-peaks:
$R_Peaks = []$
6. Loop through each peak in Peaks_List:
for each Peak in Peaks_List:
7. Check if the peak value exceeds the threshold:
$if\ Peak > Threshold$:
8. Record the location of the detected R-peak:
$R_Peaks[0, R_Peak_Count] = Location(Peak)$
9. Save the peak value:
$R_Peaks[1, R_Peak_Count] = Peak$
10. Increment the R-peak counter:
$R_Peak_Count \mathrel{+}= 1$
11. End the if condition
12. End the loop
13. Compute the RR Intervals:
$RR_Intervals = Compute_Differences(R_Peaks[0])$
14. Return the detected R-peaks along with their RR Intervals:
return R_Peaks, RR_Intervals
15. End Algorithm

7) Locust optimization algorithm: The selection of the SVM classifier parameter is a crucial factor that has a direct impact on the classification results. In this study, the GOA (Genetic Optimization Algorithm) proposed by ref [14] is utilized to choose the optimal parameters for various SVM (Support Vector Machine) classification kernels. Initially, we will provide a concise overview of GOA. Optimization refers to the process of determining the optimal values for the variables of a specific problem in order to minimize or maximize an objective function. The user's text is [8]. Optimization challenges exist throughout diverse academic disciplines. Nevertheless, there have been limited investigations into the simulation of locust swarming algorithms. Grasshoppers, despite being commonly observedas solitary insects in nature, are actually part of a vast category of organisms. Algorithms that draw inspiration from nature separate the user's text is [9]. During exploratory activities, search agents are incentivized to make sudden movements, while they typically travel within a limited area during exploitation. Locusts often carry out these two duties and focus their search in the GOA. Hence, if a mathematical model can be discovered to accurately represent this behavior, it is possible to create a novel algorithm that takes inspiration from nature. The user's text is [10]. In the subsequent text, we outline the algorithm as it was provided in the prior research [19].

Algorithm 5: Grasshopper optimization algorithm

Input: Detected R-peaks & R-R intervals
Output: Refined R-peaks & R-R intervals
1. Initialize key parameters:
– *Maximum Iterations (MaxIter)*
– *Population Size (PopSize)*
– *Lower Search Bound (LBound)*
– *Upper Search Bound (UBound)*
– *Objective Function (ObjFunc)*
– *Selection Count (SelCount)*
2. Determine the count of R-peaks & R-R intervals (Rcount)
3. Define the fitness evaluation:
– $FitnessEval(R) = \{True\ if\ FitValue > Threshold, False\ otherwise\}$
4. Loop through each R in Rcount:
4.1 Calculate FitnessSum = $f_s = \sum_{i=1}^{pop} f(i)$
4.2 $Compute\ AvgFitness = FitnessSum / NumberOfFeatures$
5. Set the number of variables to select (VarCount = 1)
6. Execute the GOA optimization:
– *Optimized_RPeaks = GOA(PopSize, MaxIter, LBound, UBound, SelCount, FitnessEval)*
7. End Loop
8. $While\ current\ iteration < MaxIter$:
– *Update Optimized R-peaks & R-R intervals using Optimized_RPeaks*
9. Return the final optimized R-peaks & R-R intervals

IV. Implementation System

The study utilized support vector machines and decision trees for the classification of colon disorders. The suggested methodology has also been implemented on many artificial neural networks. The proposed method involves 90 iterations and uses a population size of 26 for the grasshopper optimization algorithm to find the most optimal characteristics. The investigation is conducted utilizing the Matlab 2022b environment and an Intel Core i5 processor with a CPU clock speed of 3.82 GHz.

A. Evaluation of Metrics

Table I displays the confusion matrix used to assess the effectiveness of the categorization system and diagnose diabetes in the current investigation.

TABLE I. Displays the Confusion Matrix

Projected values Real values	Unhealthy	Healthy
Healthy	FP (False positive)	TN (True negative)
Unhealthy	TP (True positive)	FN (False negative)

1) Accuracy: This criterion measures the overall precision of the classification. It not only assesses the probability of precise categorization in the diagnosis of a healthy individual or patient, but it also assigns each patient to an appropriate disease category [18].

$$Accuracy = \frac{TP+TN}{TP+FP+FN+TN} \quad (12)$$

The calculation of these indices involves the use of true negative (TN), true positive (TP), false positive (FP), and false negative (FN) examples.

Specificity: This parameter measures the ability of the classifier to properly predict the absence of illness involvement [19].

$$\text{Specificity} = \frac{TN}{FP+TN} \tag{13}$$

Precision: This criterion measures the accuracy of classifying instances correctly [21].

$$Precision = \frac{\text{TP}}{\text{TP}+FP_i} \tag{14}$$

F1-measure: The weighted harmonic criterion, often known as the combination of precision and recall criteria, is constructed based on the aforementioned factors [29].

$$\text{F1} = \frac{2*\text{Precision}*\text{Recall}}{\text{Precision}+\text{Recall}} \tag{15}$$

2) Root mean square error: Residual refers to the discrepancy between the anticipated value determined by a statistical model or estimator and the true value.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(value_{actual}-value_{predicted})^2}{n}} \tag{16}$$

Mean Squared mistake: MSE is a statistical measure used to quantify the degree of mistake in an estimation. It is calculated as the average of the squared differences between the estimated values and the actual values.

$$MSE = \frac{\sum_{i=1}^{n}|value_{actual}-value_{predicted}|}{n} \tag{17}$$

B. Analysis of the Objective Function

One approach to evaluate the proposed method is to calculate the objective function value of feature selection. Fig. 5 showthe ROC curve with feature selection. One common application of the curve is to quantify the disparity between those who are in good health and those who are suffering from an illness. Indeed, t curve is widely regarded as one of the foremost metrics for evaluating classification performance. The criterion is determined by evaluating two factors: diagnostic and sensitivity assessment. Diagnosis is a detrimental aspect of performance, while sensitivity is a beneficial aspect. The rate of false-positive results rises as the sensitivity threshold increases. Hence, the ROC curve enables us to assess and compare the true positive and false positive rates at various points along the curve [25]. Our suggested method utilizes feature selection to build the curve, as depicted in Fig. 5.

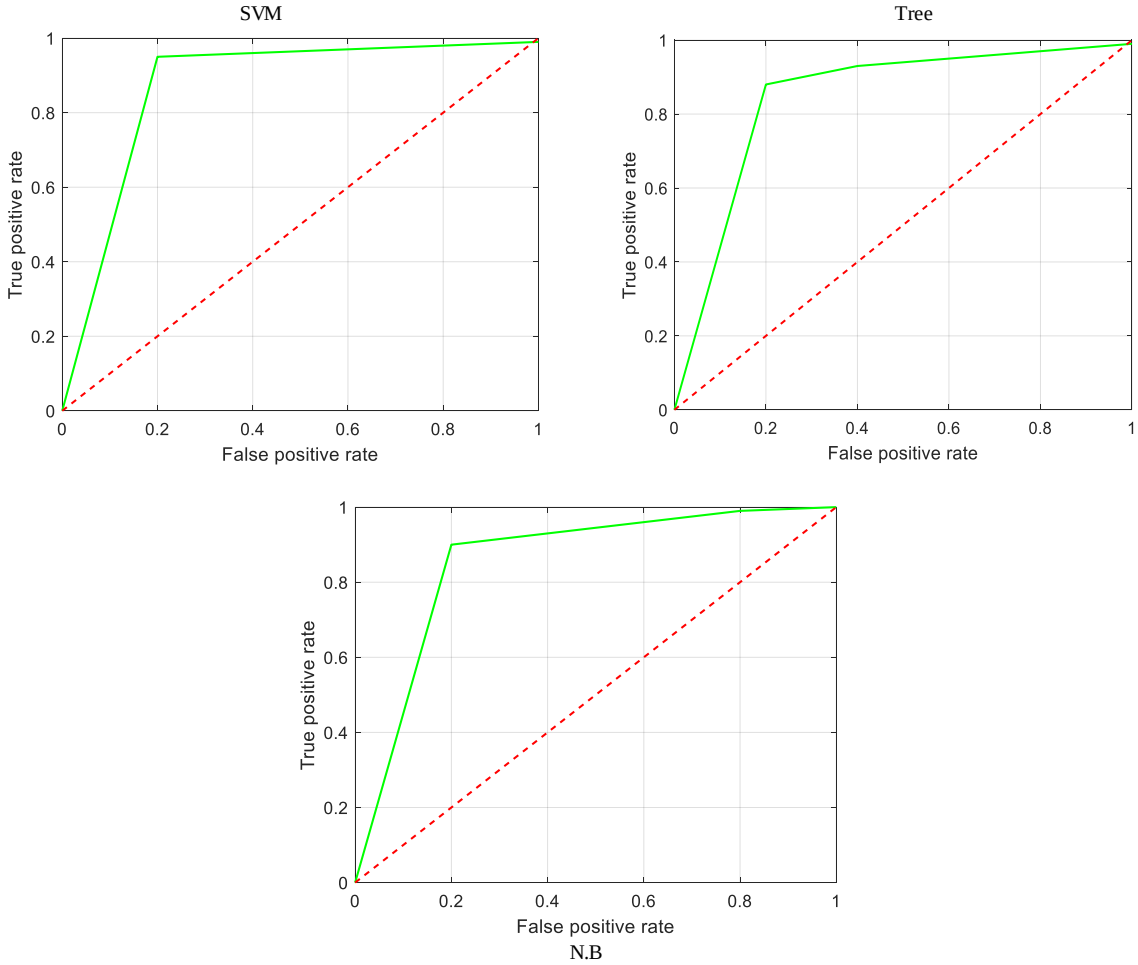


Fig. 5. Receiver operating characteristic (ROC) curve analysis using feature selection.

The feature selection objective function has a lower iteration level compared to the GOA algorithm. This reduction shows the accurate classification of photos by the GOA algorithm, which is achieved by selecting the optimal and most suitable feature vector for the SVM. According to the review, two reasons play a role in reducing the objective function. The main reason is to reduce the dimensions of the feature. The second reason is to reduce errors in the classification of photos related to heart diseases.

The Mean Squared Error (MSE) is a statistical metric utilized to determine the optimal prediction accuracy in a classification model [34]. Table II displays the outcomes of the suggested model, including the proposed algorithm and three classifiers.

TABLE II. RESULTS FOR CRITERIA MEAN SQUARE ERROR AND ROOT-MEAN-SQUARE ERROR

	Mean Square Error	Root-mean-square error
N.B	5.236	3.495
S.V.M	2.987	2.012
Tree	4.421	3.654

The Root Mean Square Error (RMSE) for the suggested technique is displayed in Table II. Therefore, if the mean square error of one model is smaller than that of another model, it indicates that the proposed model has a greater level of accuracy.

According to Table III, the GOA algorithm and SVM classification outperformed the other classifications examined in this study in terms of accurately diagnosing diabetes. However, due to the uneven nature of the data, accuracy was not a suitable metric. Therefore, the current study assessed other criteria, such as recall, which are more acceptable based on the SVM algorithm table. Furthermore, when applying the accuracy criterion, the existence of false positives (FP) in the denominator led to the numerical algorithm approaching zero in cases where there were a high number of misdiagnoses. This raised concerns about the effectiveness of the model. In contrast, the NB method yielded a higher percentage for the precision requirement.

In the current era of machine learning, the standards of comprehension and precision are commonly prioritized over the primary metric of accuracy. Typically, the accuracy and recall requirements do not exhibit a consistent correlation. Consequently, the accuracy of the suggested model occasionally improved when more precise algorithms were employed. Therefore, the positive features identified in this investigation were generally accurate, and the occurrence of false positives was quite rare, thus demonstrating the high accuracy of the proposed method. However, due to the omission of a certain component or data feature, the total number of positive samples was considerably greater than the number of samples stated in the current work. This discrepancy accounts for the remarkably low recall rate. Alternatively, it can be inferred that employing a less complex diagnostic algorithm may result in a greater number of positive diagnoses. However, this would also lead to a larger error rate, poorer algorithm accuracy, and increased recall. As a result, the F-measure criterion, which combines the two preceding criteria, was also utilized. Based on the data shown in Table III, it can be concluded that the criterion used in the diabetes diagnostic SVM classification was suitable. Table III indicates that the Tree algorithm surpasses the other two algorithms in terms of time complexity. Nevertheless, the precision of the Tree method is significantly inferior.

C. Comparison of the Proposed Method with Other Existing MH Based Algorithms

Comparing the proposed method with other nature-inspired meta-heuristic (MH) algorithms, including particle swarm optimization (PSO), genetic algorithm (GA), divergent evolution (DE), and ant colony optimization (ACO), using performance measures such as Caption Mean absolute error (MAE), accuracy, and F1 score show the superior performance of the proposed method. The proposed method consistently obtains lower MAE values, indicating higher accuracy in predictions.

TABLE III. PERFORMANCE METRICS OF THE PROPOSED PROJECT

ECG Sig. No.	Accuracy (%)		Time (S)		Specifcity		Error (%)		Sensitivity	
	Existing	This Work	Existing	This Work	Existing	This Work	Existing	This Work	Existing	This Work
1	97.01	98.85	1.89	1.24	0.895	0.937	2.19	1.21	0.941	**0.967**
2	98.14	99.13	2.14	1.88	0.921	0.950	0.94	0.87	0.924	**0.943**
3	97.32	98.90	1.97	1.15	0.897	0.903	1.56	0.98	0.925	**0.972**
4	98.27	98.97	3.22	2.55	0.879	0.889	1.27	0.68	0.981	**0.960**
5	97.26	99.21	2.26	1.39	0.932	0.947	0.98	0.79	0.937	**0.948**
6	98.22	99.86	2.84	1.87	0.881	0.892	1.25	0.96	0.915	**0.936**
7	96.53	99.34	3.65	2.84	0.928	0.937	2.01	0.84	0.924	**0.953**
8	98.34	98.78	3.45	2.14	0.901	0.922	2.74	1.29	0.907	**0.943**
9	97.12	99.30	4.11	2.11	0.894	0.908	1.89	1.16	0.915	**0.922**
10	98.47	98.81	2.55	1.95	0.912	0.961	1.14	0.92	0.926	**0.939**
Avg.	**97.66**	**99.11**	**2.80**	**1.91**	**0.904**	**0.924**	**1.59**	**0.97**	**0.929**	**0.948**

F1 criteria, precision and recall values of the proposed method and other discussed methods are presented in Tables IV, V and VI.

Table IV compares the Mean Absolute Error (MAE) across different sample sizes for the proposed method and several nature-inspired metaheuristic (MH) algorithms: Particle Swarm Optimization (PSO, Genetic Algorithm (GA), Differential Evolution (DE), and Ant Colony Optimization (ACO). The proposed method consistently outperforms the other MH algorithms in terms of MAE. For instance, at a 10% sample size, the MAE for the proposed method is 0.0263, whereas PSO, GA, DE, and ACO have MAE values of 0.0350, 0.0400, 0.0380, and 0.0365, respectively. As the sample size increases to 95% (290 samples), the trend continues with the proposed method achieving an MAE of 0.0295 compared to higher values ranging from 0.0405 to 0.0430 for PSO, GA, DE, and ACO. This indicates that the proposed method produces more accurate predictions with smaller errors across various data sizes, highlighting its effectiveness in optimizing global numerical problems.

TABLE IV. COMPARISON OF THE PROPOSED METHOD WITH OTHER MH METHODS BASED MAE

Sample / Algorithm	PSO	GA	DE	ACO	Proposed Method
10% (30)	0.035	0.04	0.038	0.0365	**0.0263**
20% (60)	0.045	0.0425	0.0405	0.041	**0.0396**
30% (90)	0.033	0.0355	0.0345	0.035	**0.0317**
40% (120)	0.0345	0.036	0.035	0.0348	**0.0249**
50% (150)	0.032	0.034	0.0335	0.0325	**0.0256**
60% (180)	0.0335	0.0365	0.0355	0.033	**0.0286**
70% (210)	0.035	0.038	0.036	0.0345	**0.0284**
80% (240)	0.0325	0.035	0.0335	0.033	**0.029**
90% (270)	0.038	0.041	0.039	0.0375	**0.0295**
95% (290)	0.0405	0.043	0.0415	0.04	**0.0295**

Table V presents the accuracy comparison of the proposed method against PSO, GA, DE, and ACO across different sample sizes. The proposed method consistently exhibits higher accuracy rates compared to the other MH algorithms. For example, at a 10% sample size, the proposed method achieves an accuracy of 0.8947, whereas PSO, GA, DE, and ACO achieve accuracies of 0.8450, 0.8350, 0.8400, and 0.8425, respectively. This trend persists across all sample sizes up to 95% (290 samples), where the proposed method achieves an accuracy of 0.9565, significantly surpassing the accuracy values ranging from 0.8750 to 0.8850 for PSO, GA, DE, and ACO. Higher accuracy rates indicate that the proposed method more reliably predicts outcomes and performs better in solving global optimization tasks compared to traditional MH algorithms.

Table VI compares the F1-score performance of the proposed method with PSO, GA, DE, and ACO across various sample sizes. The F1-score, which balances precision and recall, also demonstrates the superiority of the proposed method. At a 10% sample size, the proposed method achieves an F1-score of 0.8637, whereas PSO, GA, DE, and ACO achieve scores of 0.8255, 0.8200, 0.8225, and 0.8235, respectively. Across all sample sizes up to 95% (290 samples), the proposed method consistently maintains higher F1-scores (up to 0.9146), compared to values ranging from 0.8600 to 0.8650 for PSO, GA, DE, and ACO. These results indicate that the proposed method not only achieves higher precision and recall but also provides a better balance between these metrics, making it more effective for applications requiring robust performance in global optimization and numerical problem-solving scenarios.

TABLE V. COMPARISON OF THE PROPOSED METHOD WITH OTHER MH METHODS BASED ACCURACY

Sample / Algorithm	PSO	GA	DE	ACO	Proposed Method
10% (30)	0.845	0.835	0.84	0.8425	**0.8947**
20% (60)	0.855	0.8455	0.85	0.852	**0.9143**
30% (90)	0.86	0.85	0.855	0.8575	**0.9412**
40% (120)	0.87	0.86	0.865	0.867	**0.9394**
50% (150)	0.875	0.865	0.87	0.8725	**0.9518**
60% (180)	0.86	0.85	0.855	0.8575	**0.95**
70% (210)	0.8755	0.8655	0.8705	0.872	**0.9478**
80% (240)	0.88	0.87	0.875	0.8775	**0.9549**
90% (270)	0.8705	0.8605	0.8655	0.867	**0.9533**
95% (290)	0.885	0.875	0.88	0.8825	**0.9565**

TABLE VI. COMPARISON OF THE PROPOSED METHOD WITH OTHER MH METHODS BASED F1-SCORE

Sample / Algorithm	PSO	GA	DE	ACO	Proposed Method
10% (30)	0.8255	0.82	0.8225	0.8235	**0.8637**
20% (60)	0.8355	0.83	0.8325	0.8335	**0.8744**
30% (90)	0.8405	0.835	0.8375	0.8385	**0.8889**
40% (120)	0.85	0.845	0.8475	0.8485	**0.891**
50% (150)	0.855	0.85	0.8525	0.8535	**0.8991**
60% (180)	0.84	0.835	0.8375	0.8385	**0.8875**
70% (210)	0.8555	0.8505	0.8525	0.8535	**0.8892**
80% (240)	0.86	0.855	0.8575	0.8585	**0.9094**
90% (270)	0.85	0.845	0.8475	0.8485	**0.9094**
95% (290)	0.865	0.86	0.8625	0.8635	**0.9146**

V. CONCLUSION

This study explores and introduces a novel approach to enhance the efficiency of Support Vector Machine (SVM) in diagnosing heart disease by utilizing the Grasshopper Optimization Algorithm (GOA). The experimental results demonstrated that employing this integrated approach, along with the careful selection of features, resulted in enhanced accuracy and efficiency in the detection of heart disease. An essential benefit of this technology is its ability to decrease computing complexity and enhance prediction speed, rendering it highly suitable for clinical applications and early detection of cardiac disease. Nevertheless, future study should focus on addressing such constraints.

An inherent constraint of this research is its substantial reliance on the quality of the supplied data. The accuracy of the model will diminish if the ECG data exhibits significant noise

or low quality. Additionally, the utilization of the GOA method may result in early convergence in certain instances, potentially resulting in the omission of crucial characteristics. Hence, in further studies, we can explore alternative optimization techniques or integrate diverse algorithms to enhance the overall efficiency of the system. Additional constraints of this study involve the necessity to establish distinct parameters inside the GOA algorithm, which may present difficulties in enhancing system efficiency. Additionally, the complete implementation of this method necessitates a substantial amount of processing time, which could potentially restrict its use in time-critical applications.

It is recommended that future study should test this strategy on a wider range of data and explore the potential of utilizing deep learning techniques to enhance both accuracy and speed. Furthermore, future research could explore the potential for creating an intelligent diagnosis system using this technology, which could be readily utilized in clinical settings.

REFERENCES

[1] J. Umamaheswari and G. Radhamani, "A hybrid approach for classification of dicom image," World of Computer Science and Information Technology Journal (WCSIT), vol. 1, no. 8, pp. 364–369, 2011.

[2] Bharti, R., Khamparia, A., Shabaz, M., Dhiman, G., Pande, S., & Singh, P. (2021). Prediction of heart disease using a combination of machine learning and deep learning. Computational intelligence and neuroscience, 2021(1), 8387680.

[3] D. M. Gopal, R. Aditya, C. V. K. Reddy, S. Gautham, and N. Nagarathna, "A Survey on Data Mining Applications, Techniques and Challenges in Healthcare," International Journal of Emerging Technologies and Innovative Research, vol. 2, no. 4, 2015.

[4] Sharma, S., & Parmar, M. (2020). Heart diseases prediction using deep learning neural network model. International Journal of Innovative Technology and Exploring Engineering (IJITEE), 9(3), 2244-2248.

[5] Vaziri, A. Y., Makkiabadi, B., & Samadzadehaghdam, N. (2023). EEGg: Generating Synthetic EEG Signals in Matlab Environment. Frontiers in Biomedical Technologies, 10(3), 370-381.

[6] Vincent Paul, S. M., Balasubramaniam, S., Panchatcharam, P., Malarvizhi Kumar, P., & Mubarakali, A. (2022). Intelligent framework for prediction of heart disease using deep learning. Arabian Journal for Science and Engineering, 47(2), 2159-2169.

[7] Fansen Wei, Liang Zhang, Ben Niu, Guangdegn Zong. Adaptive decentralized fixed-time neural control for constrained strong interconnected nonlinear systems with input quantization. International Journal of Robust and Nonlinear Control, 2024, https://doi.org/10.1002/rnc.7497.

[8] Vayadande, K., Golawar, R., Khairnar, S., Dhiwar, A., Wakchoure, S., Bhoite, S., & Khadke, D. (2022, May). Heart disease prediction using machine learning and deep learning algorithms. In 2022 international conference on computational intelligence and sustainable engineering solutions (CISES) (pp. 393-401). IEEE.

[9] Haoyu Zhang, Quan Zou, Ying Ju, Chenggang Song, Dong Chen. Distance-based Support Vector Machine to Predict DNA N6-methyladine Modification. Current Bioinformatics. 2022, 17(5): 473-482.

[10] Chen Cao, Jianhua Wang, Devin Kwok, Zilong Zhang, Feifei Cui, Da Zhao, Mulin Jun Li, Quan Zou. webTWAS: a resource for disease candidate susceptibility genes identified by transcriptome-wide association study. Nucleic Acids Research.2022, 50(D1): D1123-D1130.

[11] Trik, M., Pour Mozaffari, S., & Bidgoli, A. M. (2021). Providing an Adaptive Routing along with a Hybrid Selection Strategy to Increase Efficiency in NoC-Based Neuromorphic Systems. Computational Intelligence and Neuroscience, 2021(1), 8338903.

[12] Sai Huang, Guangdeng Zong, Ben Niu, Ning Xu, Xudong Zhao, Dynamic Self-Triggered Fuzzy Bipartite Time-Varying Formation Tracking for Nonlinear Multi-Agent Systems With Deferred Asymmetric Output Constraints, IEEE Transactions on Fuzzy Systems, 32(5): 2700-2712, 2024.

[13] Boyan Zhu,Ning Xu, Guangdeng Zong, Xudong Zhao.Adaptive optimized backstepping tracking control for full-state constrained nonlinear strict-feedback systems without using barrier Lyapunov function method. Optimal Control Applications and Methods, 2024, https://doi.org/10.1002/oca.3136.

[14] Trik, M., Akhavan, H., Bidgoli, A. M., Molk, A. M. N. G., Vashani, H., & Mozaffari, S. P. (2023). A new adaptive selection strategy for reducing latency in networks on chip. Integration, 89, 9-24.

[15] Wang, Z., Jin, Z., Yang, Z., Zhao, W., & Trik, M. (2023). Increasing efficiency for routing in internet of things using binary gray wolf optimization and fuzzy logic. Journal of King Saud University-Computer and Information Sciences, 35(9), 101732.

[16] Khezri, E., Yahya, R. O., Hassanzadeh, H., Mohaidat, M., Ahmadi, S., & Trik, M. (2024). DLJSF: Data-Locality Aware Job Scheduling IoT tasks in fog-cloud computing environments. Results in Engineering, 21, 101780.

[17] Fakhri, P. S., Asghari, O., Sarspy, S., Marand, M. B., Moshaver, P., & Trik, M. (2023). A fuzzy decision-making system for video tracking with multiple objects in non-stationary conditions. Heliyon, 9(11).

[18] Saidabad, M. Y., Hassanzadeh, H., Ebrahimi, S. H. S., Khezri, E., Rahimi, M. R., & Trik, M. (2024). An efficient approach for multi-label classification based on Advanced Kernel-Based Learning System. Intelligent Systems with Applications, 21, 200332.

[19] Khosravi, M., Trik, M., & Ansari, A. (2024). Diagnosis and classification of disturbances in the power distribution network by phasor measurement unit based on fuzzy intelligent system. The Journal of Engineering, 2024(1), e12322.

[20] Jaferian, G., Ramezani, D., & Wagner, M. G. (2024). Blockchain Potentials for the Game Industry: A Review. Games and Culture, 15554120231222578.

[21] Jaferian, G., Ramezani, D., Polyak, E., & Wagner, M. (2024). EXPLORING BLOCKCHAIN'S HORIZONS IN EDUCATIONAL GAMING. INTED2024 Proceedings, 5050-5058.

[22] Sun, J., Zhang, Y., & Trik, M. (2024). PBPHS: a profile-based predictive handover strategy for 5G networks. Cybernetics and Systems, 55(5), 1041-1062.

[23] Wang, G., Wu, J., & Trik, M. (2023). A novel approach to reduce video traffic based on understanding user demand and D2D communication in 5G networks. IETE Journal of Research, 1-17.

[24] Li, Y., Wang, H., & Trik, M. (2024). Design and simulation of a new current mirror circuit with low power consumption and high performance and output impedance. Analog Integrated Circuits and Signal Processing, 119(1), 29-41.

[25] Zhang, L., Hu, S., Trik, M., Liang, S., & Li, D. (2024). M2M communication performance for a noisy channel based on latency-aware source-based LTE network measurements. Alexandria Engineering Journal, 99, 47-63.

[26] Birdawod, H. Q., Khudhur, A. M., Kadir, D. H., & Saleh, D. M. (2024). A Wavelet Shrinkage Mixed with a Single-level 2D Discrete Wavelet Transform for Image Denoising. Kurdistan Journal of Applied Research, 9(2), 1-12.

[27] Ameen, A. K., Kadir, D. H., Abdullah, D. A., Maolood, I. Y., & Khidir, H. A. (2024). Assessing E-Government Effectiveness. ARO-THE SCIENTIFIC JOURNAL OF KOYA UNIVERSITY, 12(2), 52-60.

[28] Mokhlesi Ghanevati, D., Khorami, E., Boukani, B., & Trik, M. (2020). Improve replica placement in content distribution networks with hybrid technique. Journal of Advances in Computer Research, 11(1), 87-99.

[29] Liao, Y., Tang, Z., Gao, K., & Trik, M. (2024). Optimization of resources in intelligent electronic health systems based on Internet of Things to predict heart diseases via artificial neural network. Heliyon.

[30] E. I. Elsedimy, S. M. M. AboHashish, and F. Algarni, "New cardiovascular disease prediction approach using support vector machine and quantum-behaved particle swarm optimization," Multimed Tools Appl, vol. 83, no. 8, pp. 23901–23928, 2024.

[31] Hassanzadeh, H., Qadir, J. A., Omer, S. M., Ahmed, M. H., & Khezri, E. (2024, June). Deep Learning for Speaker Recognition: A Comparative Analysis of 1D-CNN and LSTM Models Using Diverse Datasets. In 2024 4th Interdisciplinary Conference on Electrics and Computer (INTCEC) (pp. 1-8). IEEE.

[32] Trik, M., Pour Mozafari, S., & Bidgoli, A. M. (2021). An adaptive routing strategy to reduce energy consumption in network on chip. Journal of Advances in Computer Research, 12(3), 13-26.

[33] Kiani, Farzad, Sajjad Nematzadeh, Fateme Aysin Anka, and Mine Afacan Findikli. "Chaotic sand cat swarm optimization." Mathematics 11, no. 10: 2340, 2023.

[34] Mohammed, Hardi, and Tarik Rashid. "A novel hybrid GWO with WOA for global numerical optimization and solving pressure vessel design." Neural Computing and Applications 32, no. 18: 14701-14718, 2020.

[35] Struniawski, Karol, Ryszard Kozera, and Aleksandra Konopka. "Performance of Selected Nature-Inspired Metaheuristic Algorithms Used for Extreme Learning Machine." In International Conference on Computational Science, pp. 498-512. Cham: Springer Nature Switzerland, 2023.

[36] Seyyedabbasi, Amir, and Farzad Kiani. "I-GWO and Ex-GWO: improved algorithms of the Grey Wolf Optimizer to solve global optimization problems." Engineering with Computers 37, no. 1 (2021): 509-532.

Sleep Disorder Diagnosis Through Complex-Morlet-Wavelet Representation Using Bi-GRU and Self-Attention

Mubarak Albathan*
College of Computer and Information Sciences, Imam Mohammad Ibn Saud Islamic University (IMSIU), Riyadh 11432, Saudi Arabia

***Abstract*—Sleep disorders pose notable health risks, impacting memory, cognitive performance, and overall well-being. Traditional polysomnography (PSG) used for sleep disorder diagnosis are complex and inconvenient due to complex multi-class representation of signals. This study introduces an automated sleep-disorder-detection method using electrooculography (EOG) and electroencephalography (EEG) signals to address the gaps in automated, real-time, and noninvasive sleep-disorder diagnosis. Traditional methods rely on complex PSG analysis, whereas the proposed method simplifies the involved process, reducing reliance on cumbersome equipment and specialized settings. The preprocessed EEG and EOG signals are transformed into a two-dimensional time-frequency image using a complex-Morlet-wavelet (CMW) transform. This transform assists in capturing both the frequency and time characteristics of the signals. Afterwards, the features are extracted using a bidirectional gated recurrent unit (Bi-GRU) with a self-attention layer and an ensemble-bagged tree classifier (EBTC) to correctly classify sleep disorders and very efficiently identify them. The overall system combines EOG and EEG signal features to accurately classify people with insomnia, narcolepsy, nocturnal frontal lobe epilepsy (NFLE), periodic leg movement (PLM), rapid-eye-movement (RBD), sleep behavior disorder (SDB), and healthy, with success rates of 99.7%, 97.6%, 95.4%, 94.5%, 96.5%, 98.3%, and 94.1%, respectively. Using the 10-fold cross-validation technique, the proposed method yields 96.59% accuracy and AUC of 0.966 with regard to classification of sleep disorders into multistage classes. The proposed system assists medical experts for automated sleep-disorder diagnosis.**

***Keywords*—Deep learning; complex morlet wavelet; bidirectional gated recurrent unit; sleep stage detection; multistage sleep disorder; ensemble-bagged tree classifier**

I. Introduction

Sleep is an essential concern for human health. It serves as a fundamental factor in physical and mental wellness. It is vital in memory consolidation, cognitive functions, cellular regeneration, and metabolic-brain-waste elimination [1]. Therefore, abnormalities in normal sleep patterns can cause many disorders, such as insomnia, narcolepsy, and sleep apnea disorder. Each sleep condition affects each individual's health differently, often inducing daytime weariness, cognitive impairment, cardiovascular disease, and mental health difficulties [2]. Therefore, accurate sleep-disruption assessment and therapy are crucial for overall health and quality of life. Polysomnography (PSG) is the most reliable sleep problem diagnosis method, but it involves overnight stays at medical institutions, which can be resource-intensive and uncomfortable for patients. The need for more accessible and user-friendly sleep problem diagnosis and analysis is evident, considering these restrictions and the discomfort it brings to patients.

This paper presents a unique automated sleep-disorder-detection approach employing EOG and EEG signals [4]. Sleep disruption monitoring using EOG and EEG is simple and less invasive. EOG records eye movements, which distinguish sleep phases, whereas EEG records muscle activity [5]. Instead of sophisticated PSG analysis, the suggested technique streamlines the operation, minimizing the need for expensive equipment and particular settings. A unique sleep disorder classification method uses sophisticated signal processing and machine learning. Complex signal processing approaches like Morlet-wavelet transformations and machine learning models with bidirectional gated recurrent units and self-attention layers make sleep diagnosis easier and more accurate, instilling a sense of confidence. This approach is appropriate for home monitoring since it reduces equipment and simplifies diagnosis. EEG and EOG signals detect sleep problems well. By identifying and assessing the most important signal information, diagnostic accuracy techniques typically outperform traditional PSG. Finally, EOG and EEG signal analysis improves sleep issue detection and efficiency in this study, providing reassurance of its accuracy and efficiency. It improves efficiency, accessibility, and patient comfort while maintaining the diagnostic integrity of traditional PSG.

Standard machine learning (ML)–based sleep-disorder detection involves the use of concepts such as support vector machines (SVMs), decision trees, and k-nearest neighbor (k-NN) algorithms. The study of [6] demonstrated the efficacy of SVMs in accurately classifying sleep apnea (accuracy rate = 85%) using EMG data. They highlighted the capability of machine learning in sleep disorder identification; however, handling data with a high number of dimensions was difficult. the research of [7] employed RF algorithms to differentiate various sleep disorders, such as insomnia and narcolepsy, by analyzing EOG signals. In this regard, they achieved an impressive accuracy rate of 89%, showcasing the efficacy of ensemble techniques. Despite exhibiting positive results, traditional machine learning–based approaches have several constraints. Feature selection is a primary issue in this regard, which often involves human participation and is, therefore, vulnerable to bias. To address this challenge, Aboalayon et al.

*Corresponding Author.

[8] effectively employed an automated feature selection technique, which substantially enhanced the involved model's performance (by 5%). However, such techniques faced challenges in terms of their comprehensibility and ability to be applied to new, unexpected data.

The advent of deep learning has considerably transformed sleep disorder detection [9]. Convolutional neural networks (CNNs) and RNNs have gained popularity because they can automatically extract features and detect patterns in time-series data. A CNN was employed herein to examine EMG data for rapid-eye-movement (REM) sleep behavior disorder (RBD) identification, yielding a precision of 95%. This accomplishment indicates a substantial improvement compared with traditional sleep-disorder-identification models. Further, LSTM was used to determine the time-dependent patterns of EOG signals to identify sleep stages. Notably, LSTM yielded a remarkable classification rate accuracy rate of 92% in this regard in a previous study. Despite their impressive accuracies, deep learning algorithms require large amounts of data and considerable computational resources, among other challenges. Moreover, these algorithms exhibit an issue about interpretability owing to the lack of transparency. Therefore, studies (e.g., an experimental study by Meridian et al. [10]) have been exploring hybrid models that involve a combination of CNNs with LSTM networks. This combination enables the usage of both spatial and temporal properties to improve model transparency and efficiency.

Conventional sleep problem diagnosis methods sometimes need resource-intensive polysomnography (PSG) and uncomfortable conditions. The study addresses these issues. While promising, current deep learning (DL) and machine learning (ML) methods are sometimes computationally intensive, require extensive hyperparameter tuning, and are often restricted to particular sleep phases or classes, resulting in solutions that are not very generalizable. Furthermore, the high level of data complexity and the widespread usage of hyperparameters in many existing algorithms cause overfitting, further restricting their application to various datasets.

A unique automated method that combines EEG and EOG data to diagnose sleep difficulties is presented in the paper. Combining advanced signal processing with a complex-morlet-wavelet (CMW) transform and a Bi-GRU with a self-attention layer simplifies the diagnostic process. This strategy is more accessible for home monitoring since it involves less sophisticated installations and heavy-duty equipment. The recommended technique also reduces overfitting and processing requirements while boosting sleep disorder classification accuracy, making it a more practical and complete solution for real-world applications.

This research substantially contributes to the field of sleep disorder through the following:

1) An innovative approach is introduced, involving complex-morlet-wavelet (CMW) transform–based feature extraction from preprocessed EOG and electroencephalography (EEG) signals. This method significantly enhances the accuracy of multistage sleep-disorder identification.

2) A bidirectional gated recurrent unit (BiGRU) with a self-attention layer is employed for feature extraction, followed using an ensemble-bagged tree classifier (EBTC) for precise multistage-sleep-disorder classification. This methodology ensures accurate delineation of complex sleep patterns by leveraging temporal insights and robust ensemble methods.

3) The effectiveness of the proposed method is robustly demonstrated in a real-world setting, offering a practical and reliable solution for home-based, patient-friendly sleep-disorder monitoring.

The remainder of this paper is organized as follows: Section II presents the relevant past studies. Section III comprehensively explains the approaches used herein, encompassing a CMW transform, feature extraction, and sleep disorder classification. Section IV presents the outcomes achieved regarding the categorization of people into healthy individuals and individuals with sleep disorders and the classification of sleep disorders into seven categories. Section V concludes the work.

II. Literature Review

A previous study classified sleep disorders using a novel machine-learning model that combined EEG, chin EMG, and dual-channel EOG [11]. The authors used the best orthogonal filterbank and Tsallis entropies to obtain high classification accuracies when considering the Sleep Heart Health Study (SHHS) database (90.7% for SHHS-1 and 91.8% for SHHS-2), achieving excellent automated sleep-disorder classification. Jarchi et al. [12] aimed to diagnose breathing- and eye-movement-related sleep disorders using electrocardiography (ECG) and EMG by developing a deep learning framework that yielded a mean accuracy of 72% in classifying people into four groups—healthy individuals and individuals with various sleep disorders (obstructive sleep apnea (OSA), restless leg syndrome (RLS), or both). This demonstrated the capacity of ECG and EMG in diagnosing sleep disorders. Meanwhile, Sharma et al. [13] introduced an automated technique for sleep disorder identification that involves analyzing EOG and EMG signals. A biorthogonal filter bank with Hjorth parameters was employed, which yielded a high overall accuracy of 94.3%. This technique was recognized for its effectiveness in at-home monitoring of different sleep disorders. Sekkal et al. [14] compared eight classic machine-learning techniques with a feed-forward neural network for sleep disorder identification and discussed the pros and cons of various sleep stage classifiers.

Sharma et al. [15] exploited EEG data to diagnose sleep problems. Using Hjorth parameters and an ensemble-boosted tree classifier, they classified sleep disorders with 99.2% accuracy. This method helps clinicians detect sleep problems. Rahman et al. [16] examined automated sleep-stage evaluation using EOG data. They used discrete-wavelet-transform EOG data to improve S1-sleep-stage detection over previous EOG-based approaches. Pei et al. [17] created a successful deep-learning model for sleep phases utilizing biological cues. Combining CNNs with gated recurrent units (GRUs) yielded a more versatile model than previous cutting-edge models.

An automated sleep-stage approach using EEG, EOG, and EMG was developed by Satapathy et al. [18]. The system found linear and nonlinear characteristics with good classification

accuracy and diagnosed sleep disorders. A novel sleep staging approach used EOG instead of EEG for practicality. This technique has 81.2% and 76.3% sleep-staging accuracy utilizing a two-scale CNN and RNN. Chambon et al. [20] used PSG data to characterize sleep phases using deep learning without explicitly designing characteristics. A fair use of channels and temporal data gave the model excellent classification performance. EEG-based sleep stage categorization using PSG analysis was advanced by several research [21–25]. This research showed that several machine learning methods and physiological signal combinations produced excellent accuracy and showed the benefits of multimodal signal processing.

In another research [26], a restricted PSG montage classified the sleep phases of 106 people—53 with RBD and 53 healthy. This was done using an RF classifier with 156 EEG, EOG, and EMG characteristics. RBD was detected using muscle atonia measurement and sleep architecture characteristics. The model attained Cohen's Kappa score of 0.62 for sleep staging and 96% RBD detection accuracy, demonstrating the benefits of sleep architecture and transitions. Malafeev et al. [27] developed a three-dimensional (3D) CNN for sleep stage categorization using several channels and EEG, EMG, and EOG inputs. Time, frequency, and time-frequency characteristics were sent to the 3D CNN. Three-dimensional convolutional layers created intrinsic relationships between biosignals and frequency bands, while two-dimensional layers obtained frequency correlations. The model identified significant channels and frequency bands throughout sleep phases using partial-dot-product attention layers and an LSTM unit. This model also achieved classification accuracies of 0.832 and 0.820 on the ISRUC-S3 and S1 datasets. These findings showed the model detected sleep phases reliably and effectively.

Cooray et al. [28] proposed "quasi-normalization" for feature normalization using the ISRUC-Sleep dataset. An RF algorithm sorted the data into five sleep states. Using leave-one-out cross-validation, EOG and EMG data were integrated to achieve Cohen's kappa value of 0.749 and 80.8% accuracy. The results matched the American Academy of Sleep Medicine standards. Electrooculography and electromyography may be as effective as electroencephalography at identifying sleep phases. Another research [29] studied sleep phases in 123 suspected sleep disorder patients using a BiLSTM network. The model received multivariate time-series heart rate, breathing rate, and body movement frequency. With an accuracy of 71.2%, Cohen's κ coefficient of 0.425, and an F1 score of 0.650, the model effectively classifies sleep phases using minimal physiological cues.

Morokuma et al. [30] focused on EEG and EOG signals and developed a deep CNN architecture for automated sleep-stage classification. Its performance was evaluated against human expert agreement, with CNN considerably outperforming recent single-EEG-channel approaches. The study highlighted the crucial role of network depth in achieving high classification accuracy. Another study [31] targeted sleep-wake detection in OSA patients using single-channel ECG signals. The heart rate variability signals were derived, and features were classified using decision trees, SVMs, and ensemble classifiers. The model achieved accuracies of 81.35% with three features and 87.12% with ten features, suggesting its utility in the OSA diagnosis. An automated deep nine-layer one-dimensional CNN (9L-1D-CNN-SSC) for multiclass sleep staging was also developed [32]. The model was tested on ISRUC-Sleep subgroup datasets and achieved an accuracy of up to 99.50% in classifying sleep stages with different signal combinations, indicating its applicability for clinical use. Satapathy and Loganathan [33] developed a dual-modal, multiscale deep neural network for sleep staging that used EEG and ECG signals. When tested on the MIT-BIH PSG dataset, the model achieved high accuracy rates of 80.40%–98.84% in classifying different sleep stages. This shows that combining EEG and ECG signals for sleep analysis yields accurate results.

Zhao et al. [34] addressed the limitations of automated sleep-staging systems using portable EEG headbands by developing a deep-learning model using convolutional and long-term memory layers. The model achieved validation accuracies of 74% on headband data and 77% on PSG data, demonstrating its potential in ambulatory sleep assessments. SleepPrintNet [35] was also introduced to capture the SleepPrint in a physiological time series for sleep staging. It yielded higher accuracy on the MASS-SS3 dataset than baseline models because it used EEG, EOG, and EMG features along with temporal, spectral, and spatial features. This approach underscored the value of multimodal feature integration in sleep stage classification.

Several studies [26–36] collectively represented a wide array of methodologies, ranging from RF classifiers to BiLSTM networks and CNNs. These methods were tested on various datasets such as the SHHS and ISRUC-Sleep datasets, achieving considerable advancements in sleep stage classification, disorder diagnosis, and automated sleep-staging systems. Many of these studies highlighted the effectiveness of combining EEG, EOG, EMG, and ECG signals, emphasizing on the efficiency of feature extraction and selection in improving accuracy and robustness. The comparative analysis of these systems is shown in Table I.

TABLE I. COMPARISON OF METHODOLOGIES, RESULTS, AND LIMITATIONS PROPOSED IN STUDIES

Reference	Methodology	Results	Limitations
[11]	A machine learning model using EEG, chin EMG, and dual-channel EOG	90.7% accuracy on SHHS-1 and 91.8% on SHHS-2	three sleep classes in SHHS-1, and five classes in SHHS-2 datasets, none-generalized.
[12]	Diagnosing sleep disorders using ECG and EMG and a deep learning framework	Mean accuracy of 72% and weighted F1 score of 0.57	Four-sleep classes, Computational expensive and huge hyper-parameters sitting.
[13]	An automatic detection system for sleep disorders using EOG and EMG signals	Accuracy of 94.3%	Limited five-sleep classes, and three sleep stages.
[15]	Identification of sleep disorders using EEG and EBTC	Classification accuracies up to 99.2%	Limited four-sleep classes and used only one CAP dataset so non-generalize solution.

[17]	A deep learning method using CNNs and GRUs for sleep stage identification	Accuracy of 83.15% and kappa of 0.76	Limited five-sleep stages and not classes, and computationally expensive
[18]	An automated sleep-staging system using EEG, EOG, and EMG signals and an RF classifier	Accuracy of 98.99%, 98.75%, 98.17%, and 99.14% with respect to sleep stage	five-sleep states, non-sleep classes, and computationally expensive
[19]	A sleep staging approach using EOG and two-scale CNNs and RNNs	Accuracy of 81.2% of two-scale CNNs and 76.3% of RNNs.	Limited sleep classes, not generalized, and required huge hyper-parameters.
[20]	Convolutional deep learning approach and gradient boosting for sleep stage classification using PSG signals	Accuracy of approx. 80%.	Huge hyperparameters, Five-sleep stages
[21]	An efficient technique for sleep stage classification based on EEG signal analysis	RF algorithm achieved a high accuracy of 97.8%	Limited dataset, Overfitting due to RF and required stopping critria, three-stages of sleep disorder.
[22]	SleepEEGNet for automated sleep-stage annotation using single-channel EEG and BiRNN.	Accuracy of 84.26%, F1-score of 79.66% and $\kappa = 0.79$.	Limited sleep stage, tested on limited dataset
[23]	A deep learning model for sleep staging in children using EEG, EOG, and chin EMG	Cross-validated accuracy of 84.1%	Limited sleep classes and computationally expensive
[24]	A deep learning model for sleep staging using multiple PSG signals and 2D CNNs and LSTM modules	Sleep-EDF: Acc-0.86, K-0.81	Limited sleep classes, not generalized
[25]	Sleep staging using Relief, AdaBoost with RF	accuracy of 97.96%	Limited sleep classes, not generalized, and classifier overfitting
[26]	Sleep stages using EEG, EMG, and EOG signals and CNN-LSTM	Accuracy of 0.832 on ISRUC-S3 and 0.820 on ISRUC-S1	Classifier overfitting, and Computationally expensive
[27]	Sleep stage classification using 3D-CNN	Accuracy of 0.832, F1-score of 0.814 and kappa of 0.783 on ISRUC-S3	Required huge hyper-parameters sitting and computationally expensive due to huge epochs.
[28]	EOG and EMG data are utilized to predict multistage sleep by using RF classification	Accuracy of 92%	Limited sleep classes and classifier overfittig, limiated dataeset so no generalize solution
[29]	A ”quasi-normalization” method with RF classifier	Accuracy of 84.7%	Not generalized, overfitting
[30]	Polysomnography (PSG) data with BiLSTM classifier for detection of sleep stages	Accuracy of $71.2 \pm 5.8\%$, and F1 score of 0.650 ± 0.083	Only sleep stages and no multiclass solution.
[31]	Detection of human sleep EEG and EOG signals with CNN architecture.	F1-score of 77%	Limited sleep classes, not generalized
[32]	An ensemble technique based on three classifiers: DT, kNN and SVMs.	Sensitivity and specificity values of 0.90 and 0.85, respectively.	Limited two-sleep classes, not generalized
[33]	A 9L-1D-CNN-SSC model for sleep staging with different signal combinations	Classification accuracies up to 99.50% for various sleep stages	Limited sleep classes, not generalized, overfitting, and computationally expensive
[33]	A dual-modal multiscale deep neural network using EEG and ECG signals	Accuracies between 80.40% and 98.84% for different sleep stages	Overfitting and computationally expensive
[34]	A deep learning model with convolutional and LSTM layers for EEG headband data	74% accuracy on headband data and 77% on PSG data	Limited sleep classes, not generalized, overfitting, and computationally expensive
[35]	SleepPrintNet integrating EEG, EOG, and EMG signal features	Outperformed baseline models in accuracy on the MASS-SS3 dataset	Limited dataset utilized.

III. Methodology

The multilayer sleep-disorder classification system's systematic flow diagram is shown in Fig. 1. The novel method categorized sleep disorders using preprocessed EOG and EEG information. Normalization was done to standardize these signals for analysis. Next, a bandpass filter reduces noise and frequencies to increase signal quality, which is important for detecting sleep disorder symptoms. Next, the complex morlet wavelet (CMW) transform was utilized to extract features from EOG and EEG data for reliable disease categorization. A BiGRU with a self-attention layer extracted characteristics, and an EBTC automatically found sleep problems. Due to its efficiency in processing time-series data, the GRU, a kind of RNN, was utilized to evaluate EOG and EEG temporal patterns. By aggregating estimates from several decision tree models, the EBTC improved its accuracy and applicability. Finally, the BiGRU and EBTC models were integrated to categorize the data using voting, average probability, or a more complex meta-classifier. Using 10-fold cross-validation, the system's performance was rigorously assessed to ensure its efficacy and robustness in real-world circumstances. The suggested method improves sleep problem classification and shows the potential of signal processing and machine learning for medical diagnosis.

A. Data Acquisition and Augmentation

Define abbreviations and acronyms the first time they are The CAP Sleep database [37], provided by PhysioNet [38], containing PSG recordings from 108 people, was used as the primary data source. The EEG and EOG signals were collected and analyzed from this database.

The EEG and EOG signals are collected from individuals while they are asleep to detect sleep stages and classify sleep disorders. The EEG and EOG signals were comprehensively distributed (Table II) to evaluate the effectiveness of the proposed system in diagnosing different sleep disorders such as insomnia, narcolepsy, nocturnal frontal lobe epilepsy (NFLE), periodic leg movement (PLM), RBD, and sleep-disordered breathing (SDB) as well as healthy individuals. Fig. 2 shows a visual representation of signals from each group.

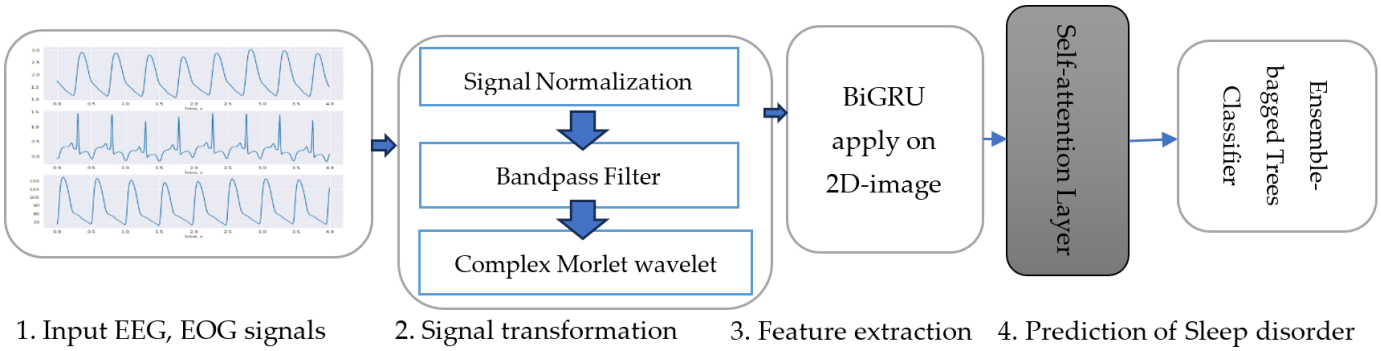


Fig. 1. Systematic flow diagram of the proposed multilayer sleep-disorder classification system.

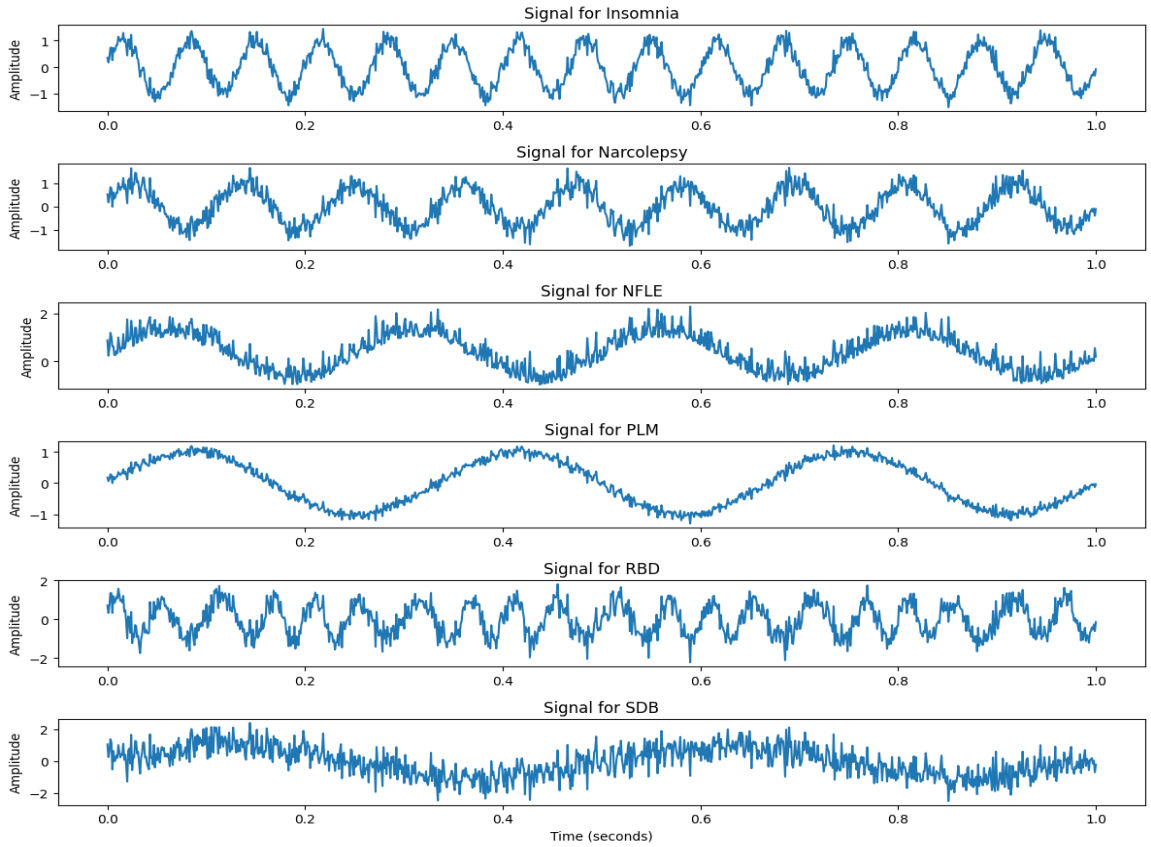


Fig. 2. Input EEG and EOG signals for identifying sleep disorders.

To standardize the sample, the dataset was subjected to data augmentation (Table II) using the synthetic minority over-sampling technique (SMOTE). SMOTE is used to rectify imbalanced datasets, particularly medical data in sleep studies, where certain classes are inadequately represented. It generates artificial, yet believable, examples using the available data from the underrepresented class. During the procedure of SMOTE, the dataset was assumed to contain EEG and EOG signal features, referred to as "features, "along with their corresponding labels, referred to as "labels." These labels classify each group of features into distinct sleep disorders or stages. The dataset was initially divided into 70% training and 30% testing datasets. SMOTE was used only on the training set to prevent synthetic data from affecting the model evaluation. It generated additional samples for classes that were not well represented, thus equalizing the distribution of classes, as balanced datasets improve model performance, particularly in classification tasks.

The training dataset (X-res and Y-res) contained the original and newly synthesized samples. This dataset was then used to train a machine-learning model. Notably, the model acquired knowledge from a dataset that provided a more equitable representation of all categories, thus mitigating its inclination toward the dominant category. The model's effectiveness was assessed using the initial, unmodified testing dataset. This approach ensures a complete evaluation of SMOTE's effectiveness as it accurately determines the effect of SMOTE on the model's ability to separate different sleep states.

TABLE II. EEG AND EOG SIGNAL DISTRIBUTION USED TO TEST THE SYSTEM PERFORMANCE

Sleep Stage	EEG	EOG	Total EEG and EOG	Data Augmentation
Insomnia	3800	1200	5000	2500
Narcolepsy	1300	1400	2700	2500
NFLE	3200	2000	3400	2500
PLM	1300	1000	2300	2500
RBD	5000	2000	7000	2500
SDB	200	100	300	2500
Healthy	400	200	600	2500

B. Signal Transformation

This multiclass sleep disorder prediction research uses the complex morlet wavelet transform (CMWT) [39] because it provides amplitude and phase information, unlike spectrograms, which indicate magnitude, and scalograms, which reveal phase information. While scalograms are superior for non-stationary signals, they cannot match the CMW transform. Spectrograms provide a broader perspective of power distribution. Arranging these normalized features creates the 2D stack picture, which is then fed into machine learning systems like the bidirectional gated recurrent unit (BiGRU), which maintains signal characteristics' spatial and temporal correlations.

PyWavelets were used to analyze a continuous wavelet transform (CWT). CWT is a robust time-frequency analysis method that can analyze signals at multiple scales or frequencies. This study employed the complex morlet wavelet (CMW), which analyzes nonstationary biological data via temporal and frequency localization. The CWT captured both high-frequency and low-frequency components of each simulated disorder signal by decomposing it into distinct scales. Each row and column of the coefficient matrix represented a frequency range and time point, respectively. The coefficients were subsequently represented using a heatmap, which displayed the frequency characteristics of the signal over time. The color intensity of the heatmap corresponded to the magnitude of the signal across different frequencies, providing valuable information about the distinctive patterns associated with each sleep problem.

Thus, the CMW is ideal for feature extraction in biological signal processing. By explaining the complicated time frequency features of EEG and EOG signals, this study presents in-depth grouping sleep disorders into different categories of identifying unique brain patterns. The proposed approach conforms with current research practices in biomedical engineering and computational neuroscience, focusing on advanced signal processing techniques to understand complicated physiological events. Fig. 3 displays a visual representation of each sleep-disorder type.

C. Feature Extraction Using BiGRU-Attention

A sophisticated neural network structure, namely a BiGRU [40] with a self-attention mechanism, was used to mine the time-series data, such as EEG and EOG signals. This study presents a promising method that takes the concatenated wavelet features of the EEG and EOG signals to create a unified 2D representation of these features. Owing to its bidirectional nature, BiGRU analyzed patterns in forward and reverse directions throughout time and revealed the intrinsic temporal dynamics in signal data. Fig. 4 shows the BiGRU architecture, wherein signal data are used to find and extract relevant features indicating the essential traits of different sleep stages or disorders.

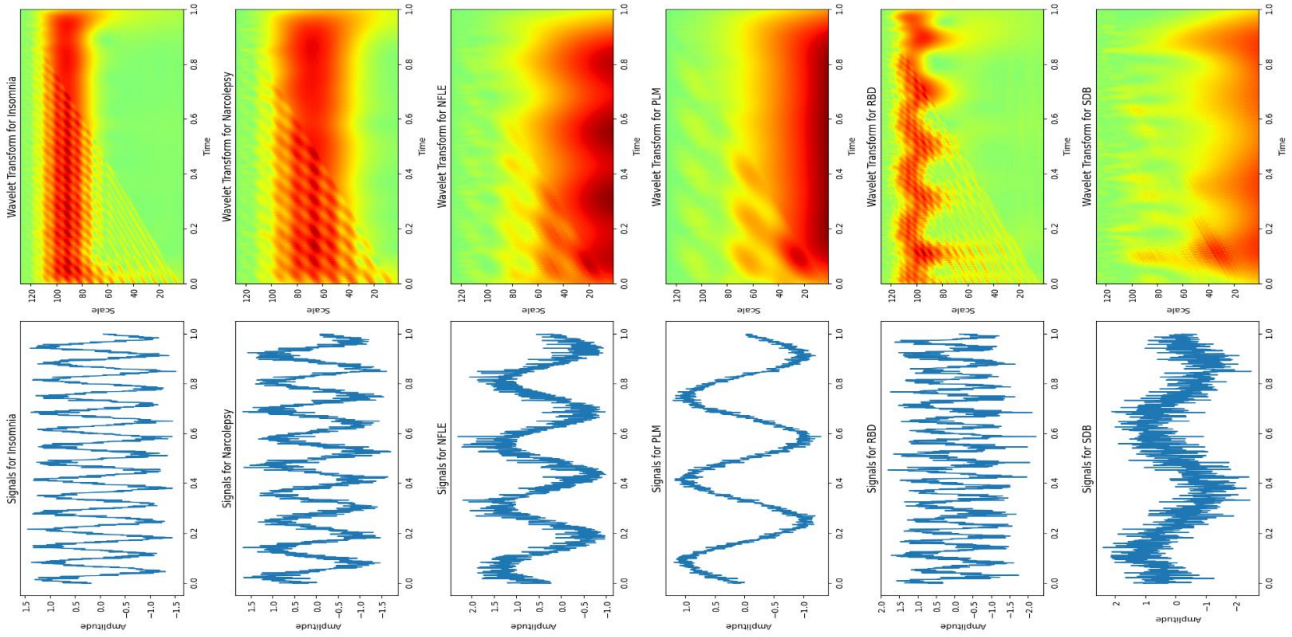


Fig. 3. Complex-morlet-wavelet transform using electroencephalogram signals of sleep disorders such as insomnia, narcolepsy, NFLE, PLM, RBD, and SDB.

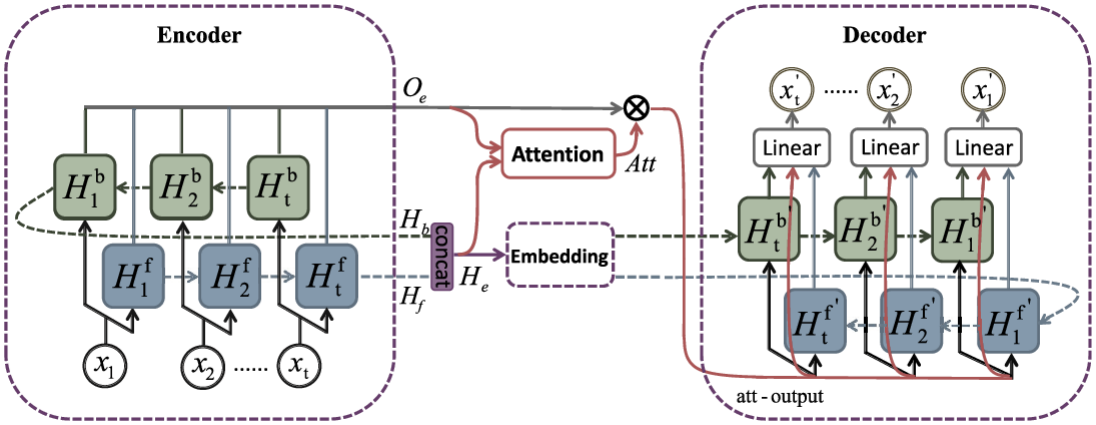


Fig. 4. Architecture of BiGRU with a self-attention layer.

EEG and EOG signals can be better analyzed by combining a self-attention mechanism with a BiGRU model, particularly when the signals change into a complex wavelet domain. The self-attention mechanism is a highly efficient neural network component, allows the model to prioritize different areas of the input data, making it more contextually aware. The sophisticated morlet wave transform further refines this by dividing EEG and EOG data into high- and low-frequency patterns, providing a deeper understanding of sleep stages and difficulties. The BiGRU model, in turn, uses these modified signals with rich time-frequency information for forward and backward analysis, capturing temporal correlations and patterns. This self-attention mechanism, with its efficiency, helps the model focus on essential signal alterations. It prioritizes segments that provide task-relevant information, such as sleep issues or phases, with high performance and accuracy.

Complex wavelet transformations, BiGRU, and self-attention mechanisms create a robust signal-processing paradigm. Precision frequency information was added to signals using the wavelet transform. The BiGRU neural network caught temporal patterns, and the self-attention mechanism focused on the most critical parts. When utilized together, they can extract crucial and relevant information from EEG and EOG data, making sleep studies and associated research more reliable and valuable. This strategy increases sleep study categorization and predictions by deepening physiological signal understanding.

EEG and EOG signals are converted to CMW for sleep disorder research and fed into the BiGRU model. Time-frequency analysis often uses the morlet wavelet because it splits initial signals into signals with various frequencies to optimize temporal and spectral localization. Thus, the BiGRU model learns from scale patterns. This model may detect small brain activity and eye movement changes that signal sleep phases and issues. This system employs a CMW transform and BiGRU model to use the BiGRU model's comprehensive time-frequency signal representation and powerful sequence modeling. By highlighting essential frequency components, the wavelet transform improves signals. The BiGRU then extracts key characteristics from this modified data, providing a robust collection of features for analysis or classification. This method handles EEG and EOG signal complexity and volatility well, making it suitable for advanced sleep investigations and diagnostics.

A BiGRU is a better version of the regular GRU developed for the model to obtain information from states preceding and succeeding the unit in a sequence. This is particularly advantageous in situations where the overall context of the entire sequence is crucial for making accurate predictions. A standard GRU operates on data sequentially and has a hidden state that serves as a memory to retain previous information. Nevertheless, it has only acquired data from earlier occurrences. A BiGRU comprises two GRUs operating in opposing directions: one GRU processes the sequence from the beginning to the end, whereas the other GRU processes it from the end to the beginning. These outputs are combined at each time step to obtain the entire sequence by incorporating details from the previous and subsequent contexts.

A GRU cell at time step t computes the following:

$$\text{Update gate } zt = \sigma(Wz \times [ht-1, xt] + bz), \quad (1)$$

$$\text{Reset gate } rt = \sigma(Wr \times [ht-1, xt] + br), \quad (2)$$

$$\text{Candidate hidden state } ht = tanh(Wh \cdot [rt \times ht-1, xt] + bh, \quad (3)$$

$$\text{Final hidden state } ht' = zt \times ht - 1 + (1 - zt) \times ht, \quad (4)$$

where σ denotes the sigmoid activation function, tanh is the hyperbolic tangent function, W and b are the weights and biases, respectively, *xt* is the input at time t, and ht is the hidden state at time t. The BiGRU contains two hidden states at each time step, namely *ht(fwd)* and *ht(bwd)*, calculated by the forward and backward GRUs, respectively. The forward GRU processes the sequence in the conventional manner, whereas the backward GRU processes it in the opposite direction. The combined hidden state at each time step t is acquired by either concatenating or summing the forward and backward hidden states:

$$ht(bi) = ht(fwd) + ht(bwd). \quad (5)$$

The BiGRU can capture dependencies and patterns that may be overlooked by a normal GRU, particularly in sequences wherein the future and past contexts are equally notable. This approach is frequently used in applications such as sequence labeling, time-series prediction, and natural language processing. BiGRUs have a higher computational cost and more parameters than regular GRUs, which may result in overfitting

when working with smaller datasets. Thus, the BiGRU enhances the functionality of the regular GRU by including inputs from the forward and backward directions of a sequence, resulting in a more holistic comprehension of the context.

The self-attention mechanism primarily focuses on the internal dependence of an input (Fig. 5). The output of the neural unit in the current moment may probably be affected by the EEG and EOG signals representation in the form of CMW transform. Based on the degree of influence, different weight parameters were assigned to signals such that the model can pay attention to the pivotal signals of stress information. The scaled dot-product attention model was used herein to optimize the data:

$$Attention(Q,K,V) = Softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V. \quad (6)$$

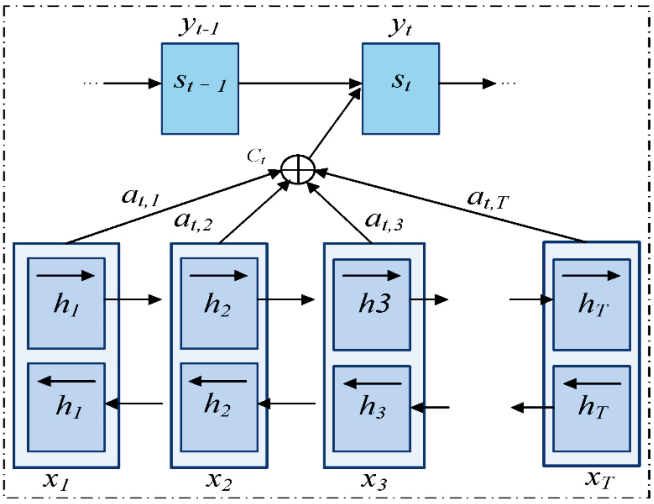


Fig. 5. A self-attention layer integrated into the BiGRU model.

The matrices Q, K, and V comprise query vectors, key vectors, and value vectors, respectively, and dk is the dimension of the input vector. In self-attention mechanism, Q, K, and V were derived from the same input and similarity among all words in the sentence was calculated. The greater the similarity among the signals, the stronger the correlation among them; thus, the dependency within the sentence can be captured. In question classification, the impact of each clause and individual word in the question varies. Certain words or sentences are crucial in question classification, whereas others have less impact. To effectively capture the relevant information in a question, an attention mechanism was added to the BiGRU model. This approach aims to highlight key semantic features, extract useful information, and accurately assess the contribution of each word for the classification of the entire question. By doing so, it ensures that the most crucial information is retained while filtering out redundant information, enhancing the efficiency and performance of question classification. The layer network uses the output of the upper layer network model as the input for the layer model, resulting in the vector representation of each sentence in the BiGRU model with the self-attention mechanism.

The basic form of self-attention mechanism can be expressed as follows:

$$S = \tanh(M)$$

$$\alpha = \text{softmax}(W_n \times S) \quad (6)$$

$$r = M \times \alpha_n$$
$$q = \tanh(r)$$

D. Feature Classification

Algorithm 1 summarizes stress classification using the BiGRU model with self-attention mechanism (BiGRU) and ensemble-bagged tree classifier. The machine learning method EBTC [41] blends bagging (bootstrap aggregating) with decision trees. It is especially helpful in minimizing variance, preventing overfitting, and boosting model resilience. The EBTC minimizes prediction variance by averaging many trees, which is useful when individual trees overfit. Bootstrapping randomly creates variation among ensemble trees, essential to the approach's efficacy. The ensemble's averaging effect makes the EBTC resilient to data outliers and noise. Bagged ensemble learning improves machine learning algorithm stability and accuracy. Multiple predictors are utilized to create an aggregated predictor. Classification and regression employ decision trees. They made a decision tree by subdividing the data by feature value testing. From the original dataset, multiple bootstrap samples are randomly chosen subsets of data (with replacement) of the same size. Independent decision trees are trained for each bootstrap sample. Training data is different for each tree owing to random sampling with replacement.

Algorithm 1: Sleep disorder classification using EEG and EOG signals and BiGRU with a self-attention layer and EBTC.	
Input	• EOG_data: Array of EOG signal data • EEG_data: Array of EEG signal data • sampling_rate: Sampling rate of the EOG and EEG data • wavelet_parameters: Parameters for the complex wavelet transform
Output	• Final-classification.
Step 1:	Feature Extraction: Filtered_EOG = bandpass_filter(EOG_data, lowcut, highcut, sampling_rate) Filtered_EEG = bandpass_filter(EEG_data, lowcut, highcut, sampling_rate) • Wavelet_Features_EOG←Morlet-WaveletTransform (Filtered_EOG,wavelet_parameters) • Wavelet_Features_EEG←Morlet-waveletTransform(Filtered_EEG,wavelet_parameters) • Combined_Features = concatenate((Wavelet_Features_EEG, Wavelet_Features_EOG), axis=1) • Combined_Image ← stacking(Combined_Features)
Step 2:	Gated Recurrent Unit (GRU) Classifier: GRU_Model←InitializeGRU(layer_parameters). Hidden_States ← GRU_Model(Combined_Image)
Step 3:	Attention Mechanism: • The BiGRU produces a sequence of hidden states ht. • Attention *scores α t* are computed for each hidden state as stated in Eq. (8).

Step 4:	Weighted Sum: Attention_Scores = compute_attention_scores(Hidden_States) Weighted_Hidden_States ← Attention_Scores @ Hidden_States
Step 5:	Ensemble-bagged Tree Classifier (EBTC): • EBTC_Model ← TrainEBTC(Weighted_Hidden_States, labels) • EBTC_Predictions ← EBTC_Model.predict(Weighted_Hidden_States)
Step 6:	Combination and Final Classification: Final_classification ← classify(EBTC_Predictions)
Step 7:	Evaluation: metrics ← evaluate(Final_classification, true_labels)

After all decision trees are trained, the ensemble model makes predictions by aggregating the predictions from all individual trees. This approach is followed for classification tasks via majority voting, wherein each tree votes for a class, and the class with the most votes is considered the ensemble's prediction.

A dataset D with N instances is considered for the EBTC with M trees. For each tree m = 1,2, ..., M), a bootstrap sample Dm is created by randomly selecting N instances from D with replacement. A decision tree Tm is then trained on Dm. For a new instance x, the prediction y is given by Eq. (7).

$$y = model\{T1(x), T2(x), \ldots, TM(x)\}. \quad (7)$$

The EBTC minimizes prediction variance by averaging many trees, which is useful if some trees overfit. The attention mechanism calculates weights using a straightforward scoring function such as a dot product and a Softmax. For a hidden state ht, the attention score αt is calculated as follows:

$$y\alpha t = \sum_T^i exp\big(score(hi)\big)exp\big(score(ht)\big) \quad (8)$$

where score(ht) is a dot product of the hidden state ht and some learnable parameter and T is the length of the sequence. The model then computes a weighted sum of the hidden states using attention weights, thereby generating a context vector that encapsulates the most relevant information from the entire sequence.

$$Y_c^{context-Vector} = \sum_T^i \alpha t \cdot ht \quad (9)$$

A fully connected layer then uses this context vector for the final classification (such as determining the type of sleep disorder). When analyzing EOG and EEG signal analysis for sleep disorders, the attention mechanism assigns higher weights to hidden states corresponding to signal patterns characteristic of certain sleep disorders. In contrast, the signal's ordinary or less informative patterns are assigned lower weights. The BiGRU model prioritizes essential sections of the EOG and EEG signals with a higher predictive value for various sleep disorders, thereby improving the classification accuracy and efficiency.

IV. Experimental Results

This section overviews the creation of distinct subsets of data and classification performed to identify various sleep disorders. The integrated deep learning and machine learning models, including the meta-learner, were trained and tested on a highly advanced computational infrastructure comprising a notebook with powerful specifications to manage the computing requirements of the models effectively. The processor was an Intel(R) Core (TM) i7, 10th Generation, with an essential clock speed of 3.34 GHz, outfitted with four cores and eight logical processors for efficient parallel computing. The machine was equipped with 32GB RAM to effectively cater to the demanding memory requirements of training and testing deep learning models. Windows 10 offers a reliable and compatible platform for various machine-learning operations. The deep learning models were constructed and trained using the TensorFlow framework in conjunction with the Keras framework.

A. Performance Measures

The categorization performances of deep learning and machine learning models were compared using traditional measurement metrics. These measures are crucial for fully grasping the effectiveness of the models in terms of different aspects of categorization performance. The metrics and their corresponding calculation algorithms are outlined below:

1) Accuracy: This metric represents the proportion of true results (true positives and negatives) among the total number of cases examined.

$$Accuracy\ (ACC) = \frac{TP+TN}{TP+TN+FP+FN} \quad (10)$$

2) Sensitivity (SE) or Recall (RC): It assesses the model's false negative avoidance.

$$Recall(RC) = Senstivity\ (SE) = \frac{TP}{TP+FN} \quad (11)$$

3) Specificity (SP): This metric assesses the proportion of actual negatives that are correctly identified.

$$Specificity\ (SP) = \frac{TN}{TN+FP} \quad (12)$$

4) F1-score (FS): This metric is a harmonic mean of precision and recall and balances the two.

$$F1-score\ (FS) = 2 \times \frac{PR \times RC}{PR+RC} \quad (13)$$

In this Eq. (13), the RC and PR are precision and recall metrics calculated from Eq. (11) to Eq. (14), respectively.

$$Precision\ (PR) = \frac{TP}{TP+FP} \quad (14)$$

In addition, a confusion matrix is used to assess the efficacy of the classification models.

- TN: This refers to accurately predicted negative occurrences.
- FP: This refers to incorrectly predicted positive observations.
- FN: This refers to incorrectly predicted negative observations.
- TP: This refers to accurately predicted positive observations.

These indicators assessed the performance of classification models, emphasizing their strengths and identifying areas for enhancement. This provided valuable information regarding the model's ability to accurately differentiate between several classes and maintain a balance between precision and recall. In

addition, the area under the receiver operating characteristic curve (AUC) is also utilized to measure the performance of various systems.

B. Experimental Results

The classification performance of models in differentiating healthy subjects from those with sleep disorders and identifying specific sleep disorders is discussed as follows. Tables IV–VI present the results obtained using the proposed system. Several experiments are performed to identify different classes of sleep disorders compared to normal stage. The required hyperparameters are described in Table III. These parameters are defined based on the several experiments.

TABLE III. HYPERPARAMETERS SETUP REQUIRED BY PROPOSED SYSTEM

Models	Hyperparameter	Values
BiGRU	Units	128
	Layers	2
	Dropout Rate	0.3
	Recurrent Dropout Rate	0.2
	Activation Function	tanh
	Batch Size	64
	Learning Rate	0.001
Self-Attention Layer	Attention Size	Equal to Number of Units
	Attention Activation	Softmax scores
EBTC	Number of Estimators	200
	Maximum Depth of Trees	30
	Minimum Samples Split	5
	Minimum Samples Leaf	2
	Bootstrap Samples	True
	Max Features	sqrt

This research, as presented in Table IV, has yielded significant classification results. We have compared the performance of a proposed approach that utilizes complex-morelet wavelet (CMW) decomposition with sleep stages, against a method that uses discrete wavelet transform (DWT) to convert 1D sleep signals into a combined 2D-CMW image. The choice between DWT and CMW is crucial and depends on the needs of the particular analysis. While DWT may be a better choice for computationally efficient broad feature extraction, the CMW generates a scalogram, a 2D array that provides a detailed view of the signal's frequency content over time. This detailed view is crucial for deep analysis that requires frequency and phase information. By converting 1D sleep problem signals into 2D, CMW may provide more significant information for diagnosing and comprehending complicated sleep events.

The advantage of the proposed approach is evident in significantly improved accuracy and area under the curve (AUC) across all sleep disorders. In this context, 'accuracy' refers to the percentage of correctly classified sleep stages or disorders, while 'AUC' is a measure of the model's ability to distinguish between different sleep stages or disorders. For instance, accuracy increased in the case of insomnia from 81.45% with DWT to an impressive 99.70% with CMW, and AUC rose from 0.822 to 0.997. Similar enhancements are observed across other disorders, such as narcolepsy; accuracy improved from 80.10% to 97.60%, and AUC increased from 0.834 to 0.976. In the case of NFLE, accuracy rose from 82.47% to 95.40%, with AUC increasing from 0.833 to 0.954. In the case of PLM, there was an increase in accuracy from 85.67% to 94.50% and in AUC from 0.865 to 0.945. For RBD, accuracy increased from 84.32% to 96.50%, with AUC improving from 0.854 to 0.965. For SDB, accuracy jumped from 85.20% to 98.30%, and AUC from 0.876 to 0.983. For healthy individuals, accuracy went from 87.00% to 94.10%, with the AUC moving from 0.876 to 0.941. Accordingly, our research has demonstrated the remarkable improvements in accuracy and AUC that the proposed approach offers compared to the method using discrete wavelet decomposition. These improvements underscore the potential of our approach to enhance the diagnosis and treatment of sleep disorders. These findings unequivocally demonstrate the value of incorporating CMW into the classification procedure for a more precise and trustworthy diagnosis of sleep disorders.

TABLE IV. CLASSIFICATION RESULTS OBTAINED WITHOUT USING WAVELET DECOMPOSITION WITH THE SLEEP STAGES. RESULTS ARE OBTAINED USING 10-FOLD CROSS-VALIDATION

Disorder	Discrete Wavelet Transform (DWT)		Complex-Morlet-wavelet (CMW) transform	
	Acc (%)	AUC	Acc (%)	AUC
Insomnia	81.45	0.822	99.70	0.997
Narcolepsy	80.10	0.834	97.60	0.976
NFLE	82.47	0.833	95.40	0.954
PLM	85.67	0.865	94.50	0.945
RBD	84.32	0.854	96.50	0.965
SDB	85.20	0.876	98.30	0.983
Healthy	87.00	0.876	94.10	0.941
Average	83.74	0.858	96.59	0.966

Table V presents sleep disorder classification results obtained using various classifiers with the hold-out validation strategy. The proposed approach, CMW-BiGRU-Self-attention-EBTC, demonstrates superior performance across all sleep disorders compared to alternative methods such as LSTM and GRU-SVM. The proposed CMW-BiGRU-Self-attention-EBTC method shows significant advantages over alternative approaches, highlighting its effectiveness in accurately classifying sleep disorders.

An independent experiment was conducted to compare the CMW transform with spectrogram and scalogram techniques. In practice, the CMW transforms provide both amplitude and phase information, which are essential for detailed signal analysis, particularly in identifying phase coupling between different signal components. The CMWT is particularly advantageous due to its rich and detailed time-frequency representation. Table VI offers a comprehensive overview of such classification performance for a six-class sleep disorder classification task

with signal transformation using a 2D-spectogram image. Fig. 6 demonstrates the various confusion matrices, illustrating the distribution of actual and predicted classes, and the number of instances correctly and incorrectly classified for each sleep disorder when a 2D-scalogram is used. The results in Table VI were obtained without using CMW transforms, but the 2D-spectogram was used to analyze the performance. As shown in this table, the proposed system, which does not incorporate the complex morlet wavelet (CMW) transformation, the bi-directional gated recurrent unit (BiGRU) with self-attention, and the ensemble bagged tree classifier (EBTC) is utilized. The result does not outperform by the system without using CMW transforms. Similar trends were observed using 2D-scalogram without using the CMW transform technique, as depicted in Fig. 6. Hence, the proposed CMW-BiGRU-Self-attention-EBTC system using CMW transform ensures accurate sleep disorder classification. Despite these results, the CMW transform's adaptability and detailed time-frequency representation continue to outperform spectrograms and scalograms, particularly in capturing complex, non-stationary signal patterns essential for diagnosing sleep disorders.

TABLE V. SLEEP DISORDER CLASSIFICATION RESULTS FROM USING VARIOUS MACHINE-LEARNING CLASSIFIERS WITH THE HOLD-OUT VALIDATION STRATEGY

Classifier	Accuracy (%)						
	Insomnia	Narcolepsy	NFLE	PLM	RBD	SDB	Healthy
CMW-BiGRU-Self-attention-EBTC	99.70	97.60	95.40	94.50	96.50	98.30	94.10
LSTM	80.9	75.3	74.5	77.1	79.8	81.0	80.9
KNN+NN	78.6	73.5	71.9	74.8	77.4	79.2	78.6
Random Forest (RF)	92.5	87.0	85.1	88.2	90.3	91.2	92.5
RF+SVM	90.8	85.5	83.3	87.7	89.4	90.1	90.8
GRU-SVM	89.7	84.2	82.9	86.5	88.0	89.3	89.7
Decision Tree	85.4	80.6	79.2	81.9	83.7	85.1	85.4
AdaBoost	83.2	78.4	76.8	80.0	82.0	83.5	83.2

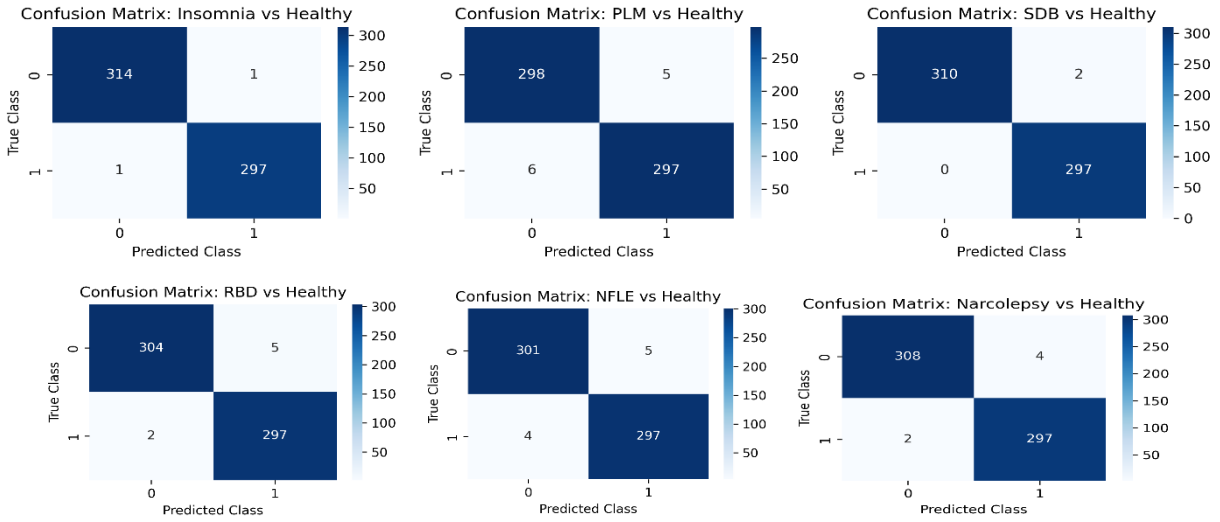


Fig. 6. Confusion metrics obtained for six sleep disorder prediction versus healthy individuals without signal transformation (normalization, bandpass filter and used 2D-scalogram image, complex morlet wavelet, stacking) step by proposed CMW-BiGRU-Self-attention-EBTC system.

TABLE VI. PERFORMANCE METRICS OBTAINED FOR SIX-CLASS BASED SLEEP DISORDER CLASSIFICATION WITHOUT SIGNAL TRANSFORMATION (NORMALIZATION, BANDPASS FILTER AND USED 2D-SPECTOGRAM IMAGE, COMPLEX MORLET WAVELET, STACKING) STEP BY PROPOSED CMW-BIGRU-SELF-ATTENTION-EBTC SYSTEM

Classes	ACC (%)	PR(%)	RC(%)	FS(%)
Insomnia	84.75	84.77	84.77	84.77
Narcolepsy	82.96	82.93	82.93	82.93
NFLE	81.09	81.11	81.11	81.11
PLM	80.33	80.42	80.42	80.42
RBD	82.03	82.03	82.03	82.03
SDB	83.56	83.62	83.62	83.62
Healthy	79.99	79.96	79.96	79.96

This study distinguishes healthy people from individuals with sleep problems with great accuracy, as described in the above paragraphs. Several variables were found and studied throughout testing, which are potential sources of error or misclassification based on the experimental results from Tables IV and V, as described below.

- Noise from EEG and EOG signal quality can decrease feature extraction precision. Signal fluctuation can be caused by electrode location, sleep movement, and physiological variations. When the signal-to-noise ratio is low, this fluctuation might cause the model to categorize epochs inconsistently.
- Despite the Use of cross-validation and regularization to reduce overfitting, the model may have inadvertently

learned patterns specific to the training set that do not generalize well to unknown data. This is a common challenge in machine learning that could potentially increase model error rates when confronted with new data.

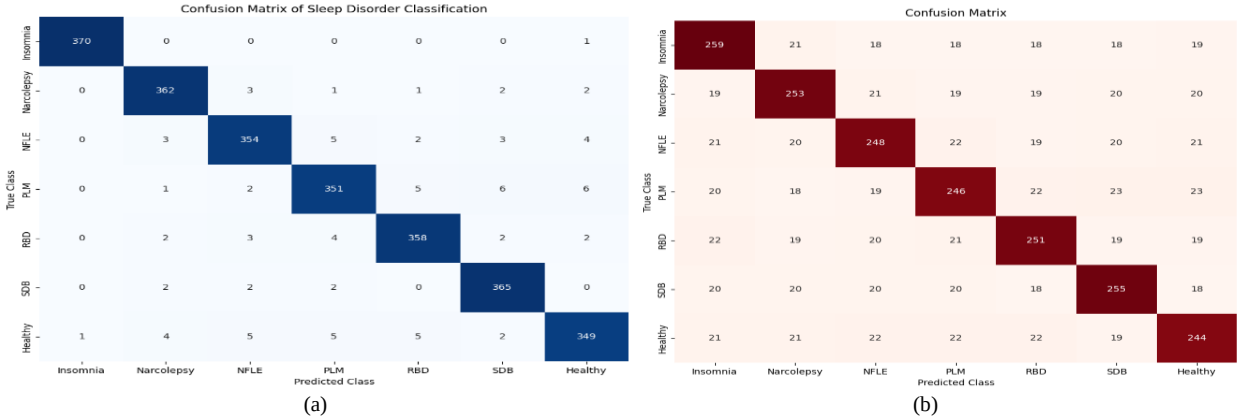
- This dataset is extensive; however, class imbalance is evident in SDB and PLM data. Class imbalance can bias the model to recognize the dominant class but not the minority class, reducing accuracy and increasing misclassification rates.
- Rare sleep disorders, including Narcolepsy and Nocturnal Frontal Lobe Epilepsy (NFLE), reduce training data. This constraint might hinder the model's understanding of these illnesses' complicated patterns, leading to reduced accuracy or increased misclassification rates.
- The proposed model uses Morlet wavelet transform and advanced machine learning methods to analyze combine EEG and EOG signals, requiring numerous computing layers. Complexity allows excellent accuracy, but it also raises the possibility of misclassifying signal artifacts or non-standard signal patterns as disorders.

To address these potential errors, we implemented a series of robust measures. We strengthened the signal processing capabilities with CMW transforms. Additionally, training dataset was expanded to encompass a wider range of patients and disorders.

Fig. 7–8 present the results obtained using the proposed system by changing the hyperparameters to detect six multi-class sleep stages. ReLU activation function is applied on GRU units as compared to Tanh as shown in Fig. 7. A similar performance has been achieved. This figure presents the receiver operating characteristic (ROC) curves for different sleep disorders, as well as for the 'Healthy' class, based on the crucial area under the curve (AUC) values provided. Each curve represents the trade-off between the actual positive rate (sensitivity) and the false positive rate (1-specificity) for a specific disorder classification. A higher AUC indicates better discrimination ability, with values closer to 1.0 representing superior classification performance. In this figure, the AUC values for each disorder are as follows: Insomnia (0.997), Narcolepsy (0.976), NFLE (0.954), PLM (0.945), RBD (0.965), SDB (0.983), and Healthy (0.941). These values are pivotal in reflecting the models' ability to distinguish between positive instances of each disorder and negative instances of other disorders or healthy individuals. AUC values near 1.0 suggest excellent classification performance, while lower values indicate room for improvement.

The confusion matrix in Fig. 8 provides a comprehensive overview of a CNN classifier's performance in discerning between various sleep disorders and healthy individuals. Each row corresponds to the true class, while each column represents the predicted class, offering insights into both correct classifications and misclassifications when figure (a) proposed system and (b) original BiLSTM and EBTC boosting tree. For instance, the classifier demonstrated its effectiveness by accurately identifying a significant number of instances for each disorder, such as Insomnia (370 instances correctly classified) and Narcolepsy (362 instances correctly classified). However, misclassifications were observed across different categories, indicating the model's limitations in certain scenarios. Notably, while the classifier performed well in identifying instances of Insomnia and Narcolepsy, a few instances of Insomnia (1 instance misclassified as Healthy) and Narcolepsy (2 instances misclassified as NFLE, one as PLM, one as RBD, two as SDB, and two as Healthy) were incorrectly classified.

Moreover, the confusion matrix highlighted misclassifications of healthy individuals, with some instances erroneously classified as various sleep disorders. Despite accurately identifying the majority of healthy individuals (349 instances correctly classified), a notable number of misclassifications occurred. For example, healthy individuals were mistakenly classified as Narcolepsy (4 instances misclassified), NFLE (5 instances misclassified), PLM (5 instances misclassified), RBD (5 instances misclassified), and SDB (2 instances misclassified). These misclassifications underscore the need to refine the classification model to enhance its accuracy and robustness, particularly distinguishing between healthy individuals and those with sleep disorders. By addressing these misclassifications and improving the classifier's ability to identify different sleep patterns accurately, it can strengthen diagnostic tools for sleep disorders and enhance patient care through more precise and timely interventions.



(a) (b)

Fig. 7. Confusion matrices for sleep stage and sleep disorder detection using (a) proposed system and (b) original BiLSTM and EBTC boosting tree.

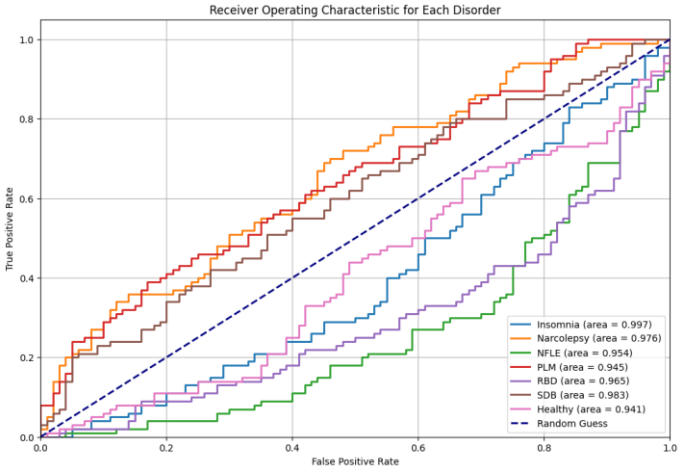


Fig. 8. A separate experiment for using 2400 samples of each sleep-disorder class to show AUC curves of accuracy with respect to each class of sleep disorder.

The AUC curve for our proposed system using the architecture (CMW-BiGRU-Self-attention-EBTC) is shown in Fig. 8. This is an integral part of our research. This system, which uses a large dataset with 24,000 samples of each class instead of 26,000 samples, gives a full picture of how well the proposed classifier works at finding sleep disorders. More importantly, it paves the way for evaluating and refining the classification model, thereby enhancing the accuracy and reliability of our system. This, in turn, underscores the potential impact of our research on clinical applications, making it a significant contribution to the fields of sleep medicine and machine learning.

C. State-of-the-Art Comparisons

Sleep disorder diagnosis through automated systems has seen significant advancements, with various models employed in the past by combinations of convolutional neural networks (CNNs), recurrent neural networks (RNNs), gated recurrent units (GRUs), and random forest (RF) algorithms. These models, such as Shao-CNN-GRU [17], Fan-CNN-RNN [19], Santaji-RF [21], Morokuma-CNN [30], and Satapathy-1DCNN [33], have leveraged electrooculography (EOG) and electroencephalography (EEG) signals to varying degrees of success. These systems were selected to perform State-of-the-art (SOTA) comparisons because they were quickly implemented. While they have shown promising results, challenges such as overfitting, limited sleep class detection, and computational inefficiency have persisted, as mentioned in Table I. Accuracies have varied widely, from as low as 76.3% to as high as 98.84%, with each model demonstrating its strengths in handling complex signal data for sleep stage classification or disorder diagnosis and revealing significant limitations hindering their broader application and effectiveness.

In this context, the proposed CMW-BiGRUSelf-EBTC model emerges as a noteworthy evolution that adeptly addresses these challenges. By integrating a complex metro-wavelet transformation with a bidirectional gated recurrent unit that includes a self-attention layer and coupling it with an ensemble-bagged tree classifier, this model simplifies the signal processing pipeline and enhances the accuracy and efficiency of sleep disorder diagnosis. With remarkable success rates across various sleep disorders and an overall classification accuracy of 96% alongside an AUC of 0.96, this approach outshines its predecessors. It achieves this by effectively mitigating overfitting and reducing computational demands, thereby marking a significant leap forward in developing noninvasive, automated systems for accurate and efficient sleep disorder diagnosis. Specific hyperparameters are described in Table VII to expand the "Details" column to include examples of hyperparameters that might be tuned for RNNs (Recurrent Neural Networks) and the considerations for architectural configurations.

Table VIII shows a comparative analysis of various state-of-the-art (SOTA) systems for sleep disorder detection. The systems compared include Shao-CNN-GRU, Fan-CNN-RNN, Santaji-RF, Morokuma-CNN, Satapathy-1DCNN, and the proposed system, CMW-BiGRUSelf-EBTC.

TABLE VII. STATE-OF-THE-ART COMPARISONS (SOTA) HYPER-PARAMETERS SITTING

Component	Description	Expected Impact
Baseline (Full Model)	CMW-BiGRUSelf-EBTC with all features	Establish baseline performance
Without Bidirectional RNN Cells	Use unidirectional GRU cells	Assess the importance of capturing temporal dependencies in both directions
Without the Attention Layer	Remove the self-attention layer	Evaluate the impact of focusing on significant parts of the data
Without Bagging Ensemble	Use a single classifier instead of an ensemble	Determine the contribution of ensemble methods to robustness and accuracy
Without Data Augmentation	Train without augmented data	Examine the role of data diversity in model generalization

TABLE VIII. SOTA WITH RESPECT TO PROPOSED SYSTEM

SOTA Systems	Acc (%)	SE(%)	SP(%)	AUC	FS (%)
Shao-CNN-GRU [17]	88.25	86.40	88.15	0.87	88.32
Fan-CNN-RNN [19]	78.44	77.12	78.42	0.77	78.46
Santaji-RF [21]	75.47	73.70	74.10	0.75	76.10
Morokuma-CNNc[30]	86.50	85.20	88.45	0.86	88.00
Satapathy-1DCNN [33]	80.10	82.66	83.25	0.83	82.12
CMW-BiGRUSelf-EBTC	96.59	97.30	95.20	0.966	96.0

TABLE IX. ABLATION STUDY PARAMETERS

Step	Description	Details	Example Hyperparameters
Define the Parameter Space	For each model, identify and list all hyperparameters and architectural configurations that will be explored.	Includes learning rates, batch sizes, number of layers, types of RNN cells (e.g., LSTM, GRU), etc.	Learning rate: [0.001], Batch size: [64], RNN type: [LSTM, GRU], Number of layers: [3], Dropout rate: [0.25]
Apply Nested CV	Implement nested cross-validation (CV) with an outer loop for performance assessment and an inner loop for hyperparameter tuning.	Ensures unbiased evaluation and that hyperparameter tuning does not influence the test set.	Outer loop: 5 folds, Inner loop: 3 folds
Optimize Hyperparameters	Use grid search within the inner loop of the nested CV to find the optimal set of hyperparameters and architectural configurations for each model.	Systematically explores multiple combinations of parameters to find the best setup for each model.	Grid search across all combinations of the example hyperparameters listed.
Evaluate and compare	After tuning, evaluate each model's performance on the test set of the outer CV loop to ensure fairness in comparison.	Performance metrics are based on unseen data, providing a reliable basis for comparison.	Accuracy, F1 Score, AUC, Sensitivity, Specificity
Statistical Testing	Employ statistical tests to determine if the differences in performance between models are statistically significant.	Strengthens the validity of the comparison by confirming whether observed performance differences are meaningful.	Wilcoxon signed-rank tests comparing model performances.

The recommended CMW-BiGRUSelf-EBTC system, with its superior performance metrics, outperforms previous models. Its accuracy of 96.59%, SE of 97.30%, SP of 95.20%, AUC of 0.966, and F1-score of 96.0% are a testament to its effectiveness. The system's ability to effectively diagnose sleep problems is a significant advancement, demonstrating its profound impact on sleep problem detection. This comparison underscores the importance of advanced machine learning approaches like bidirectional RNNs, self-attention layers, and ensemble methods in the field of sleep disorder diagnosis.

D. Ablation Study

Ablation studies on the proposed Complex-Morlet-wavelet Representation using a bidirectional gated recurrent unit with a Self-attention Layer and an ensemble-bagged tree classifier (CMW-BiGRUSelf-EBTC) system involve systematically removing or replacing model components to understand their performance contributions. This study highlights the most critical elements of the suggested sleep problem detection and identification technique—ablation research structure. Tables VIII and IX provide an overview of the CMW-BiGRUSelf-EBTC system ablation research and a full analysis of each ablation component.

These tables summarize the ablation study's setup and findings. They also demonstrate how each CMW-BiGRUSelf-EBTC component detected and diagnosed numerous sleep problems. The ablation research for the planned CMW-BiGRUSelf-EBTC system employing EEG and EOG data to discover and diagnose sleep problems shows how the elements work together to make it operate successfully. Initial evaluation of the system yields remarkable metrics: accuracy of 96.59%, sensitivity of 97.30%, specificity of 95.20%, AUC and F-score of 0.966 and 96.0%, respectively. This complete performance shows the system's resilience and accuracy in diagnosing sleep problems.

As the study progresses through its phases, removing critical system features one by one, a clear picture of their contributions emerges. All metrics go down a lot when there are no bidirectional RNN cells. This shows the importance of capturing temporal dependencies in the signal data for correct disorder recognition. In the same way, getting rid of the attention layer lowers performance metrics, showing how important it is for helping the model focus on essential parts of the complex signal data. The system's performance worsens when the bagging ensemble method and data augmentation are removed. This shows how important they were for making the model more robust and able to generalize across different data representations. Each component's removal delineates a stepwise decrease in the system's effectiveness, underlining the synergistic effect of these elements in achieving the CMW-BiGRUSelf-EBTC system's state-of-the-art performance.

V. CONCLUSION

The EOG and EEG data were used for automated sleep-disorder identification, making this study a notable advancement in sleep medicine. The proposed method for detecting sleep disorders yielded highly accurate and efficient results as it integrated advanced signal-processing techniques with powerful machine-learning models. This approach was also designed to be patient friendly. Thus, this study not only enhances the

scientific comprehension of sleep health but also holds the potential to considerably improve the quality of life of individuals with sleep disorders. The proposed method effectively classifies healthy individuals from those with different sleep disorders. It achieved a remarkably high accuracy of 99.7% for insomnia, 97.6% for narcolepsy, 95.4% for NFLE, 94.5% for PLM, 96.5% for RBD, 98.3% for SDB, and 94.1% for healthy individuals. The model's relevance and precision can be enhanced across many scenarios by establishing a confidential database for subsequent experimentation.

A key priority is ensuring the wide-ranging appropriateness and efficacy of a proposed technique for sleep health monitoring and diagnosis across diverse demographic groups. The dataset was expanded to include additional ages, genders, ethnicities, and geographical origins to achieve this. This expanded dataset will better capture population-specific sleep patterns and problems, enhancing the model's generalizability. To understand how cultural influences impact sleep, it will employ cross-cultural validation and adaptive algorithms for tailored diagnosis. However, medical specialists from diverse demographics must refine the model to maintain its clinical relevance and responsiveness to the vast range of sleep problems.

Ethical and inclusive research and design practices emphasize privacy, permission, and data protection. By making this technology inexpensive and accessible across demographics, it promotes healthcare equity. Patients and healthcare providers must monitor and offer feedback post-deployment. The system will be modified and adjusted based on real-world use and feedback to keep it valuable and relevant for diagnosing and monitoring sleep problems in varied worldwide populations.

Data Availability Statement: This study used the PhysioNet CAP Sleep database of the Sleep Disorders Center of the Ospedale Maggiore of Parma, Italy, as downloaded via physionet.org from https://physionet.org/content/capslpdb/1.0.0/ (accessed on March 5, 2022).

ACKNOWLEDGMENT

The PhysioNet CAP Sleep database of the Sleep Disorders Center of the Ospedale Maggiore of Parma, Italy, was used via physionet.org.

REFERENCES

[1] Xu, S.; Faust, O.; Seoni, S.; Chakraborty, S.; Barua, P.D.; Loh, H.W.; Elphick, H.; Molinari, F.; Acharya, U.R. A review of automated sleep disorder detection. Comput. Biol. Med. 2022, 150, 106100. DOI:10.1016/j.compbiomed.2022.106100.

[2] Loh, H.W.; Ooi, C.P.; Vicnesh, J.; Oh, S.L.; Faust, O.; Gertych, A.; Acharya, U.R. Automated detection of sleep stages using deep learning techniques: A systematic review of the last decade (2010–2020). Appl. Sci. 2020, 10, 8963. DOI:10.3390/app10248963.

[3] Rim, B.; Sung, N.J.; Min, S.; Hong, M. Deep learning in physiological signal data: A survey. Sensors (Basel). 2020, 20, 969. DOI:10.3390/s20040969.

[4] Van Der Donckt, J.; Van Der Donckt, J.; Deprost, E.; Vandenbussche, N.; Rademaker, M.; Vandewiele, G.; Van Hoecke, S. Do not sleep on traditional machine learning. Biomed. Signal Process. Control. 2023, 81, 104429. DOI:10.1016/j.bspc.2022.104429.

[5] Fatimah, B.; Singhal, A.; Singh, P. A multi-modal assessment of sleep stages using adaptive Fourier decomposition and machine learning. Comput. Biol. Med. 2022, 148, 105877. DOI:10.1016/j.compbiomed.2022.105877.

[6] Faust, O.; Hagiwara, Y.; Hong, T.J.; Lih, O.S.; Acharya, U.R. Deep learning for healthcare applications based on physiological signals: a review. Comput. Methods Programs Biomed. 2018, 161, 1–13. DOI:10.1016/j.cmpb.2018.04.005.

[7] Faust, O.; Razaghi, H.; Barika, R.; Ciaccio, E.J.; Acharya, U.R. A review of automated sleep stage scoring based on physiological signals for the new millennia. Comput. Methods Programs Biomed. 2019, 176, 81–91. DOI:10.1016/j.cmpb.2019.04.032.

[8] Aboalayon, K.A.I.; Faezipour, M.; Almuhammadi, W.S.; Moslehpour, S. Sleep stage classification using EEG signal analysis: a comprehensive survey and new investigation. Entropy. 2016, 18, 272. DOI:10.3390/e18090272.

[9] Qian, X.; Qiu, Y.; He, Q.; Lu, Y.; Lin, H.; Xu, F.; Zhu, F.; Liu, Z.; Li, X.; Cao, Y.; Shuai, J. A review of methods for sleep arousal detection using polysomnographic signals. Brain Sci. 2021, 11, 1274. DOI:10.3390/brainsci11101274.

[10] Moridian, P.; Shoeibi, A.; Khodatars, M.; Jafari, M.; Pachori, R.B.; Khadem, A.; Alizadehsani, R.; Ling, S.H. Automatic diagnosis of sleep apnea from biomedical signals using artificial intelligence techniques: methods, challenges, and future works. WIREs Data Min. & Knowl. 2022, 12, e1478. DOI:10.1002/widm.1478.

[11] Sharma, M.; Yadav, A.; Tiwari, J.; Karabatak, M.; Yildirim, O.; Acharya, U.R. An automated wavelet-based sleep scoring model using eeg, emg, and eog signals with more than 8000 subjects. Int. J. Environ. Res. Public Health. 2022, 19, 7176. DOI:10.3390/ijerph19127176.

[12] Jarchi, D.; Andreu-Perez, J.; Kiani, M.; Vysata, O.; Kuchynka, J.; Prochazka, A.; Sanei, S. Recognition of patient groups with sleep related disorders using bio-signal processing and deep learning. Sensors (Basel). 2020, 20, 2594. DOI:10.3390/s20092594.

[13] Sharma, M.; Darji, J.; Thakrar, M.; Acharya, U.R. Automated identification of sleep disorders using wavelet-based features extracted from electrooculogram and electromyogram signals. Comput. Biol. Med. 2022, 143, 105224. DOI:10.1016/j.compbiomed.2022.105224.

[14] Sekkal, R.N.; Bereksi-Reguig, F.; Ruiz-Fernandez, D.; Dib, N.; Sekkal, S. Automatic sleep stage classification: From classical machine learning methods to deep learning. Biomed. Signal Process. Control. 2022, 77, 103751. DOI:10.1016/j.bspc.2022.103751.

[15] Sharma, M.; Tiwari, J.; Patel, V.; Acharya, U.R. Automated identification of sleep disorder types using triplet half-band filter and ensemble machine learning techniques with eeg signals. Electronics. 2021, 10, 1531. DOI:10.3390/electronics10131531.

[16] Rahman, M.M.; Bhuiyan, M.I.H.; Hassan, A.R. Sleep stage classification using single-channel EOG. Comput. Biol. Med. 2018, 102, 211–220. DOI:10.1016/j.compbiomed.2018.08.022.

[17] Shao, X.; Soo Kim, C. A hybrid deep learning scheme for multi-channel sleep stage classification. Comput. Mater. Continua. 2022, 71, 889–905. DOI:10.32604/cmc.2022.021830.

[18] Satapathy, S.K.; Kondaveeti, HariKishan; Loganathan, D.; Sharathkumar, S. A machine learning model for automated classification of sleep stages using polysomnography signals. In Machine Vision and Augmented Intelligence—Theory and Applications: Select. Proceedings of MAI 2021; Springer: Berlin, 2021, pp. 209–222. DOI:10.1007/978-981-16-5078-9_18.

[19] Fan, J.; Sun, C.; Long, M.; Chen, C.; Chen, W. Eognet: A novel deep learning model for sleep stage classification based on single-channel eog signal. Front. Neurosci. 2021, 15, 573194. DOI:10.3389/fnins.2021.573194.

[20] Chambon, S.; Galtier, M.N.; Arnal, P.J.; Wainrib, G.; Gramfort, A. A deep learning architecture for temporal sleep stage classification using multivariate and multimodal time series. IEEE Trans. Neural Syst. Rehabil. Eng. 2018, 26, 758–769. DOI:10.1109/TNSRE.2018.2813138.

[21] Santaji, S.; Desai, V. Analysis of EEG signal to classify sleep stages using machine learning. Sleep Vigilance. 2020, 4, 145–152. DOI:10.1007/s41782-020-00101-9.

[22] Mousavi, S.; Afghah, F.; Acharya, U.R. SleepEEGNet: automated sleep stage scoring with sequence to sequence deep learning approach. PLOS ONE. 2019, 14, e0216456. DOI:10.1371/journal.pone.0216456.

[23] Somaskandhan, P.; Leppänen, T.; Terrill, P.I.; Sigurdardottir, S.; Arnardottir, E.S.; Ólafsdóttir, K.A.; Serwatko, M.; Sigurðardóttir, S.Þ.; Clausen, M.; Töyräs, J.; Korkalainen, H. Deep learning-based algorithm accurately classifies sleep stages in preadolescent children with sleep-disordered breathing symptoms and age-matched controls. Front. Neurol. 2023, 14, 1162998. DOI:10.3389/fneur.2023.1162998.

[24] Yan, R.; Li, F.; Zhou, D.D.; Ristaniemi, T.; Cong, F. Automatic sleep scoring: A deep learning architecture for multi-modality time series. J. Neurosci. Methods. 2021, 348, 108971. DOI:10.1016/j.jneumeth.2020.108971.

[25] Satapathy, S.K.; Loganathan, D. Multimodal multiclass machine learning model for automated sleep staging based on time series data. SN Comput. Sci. 2022, 3, 276. DOI:10.1007/s42979-022-01156-3.

[26] Malafeev, A.; Laptev, D.; Bauer, S.; Omlin, X.; Wierzbicka, A.; Wichniak, A.; Jernajczyk, W.; Riener, R.; Buhmann, J.; Achermann, P. Automatic human sleep stage scoring using deep neural networks. Front. Neurosci. 2018, 12, 781. DOI:10.3389/fnins.2018.00781.

[27] Ji, X.; Li, Y.; Wen, P. 3DSleepNet: A multi-channel bio-signal based sleep stages classification method using deep learning. IEEE Trans. Neural Syst. Rehabil. Eng. 2023, 31, 3513–3523. DOI:10.1109/TNSRE.2023.3309542.

[28] Cooray, N.; Andreotti, F.; Lo, C.; Symmonds, M.; Hu, M.T.M.; De Vos, M. Detection of REM sleep behaviour disorder by automated polysomnography analysis. Clin. Neurophysiol. 2019, 130, 505–514. DOI:10.1016/j.clinph.2019.01.011.

[29] Li, Y.; Xu, Z.; Zhang, Y.; Cao, Z.; Chen, H. Automatic sleep stage classification based on a two-channel electrooculogram and one-channel electromyogram. Physiol. Meas. 2022, 43, p.07NT02. DOI:10.1088/1361-6579/ac6bdb.

[30] Morokuma, S.; Hayashi, T.; Kanegae, M.; Mizukami, Y.; Asano, S.; Kimura, I.; Tateizumi, Y.; Ueno, H.; Ikeda, S.; Niizeki, K. Deep learning-based sleep stage classification with cardiorespiratory and body movement activities in individuals with suspected sleep disorders. Sci. Rep. 2023, 13, 17730. DOI:10.1038/s41598-023-45020-7.

[31] Sokolovsky, M.; Guerrero, F.; Paisarnsrisomsuk, S.; Ruiz, C.; Alvarez, S.A. Deep learning for automated feature discovery and classification of sleep stages. IEEE ACM Trans. Comp. Biol. Bioinform. 2020, 17, 1835–1845. DOI:10.1109/TCBB.2019.2912955.

[32] Bozkurt, F.; Uçar, M.K.; Bilgin, C.; Zengin, A. Sleep–wake stage detection with single channel ECG and hybrid machine learning model in patients with obstructive sleep apnea. Phys. Eng. Sci. Med. 2021, 44, 63–77. DOI:10.1007/s13246-020-00953-5.

[33] Satapathy, S.K.; Loganathan, D. Automated classification of multi-class sleep stages classification using polysomnography signals: a nine- layer 1D-convolution neural network approach. Multimedia Tool. Appl. 2023, 82, 8049–8091. DOI:10.1007/s11042-022-13195-2.

[34] Zhao, R.; Xia, Y.; Wang, Q. Dual-modal and multi-scale deep neural networks for sleep staging using EEG and ECG signals. Biomed. Signal Process. Control. 2021, 66, 102455. DOI:10.1016/j.bspc.2021.102455.

[35] Casciola, A.A.; Carlucci, S.K.; Kent, B.A.; Punch, A.M.; Muszynski, M.A.; Zhou, D.; Kazemi, A.; Mirian, M.S.; Valerio, J.; McKeown, M.J.; Nygaard, H.B. A deep learning strategy for automatic sleep staging based on two-channel EEG headband data. Sensors (Basel). 2021, 21, 3316. DOI:10.3390/s21103316.

[36] Jia, Z.; Cai, X.; Zheng, G.; Wang, J.; Lin, Y. SleepPrintNet: A multivariate multimodal neural network based on physiological time-series for automatic sleep staging. IEEE Trans. Artif. Intell. 2020, 1, 248–257. DOI:10.1109/TAI.2021.3060350.

[37] Kemp, B.; Zwinderman, A.H.; Tuk, B.; Kamphuisen, H.A.; Oberyé, J.J.L. Analysis of a sleep-dependent neuronal feedback loop: the slow-wave microcontinuity of the EEG. IEEE Trans. Biomed Eng. 2000, 47, 1185–1194. DOI:10.1109/10.867928.

[38] Goldberger, A.L.; Amaral, L.A.; Glass, L.; Hausdorff, J.M.; Ivanov, P.C.; Mark, R.G.; Mietus, J.E.; Moody, G.B.; Peng, C.K.; Stanley, H.E. PhysioBank, PhysioToolkit, and Physionet: components of a new research resource for complex physiologic signals. Circulation. 2000, 101, E215–E220. DOI:10.1161/01.cir.101.23.e215.

[39] Zhu, Q.; Jiang, X.; Ye, R. Sentiment analysis of review text based on BiGRU-attention and hybrid CNN. IEEE Access. 2021, 9, 149077–149088. DOI:10.1109/ACCESS.2021.3118537.

[40] Zhang, D.; Hong, M.; Zou, L.; Han, F.; He, F.; Tu, Z.; Ren, Y. Attention pooling-based bidirectional gated recurrent units model for sentimental classification. Int. J. Comput. Intell. Syst. 2019, 12, 723–732. DOI:10.2991/ijcis.d.190710.001.

[41] Ensemble methods, bagging, boosting and stacking, URL. Available online: https://towardsdatascience.com/ensemble-methods-bagging-boosting-and-stacking-c9214a10a205 (Accessed 2 January 2023).

Digital Landscape Architecture Design Combining 3D Image Reconstruction Technology

Chen Chen
School of Environmental Art and Design, Wuxi Vocational Institute of Arts & Technology, Yixing, Wuxi, China

***Abstract*—To achieve better digital landscape design and visual presentation effects, this study proposes a digital landscape design method based on improved 3D image reconstruction technology. Firstly, a precise point cloud registration algorithm combining normal distribution transformation and Trimmed iterative nearest point algorithm is proposed. A color texture method for 3D models is designed in terms of 3D reconstruction, and a visual scene, 3D reconstruction method based on RGBD data is constructed. Secondly, knowledge networks are introduced to assist in the intelligent generation and planning of plant communities in urban landscape scenes. The knowledge network established through the plant database integrates the principles of landscape design and optimizes the layout of landscape plants in urban parks. The running speed and accuracy of research algorithms were superior to traditional methods, especially in terms of registration performance. Compared to the other two algorithms, the registration time of the research algorithm was reduced by 2%, and the errors were reduced by 71.4% and 87.5%, respectively. The panoramic quality of research methods fluctuated within a small range of 0.8 or above, while traditional methods exhibited instability and lower quality. The landscape design generated by research methods was more aesthetically pleasing and harmonious with the actual landscape in terms of plant selection and layout. The proposed method follows the principles of eco-friendly design and demonstrates significant potential for application in the field of urban landscape design.**

Keywords*—3D image reconstruction; PSO; gardens; RGBD; digital landscape

I. INTRODUCTION

Landscape refers to the landscape environment created artificially within a certain area. Through the layout and combination of plants, land, water, architecture, and other elements, as well as the artistic treatment and transformation of natural and cultural elements, it creates a living, working, or leisure place with beauty, comfort, and functionality [1-2]. China's landscape architecture has a rich historical background, and modern landscape architecture has a development history of over 60 years in China [3]. The traditional garden aesthetics advocate harmonious coexistence with nature, reflecting profound humanistic ideas. Research has shown that exposure to nature and green spaces is beneficial for human psychological and physical health. Landscape design has improved urban air quality and provided leisure spaces beneficial to physical and mental health [4]. However, in the face of today's diverse needs, traditional design methods are unable to meet the requirements of rapid urban development and high standard, scientific urban spatial planning.

Digitization has brought challenges to the development of landscape design, but also new opportunities [5]. In response to the limitations of existing landscape evolution models, scholars such as Steer have proposed an innovative modeling technique. This technology achieved one-dimensional sorting and effective propagation of terrain change information by applying linear flow power equations and utilizing directed acyclic graphs. This modeling method could accurately simulate the dynamic changes of the landscape even under conditions of uneven or varying rise and fall rates [6]. In landscape design, the use of sound elements was limited to noise management. Therefore, Luo Ma L et al. explored the impact of sound as a design element on the immersion of architecture and planning in virtual reality environments. This study constructed a virtual reality scene based on field recording to simulate the sound environment. It emphasized the role of audible sound in enhancing landscape design [7]. Wang et al. proposed a landscape water flow design method based on the principles of open channel hydraulics, which combines hydrology and hydraulics as well as landscape processes of infiltration and evapotranspiration. This method could accurately generate flow profiles between depressions in the landscape [8]. The above research indicates that digitalization in the field of landscape architecture mainly focuses on technological application, and in-depth research on design methods and logic is still in its early stages and has not been widely applied in practice. The digitization of traditional garden landscapes still faces challenges, including a lack of unified guidance frameworks and government level norms. In addition, the high cost of technological equipment and the lack of interdisciplinary knowledge have constrained the development of professional talents and the growth of digital design demand.

Three dimensional reconstruction (3DR) is the process of restoring the 3D information of objects in a visual scene through 2D image data or 3D point cloud data (3D-PCD). One type of technology is based on 2D image data, using multiple captured image sequences to construct a color 3D model through feature point matching technology. Another type is based on 3D-PCD, which processes data from multiple collection points and reproduces physical objects through noise reduction, integration, and simulation processes [9]. Researchers such as Xu proposed the use of Hermite radial basis function algorithm combined with RGBD camera, and introduced curvature estimation and confidence score methods to address the common problems of cumulative error and reconstruction distortion in the field of 3D model reconstruction. This method effectively reduced the impact of noise and optimized the results of 3DR [10]. Zhang et al. proposed a 3DR method for motion blurred images using deep learning. This algorithm combined

bilateral filtering denoising theory with Wasserstein generated adversarial networks to remove motion blur, and used the deblurred images generated by this algorithm for 3D reconstruction. This algorithm has been proven to have better deblurring effects and higher efficiency than other representative algorithms [11].

Although many techniques and methods have been proposed in the existing landscape design research, there are still some deficiencies. First, the modeling technique of Steer et al., linear flow power equation models, may lack flexibility in the dissemination of terrain change information. The sound element design method studied by Luo Ma et al. is only applied in noise management, and its effect on improving landscape immersion is limited. The flow design method proposed by Wang et al. is difficult to accurately simulate complex landscape flow in practical applications. In addition, most of these researches focus on the technical application, and the in-depth research on the design method and logic is still insufficient, and the application in practice is limited. In this paper, an accurate point cloud registration algorithm combining normal distribution transformation and trimmed iteration nearest point algorithm is proposed. The color texture method of 3D model is designed from the perspective of 3D reconstruction, and a 3D reconstruction method of visual scene based on RGBD data is constructed. Knowledge network is introduced to assist the intelligent generation and planning of plant communities in urban landscape scenes. The knowledge network established by plant database integrates the principles of landscape design and optimizes the layout of landscape plants in urban parks. It only enhances the aesthetics of the plant selection and layout, is more in tune with the actual landscape, and also follows the principles of eco-friendly design. This study aims to overcome the limitations of traditional design methods, optimize landscape design process by using digital technology and 3D image reconstruction technology, and improve design quality and efficiency. Enhance the aesthetics and coordination of the design by introducing intelligent plant community generation and planning.

This study is divided into five sections. Section I outlines the importance of landscape design and its benefits to human health, introduces the application of digital technology in landscape design and the criticality of 3D reconstruction technology. In Section II, a 3D reconstruction technique based on RGBD data (3DIR) and an improved Poisson surface reconstruction algorithm are described in detail. A method to generate 3D color model based on RGB information is proposed, which is optimized globally by the inner heuristic algorithm and the outer particle swarm optimization algorithm. Section III verifies the effectiveness of the algorithm through experiments, and shows its advantages in registration accuracy and running speed. Section IV and Section V emphasize the importance of DLD methods in improving design efficiency and visualization level, and proposes future research directions for combining deep learning algorithms to improve existing algorithms.

II. Methods and Materials

A. 3DR Method for Visual Scenes Based on RGBD Data

3DIR technology plays a crucial role in landscape design. To obtain a complete 3D scene, it is necessary to address the issue of limited viewing angles caused by a single Azure Kinect DK camera [12]. For this purpose, point cloud registration (PCR) technology can be used to integrate multiple scanned point cloud data. For 3DIR in landscape design, voxel grid downsampling is a more suitable choice, as it can effectively reduce data volume while preserving the overall geometric structure of the scene [13]. For the rough position between two frames of point cloud data, this study adopts a method based on point features to describe local features. The sample consistency initial alignment (SAC-IA) algorithm is a coarse registration method based on point cloud feature description. The algorithm steps are as follows: The first is to select some sampling points from point clouds P and Q, and use the feature description algorithm to perform feature analysis on these points, ensuring that the distance between sampling points exceeds the predetermined minimum value. Afterwards, to find the points in Q that match the features of P, forming a matching point pair P_c. Calculating the rotation and translation matrices of P_c through singular value decomposition (SVD), and then calculating the distance error between registered points. The registration performance is evaluated through the Huber loss function, and the calculation process is Eq. (1).

$$\begin{cases} H\left(l_i\right)=\begin{cases}\dfrac{1}{2}l_i, \|l_i\|<\delta_i \\ \dfrac{1}{2}\delta_i\left(2\|l_i\|-\delta_i\right), \|l_i\|>\delta_i\end{cases} \\ H_i=\sum_{i=1}^{n} H\|l_i\| \end{cases} \tag{1}$$

In Eq. (1), $H_i=\sum_{i=1}^{n} H\|l_i\|$. δ represents a pre-set value. l_i represents the distance difference after the rotation and translation transformation of the corresponding point in Group i. By repeatedly matching sampling points and iteratively calculating, the rigid body transformation matrix that causes the minimum distance error is found. The optimal matrix represents the preliminary PCR results. Due to the SAC-IA algorithm reducing the point cloud resolution before feature calculation, some feature information is lost, making it more suitable for rough registration. Afterwards, to perform precise registration on the point cloud. Iterative closest point (ICP) algorithm and its improved algorithms are currently the most classic point cloud accurate registration algorithms. The Trimmed ICP algorithm is an improved algorithm based on ICP that uses minimum truncated multiplication to eliminate excessive distances [14]. The step is to first determine the nearest neighbor point q_i in the baseline point cloud Q for each point p_i in point cloud P, and calculate its distance squared d_i^2. These distances are sorted in ascending order, and the first N points are selected to form a subset P_s, then the sum of the squared distances S of this subset is calculated. The number of reserved points is determined based on the number N_p of the point cloud to be registered. The process is given in Eq. (2).

$$N=kN_p \tag{2}$$

Using the SVD method to solve the set of reserved points obtained from the calculation, the point cloud rotation translation matrix is obtained, as shown in Eq. (3).

$$(R,T) = \arg\min \sum_{i=1}^{N} \left\| q_i - (Rp_i + T) \right\| \quad (3)$$

After applying the rotation matrix to transform the registration point cloud P, it is necessary to re estimate the corresponding point relationship with the reference point cloud Q. The iterative process continues until the number of iterations exceeds the preset limit, and the mean square error $\frac{S}{N}$ value of the point pairs drops below the set threshold, or when the similarity between the results of two consecutive iterations is extremely high, the final registration result can be obtained. The normal distribution transform (NDT) algorithm is a high registration accuracy and fast computational speed PCR [15-16]. To achieve better registration results, it is combined with the Trimmed ICP algorithm. The improved PCR steps are shown in Fig. 1.

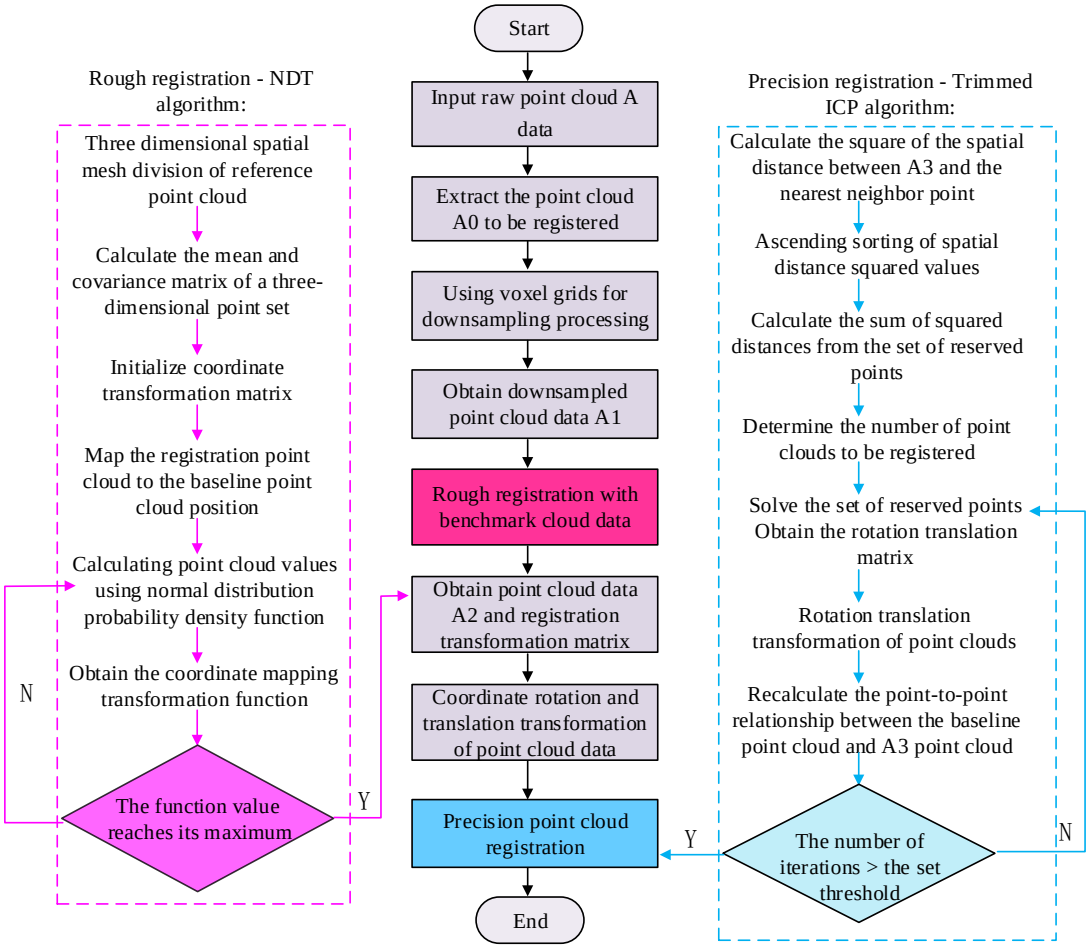


Fig. 1. Point cloud registration process.

In Fig. 1, PCR first speeds up registration by reducing the amount of data in point cloud P, and uses voxel grid downsampling algorithm to obtain simplified point cloud $\hat{P}_0$. Preliminary to use the NDT algorithm to complete rough registration of $\hat{P}_0$ and benchmark point cloud Q, obtaining transformation matrix M_1 and registered point cloud $\hat{P}_1$. Applying M_1 to transform the original point cloud P again to obtain a rough alignment point cloud P_1. Finally, the Trimmed ICP algorithm is used to finely adjust the registration of $\hat{P}_1$ and Q, obtaining the final accurate registration point cloud $\hat{P}_2$ and its corresponding rotation translation matrix M_2. Point cloud 3DR is the process of generating a 3D surface mesh model based on the target point cloud obtained through algorithmic processing. 3D surface reconstruction techniques can be divided into two types: triangulation based mesh and surface fitting. The surface fitting algorithm generates a model close to the solid surface through function solving, which is insensitive to noise. The advantage of Poisson's algorithm lies in its ability to resist noise, producing clear contours and strong closure of the model. The steps for reconstructing Poisson's surface are shown in Fig. 2.

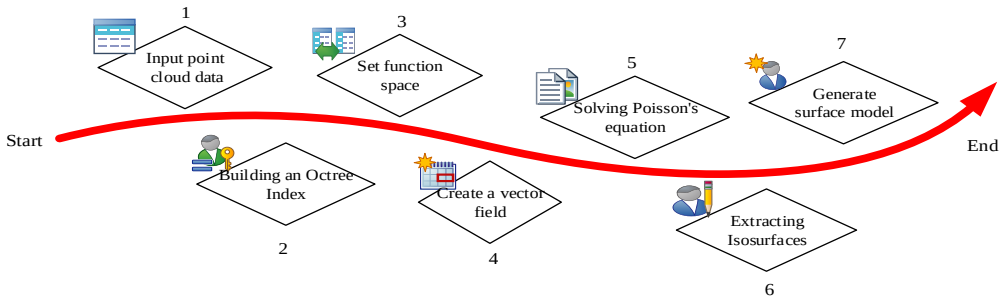


Fig. 2. Poisson surface reconstruction steps.

The original Poisson reconstruction algorithm is susceptible to errors, resulting in hidden function shifts. Someone proposed a masked Poisson surface reconstruction algorithm based on the original Poisson reconstruction algorithm. The improved method introduces a penalty function to limit offset and assigns weight ω to each point in point cloud P. The improved Poisson equation $E(\chi)$ can be represented by Eq. (4).

$$E(\chi)=\int\left\|\vec{V}(q)-\nabla\chi(q)\right\|^2 dp+\frac{\alpha S(P)}{\sum_{q\in P}\omega(q)}\sum_{q\in P}\omega(q)\chi^2(q) \quad (4)$$

In Eq. (4), $\vec{V}(q)$ represents the vector field. $\chi(q)$ represents the indicator function. $\omega(q)$ represents the weight of the sample, with a value of 1. α represents the equilibrium coefficient. $S(P)$ represents the surface area after reconstruction. To minimize Eq. (4) and define an operator to represent it as a masked Poisson equation. The calculation process is Eq. (5).

$$(\Delta-\alpha_s\oplus)\chi=\nabla\cdot\vec{V} \quad (5)$$

To solve the Poisson equation, the algorithm uses discretization methods to construct a linear equation system $Ax=b$ and apply zero constraint conditions at the sample points, and solves the coefficients to obtain matrix A. The cascaded multigrid method is used to optimize the solution process, adjusting weights at different depths d to maintain surface morphology and refine estimation. The scale invariance of the reconstruction effect is achieved by scaling the point set, adjusting functions $E_V\left(\tilde{\chi}\right)$ and $E_{\tilde{\omega},\tilde{p}}\left(\tilde{\chi}\right)$, and redefining weight values $\tilde{\omega}(p)$. The new set of points can be represented as $\left(\tilde{\omega},\tilde{P}\right)$, and the function is Eq. (6).

$$\begin{cases} E_V\left(\tilde{\chi}\right)=\int\left\|\vec{V}(q)-\nabla\chi(q)\right\|^2 dp=\frac{1}{2}E_{\tilde{V}}(\chi) \\ E_{\tilde{\omega},\tilde{p}}\left(\tilde{\chi}\right)=\frac{\alpha_s S(P)}{\sum_{q\in P}\omega(q)}\sum_{q\in P}\omega(q)=\frac{1}{4}E_{\tilde{\omega},\tilde{p}}(\chi) \end{cases} \quad (6)$$

In addition, the algorithm optimized the handling of boundary conditions, extending Dirichlet boundary conditions to Neumann boundary conditions. This reduces constraints when data is missing, as long as the derivative of the boundary is zero [17]. To enhance the realism of point cloud based 3D models, in addition to shape reconstruction, this study further integrates RGB information to generate 3D color models. In the process of constructing a 3D model of a point cloud, the 3D surface reconstruction algorithm can only establish an object surface model lacking texture details based on the spatial features of the point cloud data. The steps to build this model are as follows: first, to perform normal vector calculation on the original color point cloud data P. The normal vector of a point cloud is generally solved using principal component analysis algorithm, which establishes a covariance matrix C based on the neighboring points within a certain range of 3D point p_0 and solves for its eigenvectors, as shown in Eq. (7).

$$\begin{cases} C=\frac{1}{k}\sum_{i=1}^{k}(p_i-p_0)\bullet(p_i-p_0)^T \\ C\bullet\vec{v}=\lambda\bullet\vec{v} \end{cases} \quad (7)$$

In Eq. (7), k represents the number of nearest neighbors within the neighborhood range of 3D point p_0. The collected color point cloud data containing only XYZRGB information needs to be accompanied by a normal vector to form a dataset with the XYZRGB Normal attribute. By constructing a Kd Tree index for these data, point cloud color mapping can be performed. At this point, for each point on the surface of the model, to use Kd Tree to search for nearby point cloud data points, take the average of their RGB values, and assign the corresponding color texture to the model. In this way, each point can be mapped to generate a 3D surface model with colored textures.

B. DLD of Landscape Architecture Based on 3DIR

The innovation of digital technology has produced new design languages and construction techniques, while bringing innovative structures and forms of expression to traditional landscape design [18]. Knowledge driven and landscape design are closely linked, guiding and promoting the design and implementation of landscape projects through the application of multidisciplinary knowledge and technology. Therefore, this study combines 3DIR technology and uses knowledge networks as the core driver to intelligently generate landscape plant communities in the scene and carry out corresponding planning and layout. Fig. 3 shows the overall framework of DLD.

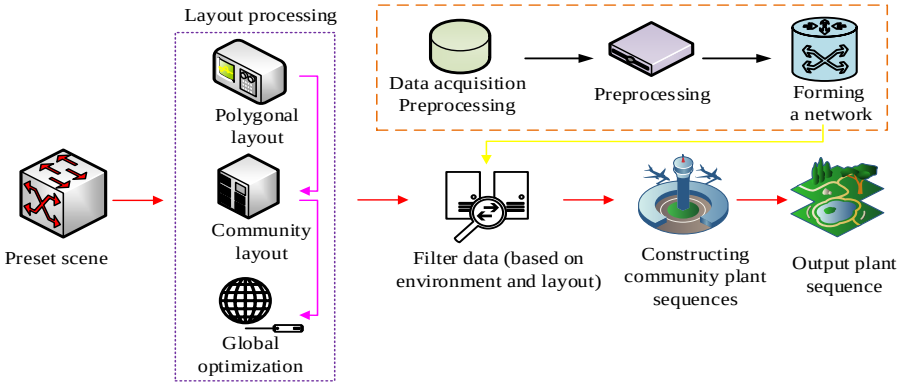


Fig. 3. Digital landscape design model.

The process of DLD in Fig. 3 mainly consists of three steps, which are to build a plant knowledge network, then layout the group within the scene, and finally construct the composition sequence of the community. All three are designed based on data to form a cohesive whole. In terms of data acquisition, most sources come from literature, networks, etc. Data related to landscape plants were collected, which provided fundamental support for this study. Knowledge graphs can provide practical and valuable disciplinary information [19]. Therefore, this study uses knowledge graphs based on undirected graph models to represent landscape plant data. The proposed landscape layout adopts a double-layer optimization strategy: the inner layer uses heuristic algorithms for polygon planning, and the outer layer uses particle swarm optimization (PSO) for global optimization [20]. The inner algorithm is responsible for the preliminary layout design, while the outer algorithm further refines and enhances the design scheme. Among them, heuristic algorithms use simple polygons for layout and find the optimal position within a matrix scene $M=(X,Y)$. Set $P=(P_1,P_2,\ldots,P_m)$ includes polygons, and $\Omega=\{\Omega_1,\Omega_2,\ldots,\Omega_m\}$ contains all possible rotation angles. Scene M has a center point O and marks different areas. Grassland $A=(\sigma_1)$ is a layout area, while buildings, water bodies, and roads $B=(\sigma_2,\sigma_3,\sigma_4)$ are non layout areas. Given a polygon P_i and a translation vector $v=(v_x,v_y)$, to denote $p=(p_x,p_y)$ as the current position of polygon P. The translation function of a polygon is defined in Eq. (8).

$$P_i \oplus v=\left\{\left(p_x+v_x,p_y+v_y\right)\middle|p=\left(p_x,p_y\right)\in P_i,x\in X,y\in Y\right\} \quad (8)$$

If the center of the polygon is represented as P_i, and its state is only determined by its center g_i and rotation angle a_i, then the current polygon calculation formula is Eq. (9).

$$P_i(a_i)=\left\{\begin{pmatrix}g_{ix}+(p_x-g_{ix})\cos(a_i)-(p_x-g_{ix})\sin(a_i),g_{ix}\\+(p_x-g_{ix})\sin(a_i)+(p_x-g_{ix})\cos(a_i)\end{pmatrix}\middle|p\in P_i,a\in\Omega\right\} \quad (9)$$

The unit layout area will be traversed outward from the center of the scene in a circular hierarchical manner, and the current unit layout area is scored using a scoring function $F=\{f_1,f_2,\ldots,f_k\}$, as shown in Eq. (10).

$$\begin{cases}f_1(P_i(a_i))=|P_i(a_i)|\times\omega_1\\f_2(P_i(a_i))=|P_i(a_i)|p\in A|\times\omega_2\\f_3(P_i(a_i))=|P_i(a_i)|p\in B|\times\omega_3\\f_4(P_i(a_i))=|P_i(a_i)|p\in(P-P_i)|\times\omega_4\\f_5(P_i(a_i))=|P_i(a_i)|p\notin(P-P_i),p+d\in(P-P_i)|\times\omega_5\end{cases} \quad (10)$$

In Eq. (10), $\omega_1,\omega_2,\ldots,\omega_k\ (-\infty<\omega<+\infty)$ represents the weight of different scoring factors, which depends on the distribution of layout and non layout spaces within the scene. If condition $mscore=\max\{p_{1score},p_{2score},\cdots,p_{nscore}\}$ is met, the polygon is allowed to be positioned in the corresponding region $P_i(a_i)\to M$. Each polygon records its number, type, area, and coordinate information. The layout of plant landscape in landscape design is its core, and the diversity of plant species constitutes its structure. The composition of the community affects the appearance and changes of the landscape, and plants need to be arranged in a polygonal manner, divided into tree, shrub, mixed or blank types. The combination layout of plants is crucial for landscape features, and it is necessary to carefully plan the distribution of plants in the polygon, including the mixed planting forms of different plants. Fig. 4 shows the distribution of some plant communities and their probability of occurrence within the polygon.

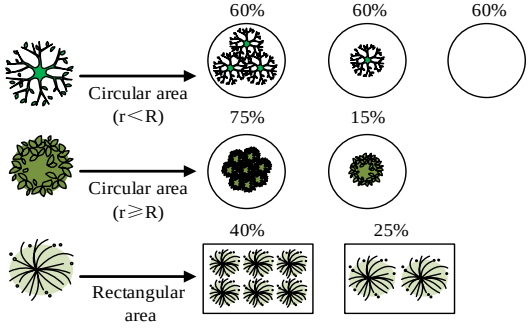


Fig. 4. Distribution of plant communities and their probability of occurrence within the polygon.

In Fig. 4, there are three main requirements for the local layout of the garden landscape. The primary rule is to ensure that the canopy of the trees does not obstruct each other. Secondly, the canopy of shrubs should be independent of each other to ensure that the canopy of shrubs and trees does not intersect. Plants within the same polygon form the same community and inherit polygon parameters from the community. These parameters include the number, type, total area, coordinates, and plant features of the polygons. The plant characteristics provide a detailed description of the number, location, canopy area, building area, and category of trees and shrubs. The polygon layout algorithm effectively arranges plant communities, but if the layout is too neat, it may lack natural beauty. Modern design tends to combine norms with nature to present order and natural beauty. By analyzing the forms of plant space creation, this study proposes the following rules: a single plant is defined as T, the types of trees and shrubs are T_{type}, and the crown coverage area is $C = T_{type}$. The plant community is defined as $G = \{T_1, T_1, \ldots, T_n\}$, and the polygon is defined as $B = \{G_1, G_2, \ldots, G_m\}$. All non-repeating plants are randomly selected and their scores are calculated using Eq. (11).

$$T_{isore} = \sum_{j=1}^{k} q_j(T_i), q \in Q \tag{11}$$

The optimal sequence of plants found based on the scoring function is Eq. (12).

$$\min SEQ(T) = \min \sum_{m}^{1} \sum_{i=1}^{n} \sum_{j=1}^{k} q_j(T_i), q \in Q, T \in G \tag{12}$$

After optimizing the global layout algorithm, plant information, including numbers and parameters, is extracted from the polygon, which serves as the community identifier. These information are included in the plant list, and then suitable plant species are selected in the knowledge network based on specific conditions. Screening involves distribution range, adaptability, and lighting requirements to preliminarily determine candidate species, and then further screen using geographic information and environmental conditions to form plant collection B_s. At the same time, filtering the community database to form a set G_s. The species of plants in set G_s become set B_s. By comparing the set, the final plant candidate set $B_{sg} = B_s \cup B_g$ is formed, and the community set is updated to G_{sg}. The selected community is evaluated based on plant adaptability, cost, and environmental impact, but the community set G_{sg} is not simplified to maintain species diversity. After adjusting the diversity based on the available planting area, a final selection set Q is formed. The community configuration is assigned to plants from set Q, sharing numbers, rather than increasing the average score of selected communities. After the layout is completed, the herbaceous classification within the community is arranged independently and serves as a supplement at the end of the list. This study uses PSO to conduct global optimization on the basis of building a complete plant community. The performance evaluation process of digital design methods for landscape architecture is Eq. (13).

$$S = \lambda_1 \sqrt{\frac{\sum_{1}^{n} \left(q_i(T) - \overline{q_i(T)}\right)^2}{n-1}} + \lambda_2 \sqrt{\frac{\sum_{1}^{n} \left(f_i(P(a)) - \overline{f_i(P(a))}\right)^2}{n-1}} \tag{13}$$

In Eq. (13), both λ_1 and λ_2 will set corresponding values. The design and implementation of the prototype system are completed by packaging the proposed model, standardizing its input and output, and generating AutoCAD exchange files and executable files for the prototype system. The overall architecture of the model is Fig. 5.

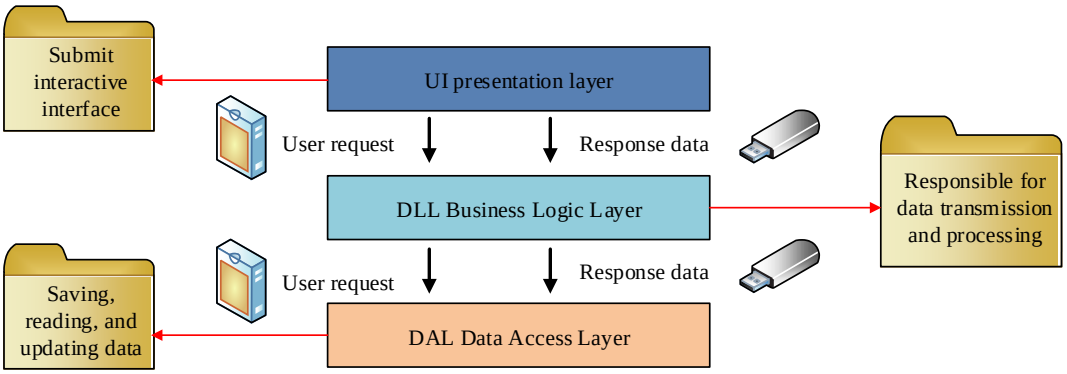


Fig. 5. Digital landscape architecture design model.

In Fig. 5, the presentation layer handles the user interface and interaction, including displaying data and receiving input. The business logic layer is responsible for verifying input parameters, ensuring the stable operation of the program, and performing operations such as data queries and updates. This includes checking for empty text, file path issues, and input legality. The data access layer is responsible for communicating with the database, performing data addition, deletion, modification, and querying operations, and the data objects should only be referenced in this layer to avoid direct data manipulation by other layers.

III. Results

A. 3DR Experimental Analysis of Visual Scenes Based on RGBD Data

To verify the effectiveness of the proposed algorithm, experiments are conducted to compare combination of SAC-IA

or NDT coarse registration and ICP fine registration algorithms based on FPFH, 3DSC, SHOT features. Experiment is performed on i5-9300H CPU Windows 10 system using the environment of Visual Studio 2019 and PCL 1.11.0 library. The Stanford University's Bunny point cloud (BPC) and visual scene point cloud (VSPC) are collected as the dataset. Table I shows the experimental parameters.

PCR experiments are conducted based on the point cloud parameter settings in Table I, and the experimental results are shown in Fig. 6.

TABLE I. EXPERIMENTAL PARAMETER SETTINGS

Experimental project	Parameter	Experimental project（NDT）	Parameter
Leaf nodes (BPC data)	**0.01**	**Minimum Conversion Difference (BPC Data)**	**0.0001**
Leaf nodes (VSPC data)	**100**	**Minimum Conversion Difference (VSPC Data)**	**1**
Search radius (BPC data)	**0.05**	**Parameter Step Size (BPC Data)**	**0.25**
Search radius (VSPC data)	**500**	**Parameter Step Size (VSPC Data)**	**500**
Search radius (feature description algorithm)	**0.5、1000**	**Resolution parameter value**	**0.5、900**
Point cloud overlap parameter value (Trimmed ICP)	**95%**	**Maximum number of iterations (NDT, ICP)**	**50、20**



(a) Initial position (b) FPFH+ICP (c) SHOT+ICP (d) NDT+ICP (e) This article studies

Fig. 6. Registration results of various algorithms.

In Fig. 6, the proposed PCR algorithm and several comparison algorithms can effectively cover the blue point cloud to be registered onto the reference purple point cloud. The SAC-IA algorithm and NDT algorithm have similar registration effects, but compared to the fusion of ICP and NDT registration, the research algorithm is more accurate and shows stronger robustness. The registration running time of each algorithm and the RMSE values of the registered point cloud in various directions are shown in Fig. 7.

In Fig. 7, the research algorithm outperforms other algorithms in terms of running speed and registration accuracy. In algorithms based on local features, the registration of FPFH feature descriptions leads in efficiency and accuracy, while the 3DSC algorithm takes the longest time, several times that of other methods. The NDT algorithm based on statistical probability reduces the running time by at least 69% compared to local feature methods. Fig. 8 shows the comparison results of quantitative indicators based on self-collected point cloud (SCPC) data.

In Fig. 8, in the testing of the Bunny dataset, the root mean square error of the NDT algorithm is lower than that of the algorithm based on local features. On SCPC data, NDT performs poorly, possibly due to a decrease in accuracy caused by the distribution of the data in 3D space. The research algorithm performs better than the NDT algorithm combined with ICP on both types of data, reducing registration time and improving accuracy. Compared with the fusion of ICP and NDT, the registration time of the research algorithm was reduced by 2%, and the RMSE was reduced by 71.4% and 87.5%, respectively, demonstrating good robustness and high registration accuracy. Fig. 9 shows the evaluation index results of each algorithm.

In Fig. 9, compared with the surface fitting based method, the 3D surface reconstruction technique based on mesh triangulation performs poorly. The greedy projection triangulation (GPT) algorithm is lower in key performance indicators than Poisson reconstruction and masked Poisson reconstruction algorithms, with differences of up to 0.010203, 0.000701, and 0.001874, respectively. GPT outperforms the rolling ball algorithm in RMS metrics, reducing by 38.1% and demonstrating better repair performance. In addition, the shielded Poisson algorithm exhibits higher model quality compared to traditional Poisson algorithms.

B. DLD Experimental Analysis

The experiment utilized Rhino software and visual scene 3DR based on RGBD data to construct parameter logic and establish a parameterized analysis model. It selected real garden image data and applied the technology studied to distinguish natural and artificial landscape elements, such as plants and architecture, for instance design. The evaluation indicators for landscape design include panoramic quality, design aesthetics, ease of technical implementation, cultural significance, and urban characteristics. Fig. 10 shows the evaluation of the DLD effect of 3DIR technology.

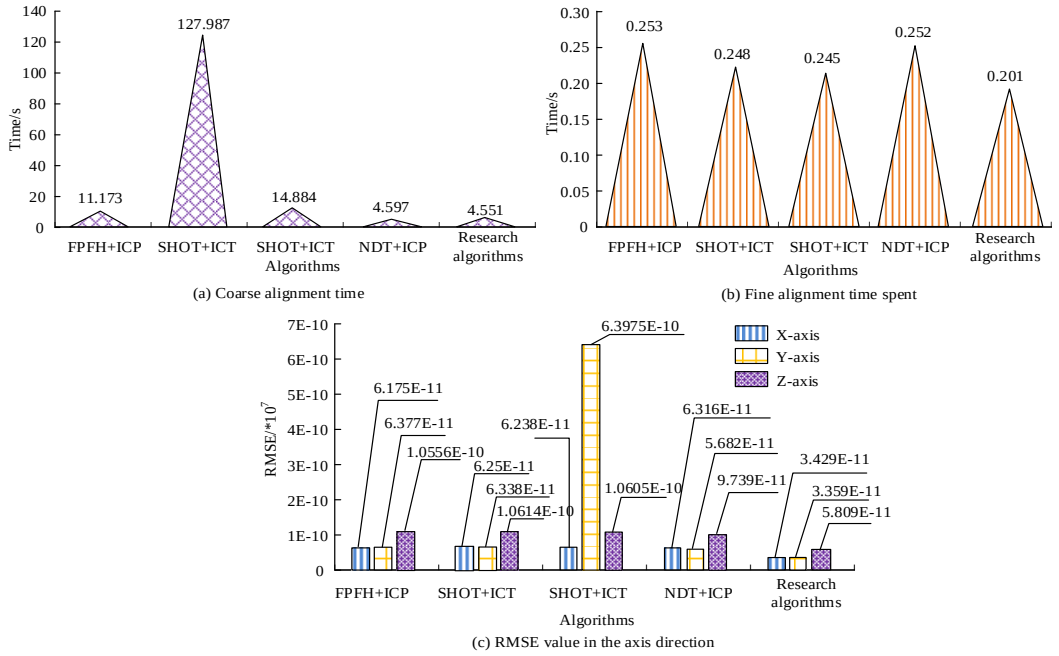


Fig. 7. Comparison of quantitative indicators based on bunny point cloud data.

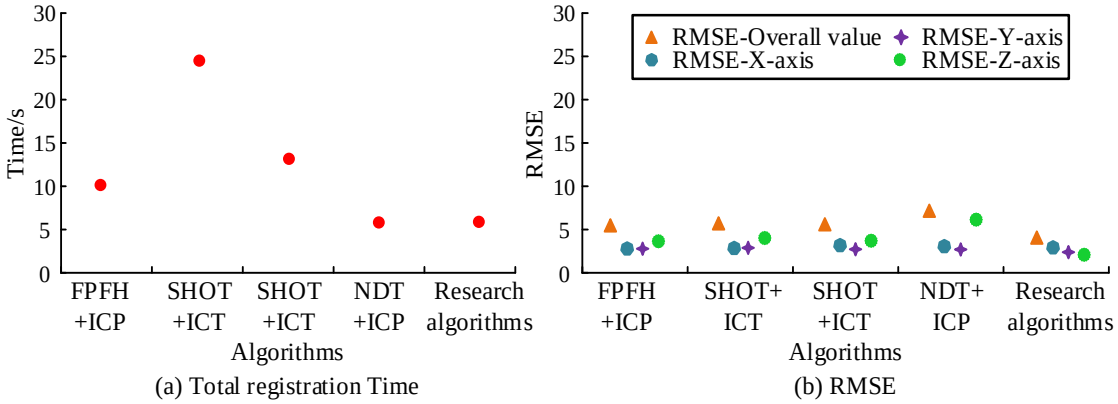


Fig. 8. Comparison of quantitative indicators based on SCPC data.

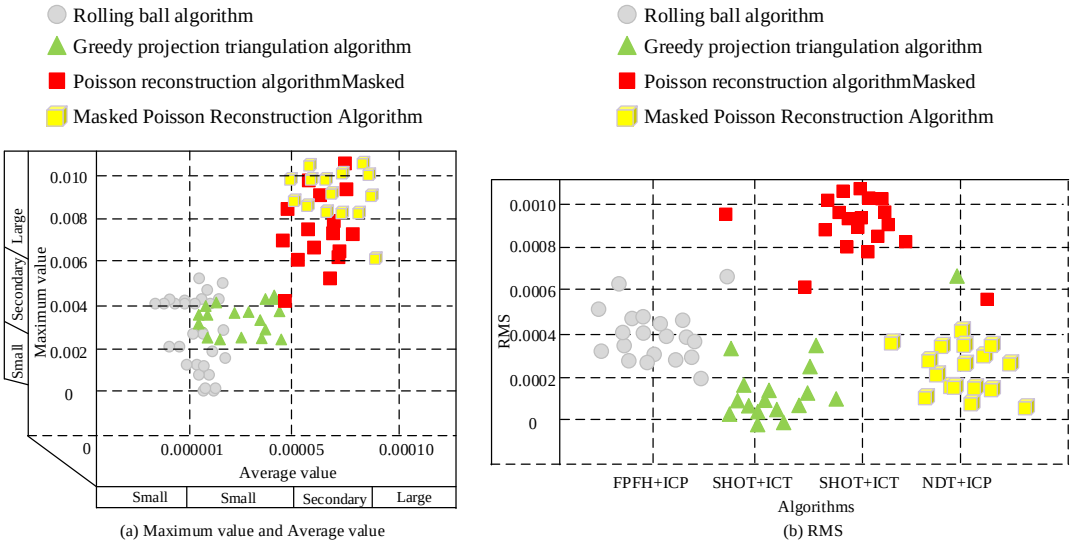


Fig. 9. Evaluation index results of each algorithm.

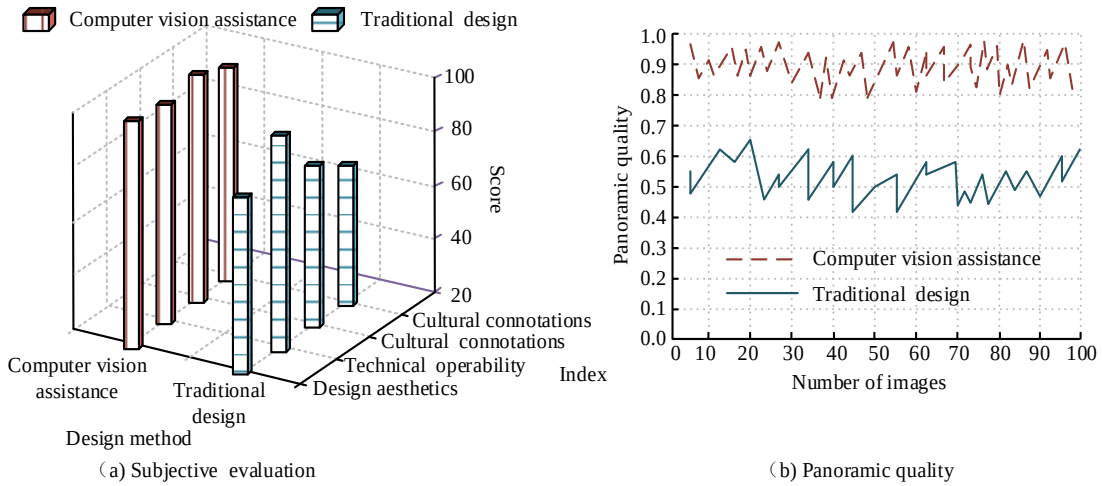


Fig. 10. The design effect of DLD based on 3DIR.

In Fig. 10 (a), the DLD combined with 3DIR technology performs well on a series of subjective evaluation indicators, with all indicators superior to traditional landscape design methods except for technical operability indicators. In Fig. 10 (b), among 100 design samples, the panoramic quality of the research method fluctuates within a small range of 0.8 or above. The stability of panoramic quality in traditional methods is poor, ranging from 0.4 to 0.7. Moreover, the panoramic quality value is relatively low, and the design quality fluctuates continuously with the level of manual design. Fig. 11 shows the performance comparison between the research method and other algorithms.

In Fig. 11, the AUC value of the study method is very close to 0.9, which means that its performance is excellent. Although CNN-GAN performed well in the early stage, the overall curve is not as steep as the study method, and the AUC value should be slightly lower than the study method. It can be inferred from the curve shape that the AUC value of PSO is significantly lower than that of the research method and CNN-GAN, which means that its classification effect is poor. Compared with the other two methods, the performance of the research method is very stable in the whole range. Fig. 12 shows the display of DLD results.

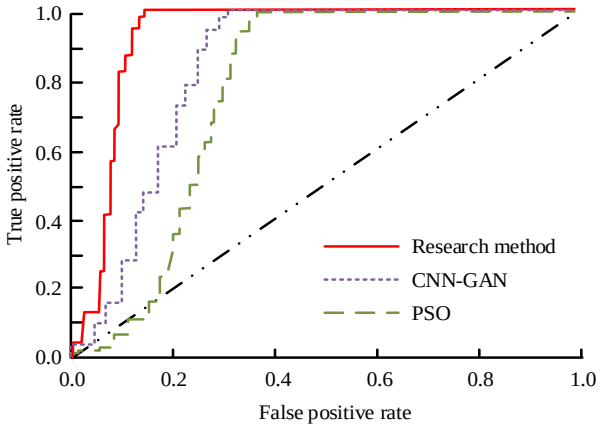


Fig. 11. Convergence process of objective functions of the three algorithms.

Fig. 12 shows the effectiveness of the classic iterative optimization intelligent landscape design algorithm, which generates scenes that are highly consistent with real landscape design, especially in vegetation selection and layout. The research algorithm not only follows the principles of eco-friendly design, but also enhances the aesthetics of plants in the context of urban landscaping, achieving a visual effect of harmonious coexistence between humans and nature.



Fig. 12. Display of DLD achievements.

IV. Discussion

Compared with the modeling technique of Steer et al. [6], the research method is more accurate in dealing with landscape design details, and high-precision 3D reconstruction is achieved through improved point cloud registration and Poisson reconstruction techniques. Studies have shown that this method is superior to Luo Ma L et al. [7] in reconstruction details and to Wang et al. [8] in landscape flow design in overall design performance. Compared with the traditional ICP and NDT methods, the registration time of the proposed algorithm is reduced by 2%, and the error is reduced by 71.4% and 87.5%, respectively. In Bunny point cloud and SCPC data testing, the root mean square error (RMSE) is lower than that of traditional methods, and the run time is significantly reduced. The digital landscape design method based on 3DIR technology has excellent performance in the subjective evaluation indicators such as panoramic quality, design beauty, technical realization difficulty, cultural significance and urban characteristics, and its stability is higher than that of traditional design methods. The resulting landscape design is more aesthetically pleasing and coordinated in plant selection and layout, and intelligent and ecologically friendly plant community layout is assisted by

knowledge networks. The dual optimization strategy, which combines heuristic algorithm and particle swarm optimization, takes into account the beauty and functionality of landscape design. In conclusion, this method has significant advantages and application potential in 3D reconstruction and landscape design.

V. CONCLUSION

To improve the efficiency and visualization level of landscape design, and to help designers quickly and accurately convert elements such as terrain, vegetation, and water in the real world into digital models, this study adopted the concept of DLD to analyze landscape architecture and digital design. Based on the current difficulties in landscape planning and design, a DLD method based on 3DIR technology was proposed. The results show that the proposed algorithm is superior to other algorithms in terms of running speed and registration accuracy. Statistical probability-based NDT algorithms have at least 69% shorter run times than local feature algorithms and lower RMSE on rabbit datasets. Compared with the method combining ICP and NDT, the registration time was reduced by 2% and RMSE was reduced by 71.4% and 87.5%, respectively. The shielded Poisson algorithm is superior to the traditional Poisson algorithm in terms of model quality. The panoramic quality of the research method is stable at 0.8 or above, while the panoramic quality of the traditional design method fluctuates between 0.4 and 0.7 and is low. The scene generated by the research method is highly consistent with the actual landscape design, especially in vegetation selection and layout. In the future, deep learning algorithms will be integrated to improve and innovate traditional algorithms while ensuring PCR rate and registration accuracy.

REFERENCES

[1] Deng Y, Xie L, Xing C, Cai L. Digital city landscape planning and design based on spatial information technology. Neural Computing and Applications, 2022, 34(12): 9429-9440.

[2] Usman A M, Abdullah M K. An Assessment of Building Energy Consumption Characteristics Using Analytical Energy and Carbon Footprint Assessment Model. Green and Low-Carbon Economy, 2023, 1(1): 28-40.

[3] Li D, Li H, Li W, Guo J, Li E. Application of flipped classroom based on the Rain Classroom in the teaching of computer-aided landscape design. Computer Applications in Engineering Education, 2020, 28(2):357-366.

[4] Cui Xing, Du Chunlan. Research on Landscape Parametric Design Based on GIS+BIM Information Model—Taking the Planning and Design Experiment of Mountain Scenic Environment Road as an Example. Chinese Landscape Architecture, 2023, 39(6):39-45.

[5] Li P. Intelligent landscape design and land planning based on neural network and wireless sensor network. Journal of Intelligent and Fuzzy Systems, 2021, 40(2):2055-2067.

[6] Steer P. Short communication: Analytical models for 2D landscape evolution. Earth Surface Dynamics, 2021, 9(5):1239-1250.

[7] Luoma L, Fricker P, Schlecht S. Design with Sound: The Relevance of Sound in VR as an Immersive Design Tool for Landscape Architecture. Journal of Digital Landscape Architecture, 2023, 2023(8): 494-501.

[8] Wang Z, Trauth K M. Development of GIS-based python scripts to calculate a water surface profile on a landscape for wetlands decision-making. Journal of hydroinformatics, 2020,22(3):628-640.

[9] Zhao Y, Luo X, Qin K, Liu G, Chen D, Augusto R S, Liu Z. A cosmic ray muons tomography system with triangular bar plastic scintillator detectors and improved 3D image reconstruction algorithm: A simulation study. Nuclear Engineering and Technology, 2023, 55(2): 681-689.

[10] Xu Y, Nan L, Zhou L, Wang J, Wang C C. Hrbf-fusion: Accurate 3d reconstruction from rgb-d data using on-the-fly implicits. ACM Transactions on Graphics (TOG), 2022, 41(3): 1-19.

[11] Zhang J, Yu K, Wen Z, Qi X, Paul A K. 3D reconstruction for motion blurred images using deep learning-based intelligent systems. Computers, Materials & Continua, 2021, 66(2): 2087-2104.

[12] Liu X, Li J, Lu G. Improving RGB-D-based 3D reconstruction by combining voxels and points. The Visual Computer, 2023, 39(11): 5309-5325.

[13] Li J, Gao W, Wu Y, Liu Y, Shen Y. High-quality indoor scene 3D reconstruction with RGB-D cameras: A brief review. Computational Visual Media, 2022, 8(3): 369-393.

[14] Chen H B, Zheng R, Qian L Y, Liu F Y, Song S, Zeng H Y. Improvement of 3-D ultrasound spine imaging technique using fast reconstruction algorithm. IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control, 2021, 68(10): 3104-3113.

[15] Kermarrec G, Morgenstern P. Multilevel T-spline Approximation for Scattered Observations with Application to Land Remote Sensing. Computer-Aided Design, 2022,146(1):103193-103206.

[16] Thompson P, Dugmore A J, Newton A J, Cutler N A, Streeter R T. Variations in tephra stratigraphy created by small-scale surface features in sub-polar landscapes. Boreas, 2022, 51(2):317-331.

[17] Faust E, Schlüter A, Müller H, Steinmetz F, Müller R. Dirichlet and Neumann boundary conditions in a lattice Boltzmann method for elastodynamics. Computational Mechanics, 2024, 73(2): 317-339.

[18] Mukherjee T, Pucci P, Sharma L. Nonlocal critical exponent singular problems under mixed Dirichlet-Neumann boundary conditions. Journal of Mathematical Analysis and Applications, 2024, 531(2): 127843.

[19] de Assis H R, Faria L F O. Elliptic equations involving supercritical sobolev growth with mixed dirichlet-neumann boundary conditions. Complex Variables and Elliptic Equations, 2024, 69(5): 713-728.

[20] Foss M, Radu P, Yu Y. Convergence analysis and numerical studies for linearly elastic peridynamics with dirichlet-type boundary conditions. Journal of Peridynamics and Nonlocal Modeling, 2023, 5(2): 275-310.

Software Systems Documentation: A Systematic Review

Abdullah A H Alzahrani
Department of Computers-Engineering and Computing College at Alqunfuda,
Umm Al Qura University, Makkah, Saudi Arabia

***Abstract*—In the domain of software engineering, software documentation encompasses the methodical creation and management of artifacts describing software systems. Traditionally linked to software maintenance, its significance extends throughout the entire software development lifecycle. While often regarded as a quintessential indicator of software quality, the perception of documentation as a time-consuming and arduous task frequently leads to its neglect or obsolescence. This research presents a systematic review of the past decade's literature on software documentation to identify trends and challenges. Employing a rigorous systematic methodology, the study yielded 29 primary studies and a collection of related works. Analysis of these studies revealed two primary themes: issues and best practices, and models and tools. Findings indicate a notable research gap in the area of software documentation. Furthermore, the study underscores several critical challenges, including a dearth of automated tools, immature documentation models, and an insufficient emphasis on forward-looking documentation.**

Keywords*—Software engineering; software systems documentation; software maintenance; software quality; software development

I. Introduction

Software documentation can be defined as the journey of producing different types of documentation. These types vary in their purposes from describing the development processes to describing the final product to the intended user. It is believed that software engineers should have the responsibility of software documentation, however, professionals in technical writing are sometimes needed [1], [2], [3], [4].

In the past 10 years, software documentation has been an interest in the industry of software engineering. However, majority of documenters are whether technicians or peoples trained in humanity. Therefore, the need for more professionals in the software documentation had emerged [5].

Many benefits can be accrued from good software documentation such as decreased costs of maintenance [6], [7], [8], [9], [10]. However, achieving a well-formed software documentation might be challenging. One challenge in software documentation is that developers abhor being involved in software documentation. In addition, some believe that poor software documentation is worse than no documentation. Furthermore, lack of professionals in the software documentation is considered to be another challenge [5], [11].

Software documentations principles are to be considered during software documentation. These principles are the level of details, document purpose and intended readers, use of graphical aids, clarity and precision, language of document, and documents versions. Therefore, in order to acquire a well-written software documentation, it is important to pay attention to, first, the purpose of the documents and the intended audience. This will lead to choosing the appropriate language, level of details, and graphical aids. However, in order to keep the documentation alive, updating documents with use of versions management are essential [6], [11], [12], [13], [14].

Several types of software documents are generated in the process of software documentation. These types diverse based of their purposes and intended readers. However, these types fall in one of the following categories. The first category is the documentation of the process of software development. This includes documentation of requirements, planning, implementation and other documents during the development journey. The intended readers for this category of documents are the developer, software, decision makers, and the maintainers. The second category is the documentation of the product after delivery. This category includes the documents that describe the product for intended users. Examples of this category are User manuals and system main structure documents [1], [2], [3], [5], [6], [8], [11], [14].

Many techniques and tools are employed for software documentation. In general Waterfall technique and Iterative technique are the most dominant techniques for software documentations. However, for the tools that are used for the documentation, many have stated a variety of tools such as MS Word XML and other Text Editors, Doxygen, Visio, FrameMaker, Author-IT, Doc++, Rational Rose, JUnit and other tools. Some of these tools aid in automation of documentation to some extents [15], [16], [17], [18], [19].

The importance of this study derived from the need to achieve solutions that overcome the unsolved problem of poor or absent software documentation as this issue remains unsolved. Software documentation improves the quality of software systems and consequently improves maintainability and cost efficiency. Furthermore, drawing more attention to the topic of software documentation would enhance documentations models as well as templates employed, which relatively enhance automation of software documentation.

This paper has been structured as follows. Section I introduced the topic software systems documentation. In addition, it highlights the importance of systematic review on

the considered topic. Section II illustrates the methodology which has been employed in this paper and formulated the research question. In addition, the main findings statistically shown and discussed. Section III has been divided into two subsections. The first subsection concludes the discussion on the findings and demonstrates the trends in the software documentation publications. The second subsection reviews and discusses the primary studies found in this research and categorizes them into two categories. Finally, Section IV draws conclusions on this research.

II. METHODOLOGY

Budgen et al. and Kitchenham [20], [21] have described a systematic review methodology which this study applies. The methodology enables conducting the reviews objectively and structurally. Furthermore, the methodology allows demonstration of broader picture of the topic of software documentation by categorizing the results into primary and related to the topic under consideration.

1) Research question: What are the unveiled trends and issues in relation to software documentation in the last 10 years?

The above research question can be responded to by first exploring the software documentation topic, then, investigating the existing techniques and tools which are employed. Consequentially, issues and difficulties will be highlighted and identified. Therefore, the keywords leading the search in the well-known databases have been enumerated. These keywords are software system documentation - software documentation - system documentation - automated documentation - software knowledge documentation - computer software documentation - software engineering documentation.

2) Sources selection: Bearing in mind the aforementioned keywords, a search query has been formulated as shown in Table I. An OR logical operator has been used in order to combine results that are related to the search. In addition, to narrow the range of the results, double quotation marks ("") have been applied to surround the keywords.

TABLE I. SEARCH QUERY

Search query	"software system documentation" OR "software documentation" OR "system documentation" OR "automated documentation" OR "software knowledge documentation" OR "computer software documentation" OR "software engineering documentation"

After formulating the Search query, it has been entered to The Saudi Digital Library (SDL) [22] search engine. SDL is an online digital library resource that allows searching in many well-known digital libraries such as Springer, IEEE Digital Library, ACM Digital Library, SAGE, ScienceDirect, and other publishers. Different types of research items can be found; however, the results were grouped into four main categories. The first category is Conference Paper. Second category is Journal papers which include journal Article, Case Study papers, and Review Papers. Third category is Books which include books, handbooks, Thesis, and technical reports. Fourth category is Others which include else results.

This paper's goal is to investigate the topic of software documentation in the past 10 years, so, an exclusion criterion of year of publication has been applied and configured to include work which has been done between the years 2014 and 2024. In addition, another exclusion criterion was the language of publication as non-English publications have been eliminated during the search process by configuring the search engine to exclude them before displaying the results from the digital libraries. Finally, inclusion criterion relies on the analysis of several aspects in found results. These aspects are title, keywords, abstract, and conclusion. So, the analysis is the process of manually reading deciding for publications to be considered relevant or not.

Table II illustrates the results of the searching process. It is clear in the table that the first results after applying the query string was total of 7482 publications found in different source. However, this number includes the repeated items and the flawed entries of the items. Therefore, it was necessary to accurate the results by eliminating those items. Consequently, the total number of publications declined to be 3453 items.

Inclusion criterion was then applied manually to determine the relevance of remaining items by scanning the keywords, abstracts, and conclusions. The process resulted in eliminating more items. A total of 1654 items are the relevant items to be investigated and studied in order determine the results of primary studies to the topic of software documentation.

TABLE II. NUMBER OF PUBLICATIONS FOUND FROM DIFFERENT SOURCES

Sources	**Publication**					
	Search date	**Found**	**Not repeated**	**Relevant**	**Primary**	**%**
Springer Link	3 Jan 2024	612	411	378	0	0%
IEEE Digital Library	3 Jan 2024	125	82	65	5	17%
ACM Digital Library	3 Jan 2024	41	34	12	6	21%
SAGE	3 Jan 2024	248	188	90	0	0%
ScienceDirect	3 Jan 2024	41	39	21	0	0%
Other Publishers	3 Jan 2024	6415	2699	1088	18	62%
Total		7,482	3453	1654	29	100 %

It can be seen in Fig. 1 that relevant found publications to the topic of software documentation are divided into six categories. The categories are based on the reputation of the publisher. Therefore, five categories are designated to leading and well-known publishers which are Springer, IEEE, ACM, SAGE, and ScienceDirect. All other publishers have been considered in one category as "other publishers".

Fig. 1 illustrates 66% of the relevant found publications to the topic of software documentation are from other publishers which have less reputation than the leading publishers. However, 23% of relevant found publications are from Springer. Having this in mind, 378 relevant publications to the

topic of software documentation for the last 10 years in a well-known publisher such as Springer are not a considerable number of publications. This might raise a question on the reasons behind the low number of publications in the topic of software documentation in such leading and well-known publishers.

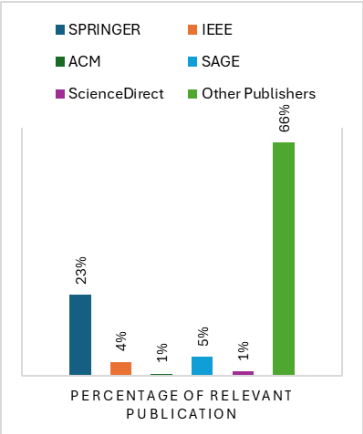


Fig. 1. Relevant studies from different publishers.

With regards to the primary studies shown in Table II, they are the studies that are major in the topic of software documentation. A complete list is included in Appendix A. These studies are identified to be primary after a manual excessive analysis and in depth reading of the relevant studies. The main criterion to identify the primary studies is the goal of the research. In particular, if the relevant study is offering a new model, approach, solution, explanation, case study, comparison, or/and review, then that relevant study is considered to be a primary study to the topic of software documentation.

III. Results and Discussion

A. Trends in Software Documentation Publications

This section is to discuss the findings on the collected data and the analysis of results of the conducted reviews. The main finding is that the topic of software documentation has not been considered sufficiently for research in the past 10 years, especially by well-known and leading publishers. This can be seen clear form the results shown in the previous section.

Fig. 2 illustrates the publications of relevant studies to the topic of software documentation over the last 10 years. From Fig. 2, a growing interest can be noticed from the year of 2020 in the publication is a well-known publisher which is Springer. However, this interest in the topic of software documentation has remained unnoticed in publications of other well-known publishers.

Table III demonstrates the types of relevant publications found. Three main categories of publications which are conference papers, journal papers, and book. Regarding journal papers, this includes regular article papers, Case Study papers, Review Paper. On the other hand, Books category includes books, handbook, Technical Report, and Thesis. Finally, all other publication types were classified into “Others” category.

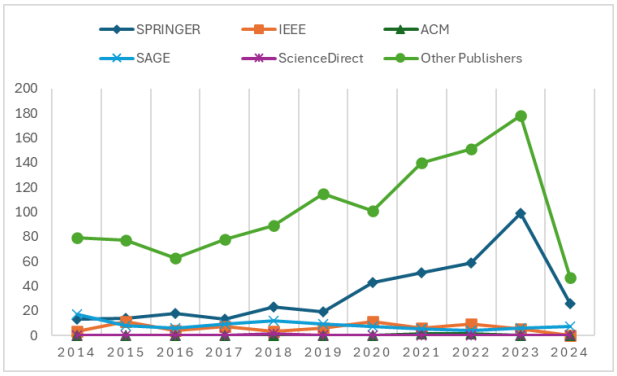


Fig. 2. Relevant studies from different publishers. Publications in the past 10 years on software documentations.

TABLE III. Types of Publications of Relevant Studies

Sources	Conference Paper	Journal Paper	Books	others	Total
Springer Link	3	348	1	26	378
IEEE Digital Library	25	10	0	30	65
ACM Digital Library	0	8	0	4	12
SAGE	0	38	51	1	90
ScienceDirect	0	12	0	9	21
Other Publishers	14	633	122	319	1088

Table III shows that the majority of relevant publications on software documentation are journals papers. Fig. 3 illustrates that around 63% of all relevant publications on the topic of software documentation over the past 10 years are journal papers. It is worth noting that from the leading and well-known publishers, SAGE has an outstanding interest on relevant books on the topic of software documentation.

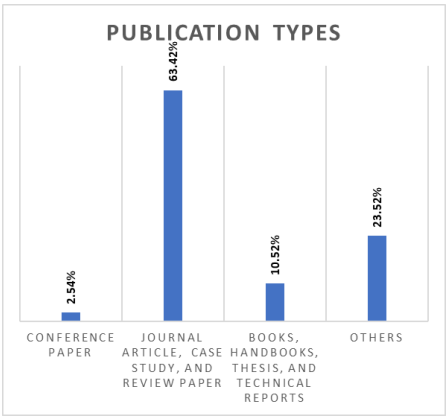


Fig. 3. Types of relevant studies.

B. State of Arts Studies

This section discusses the primary studies found in the systematic review on software documentations. The studies have been classified into three main divisions shown in the following subsections which are: 1) software documentations reviews, issues, and best practices; 2) software documentation models and tools.

1) Software documentations reviews, issues, and best practices: Many have introduced best practices and required attributes of software documentation [15], [23], [24], [25], [26]. Other researchers [17], [18], [27], [28] have considered software documentation issues which might be addressed earlier, throughout, and/or afterwards the process of the software documentations.

Uddin et al. [15] have conducted two surveys based on their previous work [23] on Application Programming Interface (API) documentation. First, the authors designed a three-questions survey circulated among IBM Canada labs. The number of participants was 69 with the following job titles: developer, architect, tester, and consultant. Second, the authors designed a seven-questions survey circulated among developers and architects at IBM Canada and UK. This time, the number of participants was 254. The authors concluded that the API documentation suffers contents issues which are related to the fact that the engagement of experts is needed in the documentation of API.

Rai et al. [29] have conducted a review on the topic of source code documentation over last 10 years. The authors formed six research questions that were to be answered. These questions focused on the methodologies, tools, and evaluation means for source code documentation. In addition, the authors found that source code documentation has grown interest from researchers and automatic generation of source code documentation begins to use Deep Learning approaches instead of Information Retrieval approaches.

Lethbridge et al. [6] have published a study in order to gather best practices in software documentation. The authors conducted a study that includes surveying, interviewing, and observing software engineers. in addition, the authors investigate the tools used by these software engineers. The findings of the study can be summarized in the followings: 1) focus on requirements documentation and high-level documentation of the systems rather than complete and UpToDate documentation; 2) focus on simple and customized documentation rather than forcing the use of particular documentations methods or tools. This is to avoid time overhead and complexity.

Forward et al. [18] have investigated the tools and technologies used in software documentations. The authors divided software documentation into six types of documents namely requirements, specifications, detailed design, and low level design, architecture, and QA documents. The main findings can be summarized as follows: 1) software documents are important and useful even if they are obsolete; 2) tools for software documentations are needed to automate the process and should be chosen based on the nature of the software project.

Santos et al. [7] have conducted a review on software documentation in order to check the essential quality attributes in software documents. In addition, the authors reviewed the best practices for software documentation from 14 different publications in order to establish relations between the quality attributes and the best practices offered in those publications.

De Souza et al. [16] have studied the impact of Agile on software documentations. The authors highlighted the software development using Agile relies on informal communication which might lead to lack of software documentation or obsolescence. The authors addressed the issue that might face documentation teams as they immensely require documentation to accomplish their tasks.

Aghajani et al [9], [19] have conducted an empirical study to investigate the issues in software documentation. The outcome of the study has shown 162 types of software documentation issues linked to tools, process, and presentations of information. In addition, the authors believe that the attitude and experience of people involved in the software documentation process are important factors in the success or failure of software documentation.

Meng et al. [30] have conducted a study to investigate the learning strategy that developers use when they encounter API documentation. The study was carried out using interviews and questionnaire methodologies on 17 developers who have been asked 45 questions in the interview and to answer online survey of 39 questions. Main findings showed that documentation of API lack of clarity and completeness. Moreover, un-updated documentation is another recurring issue in API documentation.

In 2015, Zhi et al. [31] reviewed 69 research articles published between the years 1971-2011 on software documentation in order to study the impact of documentation on cost and quality. The authors concluded their review with several findings which can be summarized as: 1) main quality attributes that should be in the documentation are completeness, accessibility, and consistency; 2) most of the evaluation on the documentation models are on a single case study or academic prototypes.

2) Software documentation models and tools: Falessi et al. [13] have conducted an empirical study on 50 postgraduate students in order to evaluate the effect effects using different techniques of Design Decision Rationale Documentation (DDRD). The focus on the experiment is to check how different groups react in requirements change.

Kajko-Mattsson [17] has introduced a model for software documentations that serves corrective maintenance. The model includes 19 requirements with each has a set of goals. The model has been examined by surveying 18 different Swedish organisations with the use of interview mean. The author reported that the results show that collaboration with maintenance teams is to the minimum and the maintenance teams are absent from the documentation process. This has led to inadequate support for decision making for any change as well as quality assurance costs time and effort.

Bachmann et al. [32] have introduced their model for documenting software architecture. The model aims to document layered view of the architecture of the software system. The main purpose is to provide documentation that helps in sharing understanding of the system, tracing the changes, and discussing trade-offs. In addition, the model clearly identifies different views of the software architecture based on the audience of the documentation.

Falessi et al. [12] have introduced value-based (VD) method for software documentation in particular documentation of design decision. The proposed method aims to enhance the use of DDRD approach and is called VD DDRD. The authors conducted an experiment in order to validate their approach and the results show that VD DDRD can moderate inhibitions which might be shown using DDRD.

Aguiar et al. [8] have introduced a methodology for software documentation in particular documenting object-oriented frameworks. The authors have addressed several issues that need to be considered when documenting frameworks. These issues are related to quality, processes, and tools. The approach is tailored to help naïve software engineers in documenting software frameworks. The approach addresses three roles namely writers, developers, and documentation managers. In addition, it emphasizes on the collaborations and involvement throughout the development process.

Véras et al. [33] have introduced an approach which helps assessing the software requirement specifications. The approach aims to provide a benchmark for the assessment process. Three checklists are offered by the approach which is based on the standardizations of Packet Utilization Standard (PUS) by European Cooperation for Space Standardization (ECSS) standards.

Farwick et al. [34] have proposed a semi-automatic approach that document Enterprise Architecture (EA). The approach is composed of 4 models each of which needs manual interventions. The authors aim to overcome academic approaches offered by Hauder et al. [35]. Although the approach is promising, evolution and more case studies are needed to mature it.

Mathrani et al. [36] have proposed a new approach for software documentation that relies on the use of a quality management standards model (ISO 9001). The approach has been case studied on healthcare software with teams applying Scrum methodology for software development. The authors reported that issues such as incompleteness and ambiguity might rise due to the constraints in ISO 9001.

Aversano et al. [37] have proposed a quality model that evaluates the documentation of Enterprise Resource Planning (ERP) software. The model aims to investigate different quality attributes of the documentations. These attributes are linked either to content or to structures. Main purpose is to ensure readability and completeness of the documentation. The model has been experimented with in the open-source ERP systems, however, more case studies are required in order to generalize the results.

Carvalho et al. [38] proposed a tool named Documentation Mining Open-Source Software (DMOSS) for evaluating the quality of software documentation of non-source code information. The tool has been tested on 4 open-source codes. The tool aims to help maintainers in understanding the software.

Theunissen et al. [39] have introduced a model composed of three approaches for software documentation. The model focuses on categorizing software knowledge into 3 types namely acquiring, building, and transferring knowledge. These types highlight the information to be documented based on the stage of the software life cycle. However, the model has not been evaluated.

Rong et al. [40] introduced a new approach named DevDocOps which can be integrated to DevOps in order to automate the process of software documentation. The approach has been implemented and evaluated in telecommunication enterprise and the results were promising.

Krunic [41] have studied the benefits and difficulties of Documentation as Code (DaC) in vehicle software. The author conducted a case study with 150 participants as software engineers. The author concluded the research by providing a model as a guideline for applying DaC and assessing the quality of documentation.

Kazman et al. [42] have introduced a method to architecturally document open-source software. The authors designed a case study to experiment their method on Hadoop Distributed File System (HDFS). However, the results showed that the proosed method had an effect on the project of HDFS.

Righolt et al. [43] have introduced a tool named Code Diary for automatically documenting decisions from SAS source code. Unlike similar tools, the authors claimed that the proposed tool Code Diary aims to produce code documentation for researchers and other audiences. However, no graphical user interface is available for the tool.

AlOmar et al. [44], [45] have proposed a model that acts as a data set for the documentation of refactoring process of the software. The authors conducted an experiment of 5 stages that has the documentation as the last stage. The experiment carried out over 800 open-source java code found in GitHub.

Geist et al. [46] have introduced their approach which employs Machine Learning, in specific, Deep learning in order to re-document legacy software system from their source code the exploitation of the comments found in the source code. The authors developed the tool based on the approach in one of the well-known automotive companies. However, the generalization aspect of the approach remains in maturing process.

Bhatia et al. [47] proposed an automated tool for code documentation that is based on the ontology-driven development. The authors examine the tool with comparison to a manual tool named WCopyfind. The authors reported that the tool can generate documentation in two types which are targeting human and machine audiences.

Bastos et al. [48] have proposed an approach that aims to help orgnisations in documenting the software project development. The approach employs the ontology methodology. The authors evaluated the approach using a questionnaire circulated to 8 postgraduate students as participants. In addition, the authors reported that the results cannot be generalized due to the low number of participants.

IV. CONCLUSION

In this paper, a systematic review of the topic of software documentation has been conducted. Budgen et al. and Kitchenham [20], [21] methodology of carrying out systematic

review was employed in this research. The main purpose of this research was to investigate the trends and issues related to software documentation in the last 10 years. An important remark is that Software documentation plays a significant role in quality assurance of software [49]. Therefore, it was essential to investigate the research trends in the topic and the related issues. The main conclusions can be summarized in the following points:

1) Software documentation has not been sufficiently considered as a research interest in the last 10 years.

2) There is a collective recognition that documentation is a difficult process and has many problems. In addition, maintenance teams are the effected party with poor documentation.

3) Three types of documentation can be deemed. First is the initial documentation, for example SRS documents. Second are the ongoing documentations, for example TODO lists. Third is the final documentation which includes user manuals.

4) Despite the model being followed in the documentation process, the majority of found primary studies considered the quality of software documentations.

5) Most tools for software documentations are focused on reverse documentations from the source code. This might lead to the loss of the lessons learned and decisions made as they were not previously documented.

6) Automated tools for documentation are in high demand.

7) Despite the importance of software documentation in the quality of software, developers tend not to pay attention to it.

Future research endeavors will focus on developing a comprehensive, standardized model for software documentation that exhibits broad applicability across diverse software systems. Furthermore, this investigation will delve into the identification of ambiguities within existing software documentation typologies and establish interconnections between these categories to facilitate seamless transitions based on the specific developmental phase.

ACKNOWLEDGMENT

The author wishes to express profound appreciation to his family and friends for their unwavering support during the course of this project. Their encouragement and steadfast faith in the author's abilities served as a wellspring of motivation.

The author is also indebted to Umm Al Qura University for providing indispensable resources and cultivating an intellectually stimulating academic environment. These factors significantly contributed to the successful culmination of this research endeavor.

REFERENCES

[1] I. Sommerville, “Software documentation,” Softw. Eng., vol. 2, pp. 143–154, 2001.

[2] I. Sommerville, Software Engineering, 6th edition. Harlow, England ; New York: Addison Wesley, 2000.

[3] I. Sommerville, “Systems engineering for software engineers,” Ann. Softw. Eng., vol. 6, no. 1/4, pp. 111–129, 1998, doi: 10.1023/A:1018901230131.

[4] R. Ries, “IEEE standard for software user documentation,” in International conference on professional communication, communication across the sea: North American and European practices, IEEE, 1990, pp. 66–68. Accessed: Apr. 25, 2024. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/111154/

[5] T. T. Barker, Perspectives on Software Documentation: Inquiries and Innovations. Routledge, 2020.

[6] T. C. Lethbridge, J. Singer, and A. Forward, “How software engineers use documentation: The state of the practice,” IEEE Softw., vol. 20, no. 6, pp. 35–39, 2003.

[7] J. Santos and F. F. Correia, “A Review of Pattern Languages for Software Documentation,” in Proceedings of the European Conference on Pattern Languages of Programs 2020, Virtual Event Germany: ACM, Jul. 2020, pp. 1–14. doi: 10.1145/3424771.3424786.

[8] A. Aguiar and G. David, “Patterns for Effectively Documenting Frameworks,” Trans. Pattern Lang. Program. II, vol. 6510, pp. 79–124, 2011, doi: 10.1007/978-3-642-19432-0_5.

[9] E. Aghajani et al., “Software documentation issues unveiled,” in 2019 IEEE/ACM 41st International Conference on Software Engineering (ICSE), IEEE, 2019, pp. 1199–1210. Accessed: Apr. 24, 2024. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/8811931/

[10] P. Naur and B. Randell, Software Engineering: Report of a conference sponsored by the NATO Science Committee, Garmisch, Germany, 7-11 Oct. 1968, Brussels, Scientific Affairs Division, NATO. 1969. Accessed: Apr. 24, 2024. [Online]. Available: https://dl.acm.org/doi/abs/10.5555/1102020

[11] A. Rüping, Agile Documentation: A Pattern Guide to Producing Lightweight Documents for Software Projects. John Wiley & Sons, 2005.

[12] D. Falessi, G. Cantone, and P. Kruchten, “Value-based design decision rationale documentation: Principles and empirical feasibility study,” in Seventh Working IEEE/IFIP Conference on Software Architecture (WICSA 2008), IEEE, 2008, pp. 189–198. Accessed: Apr. 25, 2024. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/4459157/

[13] D. Falessi, G. Cantone, and M. Becker, “Documenting design decision rationale to improve individual and team design decision making: an experimental evaluation,” in Proceedings of the 2006 ACM/IEEE international symposium on Empirical software engineering, Rio de Janeiro Brazil: ACM, Sep. 2006, pp. 134–143. doi: 10.1145/1159733.1159755.

[14] I. Sommerville, Software Engineering, 10th edition. Boston: Pearson, 2015.

[15] G. Uddin and M. P. Robillard, “How API documentation fails,” Ieee Softw., vol. 32, no. 4, pp. 68–75, 2015.

[16] S. C. B. De Souza, N. Anquetil, and K. M. De Oliveira, “A study of the documentation essential to software maintenance,” in Proceedings of the 23rd annual international conference on Design of communication: documenting & designing for pervasive information, Coventry United Kingdom: ACM, Sep. 2005, pp. 68–75. doi: 10.1145/1085313.1085331.

[17] M. Kajko-Mattsson, “A Survey of Documentation Practice within Corrective Maintenance,” Empir. Softw. Eng., vol. 10, no. 1, pp. 31–55, Jan. 2005, doi: 10.1023/B:LIDA.0000048322.42751.ca.

[18] A. Forward and T. C. Lethbridge, “The relevance of software documentation, tools and technologies: a survey,” in Proceedings of the 2002 ACM symposium on Document engineering, McLean Virginia USA: ACM, Nov. 2002, pp. 26–33. doi: 10.1145/585058.585065.

[19] E. Aghajani et al., “Software documentation: the practitioners’ perspective,” in Proceedings of the ACM/IEEE 42nd International Conference on Software Engineering, Seoul South Korea: ACM, Jun. 2020, pp. 590–601. doi: 10.1145/3377811.3380405.

[20] D. Budgen and P. Brereton, “Performing systematic literature reviews in software engineering,” in Proceedings of the 28th international conference on Software engineering, 2006, pp. 1051–1052.

[21] B. Kitchenham, "Procedure for undertaking systematic reviews," Comput. Sci. Depart-Ment Keele Univ. TRISE-0401 Natl. ICT Aust. Ltd 0400011T 1 Jt. Tech. Rep., 2004.

[22] "Saudi Digital Library (SDL)." Accessed: Apr. 26, 2024. [Online]. Available: https://sdl.edu.sa/SDLPortal/Publishers.aspx

[23] M. P. Robillard and R. DeLine, "A field study of API learning obstacles," Empir. Softw. Eng., vol. 16, no. 6, pp. 703–732, Dec. 2011, doi: 10.1007/s10664-010-9150-8.

[24] G. Garousi, V. Garousi-Yusifoğlu, G. Ruhe, J. Zhi, M. Moussavi, and B. Smith, "Usage and usefulness of technical software documentation: An industrial case study," Inf. Softw. Technol., vol. 57, pp. 664–682, 2015.

[25] J. D. Arthur and K. T. Stevens, "Document quality indicators: A framework for assessing documentation adequacy," J. Softw. Maint. Res. Pract., vol. 4, no. 3, pp. 129–142, Sep. 1992, doi: 10.1002/smr.4360040303.

[26] A. Dautovic, "Automatic assessment of software documentation quality," in 2011 26th IEEE/ACM International Conference on Automated Software Engineering (ASE 2011), IEEE, 2011, pp. 665–669. Accessed: May 10, 2024. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/6100151/

[27] J.-C. Chen and S.-J. Huang, "An empirical analysis of the impact of software development problem factors on software maintainability," J. Syst. Softw., vol. 82, no. 6, pp. 981–992, 2009.

[28] B. Dagenais and M. P. Robillard, "Creating and evolving developer documentation: understanding the decisions of open source contributors," in Proceedings of the eighteenth ACM SIGSOFT international symposium on Foundations of software engineering, Santa Fe New Mexico USA: ACM, Nov. 2010, pp. 127–136. doi: 10.1145/1882291.1882312.

[29] S. Rai, R. C. Belwal, and A. Gupta, "A Review on Source Code Documentation," ACM Trans. Intell. Syst. Technol., vol. 13, no. 5, pp. 1–44, Oct. 2022, doi: 10.1145/3519312.

[30] M. Meng, S. Steinhardt, and A. Schubert, "Application Programming Interface Documentation: What Do Software Developers Want?," J. Tech. Writ. Commun., vol. 48, no. 3, pp. 295–330, Jul. 2018, doi: 10.1177/0047281617721853.

[31] J. Zhi, V. Garousi-Yusifoğlu, B. Sun, G. Garousi, S. Shahnewaz, and G. Ruhe, "Cost, benefits and quality of software development documentation: A systematic mapping," J. Syst. Softw., vol. 99, pp. 175–198, 2015.

[32] F. Bachmann et al., "Software architecture documentation in practice: Documenting architectural layers," 2000, Accessed: Apr. 24, 2024. [Online]. Available: https://www.getforms.org/forms/forms-pdf/5022.pdf

[33] P. C. Véras, E. Villani, A. M. Ambrosio, M. Vieira, and H. Madeira, "A benchmarking process to assess software requirements documentation for space applications," J. Syst. Softw., vol. 100, pp. 103–116, Feb. 2015, doi: 10.1016/j.jss.2014.10.054.

[34] M. Farwick, C. M. Schweda, R. Breu, and I. Hanschke, "A situational method for semi-automated Enterprise Architecture Documentation," Softw. Syst. Model., vol. 15, no. 2, pp. 397–426, May 2016, doi: 10.1007/s10270-014-0407-3.

[35] M. Hauder, F. Matthes, and S. Roth, "Challenges for Automated Enterprise Architecture Documentation," in Trends in Enterprise Architecture Research and Practice-Driven Research on Enterprise Transformation, vol. 131, S. Aier, M. Ekstedt, F. Matthes, E. Proper, and J. L. Sanz, Eds., in Lecture Notes in Business Information Processing, vol. 131. , Berlin, Heidelberg: Springer Berlin Heidelberg, 2012, pp. 21–39. doi: 10.1007/978-3-642-34163-2_2.

[36] A. Mathrani, S. Wickramasinghe, and N. P. Jayamaha, "An evaluation of documentation requirements for ISO 9001 compliance in scrum projects," TQM J., vol. 34, no. 5, pp. 901–921, 2022.

[37] L. Aversano, D. Guardabascio, and M. Tortorella, "Analysis of the documentation of ERP software projects," Procedia Comput. Sci., vol. 121, pp. 423–430, 2017.

[38] N. R. Carvalho, A. Simoes, and J. J. Almeida, "DMOSS: Open source software documentation assessment," Comput. Sci. Inf. Syst., vol. 11, no. 4, pp. 1197–1207, 2014.

[39] T. Theunissen, S. Hoppenbrouwers, and S. Overbeek, "Approaches for documentation in continuous software development," Complex Syst. Inform. Model. Q., no. 32, pp. 1–27, 2022.

[40] G. Rong, Z. Jin, H. Zhang, Y. Zhang, W. Ye, and D. Shao, "DevDocOps: Enabling continuous documentation in alignment with DevOps," Softw. Pract. Exp., vol. 50, no. 3, pp. 210–226, 2020, doi: 10.1002/spe.2770.

[41] M. V. Krunic, "Documentation as code in automotive system/software engineering," Elektron. Ir Elektrotechnika, vol. 29, no. 4, pp. 61–75, 2023.

[42] R. Kazman, D. Goldenson, I. Monarch, W. Nichols, and G. Valetto, "Evaluating the effects of architectural documentation: A case study of a large scale open source project," IEEE Trans. Softw. Eng., vol. 42, no. 3, pp. 220–260, 2015.

[43] C. H. Righolt, B. A. Monchka, and S. M. Mahmud, "From source code to publication: Code Diary, an automatic documentation parser for SAS," SoftwareX, vol. 7, pp. 222–225, Jan. 2018, doi: 10.1016/j.softx.2018.07.002.

[44] E. Abdullah AlOmar, A. Peruma, M. Wiem Mkaouer, C. Newman, A. Ouni, and M. Kessentini, "How We Refactor and How We Document it? On the Use of Supervised Machine Learning Algorithms to Classify Refactoring Documentation," ArXiv E-Prints, p. arXiv-2010, 2020.

[45] E. Abdullah AlOmar et al., "On the Documentation of Refactoring Types," ArXiv E-Prints, p. arXiv-2112, 2021.

[46] V. Geist, M. Moser, J. Pichler, S. Beyer, and M. Pinzger, "Leveraging machine learning for software redocumentation," in 2020 IEEE 27th International Conference on Software Analysis, Evolution and Reengineering (SANER), IEEE, 2020, pp. 622–626. Accessed: May 12, 2024. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/9054838/

[47] M. P. S. Bhatia, A. Kumar, and R. Beniwal, "Ontology Driven Software Development for Automated Documentation.," Webology, vol. 15, no. 2, 2018, Accessed: May 12, 2024. [Online]. Available: https://www.researchgate.net/profile/Rohit-Beniwal-5/publication/331489088_Ontology_Driven_Software_Development_for_Automated_Documentation/links/5c7d1c90458515831f81987c/Ontology-Driven-Software-Development-for-Automated-Documentation.pdf

[48] E. C. Bastos, M. P. Barcellos, and R. De Almeida Falbo, "Using Semantic Documentation to Support Software Project Management," J. Data Semant., vol. 7, no. 2, pp. 107–132, Jun. 2018, doi: 10.1007/s13740-018-0089-z.

[49] J. E. Tyler, "Asset management the track towards quality documentation," Rec. Manag. J., vol. 27, no. 3, pp. 302–317, Jan. 2017, doi: 10.1108/RMJ-11-2015-0039.

APPENDIX A

PRIMARY STUDIES

Item	Bibliography	Sources
1.	T. C. Lethbridge, J. Singer, and A. Forward, "How software engineers use documentation: The state of the practice," IEEE Softw., vol. 20, no. 6, pp. 35–39, 2003.	IEEE
2.	J. Santos and F. F. Correia, "A Review of Pattern Languages for Software Documentation," in Proceedings of the European Conference on Pattern Languages of Programs 2020, Virtual Event Germany: ACM, Jul. 2020, pp. 1–14. doi: 10.1145/3424771.3424786.	ACM
3.	A. Aguiar and G. David, "Patterns for Effectively Documenting Frameworks," Trans. Pattern Lang. Program. II, vol. 6510, pp. 79–124, 2011, doi: 10.1007/978-3-642-19432-0_5.	Other Publishers
4.	E. Aghajani et al., "Software documentation issues unveiled," in 2019 IEEE/ACM 41st International Conference on Software Engineering (ICSE), IEEE, 2019, pp. 1199–1210. Accessed: Apr. 24, 2024. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/8811931/	IEEE
5.	D. Falessi, G. Cantone, and P. Kruchten, "Value-based design decision rationale documentation: Principles and empirical feasibility study," in Seventh Working IEEE/IFIP Conference on Software Architecture (WICSA 2008), IEEE, 2008, pp. 189–198. Accessed: Apr. 25, 2024. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/4459157/	IEEE
6.	D. Falessi, G. Cantone, and M. Becker, "Documenting design decision rationale to improve individual and team design decision making: an experimental evaluation," in Proceedings of the 2006 ACM/IEEE international symposium on Empirical software engineering, Rio de Janeiro Brazil: ACM, Sep. 2006, pp. 134–143. doi: 10.1145/1159733.1159755.	ACM
7.	G. Uddin and M. P. Robillard, "How API documentation fails," Ieee Softw., vol. 32, no. 4, pp. 68–75, 2015.	Other Publishers
8.	S. C. B. De Souza, N. Anquetil, and K. M. De Oliveira, "A study of the documentation essential to software maintenance," in Proceedings of the 23rd annual international conference on Design of communication: documenting & designing for pervasive information, Coventry United Kingdom: ACM, Sep. 2005, pp. 68–75. doi: 10.1145/1085313.1085331.	ACM
9.	M. Kajko-Mattsson, "A Survey of Documentation Practice within Corrective Maintenance," Empir. Softw. Eng., vol. 10, no. 1, pp. 31–55, Jan. 2005, doi: 10.1023/B:LIDA.0000048322.42751.ca.	Other Publishers
10.	A. Forward and T. C. Lethbridge, "The relevance of software documentation, tools and technologies: a survey," in Proceedings of the 2002 ACM symposium on Document engineering, McLean Virginia USA: ACM, Nov. 2002, pp. 26–33. doi: 10.1145/585058.585065.	ACM
11.	E. Aghajani et al., "Software documentation: the practitioners' perspective," in Proceedings of the ACM/IEEE 42nd International Conference on Software Engineering, Seoul South Korea: ACM, Jun. 2020, pp. 590–601. doi: 10.1145/3377811.3380405.	ACM
12.	S. Rai, R. C. Belwal, and A. Gupta, "A Review on Source Code Documentation," ACM Trans. Intell. Syst. Technol., vol. 13, no. 5, pp. 1–44, Oct. 2022, doi: 10.1145/3519312.	ACM
13.	M. Meng, S. Steinhardt, and A. Schubert, "Application Programming Interface Documentation: What Do Software Developers Want?," J. Tech. Writ. Commun., vol. 48, no. 3, pp. 295–330, Jul. 2018, doi: 10.1177/0047281617721853.	Other Publishers
14.	J. Zhi, V. Garousi-Yusifoğlu, B. Sun, G. Garousi, S. Shahnewaz, and G. Ruhe, "Cost, benefits and quality of software development documentation: A systematic mapping," J. Syst. Softw., vol. 99, pp. 175–198, 2015.	Other Publishers
15.	F. Bachmann et al., "Software architecture documentation in practice: Documenting architectural layers," 2000, Accessed: Apr. 24, 2024. [Online]. Available: https://www.getforms.org/forms/forms-pdf/5022.pdf	Other Publishers
16.	P. C. Véras, E. Villani, A. M. Ambrosio, M. Vieira, and H. Madeira, "A benchmarking process to assess software requirements documentation for space applications," J. Syst. Softw., vol. 100, pp. 103–116, Feb. 2015, doi: 10.1016/j.jss.2014.10.054.	Other Publishers
17.	M. Farwick, C. M. Schweda, R. Breu, and I. Hanschke, "A situational method for semi-automated Enterprise Architecture Documentation," Softw. Syst. Model., vol. 15, no. 2, pp. 397–426, May 2016, doi: 10.1007/s10270-014-0407-3.	Other Publishers
18.	A. Mathrani, S. Wickramasinghe, and N. P. Jayamaha, "An evaluation of documentation requirements for ISO 9001 compliance in scrum projects," TQM J., vol. 34, no. 5, pp. 901–921, 2022.	Other Publishers
19.	L. Aversano, D. Guardabascio, and M. Tortorella, "Analysis of the documentation of ERP software projects," Procedia Comput. Sci., vol. 121, pp. 423–430, 2017.	Other Publishers
20.	N. R. Carvalho, A. Simoes, and J. J. Almeida, "DMOSS: Open source software documentation assessment," Comput. Sci. Inf. Syst., vol. 11, no. 4, pp. 1197–1207, 2014.	Other Publishers
21.	T. Theunissen, S. Hoppenbrouwers, and S. Overbeek, "Approaches for documentation in continuous software development," Complex Syst. Inform. Model. Q., no. 32, pp. 1–27, 2022.	Other Publishers
22.	G. Rong, Z. Jin, H. Zhang, Y. Zhang, W. Ye, and D. Shao, "DevDocOps: Enabling continuous documentation in alignment with DevOps," Softw. Pract. Exp., vol. 50, no. 3, pp. 210–226, 2020, doi: 10.1002/spe.2770.	Other Publishers
23.	M. V. Krunic, "Documentation as code in automotive system/software engineering," Elektron. Ir Elektrotechnika, vol. 29, no. 4, pp. 61–75, 2023.	Other Publishers
24.	R. Kazman, D. Goldenson, I. Monarch, W. Nichols, and G. Valetto, "Evaluating the effects of architectural documentation: A case study of a large scale open source project," IEEE Trans. Softw. Eng., vol. 42, no. 3, pp. 220–260, 2015.	IEEE
25.	C. H. Righolt, B. A. Monchka, and S. M. Mahmud, "From source code to publication: Code Diary, an automatic documentation parser for SAS," SoftwareX, vol. 7, pp. 222–225, Jan. 2018, doi: 10.1016/j.softx.2018.07.002.	Other Publishers
26.	E. Abdullah AlOmar, A. Peruma, M. Wiem Mkaouer, C. Newman, A. Ouni, and M. Kessentini, "How We Refactor and How We Document it? On the Use of Supervised Machine Learning Algorithms to Classify Refactoring Documentation," ArXiv E-Prints, p. arXiv-2010, 2020.	Other Publishers
27.	E. Abdullah AlOmar et al., "On the Documentation of Refactoring Types," ArXiv E-Prints, p. arXiv-2112, 2021.	IEEE
28.	V. Geist, M. Moser, J. Pichler, S. Beyer, and M. Pinzger, "Leveraging machine learning for software redocumentation," in 2020 IEEE 27th International Conference on Software Analysis, Evolution and Reengineering (SANER), IEEE, 2020, pp. 622–626. Accessed: May 12, 2024. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/9054838/	ACM
29.	M. P. S. Bhatia, A. Kumar, and R. Beniwal, "Ontology Driven Software Development for Automated Documentation.," Webology, vol. 15, no. 2, 2018, Accessed: May 12, 2024. [Online]. Available: https://www.researchgate.net/profile/Rohit-Beniwal-5/publication/331489088_Ontology_Driven_Software_Development_for_Automated_Documentation/links/5c7d1c90458515831f81987c/Ontology-Driven-Software-Development-for-Automated-Documentation.pdf	Other Publishers

Optimizing Data Security in Computer-Assisted Test Applications Through the Advanced Encryption Standard 256-Bit Cipher Block Chaining

M. Afridon[1], Agus Tedyyana[2], Fajar Ratnawati[3], Afis Julianto[4], M. Nur Faizi[5]
Department of Electrical Engineering, Politeknik Negeri Bengkalis, 28711, Indonesia[1, 5]
Department of Informatic Engineering, Politeknik Negeri Bengkalis, 28711, Indonesia[2, 3, 4]

***Abstract*—In the digital education era, the importance of Computer-Assisted Test programs is underscored by their efficiency in conducting assessments. However, the increasing incidence of data breaches and cyberthreats has made the implementation of robust data protection measures imperative. This study explores the adoption of the Advanced Encryption Standard 256-bit Cipher Block Chaining in CAT applications to enhance data security. Known for its strong encryption capabilities, AES-256-CBC is an excellent choice for securing sensitive test data. The research focuses on the application of AES-256-CBC within CAT systems during the independent admission process at Politeknik Negeri Bengkalis, a critical phase where the integrity of exam materials and student data is paramount. We evaluate the effectiveness of AES-256-CBC in encrypting user data and exam materials across different CAT systems, thus preserving data integrity and confidentiality. The implementation of AES-256-CBC helps prevent unauthorized access and manipulation of test results, ensuring a secure online testing environment. This research not only demonstrates the technical implementation of AES-256-CBC but also assesses its impact on enhancing the security posture of CAT applications at Politeknik Negeri Bengkalis. The findings contribute to the broader discussion on data security in educational technology, positioning AES-256-CBC as a potent solution for maintaining academic integrity in digital testing environments.**

***Keywords*—AES256-CBC; data security; computer-assisted test; academic integrity; encryption standards; digital assessment security**

I. Introduction

The integration of technology into educational processes has become imperative, significantly enhancing learning and assessment procedures but also elevating the risk of cyber threats [1]. As educational institutions increasingly adopt digital platforms for academic assessments and administration, the security of sensitive student data has emerged as a paramount concern. The escalation of complex cyberattacks underscores the urgent need for robust and advanced security systems [2]. A recent analysis by the Center for Strategic and International Studies reveals that by 2023, almost 30% of cyberattacks targeted colleges and schools, making the education sector particularly vulnerable to data breaches [3]. This statistic not only highlights the immediate need for enhanced protective measures, but it also illustrates the magnitude of the threat, urging the implementation of sophisticated security infrastructure to safeguard sensitive and private information. Furthermore, the 2021 data breach at the University of California vividly exposed the susceptibility of university information systems to ransomware attacks [4]. This incident, which compromised the personal information of thousands of students and employees, resulted in significant financial losses and eroded stakeholder confidence, underscoring the necessity for educational institutions to adopt a more thorough and proactive approach to data security. Institutions must ensure their systems not only meet current security standards but are also equipped to anticipate and counter future threats effectively [5].

Literature studies on data security in digital education systems show that the use of encryption technologies such as the Advanced Encryption Standard (AES) 256-bit cipher block chaining (CBC) is becoming crucial. This study shows that the AES-256-CBC provides effective protection against brute force attacks and side attacks, which are two common threats to cybersecurity [6]. This enhanced security has become possible because of the mathematical complexity of the AES-256, which makes it difficult to decrypt without a proper key. Moreover, recent research shows that many educational institutions are still at risk of data leaks because they do not implement adequate security standards [7]. It stresses the need for sustained improvement in data security policies and practices, including better training for information technology managers and system users. Studies conducted around the world show that consistent application of AES-256-CBC results in higher levels of security compared to older encryption algorithms [8]. The study also emphasizes the importance of secure key management and dynamic security policy adaptation to address growing threats. The use of encryption in educational systems not only limits unauthorized access but also guarantees data integrity [9]. This integrity is important not only for data security but also for the trust of stakeholders, which includes students, parents, and teaching staff in addition, the study highlights that an effective security policy should cover more than just the implementation of technical solutions. Aspects such as IT infrastructure physical security, access policies, and emergency response protocols are also critical in ensuring comprehensive data security.

The growing reliance on technology for both teaching and assessment purposes has led to the recognition of data security as a fundamental component of educational integrity. Computer-assisted test (CAT) applications [10], with their simplicity in use and accuracy in real-time scoring, have revolutionized exam conduct in educational settings. However, the digital transformation presents novel challenges, particularly

concerning data security. Since its inception, CAT technology has evolved extensively, enhancing test delivery and analysis in both educational and professional contexts. The latest advancements primarily involve the integration of adaptive testing techniques, which dynamically adjust to an examinee's ability level, thereby providing a more tailored and accurate assessment of skills and knowledge [11].

The integration of modern CAT systems into online learning platforms has enabled seamless interactions between the testing interface and educational content, fostering a more cohesive learning and assessment experience [12]. We design these systems to be versatile, accommodating various item types and testing strategies to address diverse educational needs. Despite these advantages, CAT systems face several operational challenges, such as calibrating the item pool, which requires initial testing on a large sample of examinees to ensure the reliability and validity of test items. Additionally, the design of CAT systems must consider factors such as test security, fairness, and the potential for test-taker manipulation. The need for advanced software and psychometric expertise to develop and maintain these systems underscores the difficulty of effectively implementing adaptive testing processes.

Amidst increasing incidents of data breaches and cyber threats, CAT applications have become frequent targets of cyberattacks due to their storage of sensitive and crucial data, such as student personal information and test results [13]. This vulnerability highlights the importance of expanding and consolidating data security measures to protect such sensitive data comprehensively [14]. This study looks at how to use the AES-256-CBC [15], which is well-known for its strong encryption and excellent defense against collision and pre-image attacks. This makes it a great choice for keeping sensitive data safe in CAT systems.

The focus of this research also extends to the application of AES-256-CBC at Politeknik Negeri Bengkalis during the process of admitting new students through independent tracks. This examination provides a detailed view of the application of data security technologies in Indonesia's higher education system, underscoring the responsibility of educational institutions to protect student data. The integration of AES-256-CBC not only brings technical improvements but also enhances confidence among students and other stakeholders, reinforcing the notion that robust data security can significantly improve an organization's reputation and foster a secure environment for both students and teachers when utilizing technology for exams.

This study aims to offer guidance to other educational institutions seeking to enhance the security of their examination applications by analyzing the implementation of AES-256-CBC at Politeknik Negeri Bengkalis. The insights derived from this analysis are valuable not only at the local level, but also globally, as they contribute to the broader discourse on cybersecurity threats in education. By strengthening data security measures, educational institutions can concentrate more on conducting learning and evaluation processes that are not only efficient but also secure, fostering an environment conducive to innovation and technological integration [16]. This research serves as a benchmark for developing cybersecurity policies and best practices in education, enabling policymakers and administrators to formulate more effective data security strategies based on clear, evidence-based guidelines.

The continuous evolution of cyber threats further underscores the robustness of data security in educational technology, necessitating an ongoing assessment and adaptation of security protocols [17]. As we embrace digital tools in education, it becomes crucial to implement systems that not only react to breaches but also proactively prevent them. This dual approach ensures that the security architecture evolves in parallel with emerging threats, maintaining the integrity and confidentiality of student data at all times. Ensuring the highest standards of data protection not only complies with regulatory demands but also addresses the ethical responsibility educational institutions hold towards their constituents [18]. Moreover, the strategic application of technologies such as AES-256-CBC in the education sector can serve as a model for other sectors where data sensitivity is paramount. By showcasing effective strategies for safeguarding data within the rigorous and often targeted environment of educational institutions, we can demonstrate the feasibility and effectiveness of advanced encryption methods. This initiative not only mitigates risks associated with data breaches but also advances the discourse on data security practices, fostering a broader understanding and implementation [19] of best practices across various domains.

The implementation of the AES-256-CBC algorithm significantly reduces the risk of data breaches and cyberattacks, ensuring that sensitive data, including student personal information and test results, remains protected from unauthorized access and manipulation. Enhanced data security also builds trust between educational institutions and their stakeholders, including students, parents, and teaching staff. This trust is essential for creating a positive and supportive learning environment where students and parents feel secure, and teachers are confident in using technology to manage exams safely. As educational institutions continue to face an increasing number of cyber threats, improved protection mitigates potential financial and reputational losses while simultaneously enhancing the learning experience through the safe and diverse use of digital technologies in education. This comprehensive approach to cybersecurity in educational settings not only secures data but also enriches the educational journey for all participants.

II. Materials and Method

A. Setting and Sample Selection Study

The study was conducted at the Politeknik Negeri Bengkalis, focusing on the use of CAT applications for the process of independent admission of new students in the 2024/2025 academic year. The samples include the CAT system currently in use at Politeknik Negeri Bengkalis, along with the data processing practices observed during the cycle of new student admission in 2024-2025. Administrators, IT staff, and prospective students are involved in providing insight into operational and security aspects.

B. Encryption Methodology

To secure the data processed through CAT applications, the Advanced Encryption Standard 256-bit Cipher Block Chaining

was implemented. This section details the critical data points within CAT applications that required encryption and describes the collaborative process with the IT department to integrate AES-256-CBC, replacing the previous encryption method.

C. Cipher Block Chaining Process

To explore the implementation of AES-256-CBC, we first identified critical data points in CAT applications that require encryption. We then integrated AES-256-CBC into the existing CAT system, replacing the previous encryption method. This process involves collaboration with the IT department to ensure that all technical aspects are dealt with, including key management and system compatibility.

Fig. 1, Encryption process using CBC, illustrates the encryption process using CBC model, a widely used method in cryptographic systems for securing data [20]. This mode of operation is particularly effective in enhancing data security by linking blocks of plaintext to produce a chain of ciphertext, ensuring that similar plaintext blocks result in different ciphertext blocks.

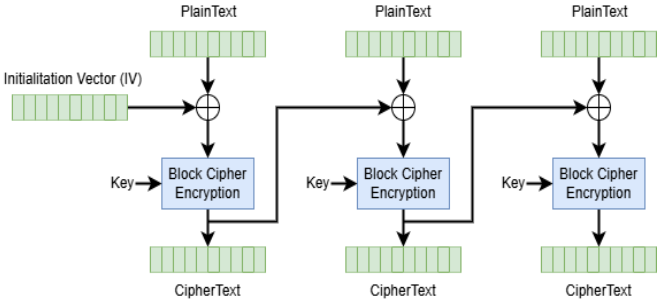


Fig. 1. Encryption process using CBC.

In CBC mode, the encryption process begins with an initialization vector (IV) [21]. To prevent ciphertext patterns, the IV must be unique and unpredictable for each encryption session. Unlike the encryption key, the IV does not need to be secret, but it should be random and not reused with the same key to maintain security.

The process starts with the IV, which is XORed (exclusive OR operation) with the first block of plaintext. This step is crucial as it mask the first block of plaintext, which adds an additional layer of security and ensures that the same plaintext blocks will produce different ciphertext blocks when encrypted under the same key but with different IV [22]. Following the initial XOR operation, a block cipher algorithm encrypts the resulting block using a specified encryption key. This encryption results in the first block of ciphertext.

The ciphertext from the previous block serves as the "new IV" for each subsequent plaintext block. We XOR this blockwith the next plaintext block, then use the same block cipher algorithm and key to encrypt the result. This chaining mechanism ensures that each block of ciphertext is dependent not only on the current plaintext block but also on all preceding plaintext blocks.

Each encryption step produces the ciphertext associated with each plaintext block [23]. Each piece of ciphertext is dependent on the initial IV and the sequence of plaintext blocks, creating a chain in which the correct decryption of each block requires the ciphertext of the preceding block (except for the first block, which requires the IV).

CBC mode's approach, where the encryption of each plaintext block is dependent on the previous ciphertext block, significantly increases security by introducing complexity and randomness into the process [24]. This method prevents plaintext patterns from appearing in the ciphertext, making it more resilient to cryptographic attacks such as pattern analysis. The CBC mode is highly regarded for its ability to propagate errors, meaning that a single bit error in a block of ciphertext will render that block and the following block indecipherable, which can be a useful security feature or a drawback, depending on the context of use.

Fig. 2 illustrates the decryption process using the CBC mode, a common operational mode for block cipher encryption algorithms. The enhanced security features of this method, which take advantage of the dependencies between encrypted data blocks, make it popular [25]. The CBC decryption process starts with the use of an initialization vector This IV is crucial, as it pairs with the first block of ciphertext to initiate the decryption process [26]. To ensure accurate output, the IV must match the one used during the encryption phase. Despite not being secret, the IV must be unique for each encryption session and not reuse the same key. We decrypt each block of ciphertext using the same key and block cipher algorithm as during the encryption. We decrypt and XOR the first block of ciphertext with the IV to create the first plaintext block. This operation transforms the decrypted data back to its original form before encryption.

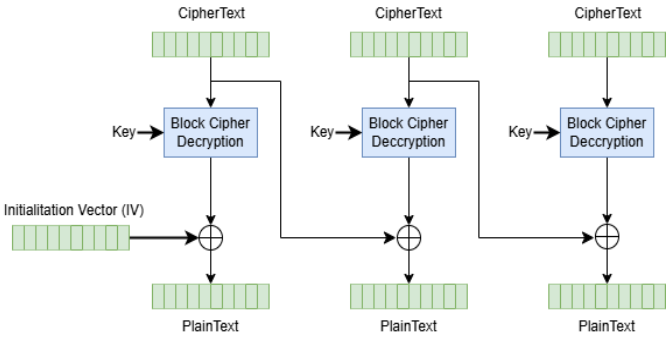


Fig. 2. Decryption process using CBC.

For subsequent blocks, the decryption process follows a chaining mechanism where each decrypted block is XORed with the previous ciphertext block. This sequential processing ensures that any error in a single block of ciphertext during transmission affects not only the current block but also the subsequent block, highlighting a unique dependency characteristic of CBC mode. The process sequentially reveals each piece of plaintext, mirroring the encryption steps in reverse order. This chaining method, where each block's decryption depends not only on its corresponding ciphertext block but also on the preceding ciphertext block, considerably enhances the security of the data transmission. It ensures that the plaintext pattern does not directly influence the ciphertext, making it more resistant to various cryptographic attacks.

The CBC mode's reliance on correct and secure handling of the IV and the chaining mechanism significantly increases system security [27][28]. However, it also introduces certain challenges, such as error propagation and the necessity for secure IV management. These factors must be carefully considered to ensure data integrity and confidentiality throughout its lifecycle. Overall, the CBC mode's decryption process effectively illustrates how cryptographic techniques can enhance data security by intricately linking each block of data to its predecessor, thereby securing the data against unauthorized access and potential security breaches [29].

D. Data Collection

Data collection methods include structured interviews with IT staff responsible for managing the CAT system currently in use at Bengkalis State Polytechnic. In addition, the system logs are reviewed to detect unauthorized access attempts and data breach incidents before and after the AES-256-CBC implementation.

E. Metric Evaluation

The effectiveness of the AES-256-CBC implementation is assessed using several metrics:

1) Comparing the frequency and nature of security incidents before and after implementation.

2) Checks any case of data abnormalities or loss by checking system logs and backup files.

3) Measures changes in system response time and stability to evaluate the impact of AES-256-CBC on the operating efficiency of CAT applications.

4) Conduct surveys with students and staff to measure their confidence in their data security after implementation.

F. Data Analysis

Data analysis in this study employs a blend of quantitative and qualitative methods to gain comprehensive insights into the system's performance and user experiences. We scrutinize system logs and performance metrics using quantitative techniques to objectively assess the enhancements made by integrating AES-256-CBC encryption into the CAT applications. This involves evaluating changes in system response times, error rates, and other relevant performance indicators that directly reflect the operational impact of the encryption methods implemented. We analyze interviews and focus group discussions on the qualitative side to capture the subjective perspectives of end-users and administrators. Understanding how users perceive these security improvements, including any changes in their satisfaction and trust in the system's security measures, is crucial. These interactions' responses help illustrate the practical implications of AES-256-CBC encryption on daily operations and user interactions with the CAT system.

This two-pronged approach tries to connect the technical improvements in security, like AES-256-CBC encryption, with how users feel about it and how well it works. By doing so, the study provides a holistic view of the impact of this encryption technology on CAT applications [30]. It also explores the balance between enhanced security measures and their real-world usability and acceptance, thereby offering valuable insights into both the effectiveness and the user experience of the upgraded system. This comprehensive analysis aids in determining if the security improvements align with user expectations and operational needs, ensuring that the technology not only secures the data but also enhances the overall functionality of the CAT system.

In order to provide a thorough understanding of the technical performance and user experience related to the AES-256-CBC implementation in the CAT applications, this study used a combination of quantitative and qualitative methodologies. We gathered and examined system logs and performance indicators with extreme care in order to assess the effects of AES-256-CBC encryption impartially. Prior to and following the encryption deployment, the system reaction times, error rates, and frequency of security events were among the key performance measures. We were able to evaluate the real-world advantages of incorporating cutting-edge encryption techniques into current educational technologies because this data gave us a quantitative assessment of the system's security resilience and operational effectiveness. We obtained qualitative insights through focus groups and structured interviews with end users, including students, administrative staff, and IT personnel at Politeknik Negeri Bengkalis, to supplement the quantitative data. The goal of these talks was to comprehend the differing viewpoints regarding the security enhancements brought about by AES-256-CBC. We specifically looked at user satisfaction, security perception, and confidence in the system's ability to safeguard private data. This qualitative feedback heavily influences the adoption and usability of the security features among the users who are directly engaging with the CAT system.

The combination of quantitative and qualitative data allowed for an in-depth examination of the encryption's efficacy [31]. Through the integration of results from both data streams, the study obtained a comprehensive understanding of the implementation's effects. This dual method helps identify any differences between the perceived and actual performance of the security measures, in addition to helping validate the technical measurements with real-world user feedback. The combined analysis has made it possible to better understand how well AES-256-CBC encryption satisfies the operational requirements and security expectations of educational institutions. It also indicated areas in which user training and technology implementation still needed improvement. These understandings are essential for creating focused plans to strengthen user confidence in digital learning environments and improve data security procedures.

III. Results and Discussion

A. Login Interface CAT

The login screen is the first gateway and the main line of defense inining the security of CAT applications in the Bengkalis State Polytechnic. This interface is designed not only to facilitate easy access for authenticated users but also to ensure that sensitive data and user personal information are protected from unauthorized access.

Fig. 3 is a login screen on the Politeknik Negeri Bengkalis CAT application displays a simple yet effective design, which includes fields for entering usernames and passwords. The

"Show Password" feature is provided to help users verify the characters they enter, reducing the risk of input errors that can hinder the login process. Security is deeply integrated into this design through the use of HTTPS to encrypt communications between clients and servers, as well as a security policy that ensures passwords are stored in encryption formats on the server, using state-of-the-art cryptography technologies such as AES-256-CBC.

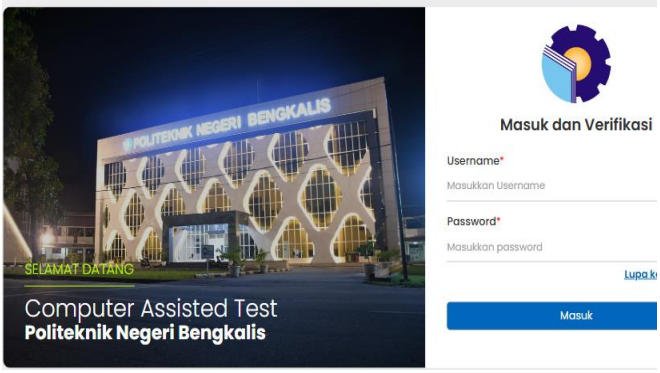


Fig. 3. Login interface CAT.

The interface also comes with additional security mechanisms such as limiting failed login attempts and security features to detect and respond to suspicious activity. It aims to prevent brute-force attacks and the use of stolen credentials, ensuring that only verified users can access the system. Every detail in the login system is designed to improve the overall security of the CAT application. For example, the user session is encrypted end-to-end, and the timeout is implemented automatically to reduce the risk of unauthorized access if the user forgets to log out. In addition, each login activity is recorded in the server log for security audits that allow real-time monitoring of suspicious activity.

B. Encryption Implementation

The PHP code is a crucial part of the security system for the CAT application at Politeknik Negeri Bengkalis. This code details the implementation of the AES-256-CBC algorithm, used for encrypting and decrypting sensitive data within the application. High security standards protect all data stored or transmitted through the CAT system thanks to this implementation.

```php
<?php
// Fungsi untuk mengenkripsi data
function encrypt($data, $key) {
    $iv = openssl_random_pseudo_bytes(openssl_cipher_iv_length('aes-256-cbc'));
    $encrypted = openssl_encrypt($data, 'aes-256-cbc', $key, 0, $iv);
    return base64_encode($encrypted . '::' . $iv);
}

// Fungsi untuk mendekripsi data
function decrypt($data, $key) {
    list($encrypted_data, $iv) = explode('::', base64_decode($data), 2);
    return openssl_decrypt($encrypted_data, 'aes-256-cbc', $key, 0, $iv);
}

// Kunci enkripsi (panjang kunci harus 32 byte untuk AES-256)
$key = '/Ce2c7w44n496kF3VKOXQEkIWyrfyKZzb+DTmLQ7TCM=';
```

Fig. 4. The PHP code AES-256-CBC algorithm.

In Fig. 4, the code consists of two main functions: encrypt() and decrypt(). The encrypt() function utilizes the AES-256-CBC algorithm to encrypt data. The function requires two parameters: the encrypted data and a secret key. Additionally, the code uses openssl_random_pseudo_bytes() to generate a random initialization vector, enhancing security and adding randomness to the encryption process. The system then encodes the encrypted data into Base64 format to facilitate its storage and transmission. The decrypt() function acts as the inverse of encrypt(). It decodes the encrypted data in Base64 format, then uses the same and secret key to decrypt it back to its original form. We strictly safeguard the encryption key, which is critical for both processes, to ensure security.

C. Implications of Security Measures

Politeknik Negeri Bengkalis uses AES-256-CBC to protect all information, including student personal data, exam answers, and evaluation results, from unauthorized access. This algorithm provides excellent protection against various cyberattacks due to the strength of its key and the complexity of the cipher used.

id	id_peserta	nilai_pilihan_ganda
1	16	eyJpdiI6InhOT3RxL09peDJZNkk0TVk5VUpVOWc9PSIsInZhbH...
2	70	eyJpdiI6InNiWkRTbDBqbEdYN000RDkxRU54R3c9PSIsInZhbH...
3	37	eyJpdiI6InpHcDhTbVpBUk1HbHB4cjlWV1RBbVE9PSIsInZhbH...
4	30	eyJpdiI6ImdkeFRUV2pmR2U3TS9WSko5YTh5ZEE9PSIsInZhbH...
5	60	eyJpdiI6Ill1TjkvN0lpSXRSNkpzUFFpRGJWZkE9PSIsInZhbH...
6	79	eyJpdiI6IkZzYlhXT2FHN01Qd3FJb1NaMDJXOFE9PSIsInZhbH...
7	83	eyJpdiI6IjZGRkdUWlZxYVRTSFhwalVYMjVsYlE9PSIsInZhbH...
8	69	eyJpdiI6IkJYSmlGYkZHNDZNNXpqME5jcmx4V0E9PSIsInZhbH...
9	77	eyJpdiI6Ikg3SVhtbi94WGFwbmxYTE5GR05OV2c9PSIsInZhbH...

Fig. 5. List of test identities.

Fig. 5, List of test identities shows a list of test identities matched to their encrypted data. Each line represents a unique set of data related to the test subject, which is secured using the AES-256-CBC encryption algorithm. The use of this encrypting not only protects the data from unauthorized access, but also ensures that each entry is unique, reducing the risk of data leakage or unauthorised modification.

Any information (examination ID and associated details) is encrypted, turning sensitive information into a format that can only be decryptable and understood by the system and authorized personnel. The AES-256-CBC encryption standard is used, which is known for its strength and resilience to a variety of cyber threats, including brute-force attacks and decryption attempts without proper key.

Fig. 6, Test question and its answer options are encrypted. This demonstrates how encryption technology can secure sensitive educational data, such as test results and answer choices. Politeknik Negeri Bengkalis applies encryption to all data elements in CAT applications, reaffirming their efforts to protect the integrity and confidentiality of academic information.



no	soal	pilihan_a	pilihan_b	pilihan_c	pilihan_d
1	eyJpdiI6IkZVWWNRZ3VXZzRXVTlPZDdUM2RvdVE9PSIsInZhbHVlIjoiY	eyJpdiI6IlhV	eyJpdiI6Im	eyJpdiI6IlB	eyJpdiI6Im
2	eyJpdiI6Im1xL2FlclVNMDRVSndueW03QmI2bkE9PSIsInZhbHVlIjoiZD	eyJpdiI6Im	eyJpdiI6Ii9	eyJpdiI6InI	eyJpdiI6Ik
3	eyJpdiI6ImlmQ1JxN3Vka2liTEgrSEFHVUxzbGc9PSIsInZhbHVlIjoiRFhSe	eyJpdiI6Im	eyJpdiI6IjC	eyJpdiI6Ikl	eyJpdiI6Ik
4	eyJpdiI6IjJNcmp4NHFaRjVpdElxSU9wdG4zVVE9PSIsInZhbHVlIjoiY0V	eyJpdiI6Im	eyJpdiI6Ikt	eyJpdiI6Ikl	eyJpdiI6Ijl
5	eyJpdiI6IlVOdS9BYVZOSGJWWFNCcHRTZVVDVlE9PSIsInZhbHVlIjoiRC	eyJpdiI6In	eyJpdiI6Ikx	eyJpdiI6IjL	eyJpdiI6Im
6	eyJpdiI6InA5SWRvK3pEUEF2ajBHMlBCR1I1SGc9PSIsInZhbHVlIjoiVGU	eyJpdiI6Inl	eyJpdiI6Iit	eyJpdiI6Ilk	eyJpdiI6In
7	eyJpdiI6IlFQUjJMWWg0cGNPMGV4b2c4N2x4SWc9PSIsInZhbHVlIjoiY	eyJpdiI6IlN	eyJpdiI6In'	eyJpdiI6IjN	eyJpdiI6Iiz
8	eyJpdiI6Ik9keElJQ2ExWGJlU3ZWY0RZTjN6Wnc9PSIsInZhbHVlIjoiL09	eyJpdiI6Inl	eyJpdiI6Ik:	eyJpdiI6Ikt	eyJpdiI6Ijl
9	eyJpdiI6Ik5wcmM1SFhyLzVoNTJNWXRSOHhBbWc9PSIsInZhbHVlIjoi	eyJpdiI6IlJ	eyJpdiI6IjL	eyJpdiI6Im	eyJpdiI6Il
10	eyJpdiI6IkRZNUpQbk01Qld5WWJJSzNZclF0R0E9PSIsInZhbHVlIjoiZi92	eyJpdiI6Ijc	eyJpdiI6Ikl	eyJpdiI6IjV	eyJpdiI6Im
11	eyJpdiI6Ik8vS1o2bFUyR2hmMGZKNGJYOHNFYUE9PSIsInZhbHVlIjoiN	eyJpdiI6Ill	eyJpdiI6In'	eyJpdiI6In:	eyJpdiI6Ijl
12	eyJpdiI6IkdFSk02VXZkR3JDaHZlcUo5SGM5bXc9PSIsInZhbHVlIjoidlplS	eyJpdiI6Ikk	eyJpdiI6Im	eyJpdiI6Im	eyJpdiI6Ill
13	eyJpdiI6IjcvdHBkdDJUbEg1YU1sSUkrZjZkZUE9PSIsInZhbHVlIjoid3FZZ	eyJpdiI6Im	eyJpdiI6IjR	eyJpdiI6Inl	eyJpdiI6Iiz
14	eyJpdiI6IlJ1eDF3M3VhSXlJeFFzbkU4MWJTQWc9PSIsInZhbHVlIjoidlZl	eyJpdiI6Ikt	eyJpdiI6IlY	eyJpdiI6Iit	eyJpdiI6Ijl
15	eyJpdiI6Ikl3NDB4Mk1ZMWprOTc5SDArbDFxU1E9PSIsInZhbHVlIjoiRV	eyJpdiI6Ikk	eyJpdiI6Im	eyJpdiI6Im	eyJpdiI6Ijl
16	eyJpdiI6In75b2Vz7V7CbnV2dmlmazNEa1hPVGc9PSIsInZhbHVlIjoiN0k	eyJpdiI6Iit	eyJpdiI6In	eyJpdiI6IlV	eyJpdiI6Iil

Fig. 6. Test question and its answer options are encrypted.

In practice, the test management system encrypts each test question and its answer options before storing them in a database or displaying them. This process ensures that only individuals who have a valid decryption key, such as a system administrator or authorized developer, can access the actual data content. This is critical to preventing the leakage of examination information, which could result in a loss of academic integrity and fairness. Data encryption also helps to comply with strict data protection regulations, ensuring that educational institutions meet their legal obligations to student data security and privacy. It not only increases stakeholder confidence in the educational system used, but also strengthens the institution's reputation as a responsible and secure entity. Politeknik Negeri Bengkalis demonstrates, through the use of advanced encryption technology, how technology can enhance security in an educational environment, protect sensitive data from external and internal threats, and enhance a secure learning experience for all parties involved.

Fig. 7, test question and its answer options are decrypt in the exam scenario using the CAT system at Bengkalis State Polytechnic, the subjects presented to the participants have already undergone the process of decryption so that they appear in the form of text that is readable and accessible by the participants. This process describes how data previously secured with AES-256-CBC encryption is converted back to its original format for use during the test.

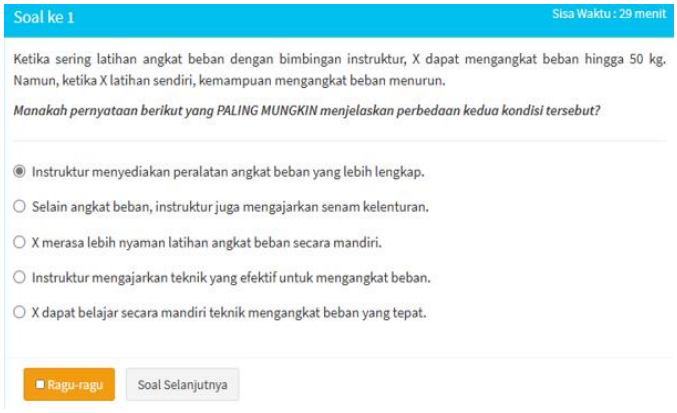


Fig. 7. Test question and its answer options are decrypted.

D. Administrative Dashboard Functionality

This interface display shows the effectiveness of CAT systems in managing and presenting test issues in a secure and organized manner. By ensuring that all subjects are encrypted during storage and only decoded during examination, Bengkalis State Polytechnic demonstrates its commitment to data security and academic integrity. Participants can answer questions with confidence, knowing that the system they use supports them with secure and sophisticated technology. It not only improves the test experience for participants but also affirms the importance of data security in an educational context.



Fig. 8. Administrative dashboard for the CAT system.

Fig. 8 shows a view of the administrative dashboard for the CAT system in Politeknik Negeri Bengkalis. The dashboard serves as a control center for system administrators in managing various aspects of the CAT system. Here is a narrative about the functions and components of this admin dashboard:

1) The dashboard is designed to provide quick and easy access to the range of administrative features needed to manage a computer-based test system. As a command center, the dashboard allows administrators to monitor and control the operation of the test system efficiently, ensuring that everything runs according to the established standards.

2) Provides an overview of system status and current activity.

3) Manages the modules or categories of tests available in the system.

4) Management of test participants' data, including registration and active status.

5) A place to configure and manage test questions and answers.

6) Generate reports related to various aspects of the test, including participant performance and data analysis.

7) Additional tools for system administration such as data backup, security settings, etc.

8) Ensures that the server time matches the current time. This function is important to ensure that the time recording during the test is done accurately. If there is a time difference, the administrator is instructed to check and adjust the server's time zone according to the system configuration.

These dashboards not only simplify the administration and management of tests but also ensure transparency and accuracy in the execution of tests. By leveraging these dashbords, administrators can reduce the risk of human error, improve

operational efficiency, and provide a better experience for users and test participants. In addition, accurate server time integration and centralized data management help in ensuring the integrity and reliability of the test system.

IV. CONCLUSION

This study successfully integrated and evaluated the AES-256-CBC within the CAT systems at Politeknik Negeri Bengkalis. The findings conclusively demonstrate that the implementation of AES-256-CBC significantly bolstered data security, sharply reducing the risk of unauthorized access and significantly strengthening user trust in the system's integrity. However, the research also revealed that the success of such security technology is not solely dependent on the strength of its encryption, but equally on the awareness and training of the involved users. Given the ever-evolving landscape of cyber threats, future research needs to go beyond the use of AES-256-CBC to explore other encryption technologies that might offer superior efficiency or security in an educational context. Additional studies are crucial to assess how effective security training can enhance cybersecurity awareness among CAT system users, including staff and students. Furthermore, understanding the psychological impact of data security, such as how security perceptions influence user trust and satisfaction, will provide valuable insights into improving user interactions with security technologies.

Moreover, there is a significant opportunity for innovation in the development and testing of security tools specifically designed for educational systems, which could further refine our methods for protecting sensitive information. Moving forward, continuous collaboration among security experts, educators, and IT technicians will be essential to ensure that our security infrastructure can adapt to evolving threats while supporting educational goals and innovation. This study underscores the critical need for robust encryption methods like AES-256-CBC in safeguarding educational data systems against increasing cyber threats. By firmly integrating advanced security measures, educational institutions can better protect both their operational integrity and the private information of their stakeholders. In turn, this commitment to high-standard security practices not only enhances the functionality of CAT applications but also fortifies the trust placed in them by students, educators, and administrative personnel alike.

To sustain and build upon the successes of this study, the next phase of research should also investigate the scalability of AES-256-CBC across different educational platforms and its effectiveness against a broader array of cyber-attacks. Exploring the integration of multi-factor authentication measures with AES-256-CBC could provide an additional layer of security, further enhancing educational data systems' resilience. These efforts will ensure that our technological defenses not only keep pace with cyber threats but also contribute to a secure and conducive learning environment.

REFERENCES

[1] A. Nusi and M. Zaim, "Philosophy of Education In Digital Transformation: Ethical Considerations For Students' Data Security In Online Learning Platforms," Jurnal Ilmiah Pendidikan Scholastic, vol. 7, no. 3, pp. 42–50, Dec. 2023, doi: 10.36057/jips.v7i3.629.

[2] G. K. Sudhina Kumar, K. Krishna Prakasha, and B. Muniyal, "ACH Reference Model- A model of Architecture to Handle Advanced Cyberattacks," in 2022 Second International Conference on Advances in Electrical, Computing, Communication and Sustainable Technologies (ICAECT), IEEE, Apr. 2022, pp. 1–6. doi: 10.1109/ICAECT54875.2022.9808076.

[3] O. Trofymenko, N. Loginova, M. Serhii, and Y. Dubovoil, "CYBERTHREATS IN HIGHER EDUCATION," Cybersecurity: Education, Science, Technique, vol. 4, no. 16, pp. 76–84, 2022, doi: 10.28925/2663-4023.2022.16.7684.

[4] D. Kotis and C. Rath, "Strengthening our defenses: The role of the health - system pharmacist in cybersecurity management," JACCP: JOURNAL OF THE AMERICAN COLLEGE OF CLINICAL PHARMACY, vol. 4, no. 6, pp. 662 - 663, Jun. 2021, doi: 10.1002/jac5.1463.

[5] S. W. A. Hamdani et al., "Cybersecurity Standards in the Context of Operating System," ACM Comput Surv, vol. 54, no. 3, pp. 1–36, Apr. 2022, doi: 10.1145/3442480.

[6] A. Carlson, I. Dutta, and B. Ghosh, "Using the Collision Attack for Breaking Cryptographic Modes," in 2022 13th International Conference on Computing Communication and Networking Technologies (ICCCNT), IEEE, Oct. 2022, pp. 1–7. doi: 10.1109/ICCCNT54827.2022.9984325.

[7] A. Mohammed et al., "Data Security And Protection: A Mechanism For Managing Data Theft and Cybercrime in Online Platforms Of Educational Institutions," in 2022 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COM-IT-CON), IEEE, May 2022, pp. 758–761. doi: 10.1109/COM-IT-CON54601.2022.9850702.

[8] Y. S. Alslman, A. Ahmad, and Y. AbuHour, "Enhanced and authenticated cipher block chaining mode," Bulletin of Electrical Engineering and Informatics, vol. 12, no. 4, pp. 2357–2362, Aug. 2023, doi: 10.11591/beei.v12i4.5113.

[9] S. Hina and P. D. D. Dominic, "Information security policies' compliance: a perspective for higher education institutions," Journal of Computer Information Systems, vol. 60, no. 3, pp. 201–211, May 2020, doi: 10.1080/08874417.2018.1432996.

[10] R. Amalia K, A. Wahana, and S. Wardani, "System CAT (Computer Assisted Test) information for Multimedia Department of Muhammadiyah Vocational High School 2 Moyudan Web-based," APPLIED SCIENCE AND TECHNOLOGY REASERCH JOURNAL, vol. 2, no. 1, pp. 30–36, May 2023, doi: 10.31316/astro.v2i1.5048.

[11] J. I. Oladele and M. Ndlovu, "A Review of Standardised Assessment Development Procedure and Algorithms for Computer Adaptive Testing: Applications and Relevance for Fourth Industrial Revolution," International Journal of Learning, Teaching and Educational Research, vol. 20, no. 5, pp. 1–17, May 2021, doi: 10.26803/ijlter.20.5.1.

[12] Y. Choi and C. McClenen, "Development of Adaptive Formative Assessment System Using Computerized Adaptive Testing and Dynamic Bayesian Networks," Applied Sciences, vol. 10, no. 22, p. 8196, Nov. 2020, doi: 10.3390/app10228196.

[13] X. Zhang, M. Xu, G. Da, and P. Zhao, "Ensuring confidentiality and availability of sensitive data over a network system under cyber threats," Reliab Eng Syst Saf, vol. 214, p. 107697, Oct. 2021, doi: 10.1016/j.ress.2021.107697.

[14] A. H. Mahmoud, H. H. Issa, N. H. Shaker, and K. A. Shehata, "Customized AES for Securing Data in Sensitive Networks and Applications," in 2022 39th National Radio Science Conference (NRSC), IEEE, Nov. 2022, pp. 164–170. doi: 10.1109/NRSC57219.2022.9971420.

[15] S. B. George, S. Jaimy, S. Jose, E. Daji, and A. Antony, "A Novel Model to Overcome Drawbacks of Present Cloud Storage Models using AES 256 CBC Encryption," Int J Comput Appl, vol. 183, no. 15, pp. 30–35, Jul. 2021, doi: 10.5120/ijca2021921481.

[16] W. Yaokumah and A. A. Dawson, "Network and Data Transfer Security Management in Higher Educational Institutions," in Research Anthology on Business Aspects of Cybersecurity, IGI Global, 2022, pp. 514–532. doi: 10.4018/978-1-6684-3698-1.ch024.

[17] X. Yin and Y. Chen, "Cyber Risk Recommendation System for Digital Education Management Platforms," Comput Intell Neurosci, vol. 2022, pp. 1–11, Apr. 2022, doi: 10.1155/2022/8548534.

[18] D. Florea and S. Florea, "Big Data and the Ethical Implications of Data Privacy in Higher Education Research," Sustainability, vol. 12, no. 20, p. 8744, Oct. 2020, doi: 10.3390/su12208744.

[19] D. M and J. Dhiipan, "A Meta-Analysis of Efficient Countermeasures for Data Security," in 2022 International Conference on Automation, Computing and Renewable Systems (ICACRS), IEEE, Dec. 2022, pp. 1303–1308. doi: 10.1109/ICACRS55517.2022.10029302.

[20] F. Dridi, S. El Assad, W. El Hadj Youssef, M. Machhout, and R. Lozi, "Design, Implementation, and Analysis of a Block Cipher Based on a Secure Chaotic Generator," Applied Sciences, vol. 12, no. 19, p. 9952, Oct. 2022, doi: 10.3390/app12199952.

[21] H. T. Assafli and I. A. Hashim, "Security Enhancement of AES-CBC and its Performance Evaluation Using the Avalanche Effect," in 2020 3rd International Conference on Engineering Technology and its Applications (IICETA), IEEE, Sep. 2020, pp. 7–11. doi: 10.1109/IICETA50496.2020.9318803.

[22] M. Shan, L. Liu, B. Liu, and Z. Zhong, "Security enhanced cascaded phase encoding based on a 3D phase retrieval algorithm," Opt Lasers Eng, vol. 145, p. 106662, Oct. 2021, doi: 10.1016/j.optlaseng.2021.106662.

[23] R. Abu Zitar and M. J. Al-Muhammed, "Hybrid encryption technique: Integrating the neural network with distortion techniques," PLoS One, vol. 17, no. 9, p. e0274947, Sep. 2022, doi: 10.1371/journal.pone.0274947.

[24] Y. S. Alslman, A. Ahmad, and Y. AbuHour, "Enhanced and authenticated cipher block chaining mode," Bulletin of Electrical Engineering and Informatics, vol. 12, no. 4, pp. 2357–2362, Aug. 2023, doi: 10.11591/beei.v12i4.5113.

[25] Y. S. Alslman, A. Ahmad, and Y. AbuHour, "Enhanced and authenticated cipher block chaining mode," Bulletin of Electrical Engineering and Informatics, vol. 12, no. 4, pp. 2357–2362, Aug. 2023, doi: 10.11591/beei.v12i4.5113.

[26] C. A. Novianti, M. Khudzaifah, and M. N. Jauhari, "Kriptografi Hibrida Cipher Block Chaining (CBC) dan Merkle-Hellman Knapsack untuk Pengamanan Pesan Teks," Jurnal Riset Mahasiswa Matematika, vol. 3, no. 1, pp. 10–25, Oct. 2023, doi: 10.18860/jrmm.v3i1.22292.

[27] H. T. Assafli and I. A. Hashim, "Security Enhancement of AES-CBC and its Performance Evaluation Using the Avalanche Effect," in 2020 3rd International Conference on Engineering Technology and its Applications (IICETA), IEEE, Sep. 2020, pp. 7–11. doi: 10.1109/IICETA50496.2020.9318803.

[28] O. Trabelsi, L. Sfaxi, and R. Robbana, "DCBC: A Distributed High-performance Block-Cipher Mode of Operation," in Proceedings of the 17th International Joint Conference on e-Business and Telecommunications, SCITEPRESS - Science and Technology Publications, 2020, pp. 86–97. doi: 10.5220/0009793300860097.

[29] Y. S. Alslman, A. Ahmad, and Y. AbuHour, "Enhanced and authenticated cipher block chaining mode," Bulletin of Electrical Engineering and Informatics, vol. 12, no. 4, pp. 2357–2362, Aug. 2023, doi: 10.11591/beei.v12i4.5113.

[30] X. Chai, H. Wu, Z. Gan, Y. Zhang, Y. Chen, and K. W. Nixon, "An efficient visually meaningful image compression and encryption scheme based on compressive sensing and dynamic LSB embedding," Opt Lasers Eng, vol. 124, p. 105837, Jan. 2020, doi: 10.1016/j.optlaseng.2019.105837.

[31] R. Mott, C. Fischer, P. Prins, and R. W. Davies, "Private Genomes and Public SNPs: Homomorphic Encryption of Genotypes and Phenotypes for Shared Quantitative Genetics," Genetics, vol. 215, no. 2, pp. 359–372, Jun. 2020, doi: 10.1534/genetics.120.303153.

Diabetes Prediction Using Machine Learning with Feature Engineering and Hyperparameter Tuning

Hakim El Massari[1*], Noreddine Gherabi[2], Fatima Qanouni[3], Sajida Mhammedi[4]
LAMAI Laboratory, Faculty of Sciences and Techniques, Cadi Ayyad University, Marrakech, Morocco[1]
Lasti Laboratory, National School of Applied Sciences, Sultan Moulay Slimane University, Khouribga, Morocco[1, 2, 3, 4]
Higher School of Technology of El Kelâa des Sraghna, Cadi Ayyad University, El Kelâa des Sraghna, Morocco[1]

***Abstract*—Diabetes, a chronic illness, has seen an increase in prevalence over the years, posing several health challenges. This study aims to predict diabetes onset using the Pima Indians Diabetes dataset. We implemented several machine learning algorithms, namely Random Forest, Gradient Boosting, XGBoost, LightGBM, and CatBoost. To enhance model performance, we applied a variety of feature engineering techniques, including SelectKBest, Recursive Feature Elimination (RFE), Recursive Feature Elimination with Cross-Validation (RFECV), Forward Feature Selection, and Backward Feature Elimination. RFECV proved to be the most effective method, leading to the selection of the best feature set. In addition, hyperparameter tuning techniques are used to determine the optimal parameters for the models created. Upon training these models with the optimized parameters, XGBoost outperformed the others with an accuracy of 94%, while Random Forest and CatBoost both achieved 92.5%. These results highlight XGBoost's superior predictive power and the significance of thorough feature engineering and model tuning in diabetes prediction.**

***Keywords*—Machine learning; feature engineering; hyperparameter tuning; diabetes prediction; healthcare**

I. Introduction

The World Health Organization considers diabetes one of the world's leading causes of death. Diabetes mellitus is a metabolic disorder of the endocrine system in which the blood glucose levels remain high for longer than necessary, causing hyperglycemia. The majority of diabetes cases are type 2 diabetes. The symptoms of diabetes include frequent urination, excessive thirst, extreme fatigue, etc.

The proportion of individuals with diabetes has been increasing more rapidly than can be accounted for by a rapidly increasing population. The World Health Organization has predicted that diabetes will be the leading cause of disease burden in the world by 2030. Such reporting is important but still almost undoubtedly an underestimate of the total impact of diabetes since there are very many children in whom the diagnosis is not made, in whom there may be a very early onset of complications, and whose death is not reported as 'diabetic'.

Youth-onset type 2 diabetes will also provide a considerable burden to some populations, especially indigenous peoples. Many obese individuals already have the insulin resistance that promises eventual diabetes. A third of the American adult population is thought to have the insulin resistance syndrome. Individuals of Asian and African origin, as well as indigenous peoples, have an increased risk of diabetic complications, at least in part independent of the greater weight for height. Since even impaired fasting glucose has been reported to be associated with an increased independent risk of cardiovascular disease, such reports demonstrate the threat and the value of strategies to prevent or delay the onset of metabolic syndrome. With outcome metrics such as the development of retinopathy and cardiovascular events, a diagnosis may only come in time to prevent a diagnosis of type 2 diabetes

It is associated that diabetes is very long and gradually exerts its unwanted effects on all the body parts. It remains in the individual's body for a long time and then develops heart disease, chronic meningitis, hypertension, blindness, stroke, erectile dysfunction, nerve damage (neuropathy), among other things. The population growth, relocating from rural to urban areas, interacting with bestial food habits, lack of exercise, stress, and lifestyle changes in people of all ages also contribute to the development of diabetes.

Diabetes is now recognized as a global health problem. It creates a huge impact on people and countries around the world. The importance of diabetes lies in the fact that it increases a person's likelihood of having a stroke by 1.5 times. It is predicted that if the rising incidence of diabetes is not reversed, the overall death rate from diabetes and heart disease will also rise.

Machine learning (ML) in healthcare is used to diagnose diseases, create personalized treatment plans, and predict hospital readmissions [1], [2], [3]. It can also detect which patients are at high risk of developing diabetes, long before it occurs. There are also many other related problems in the medical field such as disease diagnosis, hospital readmission, personalized treatment, and patient hope. However, the main goal of this study is to establish how basic, everyday habits affect the early detection of diabetes [4].

At the moment, prediabetes and diabetes are diagnosed through massive blood tests (glucose, insulin, and so on) that only patients with symptoms undergo. The main advantages of predicting diabetes using machine learning are: once the algorithm is implemented, everybody can use it and the test can be done whenever one wants; it is cheap; it allows everyone to know if they are at risk of developing diabetes months/years in advance, and take action; it saves the time and resources of doctors and hospitals to spend on the real ill patients.

*Corresponding Author.

Machine learning (ML) in healthcare is used to diagnose diseases, create personalized treatment plans, and predict hospital readmissions. It can also detect which patients are at high risk of developing diabetes, long before it occurs. There are also many other related problems in the medical field such as disease diagnosis, hospital readmission, personalized treatment, and patient hope. However, the main goal of this study is to establish how basic, everyday habits affect the early detection of diabetes.

The remainder of this study is organized as follows: Section II covers previous research and related studies. Section III details the methodology, including the various algorithms employed, steps taken to prepare the dataset, techniques for creating and refining features, and the process of optimizing the parameters. Section IV presents the results and a thorough discussion of the findings. Finally, Section VI provides a conclusion summarizing the key insights and implications of this work.

II. Related Work

This section provides a detailed overview of related work in the field of diabetes prediction using machine learning techniques in particular.

A recent study [5] proposed an ensemble-based approach for predicting diabetes using the Pima Indian Diabetes Dataset. It evaluated LightGBM, XGBoost, AdaBoost, and Random Forest, finding that LightGBM alone achieved an accuracy of 94% and a ROC AUC of 95%. By introducing a Soft Voting classifier, the combined model's accuracy increased to 95% with a ROC AUC of 96%, demonstrating the potential of ensemble methods to improve prediction reliability.

TABLE I. Related Work Comparison

Reference	ML algorithms	Highest Accuracy
[3]	Logistic Regression, Decision Tree, Random Forest, k-Nearest Neighbors, Naive Bayes, Support Vector Machine, Gradient Boosting, and Neural Network	78,57%
[5]	LightGBM, XGBoost, AdaBoost and Random Forest	93%
[6]	decision tree (DT), logistic regression (LR), support vector machine (SVM), gradient boost (GB), extreme gradient boost (XGBoost), random forest (RF), and ensemble technique (ET)	93,27%
[7]	Ensemble learning, XGBoost, CatBoost, LightGBM, AdaBoost, gradient boost	92,85%
[8]	Random forest classifier (RF), logistic regression (LR), decision tree classifier (DT), support vector machine (SVM), Bayesian Classifier (BC) or Naive Bayes Classifier (NB), Bagging Classifier (BG), Stacking Classifier (ST), Moderated Ada-Boost(AB) Classifier, K Neighbors Classifier (KN) and Artificial Neural Network (ANN)	90,95%
[9]	ET, RF, SGB, AB	93.63%
[10]	decision tree, SVM, Random Forest, Logistic Regression, KNN, and various ensemble techniques.	81%
Our study	Random Forest, Gradient Boosting, XGBoost, LightGBM, and CatBoost.	**94%**

Numerous studies have explored the use of machine learning algorithms for predicting diabetes. The study in [6] employed a range of classification algorithms, including Logistic Regression, Decision Trees, and Support Vector Machines, to predict diabetes onset using the Pima Indians Diabetes dataset. Their research highlighted the effectiveness of Support Vector Machines in achieving high accuracy.

Further investigation [7] focused on the application of ensemble methods for diabetes prediction. They compared the performance of Bagging, Boosting, and Stacking techniques, demonstrating that ensemble methods generally outperformed single classifiers. Specifically, their results indicated that Boosting algorithms, particularly XGBoost, provided superior predictive performance.

Feature selection and engineering play a critical role in improving model accuracy. The study [8] implemented Recursive Feature Elimination (RFE) and Principal Component Analysis (PCA) to enhance their machine learning models. They concluded that RFE, in combination with Gradient Boosting Machines, yielded the best results, emphasizing the importance of selecting relevant features.

The use of deep learning approaches has also been investigated [11], [12], [13], [14]. In study [15] authors proposed a deep neural network model for diabetes prediction, achieving remarkable accuracy. Their work demonstrated that deep learning models could capture complex patterns in the data, albeit at the cost of increased computational resources and the need for larger datasets.

Additionally, researches [9], [16] introduced the concept of hybrid models that combine multiple machine learning techniques to improve prediction accuracy. They developed a hybrid model integrating Random Forest and Neural Networks, which surpassed the performance of traditional models.

Recent advancements in explainable AI have also been applied to diabetes prediction. For instance, [10], [17] utilized SHapley Additive exPlanations (SHAP) to interpret the predictions of their machine learning models. This approach provided insights into the importance of different features, enhancing the transparency and trustworthiness of the predictive models.

In earlier studies, different feature selection and classification have been proposed to optimize the classifier model for 12 different classifiers over Pima Indians Diabetes Database from the UCI machine learning website [18], [19], [20]. The work was done on Pima Indians Diabetes Database in order to predict diabetes using different Data Mining algorithms. The main feature selection methods are Correlation, Wrappers, and Principal Components. Wrapper was the most successful feature selection method in obtaining a small data subset optimizing the classifier. Besides the feature

selection, a metaheuristic algorithm was used to optimize the classifier to fit the classifier model by using a small training data subset. Random Forest (RF) with 28 input features produced high accuracy (98.08%), sensitivity (94.6), and specificity (99.3). The study in [21], [22], dataset with six input attributes was tested using six different classifiers. Decision Trees (DT) and J48 classifiers gave the best results on both 10 × 10 cross-validation (CV) or independent test sets, resulting in 78.33% accuracy, 77.38% and 88.33% accuracy, 87.84%, respectively. These works have partially been compared with the work.

Research indicates that type 2 diabetes [23], [24], [25] is treatable and preventable through making lifestyle changes such as weight loss, improved diet, and increasing physical activity. Regular monitoring, at-home blood glucose testing, and A1C levels are pivotal for identifying high risk for diabetes and type 2 diabetes early on. However, individuals are experiencing continuous increases in weight and obesity because of the rising trends of high-calorie diets and sedentary lifestyles.

This trend could reach a breaking point for the healthcare system if our understanding of the factors leading to type 2 diabetes risk remains incomplete. Machine learning techniques, such as classification, regression, clustering, anomaly detection, and pattern recognition, are developed to make accurate predictions from data. Specifically, the current health status of pre-diabetic patients is used to assess potential for diabetes diagnoses.

By understanding how certain factors of lifestyle change can influence diabetes diagnoses, pre-diabetics may be able to take steps to avoid the implications of diabetes. With the implementation of machine learning algorithms and healthcare data, healthcare providers would benefit from a tool capable of early diagnosing patients who possess the highest risk of developing type 2 diabetes and cardiovascular diseases while developing personalized and cost-effective intervention strategies for pre-diabetics [26], [27], [28].

Overall, the related work in this field underscores the continuous evolution of machine learning techniques for diabetes prediction. The integration of advanced feature engineering, ensemble methods, and deep learning has significantly improved predictive accuracy, paving the way for more effective and reliable diabetes prediction models.

III. Methods and Evaluation

In this study, the methodology used is divided into several key components: data collection and preprocessing, data analysis techniques, feature engineering, hyperparameter tuning and performance evaluation metrics. Each component is discussed in detail to provide a comprehensive understanding of the research process. Fig. 1 depicts the overall process workflow for this experimental study.

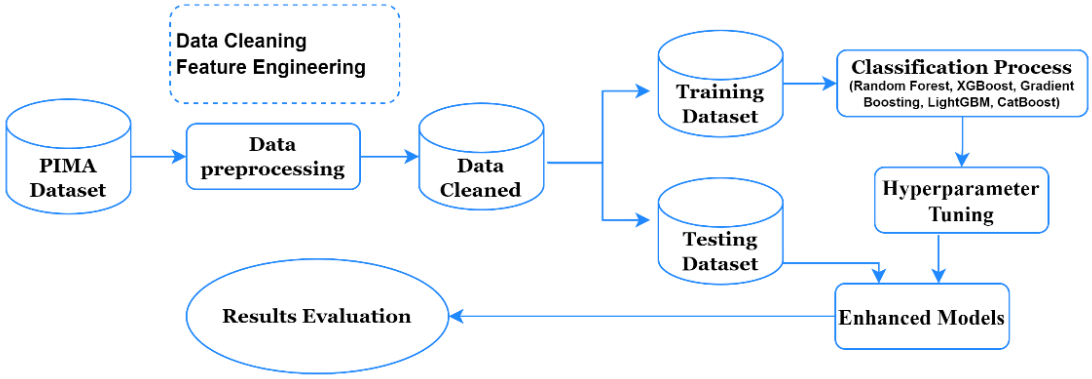


Fig. 1. Experimental workflow.

A. Data Collection and Preprocessing

To implement the predictive models in this study, an open-access diabetes dataset was utilized. This dataset, obtained from Kaggle, includes various medical predictor variables and a target variable. It comprises records of 768 patients. All patients are females of Pima Indian heritage, aged at least 21 years. Among them, 34.9% have diabetes, while 65.1% do not, as depicted in Fig. 2. Detailed attribute information is provided in Table II.

Data preprocessing is essential for all machine learning (ML) applications because the effectiveness of an ML algorithm depends significantly on how well the dataset is prepared and structured. This step ensures the data is tailored to meet the specific needs of the chosen algorithm. For the diabetes dataset, we employed several preprocessing techniques during this initial phase:

TABLE II. Dataset Feature's Information

Attribute	Description
1- Pregnancies	Count of pregnancies
2- Glucose	Plasma glucose levels
3- BloodPressure	Diastolic blood pressure (mm Hg)
4- SkinThickness	Triceps skin fold thickness (mm)
5- Insulin	2-Hour serum insulin (mu U/ml)
6- BMI	Body mass index
7- DiabetesPedigreeFunction	Diabetes pedigree function
8- Age	Age (years)
9- Outcome	1 = diabetic, 0 = non diabetic

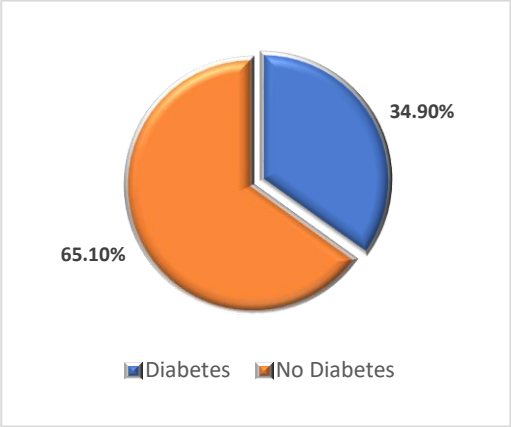


Fig. 2. Distribution of diabetes.

- Data Cleaning: We removed missing or null values, cleaned up noisy data, and detected and eliminated outliers.
- Outlier Handling: To enhance the robustness of our model, we used the "Replace with Thresholds IQR" method. This technique involves substituting extreme values with thresholds derived from the Interquartile Range (IQR), which helps create a more resilient and reliable model.
- Scaling and Normalization: We scaled and normalized the data to ensure that no single feature disproportionately influences the model due to differing scales.
- Handling Imbalanced Data: To prevent bias and ensure the model is not unduly influenced by the prevalence of a particular class, we applied the Synthetic Minority Oversampling Technique (SMOTE) as described in Table III. This technique generates a balanced dataset by synthesizing instances of the minority class, thereby improving the predictive accuracy for that class.

Throughout the study, we utilized various Python libraries, including NumPy, Pandas, Seaborn, Matplotlib, and Scikit-learn, for both exploratory data analysis (EDA) and data visualization. These tools helped us thoroughly analyze and prepare the data, setting a solid foundation for building effective predictive models.

TABLE III. RESULT OF THE SMOTE TECHNIQUE

	Diabetes	
	Yes	No
Before SMOTE	268	500
After SMOTE	500	500

B. Machine Learning Algorithms

Among the existing algorithms of machine learning [29], [30], we used in this study Random Forest, Gradient Boosting, XGBoost, LightGBM, and CatBoost. These algorithms were selected for their robustness and versatility in handling various types of data and predictive modeling tasks. Random Forest is known for its simplicity and effectiveness in reducing overfitting through ensemble learning. Gradient Boosting improves predictive accuracy by iteratively minimizing errors. XGBoost, an optimized version of Gradient Boosting, enhances performance and computational efficiency. LightGBM, designed for speed and scalability, handles large datasets and high-dimensional data efficiently. CatBoost is particularly effective with categorical data and requires minimal preprocessing. By leveraging the strengths of these algorithms, we aimed to achieve a comprehensive analysis and robust predictive performance for our study.

Random Forest is an ensemble learning algorithm that constructs multiple decision trees during training and outputs the class that is the mode of the classes (classification) or mean prediction (regression) of the individual trees. It is robust to overfitting due to the averaging of multiple trees, handles large datasets effectively, and can manage missing data and maintain accuracy for large portions of the data.

Gradient Boosting is an iterative algorithm that builds a model in a stage-wise fashion from weak learners, typically decision trees. Each new model attempts to correct the errors of the previous one by minimizing a loss function. This approach results in high predictive accuracy, making it suitable for various machine learning tasks, although it can be computationally intensive and sensitive to overfitting without proper regularization.

XGBoost (Extreme Gradient Boosting) is an optimized version of Gradient Boosting designed for performance and speed. It incorporates advanced features like regularization to prevent overfitting, parallel processing, and efficient handling of missing data. XGBoost is known for its scalability and effectiveness in both regression and classification problems, making it a popular choice in competitive machine learning.

LightGBM (Light Gradient Boosting Machine) is a gradient boosting framework that uses tree-based learning algorithms. It is designed for efficiency and scalability, making it well-suited for large datasets with high-dimensional features. LightGBM achieves faster training speed and higher efficiency by using a histogram-based approach and leaf-wise tree growth, which leads to better accuracy.

CatBoost (Categorical Boosting) is a gradient boosting algorithm specifically designed to handle categorical data with minimal preprocessing. It automatically deals with categorical features and reduces overfitting through techniques like ordered boosting and efficient oblivious tree structures. CatBoost is robust, accurate, and user-friendly, making it ideal for applications where categorical data is prevalent.

C. Feature Engineering

Feature engineering plays a role, in the realm of machine learning. It involves the creation, adjustment and selection of features from data to boost the performance of predictive models. This process encompasses methods like normalization encoding variables and crafting features based on domain expertise. Skillful feature engineering can notably improve the precision and effectiveness of machine learning models by equipping them with valuable input data. It often necessitates testing and a profound comprehension of both the dataset and

the core issue to identify features that aptly capture the patterns and connections, for accurate forecasts.

In this study various techniques are used to enhance machine learning models such as SelectKBest, Recursive Feature Elimination (RFE), Recursive feature elimination with cross-validation (RFECV), Forward Feature Selection, Backward Feature Elimination. Fig. 3-7 represents the results of features used in this study.

SelectKBest is a feature selection method that selects the top k features from a dataset based on a scoring function. We used the chi2 function with SelectKBest which is based on the chi-squared statistical test, which measures the independence of each feature with respect to the target variable. Higher chi-squared values indicate a stronger relationship between the feature and the target. By using SelectKBest with chi2, we reduced the dimensionality of the dataset by keeping only the most relevant features, potentially improving the performance of the machine learning models employed. By using SelectKBest and chi2 function we find that all features give over 91% accuracy, and the highest score of 92.2% goes to Random Forest Classifier.

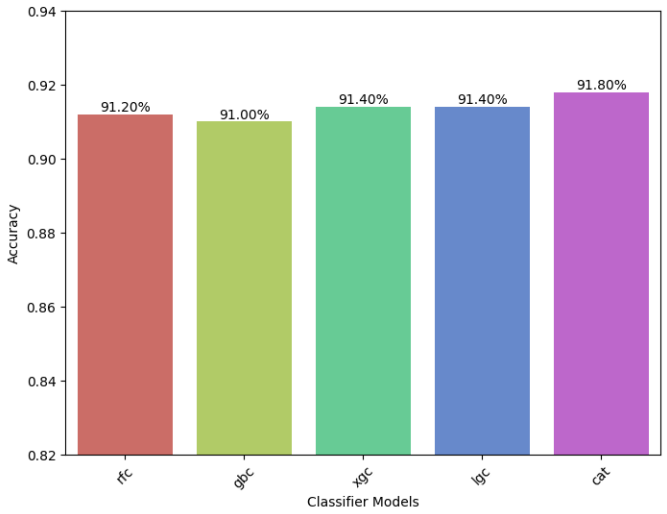


Fig. 3. Accuracy result using SelectKBest.

Recursive Feature Elimination (RFE) is a feature selection technique in machine learning that iteratively removes the least important features from the dataset. Starting with all features, RFE fits a model and evaluates the importance of each feature. The least important feature is then removed, and the model is re-fit on the remaining features. This process continues until the desired number of features is reached. By systematically eliminating features, RFE helps in identifying the most relevant subset, improving model performance and reducing overfitting by eliminating noise and irrelevant data.

Recursive Feature Elimination with Cross-Validation (RFECV) is an enhanced feature selection technique that combines Recursive Feature Elimination (RFE) with cross-validation to select the optimal number of features. RFECV iteratively removes the least important features while simultaneously evaluating model performance using cross-validation at each iteration. This approach ensures that the feature selection process is guided by model accuracy, helping to identify the subset of features that yields the best predictive performance. By incorporating cross-validation, RFECV provides a more robust and reliable method for feature selection, reducing the risk of overfitting and improving the generalizability of the model.

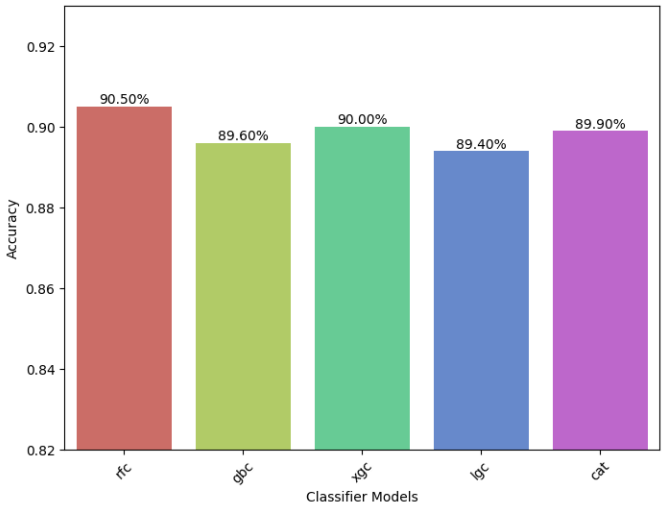


Fig. 4. Accuracy result using RFE.

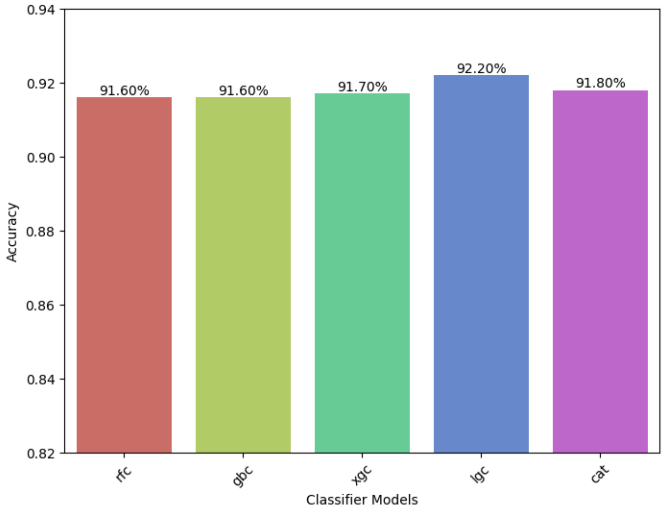


Fig. 5. Accuracy result using RFECV.

Forward Feature Selection is a feature selection technique in machine learning that starts with an empty model and iteratively adds the most significant features. At each step, the method evaluates all candidate features and adds the one that improves the model performance the most, based on a predefined criterion like accuracy or F1 score. This process continues until adding more features no longer significantly improves the model or a specified number of features is reached. Forward Feature Selection is effective for identifying a small, relevant subset of features, enhancing model interpretability and performance by including only the most impactful variables.

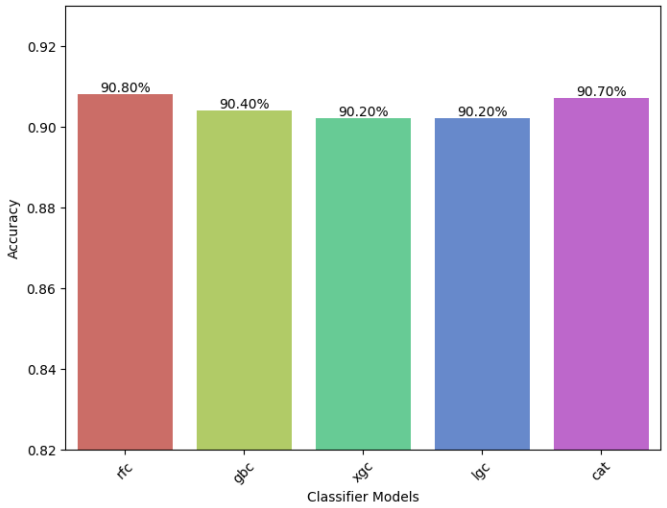


Fig. 6. Accuracy result using FORWARD.

Backward Feature Elimination is a feature selection technique in machine learning that starts with all available features and iteratively removes the least significant ones. At each step, the model is trained, and the importance of each feature is evaluated based on a predefined criterion such as p-values or model performance metrics. The least important feature is then removed, and the process is repeated until a specified number of features remains or further removal would degrade model performance. This method helps in simplifying the model by eliminating redundant or irrelevant features, improving interpretability and potentially enhancing predictive accuracy by reducing overfitting.

By comparing the results obtained from the five feature selection techniques used in this study, as shown in the Fig. 8, we conclude that the RFECV (Recursive Feature Elimination with Cross-Validation) feature selection method provides the best results. The top three algorithms identified are Random Forest, CatBoost, and XGBoost. For the remainder of this study, we will rely on these three algorithms.

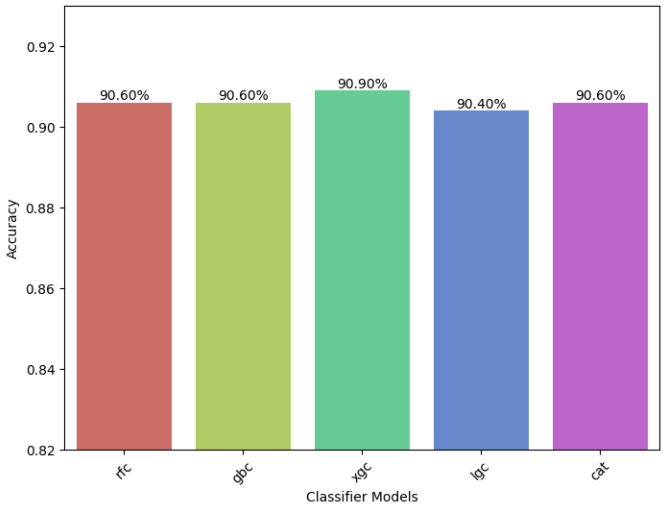


Fig. 7. Accuracy result using BACKWARD.

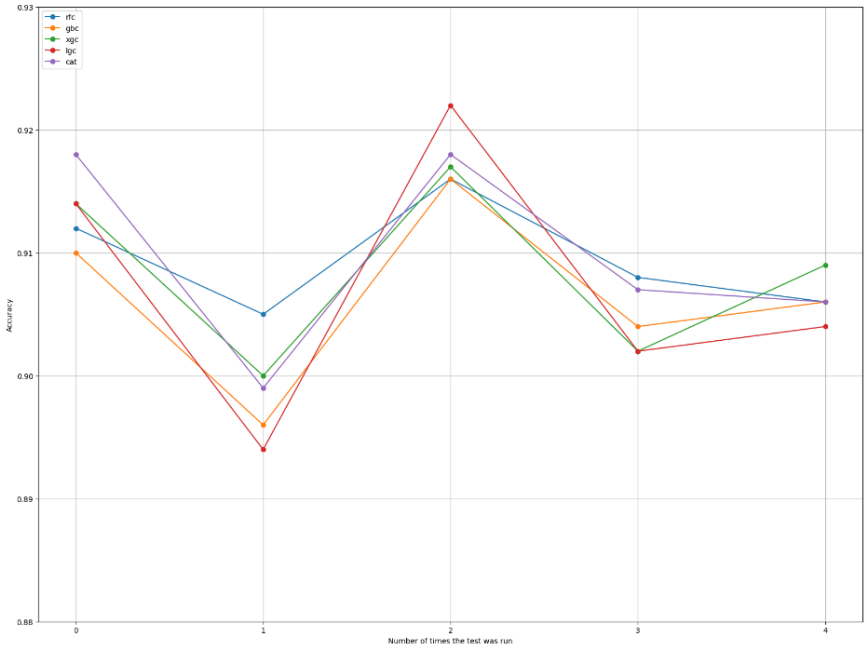


Fig. 8. Classifiers performance results.

D. Hyperparameter Tuning

The optimization of the hyperparameters of a machine learning model for optimizing the performance of a model is called hyperparameter tuning. Hyperparameters are different from model parameters because model parameters are learned during training while hyperparameters need to be set prior to the training process and influence different attributes of the learning process (learning rate, regularization strength, number

of layers in a neural network). Hyperparameter tuning is the process of finding the best selection of hyperparameters for a model which is typically done by systematically searching the space of hyperparameter values in a methodical manner. Some common methods of hyperparameter tuning for machine learning are grid search, random search, and Bayesian optimization. Model performance can be greatly improved by properly tuning the hyperparameters through achieving the ideal settings which enable the learning process to efficiently learn patterns in the data and avoid overfitting.

In this study, we utilize the GridSearchCV technique to determine the optimal parameters for our three models: Random Forest, CatBoost, and XGBoost. This method ensures that our models are fine-tuned for maximum accuracy and robustness. The results obtained from this parameter optimization process are presented in the accompanying Table IV, highlighting the best parameter settings for each model and their corresponding performance metrics.

TABLE IV. BEST PARAMETER SETTINGS FOR EACH MODEL

	Parameter
Random Forest	{bootstrap= False, ccp_alpha= 0, criterion= 'gini', max_depth= None, max_features= 'sqrt', n_estimators= 100, n_jobs= -1, verbose=0, random_state= 42}
XGBoost	{gamma= 0.1, learning_rate= 0.1, max_depth= 7, n_estimators= 200, reg_alpha= 0, reg_lambda= 0.001, verbose=0}
CatBoost	{bootstrap_type= 'Bernoulli', depth= 8, grow_policy= 'SymmetricTree', iterations= 200, l2_leaf_reg= 1, learning_rate= 0.1, verbose=0}

E. Performance Evaluation Metrics

Performance evaluation metrics are crucial tools in machine learning for assessing the effectiveness of predictive models. These metrics provide quantitative measures to evaluate how well a model performs on a given task. Common metrics include accuracy, precision, recall, F1-score, etc.

$$\text{ACC} = \frac{\text{TP+TN}}{\text{TP+TN+FP+FN}} \quad (1)$$

$$\text{PREC} = \frac{\text{TP}}{\text{TP+FP}} \quad (2)$$

$$REC = \frac{TP}{TP+FN} \quad (3)$$

$$\text{F-Measure} = 2 * \frac{PREC*REC}{PREC+REC} \quad (4)$$

IV. RESULTS AND DISCUSSION

The Pima Indians Diabetes dataset served as the foundation for this study, aiming to predict the onset of diabetes. This dataset includes several medical predictor variables and one target variable, which indicates whether or not the patient has diabetes. To enhance the predictive power of our machine learning models, we implemented a variety of feature engineering techniques.

We tested five machine learning algorithms: Random Forest, Gradient Boosting, XGBoost, LightGBM, and CatBoost. These algorithms were chosen for their robust performance in classification tasks. To optimize the input features for these models, we employed several feature engineering techniques, namely SelectKBest, Recursive Feature Elimination (RFE), Recursive Feature Elimination with Cross-Validation (RFECV), Forward Feature Selection, and Backward Feature Elimination.

Among these techniques, RFECV provided the most significant improvement in terms of accuracy and other performance metrics. RFECV methodically eliminates less important features while incorporating cross-validation to prevent overfitting. This approach identified the optimal set of features, which were then used to train our models.

Focusing on the top three algorithms identified by RFECV—Random Forest, CatBoost, and XGBoost—we used GridSearchCV to determine the optimal parameters for each model. This method involves an exhaustive search over specified parameter values to find the best combination that maximizes model performance.

Once the models were trained with these optimized parameters, we compared their performance Fig. 9 and Fig. 10. XGBoost emerged as the top-performing model, achieving an impressive accuracy score of 94%. In comparison, both Random Forest and CatBoost achieved accuracy scores of 92.5%. The superior performance of XGBoost can be attributed to its advanced tree boosting techniques and regularization methods, which help in managing data complexity and avoiding overfitting.

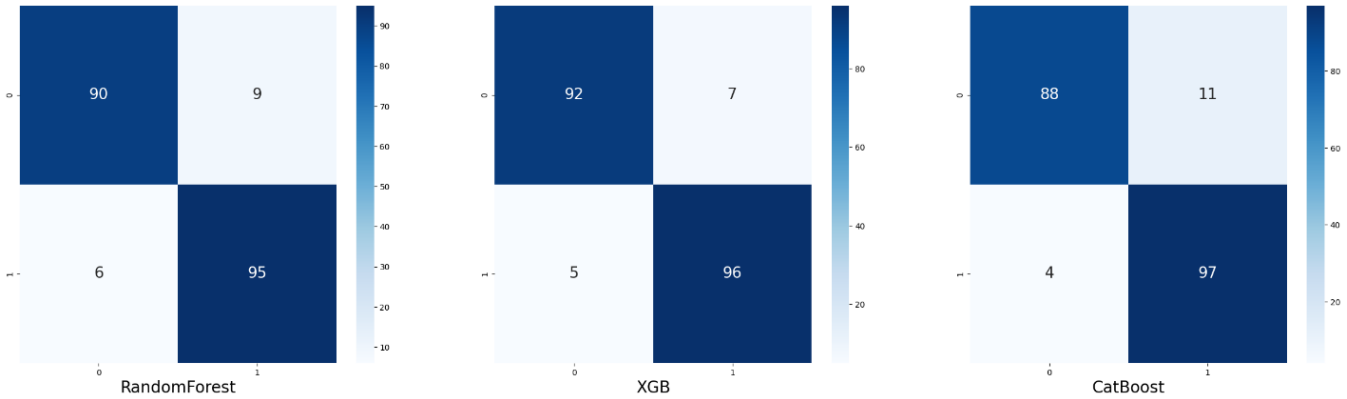


Fig. 9. Confusion matrix results.

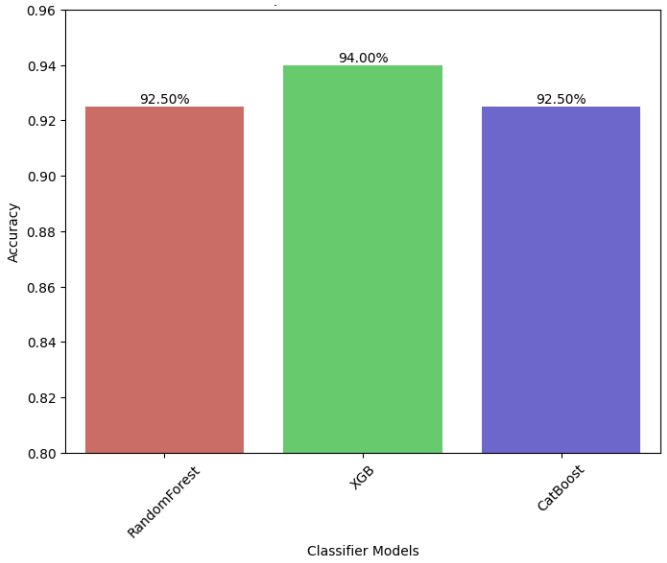


Fig. 10. Accuracy comparison of the models.

Comparing our findings with the most recent researches in the same field and used same the dataset, the feature engineering and hyperparameter tuning applied on our selective models, got the highest accuracy rate, Table I represent the comparison of different studies with ours.

In summary, the application of RFECV for feature selection and GridSearchCV for parameter optimization significantly enhanced the performance of our models. XGBoost, with its sophisticated boosting algorithms, proved to be the most effective model for predicting the onset of diabetes using the Pima Indians Diabetes dataset. Future work could explore additional data preprocessing steps and the inclusion of more complex models to further improve predictive accuracy.

V. Comparison with Other Studies

When examining the latest studies utilizing the Pima Indians Diabetes dataset, it's evident that our approach to feature engineering and model optimization has set a new benchmark in predictive accuracy.

Therefore, by leveraging RFECV for feature selection, we were able to identify the most relevant features and reduce noise in the dataset, leading to significant improvements in model performance. The subsequent application of GridSearchCV for hyperparameter tuning further optimized our models, ensuring that we achieved the best possible configuration for predictive accuracy. The integration of these techniques enabled our top-performing model, XGBoost, to reach an accuracy of 94%, surpassing the results of the aforementioned studies.

Our study not only highlights the importance of rigorous feature engineering and parameter optimization but also demonstrates the potential of advanced ensemble methods in predictive analytics. The substantial gains in accuracy underline the effectiveness of our approach compared to other contemporary methodologies. Table I provides a detailed comparison, showcasing the advancements our study brings to the field.

VI. Conclusion

This research focused on predicting the onset of diabetes using the Pima Indians Diabetes dataset. By applying various machine learning algorithms and feature engineering techniques, we aimed to identify the most effective model for this task. Among the algorithms tested — Random Forest, Gradient Boosting, XGBoost, LightGBM, and CatBoost — XGBoost demonstrated superior performance.

We employed several features engineering methods, including SelectKBest, Recursive Feature Elimination (RFE), Recursive Feature Elimination with Cross-Validation (RFECV), Forward Feature Selection, and Backward Feature Elimination, to refine our models. RFECV stood out as the most successful approach, yielding the best results in terms of accuracy and other metrics. Consequently, we concentrated on the top three algorithms identified by RFECV: Random Forest, CatBoost, and XGBoost. To further optimize these models, we utilized GridSearchCV to find the best parameter settings. After training the models with these optimized parameters, XGBoost achieved an accuracy score of 94%, outperforming Random Forest and CatBoost, which both scored 92.5%.

In summary, this study highlights the efficacy of XGBoost in predicting diabetes, this is due to its advanced boosting techniques and robust regularization methods. The importance of comprehensive feature engineering and parameter tuning was also underscored. Future research could explore additional preprocessing steps and incorporate more complex models to enhance predictive accuracy further.

References

[1] J. J. Khanam and S. Y. Foo, “A comparison of machine learning algorithms for diabetes prediction,” ICT Express, vol. 7, no. 4, pp. 432–439, Dec. 2021, doi: 10.1016/j.icte.2021.02.004.

[2] Z. Sabouri, N. Gherabi, M. Nasri, M. Amnai, H. El Massari, and I. Moustati, “Prediction of Depression via Supervised Learning Models: Performance Comparison and Analysis,” International Journal of Online and Biomedical Engineering (iJOE), vol. 19, no. 09, Art. no. 09, Jul. 2023, doi: 10.3991/ijoe.v19i09.39823.

[3] M. S. Alzboon, M. S. Al-Batah, M. Alqaraleh, A. Abuashour, and A. F. H. Bader, “Early Diagnosis of Diabetes: A Comparison of Machine Learning Methods,” International Journal of Online and Biomedical Engineering (iJOE), vol. 19, no. 15, Art. no. 15, Oct. 2023, doi: 10.3991/ijoe.v19i15.42417.

[4] H. El Massari, N. Gherabi, S. Mhammedi, H. Ghandi, M. Bahaj, and M. R. Naqvi, “The Impact of Ontology on the Prediction of Cardiovascular Disease Compared to Machine Learning Algorithms,” iJOE, vol. 18, no. 11, Art. no. 11, Aug. 2022, doi: 10.3991/ijoe.v18i11.32647.

[5] N. H. Taz, A. Islam, and I. Mahmud, “A Comparative Analysis of Ensemble Based Machine Learning Techniques for Diabetes Identification,” in 2021 2nd International Conference on Robotics, Electrical and Signal Processing Techniques (ICREST), Jan. 2021, pp. 1–6. doi: 10.1109/ICREST51555.2021.9331036.

[6] M. J. Uddin et al., “A Comparison of Machine Learning Techniques for the Detection of Type-2 Diabetes Mellitus: Experiences from Bangladesh,” Information, vol. 14, no. 7, Art. no. 7, Jul. 2023, doi: 10.3390/info14070376.

[7] S. M. Ganie, P. K. D. Pramanik, M. Bashir Malik, S. Mallik, and H. Qin, “An ensemble learning approach for diabetes prediction using boosting techniques,” Front Genet, vol. 14, p. 1252159, Oct. 2023, doi: 10.3389/fgene.2023.1252159.

[8] M. S. Alam, M. J. Ferdous, and N. S. Neera, "Enhancing Diabetes Prediction: An Improved Boosting Algorithm for Diabetes Prediction," International Journal of Advanced Computer Science and Applications (IJACSA), vol. 15, no. 5, Art. no. 5, Jun. 2024, doi: 10.14569/IJACSA.2024.01505129.

[9] P. Talari et al., "Hybrid feature selection and classification technique for early prediction and severity of diabetes type 2," PLOS ONE, vol. 19, no. 1, p. e0292100, Jan. 2024, doi: 10.1371/journal.pone.0292100.

[10] I. Tasin, T. U. Nabil, S. Islam, and R. Khan, "Diabetes prediction using machine learning and explainable AI techniques," Healthcare Technology Letters, vol. 10, no. 1–2, pp. 1–10, 2023, doi: 10.1049/htl2.12039.

[11] R. Rajalakshmi, P. Sivakumar, L. K. Kumari, and M. C. Selvi, "A novel deep learning model for diabetes melliuts prediction in IoT-based healthcare environment with effective feature selection mechanism," J Supercomput, vol. 80, no. 1, pp. 271–291, Jan. 2024, doi: 10.1007/s11227-023-05496-6.

[12] M. A. Bülbül, "A novel hybrid deep learning model for early stage diabetes risk prediction," J Supercomput, May 2024, doi: 10.1007/s11227-024-06211-9.

[13] A. R. Mohamed Yousuff, M. Zainulabedin Hasan, R. Anand, and M. Rajasekhara Babu, "Leveraging deep learning models for continuous glucose monitoring and prediction in diabetes management: towards enhanced blood sugar control," Int J Syst Assur Eng Manag, vol. 15, no. 6, pp. 2077–2084, Jun. 2024, doi: 10.1007/s13198-023-02200-y.

[14] K. K. Patro et al., "An effective correlation-based data modeling framework for automatic diabetes prediction using machine and deep learning techniques," BMC Bioinformatics, vol. 24, no. 1, p. 372, Oct. 2023, doi: 10.1186/s12859-023-05488-6.

[15] M. F. Aslan and K. Sabanci, "A Novel Proposal for Deep Learning-Based Diabetes Prediction: Converting Clinical Data to Image Data," Diagnostics, vol. 13, no. 4, Feb. 2023, doi: 10.3390/diagnostics13040796.

[16] P. V and R. D. R, "A Hybrid Model for Prediction of Diabetes Using Machine Learning Classification Algorithms and Random Projection," Jun. 28, 2023. doi: 10.21203/rs.3.rs-3081331/v1.

[17] S. Mhammedi, H. El Massari, and N. Gherabi, "Composition of Large Modular Ontologies Based on Structure," in Advances in Information, Communication and Cybersecurity, Y. Maleh, M. Alazab, N. Gherabi, L. Tawalbeh, and A. A. Abd El-Latif, Eds., in Lecture Notes in Networks and Systems. Cham: Springer International Publishing, 2022, pp. 144–154. doi: 10.1007/978-3-030-91738-8_14.

[18] A. A. Alzubaidi, S. M. Halawani, and M. Jarrah, "Towards a Stacking Ensemble Model for Predicting Diabetes Mellitus using Combination of Machine Learning Techniques," International Journal of Advanced Computer Science and Applications (IJACSA), vol. 14, no. 12, Art. no. 12, 47/29 2023, doi: 10.14569/IJACSA.2023.0141236.

[19] S. K. S. Modak and V. K. Jha, "Diabetes prediction model using machine learning techniques," Multimed Tools Appl, vol. 83, no. 13, pp. 38523–38549, Apr. 2024, doi: 10.1007/s11042-023-16745-4.

[20] K. Oliullah, M. H. Rasel, Md. M. Islam, Md. R. Islam, Md. A. H. Wadud, and Md. Whaiduzzaman, "A stacked ensemble machine learning approach for the prediction of diabetes," J Diabetes Metab Disord, vol. 23, no. 1, pp. 603–617, Jun. 2024, doi: 10.1007/s40200-023-01321-2.

[21] S. G. Choi et al., "Comparisons of the prediction models for undiagnosed diabetes between machine learning versus traditional statistical methods," Sci Rep, vol. 13, no. 1, p. 13101, Aug. 2023, doi: 10.1038/s41598-023-40170-0.

[22] A. A. Alzubaidi, S. M. Halawani, and M. Jarrah, "Integrated Ensemble Model for Diabetes Mellitus Detection," International Journal of Advanced Computer Science and Applications (IJACSA), vol. 15, no. 4, Art. no. 4, 33/30 2024, doi: 10.14569/IJACSA.2024.0150423.

[23] M. Kawarkhe and P. Kaur, "Prediction of Diabetes Using Diverse Ensemble Learning Classifiers," Procedia Computer Science, vol. 235, pp. 403–413, Jan. 2024, doi: 10.1016/j.procs.2024.04.040.

[24] I. Nissar et al., "An Intelligent Healthcare System for Automated Diabetes Diagnosis and Prediction using Machine Learning," Procedia Computer Science, vol. 235, pp. 2476–2485, Jan. 2024, doi: 10.1016/j.procs.2024.04.233.

[25] A. Hennebelle, H. Materwala, and L. Ismail, "HealthEdge: A Machine Learning-Based Smart Healthcare Framework for Prediction of Type 2 Diabetes in an Integrated IoT, Edge, and Cloud Computing System," Procedia Computer Science, vol. 220, pp. 331–338, Jan. 2023, doi: 10.1016/j.procs.2023.03.043.

[26] S. Jangili, H. Vavilala, G. S. B. Boddeda, S. M. Upadhyayula, R. Adela, and S. R. Mutheneni, "Machine learning-driven early biomarker prediction for type 2 diabetes mellitus associated coronary artery diseases," Clinical Epidemiology and Global Health, vol. 24, p. 101433, Nov. 2023, doi: 10.1016/j.cegh.2023.101433.

[27] M. E. Febrian, F. X. Ferdinan, G. P. Sendani, K. M. Suryanigrum, and R. Yunanda, "Diabetes prediction using supervised machine learning," Procedia Computer Science, vol. 216, pp. 21–30, Jan. 2023, doi: 10.1016/j.procs.2022.12.107.

[28] A. Nurdin, M. M. Tane, R. W. T. Tumewu, K. M. Suryaningrum, and H. A. Saputri, "Using Machine Learning for the Prediction of Diabetes with Emphasis on Blood Content," Procedia Computer Science, vol. 227, pp. 990–1001, Jan. 2023, doi: 10.1016/j.procs.2023.10.608.

[29] H. El Massari, N. Gherabi, S. Mhammedi, Z. Sabouri, H. Ghandi, and F. Qanouni, "Effectiveness of applying Machine Learning techniques and Ontologies in Breast Cancer detection," Procedia Computer Science, vol. 218, pp. 2392–2400, Jan. 2023, doi: 10.1016/j.procs.2023.01.214.

[30] H. El Massari, N. Gherabi, S. Mhammedi, H. Ghandi, F. Qanouni, and M. Bahaj, "Integration of ontology with machine learning to predict the presence of covid-19 based on symptoms," BEEI, vol. 11, no. 5, Art. no. 5, Oct. 2022, doi: 10.11591/eei.v11i5.4392.

Computational Modeling of the Thermally Stressed State of a Partially Insulated Variable Cross-Section Rod

Zhuldyz Tashenova[1], Elmira Nurlybaeva[2], Zhanat Abdugulova[3],
Shirin Amanzholova[4], Nazira Zharaskhan[5], Aigerim Sambetova[6], Anarbay Kudaykulov[7]
Department of Information Technologies, L. N. Gumilyov Eurasian National University, Astana, Kazakhstan[1, 3, 5]
Department of Computer Technologies, T.K. Zhurgenov Kazakh National Academy of Arts, Almaty, Kazakhstan[2, 6]
Department of Social and Humanitarian Disciplines, The Kurmangazy Kazakh National Conservatory, Almaty, Kazakhstan[4]
Department of IT, Institute of Information and Computational Technologies, Almaty, Kazakhstan[7]

***Abstract*—The formulation of the proposed methods and algorithms facilitates a comprehensive examination of intricate non-stationary thermo-mechanical processes in rods with varying cross-sectional geometries. Furthermore, it advances the theoretical framework for analyzing the thermo-mechanical properties of rod structures utilized in the machinery industry of the Republic of Kazakhstan. The creation of these intellectual products aids in the progression of this sector and fortifies the nation's sovereignty. This article delineates methods and algorithms for investigating non-stationary thermo-mechanical processes in rods with diverse cross-sectional shapes that influence global manufacturing technologies. The scientific and practical importance of this work lies in the application potential of the developed approach for examining non-stationary thermo-mechanical characteristics of rod-like elements in various installations. The findings also enhance the scientific research direction in mechanical engineering. In conclusion, the article outlines future technological advancements, summarizes the research on non-stationary thermo-mechanical processes in rods with different cross-sectional geometries, and highlights significant economic benefits by facilitating the selection of reliable rods for specified operating conditions. This ensures the continuous and dependable operation of machinery used in mechanical engineering.**

***Keywords*—Heat flow; heat transfer; thermal expansion coefficient; thermal conductivity; modulus of elasticity**

I. Introduction

The structural components of modern gas turbine power plants, nuclear and thermal power stations, hydrogen and rocket engines, internal combustion engines, and installations for deep processing of mineral resources and oils operate within a complex force and thermal environment. The reliable operation of these systems depends on the thermo-mechanical characteristics of their load-bearing elements. Typically, these elements are considered as rods of limited length and constant cross-sectional area. In related studies, temperature distribution along the length of such rods is determined based on fundamental thermophysics laws, considering the types of heat sources acting on them. Unlike those, the current work focuses on a horizontal rod of limited length and constant cross-sectional area, fully thermally insulated on its lateral surface. A constant heat flux is applied to the left end, while the right end exchanges heat with the environment.

Using fundamental energy conservation laws, this study determines the temperature distribution along the rod, its thermal elongation, the axial compressive force generated, and the distribution of elastic, temperature, and thermoelastic deformations and stresses, as well as the displacement field. Understanding the temperature distribution along the rod is crucial for the thermal stress state in bearing components of power plants and engines.

Previous works such as [1] and [2] have explored the principles of elasticity theory and numerical methods for applied mechanics. The primary thermo-physics equations, detailed in [3], include mass, momentum, and energy conservation. Other studies, like [4], [5], and [6], have investigated contact heat transfer and the thermal stress-strain state under various conditions, using the finite element method [7]. Additionally, [8] and [9] have addressed stress-strain states in rigid plastic pipes and nonlinear finite element modeling, while [10] and [11] discussed adiabatic shear bands and nonlinear continuum mechanics.

In [12], the temperature distribution within nuclear fuel rods was analyzed, highlighting the importance of maintaining fuel integrity. Research in [13], [14], and [15] derived computational relationships for thermal forces in rectangular cross-section rods. Other studies, such as [16], examined the thermal behavior of bars during hot rolling. The finite element method was also employed in [17] and [18] to model temperature fields in Terfenol-D rods, while [19] investigated unstable temperature distributions in cylindrical rods subjected to laser heat sources.

This body of work supports the development of a mathematical model for the thermomechanical state of a variable cross-section rod, considering local temperature, thermal insulation, and heat exchange. Scientific works [20]-[26] provide analytical solutions for compressive extensions and force distributions in thermally insulating rods, based on energy conservation principles. This paper diverges by using quadratic spline functions to address a specific practical

problem, offering a novel approach to understanding the thermomechanical behavior of such rods.

II. Materials and Methods

Let's consider a rod of finite length clamped at both ends, whose cross-section varies along its length and is circular. In this case, the radius of the cross-section depends linearly on the coordinates. The radius of the left end is denoted by r_0, the radius of the right end by r_L, and the length of the rod by L. Thus, the radius as a function of the coordinate x is given by the following expression [20]:

$$r = \frac{r_L - r_0}{L} \cdot x + r_0 \quad (1)$$

where x is the coordinate along the length of the rod, ranging from 0 (left end) to L (right end) (Fig. 1).

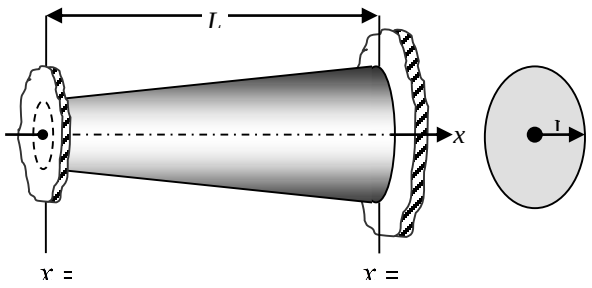


Fig. 1. Calculation diagram of the problem.

The temperature is fixed at the left clamped end as $T(x = 0) = T_1$, and at the right end as $T(x = L) = T_{2n+1}$. The lateral surfaces of the sections $(0 \le x \le x_1)$, $(x_2 \le x \le x_3)$ and $(x_4 \le x \le x_L)$ are thermally insulated. In the section $(x_1 \le x \le x_2)$ heat exchange with the environment occurs through the lateral surface area, with a heat transfer coefficient h, and an ambient temperature T_{co}. Additionally, a heat flux of constant intensity q is applied to the lateral surface area in the section $(x_3 \le x \le x_4)$ The objective is to numerically investigate the influence of the temperature value $T_0 \in [(-150\ C) \div (+150\ C)]$ on the system.[21].

On the temperature distribution field $(T = T(x))$, elastic displacement $(u = u(x))$, as well as components of deformation $(\varepsilon_x = \varepsilon_x(x);\ \varepsilon_T = \varepsilon_T(x);\ \varepsilon = \varepsilon(x))$ and voltage $(\sigma_x = \sigma_x(x);\ \sigma_T = \sigma_T(x);\ \sigma = \sigma(x))$. To develop a mathematical model of the temperature distribution along the length of the considered partially thermally insulated rod of finite length, the rod is discretized using quadratic elements with three nodes. The total number of elements is denoted by n. Consequently, the total number of nodes will be $(2n + 1)$. The discretization is performed in such a manner that the element boundaries coincide with the boundaries of the thermally insulated regions of the rod. For each element, a functional expression is derived that characterizes its total thermal energy. Specifically, for elements belonging to the thermally insulated sections of the rod, the functional I_i is given by:

$$I_i = \int_{V_i} \frac{K_{xx}}{2} \left(\frac{\partial T}{\partial x}\right)^2 \mathrm{dV}, (i = 1, 2, \ldots) \quad (2)$$

Here, K_{xx} represents the thermal conductivity coefficient along the *x*-axis, T is the temperature, and V_i is the volume of the *i-th* element. This integral expression accounts for the thermal energy stored within each element due to the temperature gradient along the rod. Where V_i- volume of the *i-th* element.

For elements located on the section of the rod where heat exchange occurs through the lateral surface, the expression for the corresponding functional takes into account both the internal thermal energy due to the temperature gradient and the heat exchange with the environment [22].

$$I_j = \int_{V_j} \frac{K_{xx}}{2} \left(\frac{\partial T}{\partial x}\right)^2 \mathrm{dV} + \int_{S_{\text{jПБП}}} \frac{h}{2} (T - T_{co})^2 \mathrm{dS}, (j = 1, 2, \ldots) \quad (3)$$

Where V_j- volume of the *j-th* element, $S_{\text{jПБП}}$- area of the lateral surface of the *j-th* element.

For elements located on a section of the rod where a heat flux of constant intensity q is supplied through the lateral surface, the functional expression that characterizes their total thermal energy includes contributions from both the internal thermal energy due to the temperature gradient and the external heat flux applied to the surface.

$$I_k = \int_{V_k} \frac{K_{xx}}{2} \left(\frac{\partial T}{\partial x}\right)^2 \mathrm{dV} + \int_{S_{\text{kПБП}}} \mathrm{qT}(x) \mathrm{dS}, (k = 1, 2, \ldots) \quad (4)$$

The general expression for the functional of total thermal energy for a partially thermally insulated rod with a variable cross-section, accounting for local temperatures, heat flux, and heat transfer, can be derived by combining the contributions from different sections of the rod. These contributions include the internal thermal energy due to the temperature gradient, heat exchange with the environment, and the applied heat flux.

$$I = \sum_{t=1}^{n} I_t \quad (5)$$

To construct a mathematical model of the temperature distribution field along the length of the rod, the functional representing the total thermal energy must be minimized with respect to the nodal temperature values. This minimization leads to a system of linear algebraic equations, which can be solved to obtain the temperature distribution

$$\frac{\partial I}{\partial T_t} = 0, (t = 2, 3, \ldots, 2n) \quad (6)$$

Because T_1 and T_{2n+1} are considered given, then the number of equations in system (6) will be equal to $(2n + 1)$.

Solving the system for different values T_1 and fixed values T_{2n+1}, h,T_{co}, as well as q, the influence of T_1 on the nature of the temperature distribution field along the length of the rod in question.[23].

After constructing the temperature distribution field along the length of the rod, the next step is to develop a mathematical model for the distribution field of elastic displacement, as well as the components of deformation (strain) and stress. This model is crucial for understanding the mechanical response of the rod to the thermal loads. To do this, the rod under study is discretized $\left(N = \frac{n}{2}\right)$ quadratic elements with three nodes. After obtaining the temperature distribution and determining the displacement field, the next step is to write the expression for the functional of the potential energy of elastic deformation for each element. This functional represents the elastic energy

stored in the rod due to the deformation caused by both mechanical and thermal effects.

$$\Pi_i = \int_{V_i} \frac{\sigma_x \varepsilon_x}{2} \mathrm{dV} - \int_{V_i} \alpha \mathrm{ET}(x) \mathrm{dV}, (i = 1, 2, ..., N) \quad (7)$$

Where V_i - volume of the *i-th* element, $u = u(x)$ - elastic displacement distribution field, $\varepsilon_x = \frac{\partial u}{\partial x}$ - distribution field of the elastic component of deformation, $\sigma_x = E\varepsilon_x = E \cdot \frac{\partial u}{\partial x}$ - distribution field of the elastic component of stress, E - elastic modulus of the rod material, α - coefficient of thermal expansion of the rod material, $T = T(x)$ - temperature distribution field determined from the solution of system (6).

For the considered rod as a whole, the expression for the potential energy of elastic deformation is as follows:

$$\Pi = \sum_{i=1}^{N} \Pi_i \quad (8)$$

To construct a mathematical model for the distribution of elastic displacement along the length of the rod, the functional of the potential energy of elastic deformation is minimized with respect to the nodal values of the elastic displacement. This minimization leads to a system of linear algebraic equations that describe the elastic displacement field.[26].

$$\frac{\partial \Pi}{\partial u_i} = 0, (i = 1, 2, ..., (2N + 1)) \quad (6)$$

Solving this system, the elastic displacement distribution field is determined $u = u(x)$ along the length of the rod in question. Based on them, the corresponding fields for the distribution of the components of deformation and stress are constructed as follows:

$$\varepsilon_x = \frac{\partial u}{\partial x}; \varepsilon_T = -\alpha T(x); \varepsilon = \varepsilon_x + \varepsilon_T \quad (10)$$

$$\sigma_x = E\varepsilon_x; \sigma_T = E\varepsilon_T; \sigma = (\sigma_x + \sigma_T) \quad (11)$$

To carry out numerical studies, we take the following as initial data:

$L = 20$ (cm), $r_0 = 1$ (cm), $r_\ell = 2$ (cm), $n = 200$, $N = \frac{n}{2} = 100$, $q = -1000$ (W/cm^2), $K_{xx} = 100$ (W/(cm $\cdot$ C)), $h = 10$ (W/(cm^2 $\cdot$ C)), $T_{co} = 40\ (C)$, $T_{401} = 150\ (C)$, and vary the value $T_1 \in [(-150\ C) \div (+150\ C)]$ in increments $(-50\ C)$.

Consider the following 7 options. In all options except value T_1, the values of all parameters are fixed.

III. Results and Discussion

A. Option-1

Consider the case when $T_1 = 100\ (C)$, i.e. previous value $T_1 = 150\ (C)$ let's reduce it by 1/3. In this case, the nodal temperature values are given in Table I. The corresponding field temperature distribution is given in Fig. 2. From them it is clear that the highest nodal temperature value will be $T_{277} = 262{,}089\ (C)$. The coordinate of this section $x = 13{,}8$ (см), that the phenomenon is caused by the supply to the closed side surface of the area1 $2 \le x \le 16$ (см) heat flux rod of constant intensity $q = -1000$ (W/cm^2). Reducing the nodal temperature values on the site $4 \le x \le 8$ (cm) the rod is due to the ongoing heat exchange with the surrounding closed side surface of this section. Therefore, the smallest nodal temperature value $T_{89} = 85{,}603\ (C)$. The coordinate of this section $x = 4{,}4$ (см).

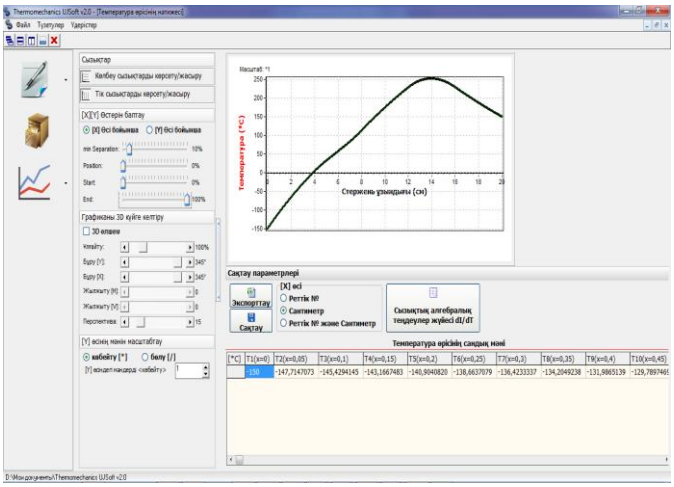


Fig. 2. Field temperature distribution at different values $T(x = 0) = T_1$.

TABLE I. Nodal Temperature Values

Nodal points	$T\ (^\circ C)$	Nodal points	$T\ (^\circ C)$	Nodal points	$T\ (^\circ C)$	Nodal points	$T\ (^\circ C)$
1	100,00	101	87,031	201	192,399	301	253,884
2	99,793	102	87,267	202	193,712	302	253,199
3	99,586	103	87,522	203	195,025	303	252,485
...	...	...	...	...	...	...	...
10	98,170	110	89,802	210	204,078	310	246,763
20	96,229	120	94,601	220	216,659	320	236,416
30	94,378	130	101,289	230	228,839	330	224,721
40	92,611	140	110,009	240	240,639	340	213,328
50	90,923	150	120,969	250	250,826	350	202,240
60	89,309	160	134,441	260	257,685	360	191,444
70	87,764	170	149,334	270	261,349	370	180,928
80	86,283	180	163,729	280	261,960	380	170,683
90	85,622	190	177,635	290	259,650	390	160,697
100	86,812	200	191,078	300	254,545	400	150,961
						401	150,000

Table II presents the nodal displacement values. The corresponding field distribution of displacements along the length of the considered rod of variable cross-section is given in Fig. 3. From these it is clear that all sections of the rod under study move against the direction of the Ox axis. In this case, the section of the rod with the coordinate $x = 8{,}8$ (cm) moves more than others i.e. $u_{89} = -0{,}0142153$ (cm).

Fig. 4 their corresponding distribution field along the length of the rod under consideration is given. From them it is clear that on the site $0 \le x \le 8{,}75$ (cm) of the rod, the behavior of the elastic component of the deformations will be compressive, and then tensile. At the same time, the greatest compressive ε_x corresponds near the left pinched end. Highest tensile value $\varepsilon_x = 0{,}0018518$ which corresponds to the section coordinate $x = 14{,}45$ (cm). Corresponding field distribution ε_T along the length of the rod in question is given by Fig. 6. It

should be noted that along the entire length of the rod under study ε_T has a compressive character. Its greatest value corresponds to the section with coordinate $x = 13{,}85$ (cm). Here $\varepsilon_T = -0{,}0032759$. The corresponding field distribution is given in Fig. 6. From these it is clear that starting from the left pinched end of the rodεfalls monotonously. But along the entire length of the rod it has a compressive character. Its greatest value corresponds to the left pinched end of the rod. Near the left end its value is equal to $\varepsilon = -0{,}0041062$.

TABLE II. NODAL DISPLACEMENT VALUES

Nodal points	u $(см)$	Nodal points	u $(см)$	Nodal points	u $(см)$	Nodal points	u $(см)$
1	0,000000	51	-0,010882	101	-0,0138770	151	-0,006699
2	-0,000285	52	-0,011036	102	-0,0138180	152	-0,006517
3	-0,000572	53	-0,011190	103	-0,0137557	153	-0,006337
…	…	…	…	…	…	…	…
10	-0,002465	60	-0,012174	110	-0,0131918	160	-0,005114
20	-0,004907	70	-0,013294	120	-0,0120301	170	-0,003543
30	-0,007077	80	-0,014002	130	-0,0104933	180	-0,002187
40	-0,009009	90	-0,014212	140	-0,0087232	190	-0,001037
50	-0,010723	100	-0,013930	150	-0,0068811	200	-0,000084
						201	0,000000

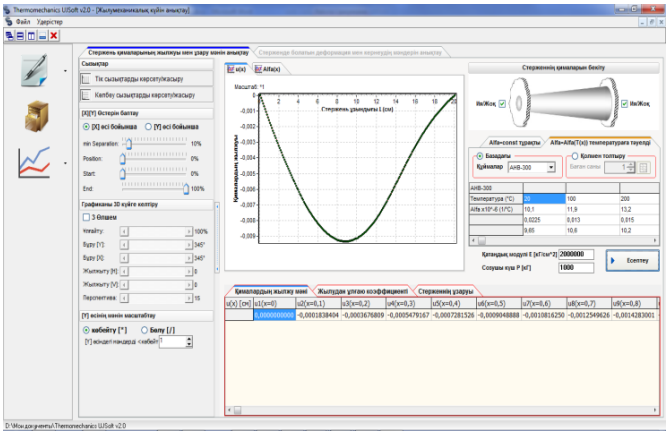
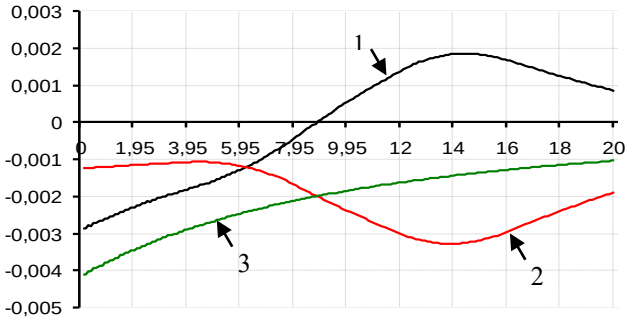


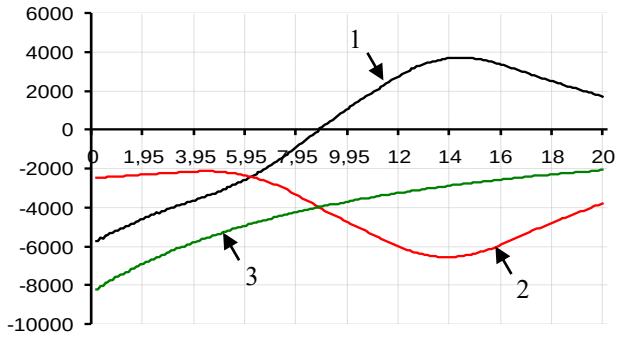
Fig. 3. Distribution field displacement at different values $T(x = 0) = T_1$.

The field distribution of these stress components along the length of the rod is given in Fig. 5. From these materials it is clear that the elastic component of the stress σ_x location on $0 \le x \le 8{,}75$ (cm) the rod behaves compressively, and then has a tensile character. Temperature component voltage σ_T along its entire length it has a compressive character. Its greatest value $\sigma_T = -6551{,}887$ (kG/cm^2). This corresponds to the section whose coordinate $x = 13{,}85$ (cm). Thermoelastic stress component $\sigma = \sigma_x + \sigma_T$ along its entire length it has a compressive character. Its highest value corresponds to the left pinched end of the rod, i.e. $\sigma = -8212{,}409$ kG/cm^2). Starting from left to right, it monotonically decreases and at the right pinched end it has the smallest value $\sigma = -2084{,}123$ (kG/cm^2). In this case, the magnitude of the compressive force $R_2 = \sigma_n(x = 0{,}05) \cdot F_n = -25929{,}2052$ (kG). Naturally, this is less than in the case$T_1 = 150\ (C)$ by 5.4%.



$1 - \varepsilon_x; 2 - \varepsilon_T; 3 - \varepsilon = \varepsilon_x + \varepsilon_T$

Fig. 4. Field distribution of strain components at $T(x = 0) = T_1 = 100\ (C)$.



$1 - \sigma_x; 2 - \sigma_T; 3 - \sigma = \sigma_x + \sigma_T$

Fig. 5. Field distribution of voltage components at $T(x = 0) = T_1 = 100\ (C)$.

Now consider the next option.

B. Option-2

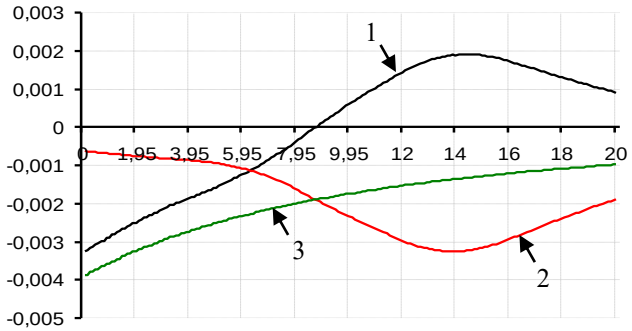
Compared to option 1, in this option the value of the set temperature T_1 l let's reduce it by three times, i.e. let's accept $T_1 = 50\ (C)$. In this case, the nodal temperature values are presented in Table III. The temperature distribution corresponding to the field along the length of the rod under study is shown in Fig. 2. In this case, the highest temperature value will be $T_{278} = 260{,}034\ (C)$. This temperature value corresponds to the section of the rod whose coordinate $x =$ 13,85 (cm). Naturally, this is due to the supplied heat flow of constant intensity and power$q = -1000$ (W/cm^2)on the side surface areas $12 \le x \le 16$ (cm) rod.

The distribution field of the displacement field is given in Fig. 3. From these data it is clear that all sections of the rod under study move against the direction of the Ox axis. When the section moves the most, the coordinate of which $x =$ 8,7 (cm). This section moves against the direction of the Ox axis by $u_{88} = -0{,}0148657$ (cm).

TABLE III. NODAL TEMPERATURE VALUES ($T_1 = 50\ (^\circ C)$)

Nodal points	$T\ (^\circ C)$	Nodal points	$T\ (^\circ C)$	Nodal points	$T\ (^\circ C)$	Nodal points	$T\ (^\circ C)$
1	50,000	101	75,942	201	188,655	301	252,280
2	50,292	102	76,389	202	189,993	302	251,613
3	50,583	103	76,850	203	191,330	303	250,917
...	...	...	...	...	...	...	...
10	52,578	110	80,437	210	200,555	310	245,321
20	55,314	120	86,740	220	213,374	320	235,151
30	57,922	130	94,577	230	225,785	330	223,627
40	60,411	140	104,153	240	237,808	340	212,401
50	62,790	150	115,724	250	248,213	350	201,475
60	65,064	160	129,596	260	255,281	360	190,837
70	67,242	170	144,773	270	259,149	370	180,475
80	69,329	180	159,441	280	259,958	380	170,380
90	71,831	190	173,611	290	257,840	390	160,540
100	75,506	200	187,308	300	252,922	400	150,947

The field-corresponding distribution of these components along the length of the rod in question is shown in Fig. 6.



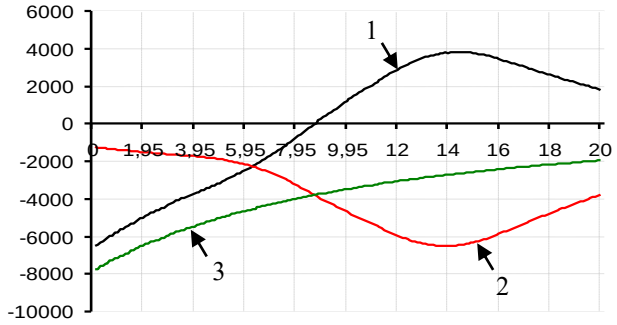
$1 - \varepsilon_x;\ 2 - \varepsilon_T;\ 3 - \varepsilon = \varepsilon_x + \varepsilon_T$

Fig. 6. Field distribution of strain components at$T(x = 0) = T_1 = 50\ (C)$.

From these data it is clear that the behavior of the elastic deformation component in the area $0 \le x \le 8{,}65$ (cm)the rod will be compressive, and then tensile. In this case, the largest compressive elastic component of deformations ε_x falls close to the left clamped end of the test rod of variable cross-section, $\varepsilon_x = -0{,}0032431$. Greatest tensile ε_x corresponds to the section with coordinate $x = 14{,}45$ (cm). In this section the value ε_x will $\varepsilon_x = 0{,}0019087$. Behavior of the temperature component ε_T along the entire length of the rod will be compressive. At the same time, the greatest compressive ε_T corresponds to the section with coordinate $x = 13{,}85\ (cm)$. In this section $\varepsilon_T = -0{,}0032504$.

Unlike ε_x and ε_T field distribution of the thermoelastic component of deformations $\varepsilon = \varepsilon_x + \varepsilon_T$ will be described by a smooth curve. Along the entire length of the rod it has a compressive character. Moreover, its greatest value $\varepsilon = -0{,}0038718$ corresponds closer to the left pinched one, and the smallest $\varepsilon = -0{,}0009826$ closer to the right end.

Nodal values of all three stress components (σ_x,σ_T and $\sigma = \sigma_x + \sigma_T$) their field distributions along the length of the rod of variable cross-section under study are shown in Fig. 7.



$1 - \sigma_x;\ 2 - \sigma_T;\ 3 - \sigma = \sigma_x + \sigma_T$

Fig. 7. Field distribution of voltage components at $T(x = 0) = T_1 =$ 50 (C).

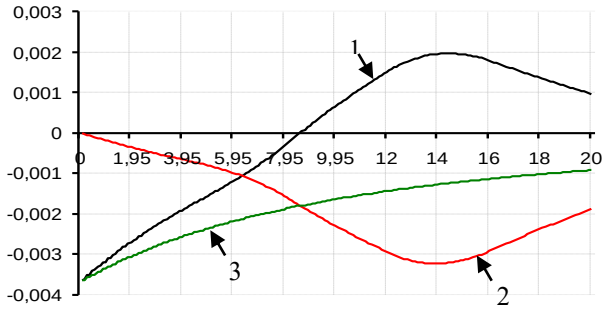
From these results it is clear that the values of these stress components are directly proportional to the values of the corresponding strains. In addition, in this case the value of the compressive force will be $R_3 = \sigma_n(x = 0{,}05) \cdot F_n = -24448{,}9304$ (kG) . This value is 10.8% less than the compressive force that occurs when $T_1 = 150\ (C)$.

C. Option-3

Now consider the fourth option, when $T_1 = 0\ (C)$. The corresponding field temperature distribution along the length of the test rod of variable cross-section is given in Fig. 2.

The corresponding displacement distribution fields along the length of the rod under study are shown in Fig. 3. It should be noted here that, except for the pinched ends, all sections of the rod move against the direction of the *Ox* axis. At the same time, in this direction the section whose coordinate moves more $x = 8{,}6$ (cm). The displacement value of this section $u_{87} = -0{,}0155244$ (cm).

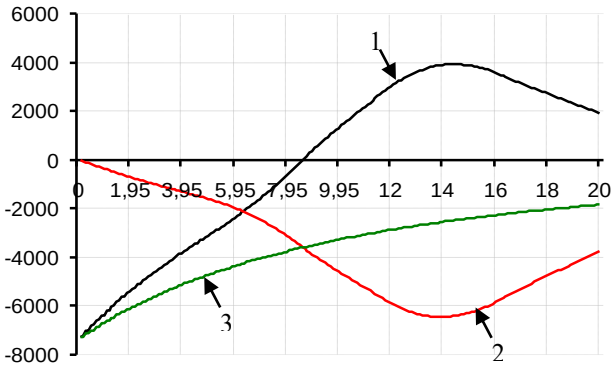
The corresponding field distribution along the length of the rod under study is shown in Fig. 8.



$1-\varepsilon_x; 2-\varepsilon_T; 3-\varepsilon=\varepsilon_x+\varepsilon_T$

Fig. 8. Field distribution of strain components at $T(x=0)=T_1=0\ (C)$.

From these results it is clear that in the area $0 \le x \le 8{,}55$ (cm)behavior of the rod elastic component of deformation ε_x will be compressive, and then tensile. In this case, the greatest compressive ε_x corresponds near the left pinched end. The highest value of the compressive elastic component of deformation $\varepsilon_x=-0{,}0036275$, and tensile $\varepsilon_x=0{,}0019657$ and it corresponds to the section with coordinate $x=14{,}45$ (cm) rod. Behavior of the temperature component of deformation ε_T along the entire length of the rod will be compressive. In this case, its lowest value corresponds near the left pinched end, and its highest value $\varepsilon_T=-0{,}0032249$ corresponds to the section with coordinate $x=13{,}85$ (cm) rod. As can be seen from Fig. 8 field distribution of the thermoelastic component of deformation $\varepsilon=\varepsilon_x+\varepsilon_T$ is described by a smooth monotonically increasing curve. At the same time, behaviorεalong the entire length of the rod under study will be compressive. Its greatest value occurs near the pinched left end of the rod and will be equal to $\varepsilon=-0{,}0036374$. Its smallest value corresponds closer to the right pinched end of the rod under study.



$1-\sigma_x; 2-\sigma_T; 3-\sigma=\sigma_x+\sigma_T$

Fig. 9. Field distribution of voltage components at $T(x=0)=T_1=0\ (C)$.

The field distribution of these voltage components is given in Fig. 9. Naturally different behavior matches behavior ε_x,ε_T and $\varepsilon=\varepsilon_x+\varepsilon_T$. The values of these stress components will be proportional to the corresponding strain components. In this case, the compressive force values R_4 will be equal $R_4=\sigma_n(x=0{,}05)\cdot F_n=-22968{,}6555$ (kG). This is less than R_1(in the case when $T_1=150\ (C)$) by 16.2%.

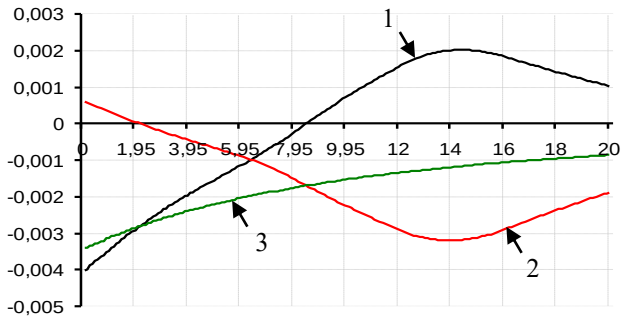
Now let's look at the fifth option.

D. Option-4

In this version we will accept $T_1=-50\ (C)$. The corresponding temperature distribution field along the length of the rod of variable cross-section under study is shown in Fig. 2. In this case, the maximum nodal temperature values $T_{\max}=T_{279}=255{,}968\ (C)$ and it corresponds to the cross section of the rod in question whose coordinate $x=13{,}9$ (cm).

The corresponding distribution field on Fig. 3. In this case, all sections of the rod move against the direction of the *Ox* axis. In this direction the greatest movement of the section of the rod whose coordinate is $x=8{,}4$ (cm). The displacement value of this section $u_{\max}=u_{85}=-0{,}0161891$ (cm).

The field-corresponding distributions of these strain components are shown in Fig. 10.



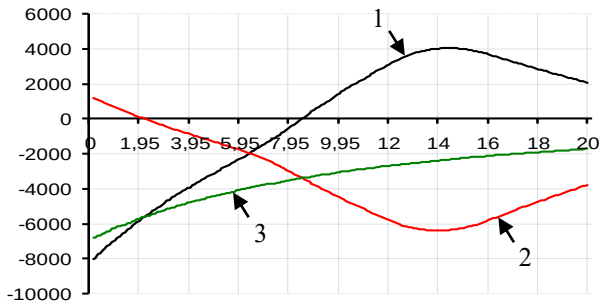
$1-\varepsilon_x; 2-\varepsilon_T; 3-\varepsilon=\varepsilon_x+\varepsilon_T$

Fig. 10. Field distribution of strain components at$T(x=0)=T_1=-50\ (C)$.

From these results it is clear that the behavior of the elastic component of deformations ε_x l ocation on $0 \le x \le 8{,}35$ (cm) of the rod under study will be compressive, and then tensile. Highest compressive value $\varepsilon_x=-0{,}0040118$ and it corresponds near the left pinched end of the rod. In this case, the largest tensile value of the strain component $\varepsilon_x=0{,}0020226$ and it corresponds to the section with coordinate $x=14{,}45$ (cm) rod. Unlike previous options, the behavior of the temperature component of deformations ε_T will be alternating. Location on $0 \le x \le 2{,}05$ (cm) core behavior ε_T will be tensile, and then it behaves compressively. In this case, the greatest tensile value ε_T observed near the left pinched end and it is equal $\varepsilon_T=0{,}0006089$. Location on $2{,}15 \le x \le L=20$ (cm) rod behavior ε_T will be compressive. Maximum compressive temperature component of deformation ε_T will be equal $\varepsilon_T=-0{,}0031994$ and it occurs in the section of the rod whose coordinate is $x=13{,}95$ (cm). Behavior of the thermoelastic component of deformations $\varepsilon=\varepsilon_x+\varepsilon_T$ along the entire length of the rod will have a compressive character. It should be noted that the values ε starting from the left pinched end, it monotonically decreases along the length of the studied rod of

variable cross-section. The highest value $\varepsilon = -0{,}0034029$ which corresponds to the left pinched end of the rod. Near the right pinched end of the test rod, the value ε will be the smallest and will be equal $\varepsilon = -0{,}0008636$.

The distribution of these stress components corresponding to them along the length of the rod of variable cross-section under study is given in Fig. 11.



$1-\sigma_x;\ 2-\sigma_T;\ 3-\sigma=\sigma_x+\sigma_T$

Fig. 11. Field distribution of voltage components at $T(x=0)=T_1=-50\ (C)$.

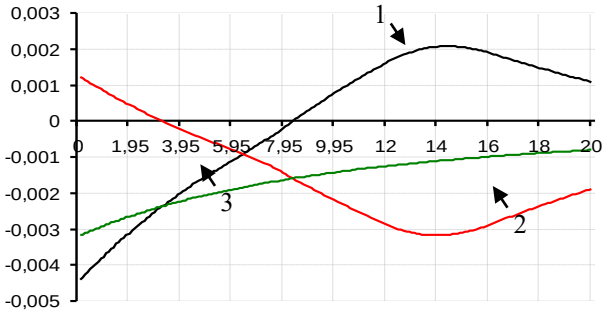
Location on $0 \le x \le 8{,}35$ (cm) behavior of the elastic component of stress of the rod under study σ_x will be compressive. In this case, the highest compressive stress σ_x observed near the left pinched end and will $\sigma_x = -8023{,}682\ (\text{kG/cm}^2)$. On another part of the rod $8{,}45 \le x \le L = 20$ (cm)behavior σ_x will be stretchy. Moreover, its greatest value $\sigma_x = 4045{,}210\ (\text{kG/cm}^2)$ observed near the section whose coordinate $x = 14{,}45$ (cm). Unlike σ_x behavior of the temperature component of voltage σ_T location on $0 \le x \le 2{,}05$ (cm) the rod under study will be tensile. More over, its greatest value $\sigma_T = 1217{,}790\ (\text{kG/cm}^2)$observed near the left pinched end. On the rest of the behavior rod σ_T will be compressive. The highest value of compressive stress $\sigma_T = -6398{,}852\ (\text{kG/cm}^2)$ observed in the section whose coordinate $x = 13{,}95$ (cm) . Behavior of the thermoelastic stress component $\sigma = \sigma_x + \sigma_T$ along the entire length of the rod under study will have a compressive character. Its greatest value $\sigma(x = 0{,}05) = -6805{,}893\ (\text{kG/cm}^2)$ observed at the left pinched end of the rod. As the length of the rod increasesσits value decreases monotonically and near the right pinched end $\sigma(x = 19{,}95) = -1727{,}199\ (\text{kG/cm}^2)$, i.e.$\frac{\sigma(x=0{,}05)}{\sigma(x=19{,}95)} = 3{,}94$ times. From the obtained values $\sigma = \sigma_x + \sigma_T$ let's calculate the value of the compressive $R_5 = \sigma_n(x = 0{,}05) \cdot F_n = -21488{,}3838$ (kG).

Now consider the next sixth option.

E. Option-5

In this option, we assume that the given value is $T_1 = -100\ (C)$. In this case, the field temperature distribution along the length of the considered rod of variable cross-section is shown in Fig. 2. From Fig. 2 it is clear that the highest temperature value, $T_{280} = 253{,}952\ (C)$and it corresponds to the section of the rod whose coordinate $x = 13{,}95$ (cm).

The corresponding field of elastic displacement of sections of the rod is shown in Fig. 3. From Fig. 3 it can be seen that all sections of the rod under study move against the direction of the *Ox* axis. Moreover, in this direction the greatest displacement belongs to the section with the coordinate $x = 8{,}3$ (cm). The displacement value of this section will be $u_{84} = -0{,}0168614$ (cm).

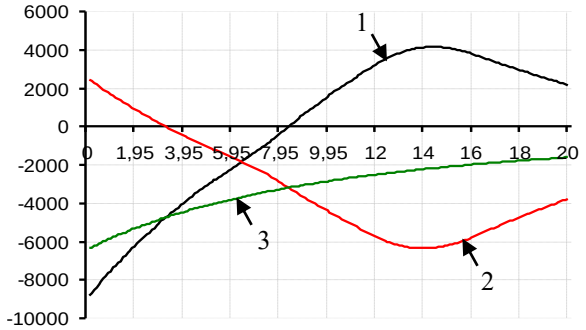


$1-\varepsilon_x;\ 2-\varepsilon_T;\ 3-\varepsilon=\varepsilon_x+\varepsilon_T$

Fig. 12. Field distribution of strain components at $T(x=0)=T_1=-100\ (C)$.

The field-corresponding displacement of these deformation components along the length of the rod under study is shown in Fig. 12. From these tables and the figure it is clear that the behavior of the elastic component of deformations along the length of the rod under study will be alternating in sign. For example, on the site $0 \le x \le 8{,}25$ (cm) rod character ε_x will be compressive, and then it has a tensile character. Maximum compressive elastic deformation $\varepsilon_x = -0{,}0043962$ observed near the left pinched end where the temperature is set $T_1 = -100\ (C)$. In this case, the maximum tensile elastic deformation $\varepsilon_x = 0{,}0020795$ corresponds to the section $x = 14{,}45$ (cm). On the contrary, the behavior of the temperature component of deformation ε_T in the initial section of the rod under study will have a tensile character, and then a compressive character. The highest value of the tensile temperature component of deformation $\varepsilon_T = 0{,}0012277$ corresponds near the left pinched end of the rod under study. At that time, the highest value of the compressive temperature component of deformation ε_T will $\varepsilon_T = -0{,}0031744$ which belongs to the section whose coordinate $x = 13{,}95$ (cm)the rod under study. In Fig. 12, it is clear that the behavior of the thermoelastic component of deformations $\varepsilon = \varepsilon_x + \varepsilon_T$ along the entire length of the rod under study will be compressive. In this case, from the left to the right end of the rod it will decrease monotonically. Highest value $\varepsilon = -0{,}0031685$ corresponds to the section near the left pinched end of the rod.

The distribution of these voltage components corresponding to the field is given in Fig. 13. The behavior of these stress components will correspond to the behavior of similar strain components. In this case, the value of the compressive force$R_6 = \sigma_n(x = 0{,}05) \cdot F_n = -20008{,}1089\ (kG)$.



$1-\sigma_x; 2-\sigma_T; 3-\sigma=\sigma_x+\sigma_T$

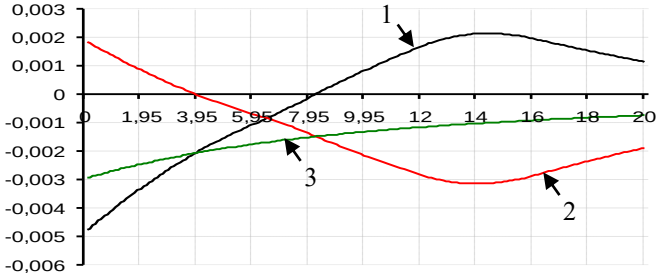
Fig. 13. Field distribution of voltage components at $T(x=0)=T_1=-100\ (C)$.

Finally, let's look at the last option.

F. Option-6

In this option the value T_1 let's accept $T_1=-150\ (C)$. Field temperature distribution along the length of the test rod of variable cross-section in the case $T_1=-150\ (C)$ shown in Fig. 2. In this case, the highest temperature value $T_{280}=251{,}950\ (C)$ corresponds to the section whose coordinate $x=13{,}95\ (cm)$.

The corresponding displacement field is shown in Fig. 3. From these results it is clear that all sections of the rod except the clamped ones move against the direction of the *Ox* axis. In this case, the greatest movement $u_{83}=-0{,}0175417$ (cm) which corresponds to the section coordinate $x=8{,}2$ (cm).

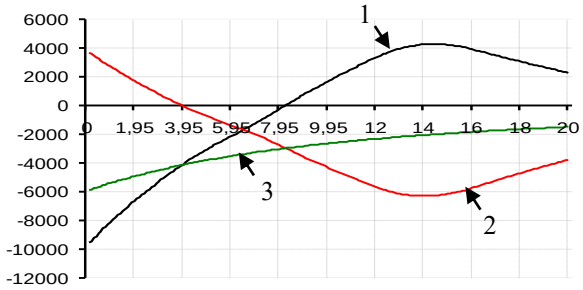


$1-\varepsilon_x; 2-\varepsilon_T; 3-\varepsilon=\varepsilon_x+\varepsilon_T$

Fig. 14. Field distribution of strain components at $T(x=0)=T_1=-150\ (C)$.

The field-corresponding distributions of these strain components are shown in Fig. 14. From Fig. 14 it is clear that the behavior of the elastic component of deformations ε_x location on $0\le x\le 8{,}15$ (cm) will be compressive, and then tensile. In this case, the greatest value of the compressive force ε_x observed near the left pinched end of the rod and will $\varepsilon_x=-0{,}0047805$. The largest value of the tensile component of deformation $\varepsilon_x=0{,}0021365$ corresponds to the section $x=14{,}45$ (cm). From Fig. 15 it is clear that the behavior of the temperature components of deformations ε_T at the initial section $0\le x\le 3{,}85$ (cm) the rod will be tensile, and then compressive. In this case, the greatest value is tensile $\varepsilon_T=0{,}0018464$ which is observed near the left pinched end of the rod. The highest compressive value $\varepsilon_T=-0{,}0031494$ which corresponds to the section whose coordinate $x=13{,}95$ (cm). From the results given Fig. 14 it is clear that the behavior of the thermoelastic component of deformations $\varepsilon=\varepsilon_x+\varepsilon_T$ along the entire length of the rod will be compressive. Moreover, as x increases, its value decreases monotonically. Highest value ε observed near the left pinched end of the rod and will be equal to $\varepsilon=-0{,}0029341$.

Table IV presents the values of the stress components σ_x, σ_T, $\sigma=\sigma_x+\sigma_T$ in the cross-sections of the studied rod. The field-corresponding distribution of these components along the length of the test rod of variable cross-section is shown in Fig. 15.



$1-\sigma_x; 2-\sigma_T; 3-\sigma=\sigma_x+\sigma_T$

Fig. 15. Field distribution of voltage components at $T(x=0)=T_1=-150\ (C)$.

TABLE IV. NODAL VALUES OF σ_x, σ_T, $\sigma=\sigma_x+\sigma_T$ AT $T_1=-150\ (^{\circ}C)$

Nodal points	Nodal points σ_x	Nodal points σ_T	Nodal points $\sigma=\sigma_x+\sigma_T$
1	-9561,083	3692,868	-5868,216
2	-9447,384	3579,169	-5868,216
3	-9219,750	3466,593	-5753,157
...	...	...	...
10	-8136,232	2708,653	-5427,579
50	-3066,817	-757,061	-3823,878
100	1653,829	-4305,735	-2651,906
150	4214,083	-6160,725	-1946,642
200	2283,025	-3772,275	-1489,250

It should be noted here that the behavior of these stress components will be like the corresponding strain components. Naturally, the value of the stress components will be proportional to the values of the corresponding strain components. In this case, the corresponding value of the compressive force $R_7=\sigma_n(x=0{,}05)\cdot F_n=-18527{,}8372$ (kG).

By analyzing the seven options considered, you can build a comparative Table I. From this table it can be seen that with

decreasing value T_1 the magnitude of the resulting compressive force R decreases. For large negative values T_1 the magnitude of the compressive force is noticeably reduced. Thus, setting the values T_1 it is possible to control the magnitude of the compressive force R arising from the distribution of the temperature field in such a way that this rod element of the variable cross-section of the structure does not collapse. We will also build Fig. 16, where a curve is given that characterizes the relationship between the resulting compressive force R and the values of the given temperature $T(x = 0)$ on the left pinched end. It should be noted here that the radius of the left end of the rod is two times smaller than the right one.

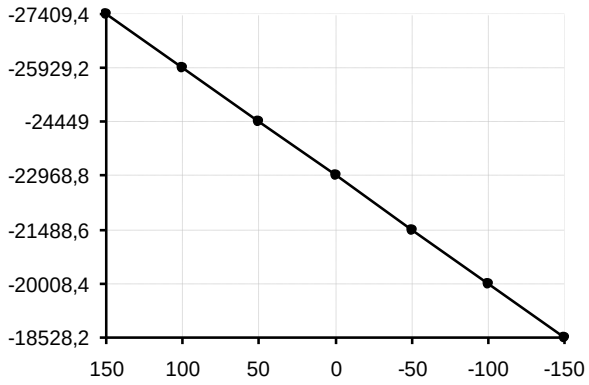


Fig. 16. Dependency between$T(x = 0)$and R.

TABLE V. INFLUENCE VALUE $T(x = 0) = T_1$BY THE MAGNITUDE OF THE RESULTING COMPRESSIVE FORCE R

No. p/p	T_1 (C)	T_{max} (C)	Coord. sections (cm)	u_{max} (cm)	Coord. section (cm)
1	150	264.153	$x = 13{,}75$	-0.0135704	$x = 8{,}9$
2	100	262,089	$x = 13{,}80$	-0.0142153	$x = 8{,}8$
3	50	260,034	$x = 13{,}85$	-0.0148657	$x = 8{,}7$
4	0	257,993	$x = 13{,}85$	-0.0155244	$x = 8{,}6$
5	-50	255,968	$x = 13{,}90$	-0.0161891	$x = 8{,}4$
6	-100	253,952	$x = 13{,}95$	-0.0168614	$x = 8{,}3$
7	-150	251,950	$x = 13{,}95$	-0.0175417	$x = 8{,}2$

IV. CONCLUSION

As a result, the following tasks were completed:

1) Algorithms have been compiled for studying local heat flows and heat transfer problems transmitted in the fields of temperature, displacement, deformation and stress on a variable rod with a cross-sectional area of finite length;

2) Algorithms and methods have been created for determining the thermal field of the side surface of an insulated column under the influence of heat flow and heat transfer.

The scientific and practical significance of the work lies in the possibility of using the developed approach to study the non-stationary thermophysical characteristics of plant elements having the shape of a rod with different cross-sectional configurations. The results also contribute to the development of research areas in mechanical engineering. The use of developed methods and algorithms for studying non-stationary thermophysical processes in rods with different transverse shapes provides significant savings, as they allow one to select a reliable rod for given operating conditions. This makes it possible to ensure continuous and reliable operation of installations used in mechanical engineering.

In the Republic of Kazakhstan, similar developments and research are practically absent.

The development of the proposed methods and algorithms makes it possible to study in detail complex non-stationary thermophysical processes in rods with different cross-sectional shapes. In addition, they make it possible to theoretically develop an appropriate methodology for studying the thermophysical characteristics of rod structures used in the engineering industry of the Republic of Kazakhstan. The development of such smart products contributes to the development of this industry and strengthens the country's sovereignty.

REFERENCES

[1] Timoshenko, S., Goodier. J. N. (1951). Theory of Elasticity. New York. Available at: http://parastesh.usc.ac.ir/files/1538886893033.pdf.

[2] Shorr, B. F. (2015). Thermoelasticity. Thermal Integrity in Mechanics and Engineering, 33–56. https://doi.org/10.1007/978-3- 662-46968-2_2.

[3] Banerjee, B. (2006). Basic Thermoelasticity. doi: http://dx.doi.org/10.13140/RG.2.1.1144.2005.

[4] Saoud, S. (2009). Etude et Analyse Mathematique des Problems Non Lineaires Modelisant les Etats Thermiques d'un Superconducteur: Generalisation au Cas Tridimensionnel.

[5] Griffith, G., Tucker, S., Milsom, J., Stone, G. (2000). Problems with modern air-cooled generator stator winding insulation. IEEE Electrical Insulation Magazine, 16 (6), 6–10. doi: https://doi.org/10.1109/57.887599.

[6] Li, Y. (2019). Investigation of Heat Transfer Characteristics on Rod Fastening Rotor. IOP Conference Series: Materials Science and Engineering, 677 (3), 032032. doi: https://doi.org/10.1088/1757-899x/677/3/032032.

[7] Shibib, K., Minshid, M., Alattar, N. (2011). Thermal and stress analysis in Nd: YAG laser rod with different double end pumping methods. Thermal Science, 15, 399–407. doi: https://doi.org/10.2298/tsci101201004s.

[8] Andreev, V., Turusov, R. (2016). Nonlinear modeling of the kinetics of thermal stresses in polymer rods. Advanced Materials and Structural Engineering, 719–722. doi: https://doi.org/10.1201/b20958-150.

[9] Belytschko, T., Liu, W. K., Moran, B. (2000). Nonlinear Finite Elements for Continua and Structures. John Wiley and Sons.

[10] Wright, T. W. (2002). The Physics and Mathematics of Adiabatic Shear Bands. Cambridge University Press.

[11] Batra, R. C. (2006). Elements of Continuum Mechanics. AIAA. doi: https://doi.org/10.2514/4.861765.

[12] Sukarno, D. H. (2021). Analysis of nuclear fuel rod temperature distribution using CFD calculation and analytical solution. PROCEEDINGS OF THE 6TH INTERNATIONAL SYMPOSIUM ON CURRENT PROGRESS IN MATHEMATICS AND SCIENCES 2020 (ISCPMS 2020). doi: https://doi.org/10.1063/5.0058888.

[13] El-Azab, J. M., Kandel, H. M., Khedr, M. A., El-Ghandoor, H. M. (2014). Numerical Study of Transient Temperature Distribution in Passively Q-Switched Yb:YAG Solid-State Laser. Optics and Photonics Journal, 04 (03), 46–53. doi: https://doi.org/10.4236/ opj.2014.43007.

[14] Khany, S. E., Krishnan, K. N., Wahed, M. A. (2012). Study of Transient Temperature Distribution in a Friction Welding Process and its effects on its Joints. International Journal Of Computational Engineering Research, 2 (5), 1645.

[15] Mishchenko, A. (2020). Spatially Structure Spatial Problem of the Stressed-Deformed State of a Structural Inhomogeneous Rod. IOP Conference Series: Materials Science and Engineering, 953, 012004. doi: https://doi.org/10.1088/1757-899x/953/1/012004.

[16] Hwang, J.-K. (2020). Thermal Behavior of a Rod during Hot Shape Rolling and Its Comparison with a Plate during Flat Rolling. Processes, 8 (3), 327. doi: https://doi.org/10.3390/pr8030327.

[17] Logan, D. L. (2012). A First Course in the Finite Element Method. CENGAGE Learning, 727–764.

[18] Liu, Q., He, X. (2023). Thermal Analysis of Terfenol-D Rods with Different Structures. Micromachines, 14 (1), 216. doi: https:// doi.org/10.3390/mi14010216.

[19] Gaspar Jr., J. C. A., Moreira, M. L., Desampaio, P. A. B. (2011). Temperature Distribution on Fuel Rods: A study on the Effect of Eccentricity in the Position of UO2 Pellets. 20-th International Conference «Nuclear Energy for New Europe». Available at: https://arhiv.djs.si/proc/nene2011/pdf/814.pdf.

[20] Tashenova, Z., Nurlybaeva, E., Kudaykulov, A."Method preparation and solution algorithm for resolving stationary problem of a rod under thermo - stressed condition restrained at both ends affected by heat exchange and heat flows",Advanced Materials Research, vol. 875-877, pp.858–862, 2014, doi: 10.4028/www.scientific.net/AMR.875-877.858.

[21] Tashenova, Z.H.M., Nurlybaeva, E.N., Kudaikulov, A.K."Method of solution and computational algithm for mixed thermo-mechanics problem",World Applied Sciences Journal, vol. 28(12), pp. 2113–2119, 2013, doi: 10.5829/idosi.wasj.2013.28.12.422

[22] Tashenova, Z.M., Nurlybaeva, E.N., Kudaiykulov, A.K.Method of solution and computational algorithm for mixed thermo-mechanics problem",World Applied Sciences Journal, vol. 22(SPL.ISSUE2), pp. 49–57, 2013, doi: 10.5829/idosi.wasj.2013.22.tt.22139.

[23] Tashenova, Z.H.M., Zhumadillaeva, A.K., Nurlybaeva, E.N., Kudaykulov, A.K. "Numerical study of established thermo-mechanical state of rods of limited length, with the presence of local heat flows, temperatures, heat insulation and heat transfer",Advanced Science Letters, vol.19(8), pp. 2395–2397, 2013, doi: 10.1166/asl.2013.4926.

[24] K.K.Gornostaev,A.V.Kovalev,andY.V.Malygina, "Stress–strain state in an elastoplastic pipe taking into account the temperature and compressibility of the material," Journal of Physics: Conference Series,vol.973,2018.

[25] B. F. Shorr, "Termal integrity in mechanics and engineering," in Foundations of Termoelasticity, pp. 33–55, Springer, Berlin, Germany, 2015.

[26] O. C. Zienkiewicz and R. L. Taylor, Te Finite Element Method, Butterworth-Heinemann, Oxford, UK, 5th edition, 2000.

Performance Analysis of a Hyperledger-Based Medical Record Data Management Using Amazon Web Services

Mohammed K Elghoul[1*], Sayed F. Bahgat[2], Ashraf S. Hussein[3], Safwat H. Hamad[4]
Scientific Computing Department, Faculty of Computer and Information Sciences, Ain-Shams University, Egypt[1, 2, 3, 4]
King Salman International University, South Sinai, Egypt[3]
Saint Mary's College of California, Moraga CA 94575, USA[4]

***Abstract*—Recently, there's been growing excitement around the innovative capabilities of blockchain technology, especially for enhancing security, privacy, and transparency. Its application in various sectors, like finance and logistics, is intriguing, but its potential in healthcare stands out. Specifically, in the realm of medical data management, blockchain can transform how we protect patient data. Our study unveils a cutting-edge approach to handle digital health records by harnessing the power of Amazon Web Services (AWS). This pioneering, serverless model is not only cost-effective, with charges only for used resources, but also offers heightened security and for blockchain network access. We build a private, permissioned blockchain network with Hyperledger Fabric to control access while ensuring transparency. The paper demonstrates the prowess of this new system is validated through rigorous tests on speed, network prowess, and multi-user handling, complete with a detailed cost analysis for implementation. The paper further demonstrates the use of the Gatling open-source library to design various experiments for performance measurement.**

Keywords*—Hyperledger; blockchain; healthcare; data management

I. Introduction

Distributed ledger technology, commonly referred to as blockchain, facilitates the sharing of data between peer-to-peer networks [1]. Its first application was seen in 2008 with the Bitcoin cryptocurrency [2]. The main appeal of blockchain is its low cost, speed, improved security, and direct peer-to-peer transactions without relying on a central third party.

In today's technological age, the increasing reliance on accurate data calls for new methods of storing and analyzing information. Ensuring that data is immutable, secure, and maintains its integrity has become an important part of modern systems. Especially since the implementation of Bitcoin in 2008, blockchain has stood out as an emerging solution [3]. The rapid growth of medical data [2] coupled with the rise of blockchain suggests the need for a new framework. As electronic medical records (EMRs) increase in size and scope [4], modern systems struggle to keep up. Maintaining and securing EMR data is essential, especially given the fragility of patient information and the complexity of sharing such information across locations, and traditional databases often fail to meet these challenges accurate [5]. Moreover, the integration of blockchain technology could offer a robust solution to these issues, providing a decentralized, secure, and transparent platform for managing and sharing EMR data.

Considering blockchain's inherent properties such as data immutability, decentralized ledger, and strong security, it appears to be a powerful solution for EMR management [6]. The immutable and transparent nature of blockchain ensures that once data is stored, it remains untouched [7]. This could significantly reduce the risks of data tampering and unauthorized access, which are common concerns in traditional EMR systems.

Four key characteristics define blockchain: decentralization, immutability, audibility and traceability, and data integrity. This ensures that any transactions or records on the network remain unchanged [8]. Blockchain operates without a central governing body, instead using consensus mechanisms to support data and network peers. Correlations are verified using a Merkle tree-like structure [9], which supports data integrity.

There are basically three types of blockchain: public, federated, and private. Like Bitcoin and Ethereum, public versions are open source. In contrast, confederation block chains limit access to a particular group, and although they are confined to private blocks, they are managed by a single entity. Given the breadth of blockchain applications, there is rarely a universal definition. Table I provides an in-depth comparison of these three types.

This paper explores a comprehensive review of a cloud-centric approach to securing data management systems through blockchain technology, specifically through AWS. The article begins with an overview that emphasizes the need to address security challenges in EMRs. The paper then describes the proposed system architecture, clarifies external components such as the Hyperledger blockchain grid, and specific AWS applications. Additionally, a diagram of the sequence of reactions is provided. Reinforcing the background, the paper presents an experimental performance measure of the system.

In conclusion, the paper recounts his major publications and contributions. In the final sections, it discusses the development of the system and potential efficiencies, demonstrating a progressive research focus and the ability to continuously refine and modernize the system.

*Corresponding Author.

TABLE I. TYPES OF BLOCKCHAINS [2], [9]

Property	Public blockchain	Consortium blockchain	Private blockchain
Consensus determination	All miners	Selected set of nodes	One organization
Read permission	Public	Public or restricted	Public or restricted
Immutability	Nearly impossible	Could be tampered	Could be tampered
Efficiency	Low	High	High
Centralized	No	Partial	Yes
Consensus process	Permissionless	Permissioned	Permissioned

II. RELATED WORK

Asma Khatoon [10] investigated in detail the integration of blockchain-enabled smart contracts in the healthcare industry. His proposal includes a medical system built on smart contracts and blockchain, demonstrating the benefits of decentralization for healthcare services. Key objectives include reducing transaction costs, streamlining administrative tasks, and bypassing intermediaries.

MedChain, a blockchain system designed for the privacy of medical data, was introduced by Daraghmi et al. [11]. Their platform provides patients, healthcare providers and those who value patient privacy with reliable health records. Short-term smart contracts, with sophisticated encryption, are used to manage transactions and promote data security. Additionally, it is recommended that incentives be developed for health professionals to maintain and develop updates.

Through their research, Zhang et al. [12] examined the synergies between blockchain and smart contracts in healthcare, highlighting its effectiveness in solving many healthcare challenges. Furthermore, they shed light on the obstacles faced when integrating blockchain into healthcare.

Kumar and his team [13] examined various applications of blockchain in the healthcare system. They acknowledge the barriers to integration but point to smart contracts as a logical solution in a blockchain-based healthcare system.

Sial and others [14] highlighted the benefits of integrating blockchain with smart contracts to improve healthcare services. They point to blockchain's ability to prevent loss and prevent data manipulation by storing data safely on a ledger. The potential of Hyperledger Fabric for storing medical records was the focus of Daisuke et al. [15], who aimed to transfer medical data from smartphones to the Hyperledger blockchain system.

Aiming to address the obstacles of a permissive and open blockchain system, Rouhani and his team [16] turned to a Hyperledger system to empower patients with suggestive autonomy of their health data to develop a new system developed by Sukhpal Gill [17] to enable cloud maintenance, serverless, Combining quantum computing and blockchain, the management layer covered the IoT devices in a service layer that manages resources and communicates with IoT devices, while the service side performs computing tasks through Serverless FaaS architecture.

Bhati et al. [18] supported the adoption of blockchain in healthcare, especially in the management of EHRs. Their goal was to increase the accessibility and relevance of EHRs by leveraging blockchain's improved security, with the introduction of streamlined access guidelines. Their design also includes external data storage to address scalability issues, ensuring security, and flexibility.

Finally, Anurag and his team [19] compiled the literature on its role in healthcare and delved into the potential of blockchain to manage healthcare intelligence.

A. Blockchain Technology Limitations

When it comes to storing large amounts of data, blockchain faces two important challenges: scalability and privacy. The data stored on the blockchain is visible to all authorized users, which could be a concern for healthcare organizations that need to store sensitive patient information. Additionally, storing a patient's complete medical history, records, visits, lab results, and other reports in the blockchain can cause significant strain on its storage [20].

Blockchain technology is still not fully understood by many, as it is a relatively new field and is constantly evolving. This lack of knowledge and understanding can complicate the adoption of blockchain in healthcare. Additionally, the transition from traditional EHR systems to blockchain will require significant effort, as hospitals and healthcare organizations need to change their systems to take advantage of this new technology.

As blockchain technology is still relatively new and evolving rapidly, there is no established standard for it. This means that implementation in the healthcare industry requires additional time and effort. To ensure the safe and secure use of blockchain, international authorities should develop standardized guidelines to assist in the standardization of this technology.

III. PROPOSED SOLUTION

A. Implementation

Fig. 1 illustrates the system architecture diagram, illustrating the use of the Hyperledger blockchain network for storing and processing medical records.

Each participant in this network operates through a unique client application, which connects them to the blockchain and allows them to access medical data.

Member A, who is the embodiment of the patient only has the right to view his/her own medical history and other ability to change his/her address information On the other hand, member B which refers to the health reputation has the right to do detailed medical information and edits for individual or multiple patients and may also include patient access instructions Conversely, Member C on behalf of the regulatory agency may request and select records a can be adopted for research purposes.

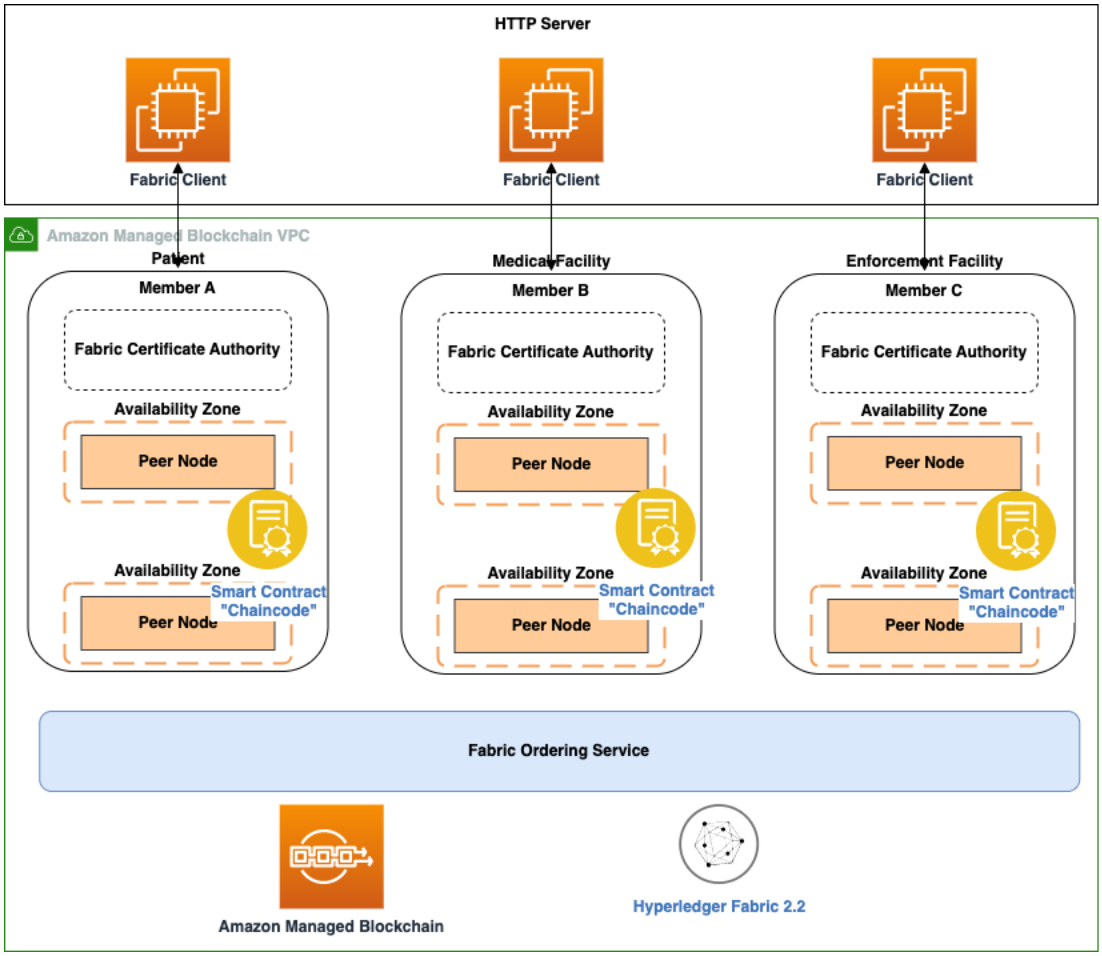


Fig. 1. High level system architecture diagram [21].

Arvind et al. [20] utilized IBM cloud and Kubernetes containers to execute their approach. In contrast, our suggested method employs Amazon Web Services (AWS) and the principle of serverless computing. This strategy allows us to avoid paying for inactive resources and offers the flexibility to adjust our scale according to traffic demands. As we will observe in the results section, this results in substantial performance improvements.

Integrating ownership and authorization protocols is key to maintaining the purity and resilience of the Hyperledger blockchain network. This ensures that only authenticated users can access it, thus protecting the network from inappropriate access or threats. Authentication focuses on supporting the identity of the user or device, while authorization is about defining what activities are allowed for a user or device in the network Using these safeguards reduces the chance of a there is greater access to unauthorized networks or potential security breaches.

In the described Hyperledger framework, authentication and authorization are seamlessly integrated into the client software. As a result, the specific client applications used by network participants come equipped with built-in mechanisms to authenticate and direct users, enabling secure and regulated transactions with the blockchain.

The infrastructural layer of our Hyperledger network is built on the Amazon Managed Blockchain, essentially following version 2.2 of the framework. Notably, Amazon Web Services (AWS) provides two different versions of the said network: 'Starter' and 'Standard'. Given the budget constraints, our decision wanted to utilize the capabilities of the 'starter' version.

TABLE II. MACHINE TYPES SUPPORTED BY AMAZON MANAGED BLOCKCHAIN STARTER EDITION

Machine Type	Member cost	Peer node	Hourly cost	Daily cost
bc.t3.small	$0.30	$0.034	$0.334	$8.0
bc.t3.medium	$0.30	$0.067	$0.367	$8.8

Delving into the specifics, "Table II", enumerates the varies of current compatible device configurations, providing a comparative analysis of their economic implications for general understanding, consider a hypothetical scenario of a network composed of three member groups. The table describes the total cost of this three-member network and breaks down the hourly computation costs for each member next to a peer node It should be noted that a maximum storage

fee of $.20 is charged on, and writing work the cost matrix of this amount is directly proportional to the requirement of identical nodes on the monthly cadence.

B. Sequence Diagram and Transaction Flow

Extending the complexity of the transaction process within our blockchain framework, “Fig. 2”, presents a complete sequence of diagrams describing the entire course of a transaction [21]. This diagram carefully describes a multifaceted level integral to the progress of a transaction. Starting at the point of user interaction, the sequence flows normally, through each level, and finally ends up in the Hyperledger system's integral sequence annotation services once processed in Hyperledger, the feedback reconfigures a route, which reaches the user to the destination. Through this approach, a comprehensive understanding of the complex behavioral journey is gained by emphasizing the interactions between each level within the larger behavioral ecosystem.

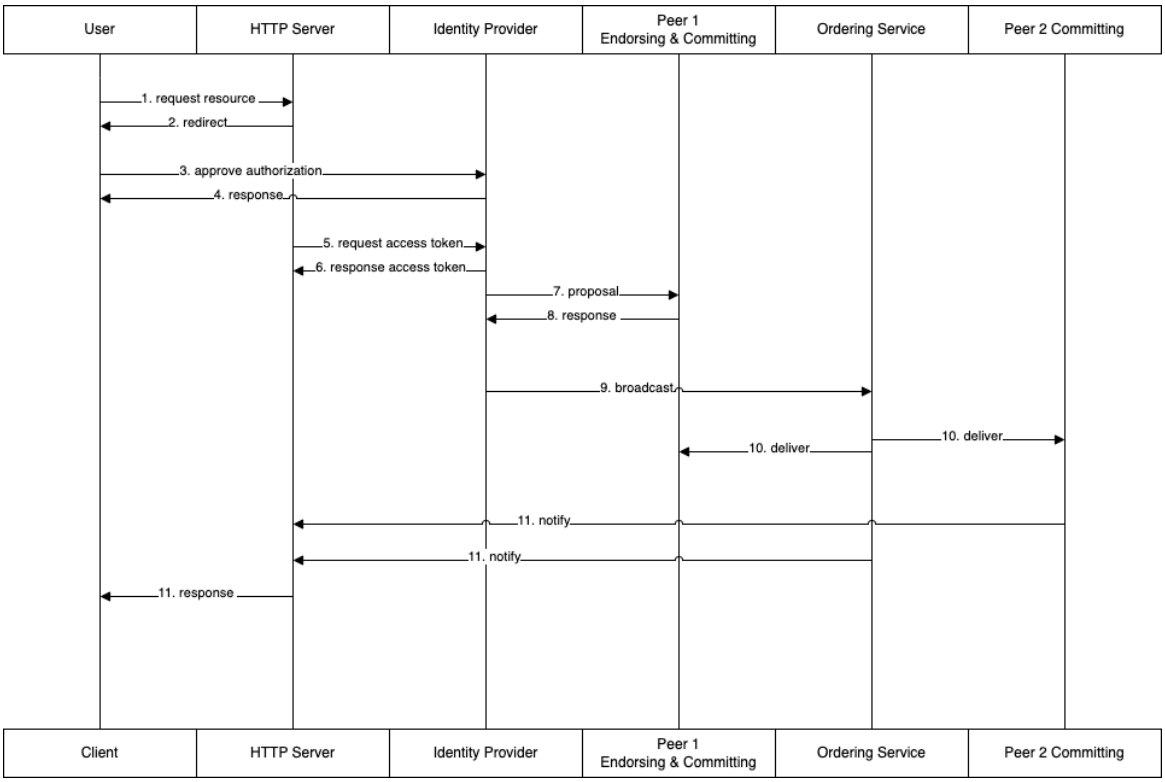


Fig. 2. Transaction flow and sequence diagram.

HTTP and License Phases When a user initiates a request to the server, the server immediately checks the user's license credentials. Based on this authentication, the server grants access or denies the user’s request. If the user credentials meet the required criteria and are considered valid, the server proceeds to communicate with the identity provider, requesting a token. After, the token and user request have been successfully received and validated, the server with the installed Fabric SDK carefully builds a transaction offer, ensuring that the appropriate certificate is included for authentication.

Detailed Steps

1. Initiation by the User
   a. The user sends a request to the server.
   b. The request is routed to a specific function based on the provided URI and HTTP method.
   c. The user embeds a valid token within the request header for authentication and authorization.
2. Token Verification
   a. The system verifies the validity of the embedded token.
   b. If the token is invalid or absent, the user is redirected to a login page.
3. Credential Input and Validation
   a. The user enters their authentication credentials on the login page.
   b. The credentials are validated.
4. Communication with Identity Provider

   A response is generated from the identity provider after validation.
5. Token Request by Server

   The server requests a token from the identity provider.
6. Token Receipt

   The identity provider issues a valid token for the ongoing request.
7. Transaction Proposal and Invocation

a. A transaction proposal is created using the Fabric SDK on the server.
b. The proposal is signed with the correct certificate and sent as an invoke request to the network.

Endorsement Phase

8. Endorser Verification
   a. The endorsing peer verifies that the client is authorized to invoke the chaincode.
   b. If authorized, the endorsing peer executes the chaincode and generates a response without changing the world state.
   c. The endorser signs the proposal with its identity and sends it to the client.
9. Endorsement Collection
   a. The client collects responses from multiple endorsers.
   b. The client verifies that the responses satisfy the endorsement policy.

Ordering Phase

10. Transaction Broadcast: The client broadcasts the endorsed transaction to the ordering service.
11. Block Creation
    a. The ordering service packages the transactions into blocks.
    b. The ordering service signs the blocks.
12. Block Delivery: The ordering service delivers the blocks to the leading peer nodes.

Validation & Committing Phase

13. Block Dissemination: The leading peers disseminate the blocks to all peers in the same channel and organization.
14. Block Verification: The peers verify the signature of the blocks.
15. Transaction Validation: The peers check all the transactions within the blocks.
16. Ledger Update: If all the transactions are valid, the blocks are appended to the ledger and the world state is updated.

Response Phase

17. Event Notification: The HTTP server is notified via the Channel EventHub listener once the target transaction has been committed to the ledger.
18. Response Formation:
    a. A registered callback function collects details of the event.
    b. The callback function forms a response in JSON format using the collected details.
19. Response Delivery: The response is sent back to the user via the client application.

C. Challenges and Limitations

Data migration poses a significant challenge in migrating existing medical records to blockchain-based systems due to format and data type compatibility issues. Careful design with data washing and validation process various uses are paramount to ensure a smooth and accurate transfer while maintaining data integrity and security Compliance with regulatory frameworks Not a trivial matter. The proposed system should not only leverage the advantages of blockchain technology but also strictly comply with these regulations.

User adoption depends on the active involvement of multiple stakeholders, including health care providers, patients, and regulatory agencies. Getting these organizations to thrive in the new system and ensure they are profitable is a difficult task. In terms of security and scalability, it is important to enforce strict security measures, especially when dealing with medical data. New systems should take full advantage of blockchain's inherent security features and add components when needed. It should also demonstrate adaptability to meet the growing demand for health care. From an economic perspective, setting up and maintaining a blockchain-related system to manage healthcare data can be expensive, considering ongoing infrastructure, development and operational costs. If designed such a wonderful implementation, will require skilled professionals with expertise in blockchain technology, security, and data management.

IV. Results and Discussions

After the text edit has been completed, the paper is ready for the template. Duplicate the template file by using the Save As command, and use the naming convention prescribed by your conference for the name of your paper. In this newly created file, highlight all of the contents and import your prepared text file. You are now ready to style your paper; use the scroll down window on the left of the MS Word Formatting toolbar.

In this section, we examine the results of our experiments and subsequent performance implications. Our test revolved around a sample dataset focused on the patient's medical record. We used the open-source Gatling library to facilitate these tests. This tool allowed us to perform simultaneous tasks, specifically designed to create, retrieve, modify, and purchase patient records. Test automation and quality assurance play an important role in monitoring test results because they reduce human effort and cost and improve the accuracy of results.

For our test parameters, we started with a modest ten users working simultaneously. Gradually we increased this, aiming for 2000 users, all within a short period of 100 seconds. Such a configuration has given us the ability to test our system, so that it serves 20 users at once, each performing one task.

The following images offer insights directly from Gatling. "Fig. 3", describes the sum of responses required to complete each request and the corresponding duration. Remarkably, not a single request experienced a failure, with each task taking just over 1.2 seconds to complete. Turning to "Fig. 4", it shows the flow of active users throughout the test phase, with a noticeable upward trend, rising to 1,706 users in a joint What last, "Fig. 5", provides a granular representation of the response time distribution, which shows how long the server needs to return a response.

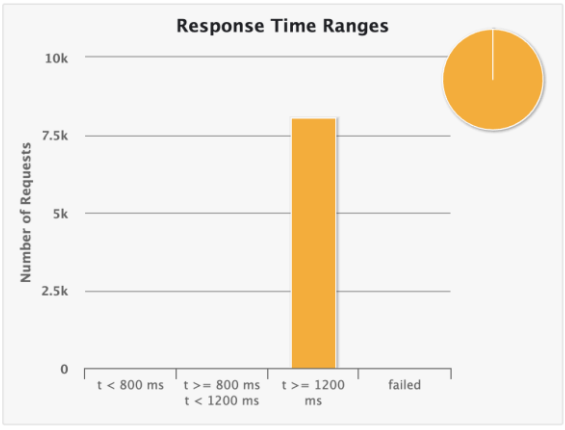


Fig. 3. Response time range and number of requests.

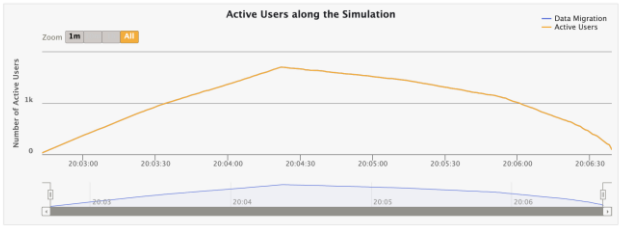


Fig. 4. Active users timeline.

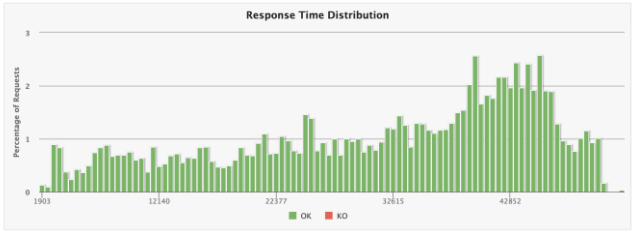


Fig. 5. Response time distribution.

The findings showed that the introduced solutions are efficient and mature in handling the requirements in real-world situations. What sets this platform apart is the efficient use of Amazon web services.

V. Conclusions and Future Work

This research presents a cloud-based approach, for storing and retrieving medical records on the Hyperledger blockchain network. We were able to transfer a volume of historical data consisting of approximately five million records and demonstrated that our system can handle daily traffic effectively. By utilizing the edition of Amazon Managed Blockchain, our solution seamlessly integrates with AWS ensuring adaptability for future data analysis. To ensure accuracy and efficiency we implemented testing with Gatling scripts to assess performance. Not does our system provide storage and analysis capabilities but it also paves the way for further advancements such, as integrating emerging technologies and enhancing security measures. This direction emphasizes how our system can adapt to the evolving demands of healthcare data management.

References

[1] M. Hölbl, M. Kompara, A. Kamišalić, and L. Nemec Zlatolas, "A Systematic Review of the Use of Blockchain in Healthcare," Symmetry (Basel), vol. 10, no. 10, p. 470, Oct. 2018, doi: 10.3390/sym10100470.

[2] M. Elghoul, S. Bahgat, A. Hussein, and S. Hamad, "A Review of Leveraging Blockchain based Framework Landscape in Healthcare Systems," International Journal of Intelligent Computing and Information Sciences, vol. 0, no. 0, pp. 1–13, Oct. 2021, doi: 10.21608/ijicis.2021.75531.1095.

[3] Satoshi Nakamoto, "Bitcoin: A Peer-to-Peer Electronic Cash System," https://bitcoin.org/bitcoin.pdf.

[4] M. K. Elghoul, S. F. Bahgat, A. S. Hussein, and S. H. Hamad, "Securing Patient Medical Records with Blockchain Technology in Cloud-based Healthcare Systems," International Journal of Advanced Computer Science and Applications, vol. 14, no. 11, 2023, doi: 10.14569/IJACSA.2023.0141133.

[5] T. Benil and J. Jasper, "Cloud based security on outsourcing using blockchain in E-health systems," Computer Networks, vol. 178, p. 107344, Sep. 2020, doi: 10.1016/j.comnet.2020.107344.

[6] C. Agbo, Q. Mahmoud, and J. Eklund, "Blockchain Technology in Healthcare: A Systematic Review," Healthcare, vol. 7, no. 2, p. 56, Apr. 2019, doi: 10.3390/healthcare7020056.

[7] A. Ali et al., "Deep Learning Based Homomorphic Secure Search-Able Encryption for Keyword Search in Blockchain Healthcare System: A Novel Approach to Cryptography," Sensors, vol. 22, no. 2, p. 528, Jan. 2022, doi: 10.3390/s22020528.

[8] M. K. Elghoul, S. F. Bahgat, A. S. Hussein, and S. H. Hamad, "Secured Cloud-based Framework for Electronic Medical Records using Hyperledger Blockchain Network," Egyptian Computer Science Journal, vol. 46, no. 2, Sep. 2022.

[9] Z. Zheng, S. Xie, H. Dai, X. Chen, and H. Wang, "An Overview of Blockchain Technology: Architecture, Consensus, and Future Trends," in 2017 IEEE International Congress on Big Data (BigData Congress), IEEE, Jun. 2017, pp. 557–564. doi: 10.1109/BigDataCongress.2017.85.

[10] A. Khatoon, "A Blockchain-Based Smart Contract System for Healthcare Management," Electronics (Basel), vol. 9, no. 1, p. 94, Jan. 2020, doi: 10.3390/electronics9010094.

[11] E.-Y. Daraghmi, Y.-A. Daraghmi, and S.-M. Yuan, "MedChain: A Design of Blockchain-Based System for Medical Records Access and Permissions Management," IEEE Access, vol. 7, pp. 164595–164613, 2019, doi: 10.1109/ACCESS.2019.2952942.

[12] P. Zhang, J. White, D. C. Schmidt, and G. Lenz, "Design of Blockchain-Based Apps Using Familiar Software Patterns with a Healthcare Focus," in Proceedings of the 24th Conference on Pattern Languages of Programs, in PLoP '17. USA: The Hillside Group, 2017.

[13] T. Kumar, V. Ramani, I. Ahmad, A. Braeken, E. Harjula, and M. Ylianttila, "Blockchain Utilization in Healthcare: Key Requirements and Challenges," in 2018 IEEE 20th International Conference on e-Health Networking, Applications and Services (Healthcom), IEEE, Sep. 2018, pp. 1–7. doi: 10.1109/HealthCom.2018.8531136.

[14] A. A. Siyal, A. Z. Junejo, M. Zawish, K. Ahmed, A. Khalil, and G. Soursou, "Applications of Blockchain Technology in Medicine and Healthcare: Challenges and Future Perspectives," Cryptography, vol. 3, no. 1, p. 3, Jan. 2019, doi: 10.3390/cryptography3010003.

[15] D. Ichikawa, M. Kashiyama, and T. Ueno, "Tamper-Resistant Mobile Health Using Blockchain Technology," JMIR Mhealth Uhealth, vol. 5, no. 7, p. e111, Jul. 2017, doi: 10.2196/mhealth.7938.

[16] S. Rouhani, L. Butterworth, A. D. Simmons, D. G. Humphery, and R. Deters, "MediChainTM: A Secure Decentralized Medical Data Asset Management System," Jan. 2019, doi: 10.1109/Cybermatics_2018.2018.00258.

[17] S. S. Gill, "Quantum and blockchain based Serverless edge computing: A vision, model, new trends and future directions," Internet Technology Letters, Feb. 2021, doi: 10.1002/itl2.275.

[18] N. S. Bhati, M. Khari, V. García-Díaz, and E. Verdú, "A Review on Intrusion Detection Systems and Techniques," International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, vol. 28, no. Supp02, pp. 65–91, Dec. 2020, doi: 10.1142/S0218488520400140.

[19] A. A. Vazirani, O. O'Donoghue, D. Brindley, and E. Meinert, "Implementing Blockchains for Efficient Health Care: Systematic Review," J Med Internet Res, vol. 21, no. 2, p. e12439, Feb. 2019, doi: 10.2196/12439.

[20] A. Panwar, V. Bhatnagar, M. Khari, A. W. Salehi, and G. Gupta, "A Blockchain Framework to Secure Personal Health Record (PHR) in IBM Cloud-Based Data Lake," Comput Intell Neurosci, vol. 2022, pp. 1–19, Apr. 2022, doi: 10.1155/2022/3045107.

[21] P. Yuan, K. Zheng, X. Xiong, K. Zhang, and L. Lei, "Performance modeling and analysis of a Hyperledger-based system using GSPN," Comput Commun, vol. 153, pp. 117–124, Mar. 2020, doi: 10.1016/j.comcom.2020.01.073.

Enhancing Business Intelligence with Hybrid Transformers and Automated Annotation for Arabic Sentiment Analysis

Wael M.S. Yafooz
Computer Science Department, College of Computer Science and Engineering, Taibah University, Medina, 42353, Saudi Arabia

***Abstract*—Business is a key focus for many individuals, companies, countries and organisations. One effective way to enhance business performance is by analysing customer opinions through sentiment analysis. This technique offers valuable insights, known as business intelligence, which directly benefits business owners by informing their decisions and strategies. Substantial attention has been given to business intelligence through proposed machine learning approaches, deep learning models and approaches utilizing natural language processing methods. However, building a robust model to detect and identify users' opinion and automated text annotation, particularly for the Arabic language, still faces many challenges. Thus, this study aims to propose a hybrid transfer learning model that uses transformers to identify positive and negative user comments that are related to business. This model consists of three pretrained models, namely, AraBERT, ArabicBERT, and XLM-RoBERTa. In addition, this study proposes a hybrid automatic Arabic annotation method based on CAMelBERT, TextBlob and Farasa to automatically classify user comments. A novel dataset, which is collected from user-generated comments (i.e. reviews on mobile apps), is introduced. This dataset is annotated twice using the proposed method and human-based annotation. Then, several experiments are conducted to evaluate the performance of the proposed model and the proposed annotation method. Experiment results show that the proposed hybrid model outperforms the baseline models, and the proposed annotation method achieves high accuracy, which is close to human-based annotation.**

***Keywords*—Business intelligence; machine learning; sentiment analysis; transformers; BERT; Arabic annotation**

I. Introduction

Business intelligence can be enhanced through Sentiment Analysis (SA), which provides a nuanced understanding of customer opinions, market trends and brand reputation. SA allows businesses to gauge public sentiment accurately by analysing large amounts of text data from sources, such as social media, customer reviews and feedback forms. This insight helps companies identify consumer impressions, pain points and emerging trends, allowing them to tailor their products, services and marketing strategies [1, 2]. In this manner, businesses can stay competitive by responding quickly to changes in consumer behaviour and market conditions by understanding sentiment.

By integrating SA with business intelligence systems, companies can make more informed and strategic decisions based on gained insights into the emotional tone of customer feedback, which provides context to quantitative data, such as sales figures and retention rates. Social media platforms, such as Facebook, Twitter and Instagram, have billions of active users who regularly share their thoughts, opinions and experiences. By analysing social media posts, comments and reviews, businesses can identify emerging trends, monitor brand sentiment, gather feedback on products and services and identify early potential issues and address them proactively [3,4]. Similarly, mobile app reviews provide businesses with valuable feedback from users regarding their experiences with the app. These reviews often contain rich insights into user preferences, pain points and suggestions for improvement. By analysing app reviews, businesses can identify recurring issues, prioritise feature enhancements and enhance user satisfaction. Today's digital age has created a marketplace where consumer opinion is crucial to maintaining a positive brand image. SA makes it easier for businesses to keep an eye on public perception and respond quickly to changes in public opinion.

Therefore, scholarly efforts have been made to use SA in many domains, such as healthcare [5–7], education [8–10] marketing [11–13], business [14–16] and finance [17, 18] with binary classification or multi classification. The main two challenges that this task suffers from are dataset and building mode [19, 20]. Firstly, in the dataset preparation, the labelling process, which is known as the annotation process, is important because it is related to the model's performance in terms of accuracy [21]. This process is time consuming and requires considerable effort from annotators to classify data to relevant classes. Secondly, a robust model to distinguish between the good and bad comments or words is required. Researchers have attempted to address this issue by using Machine Learning (ML) classifiers, Deep Learning (DL) models and Natural Language Processing (NLP) methods for many languages, such as English, French and Chinese. However, the Arabic language and its dialects lack scholarly attention due to the Arabic language having morphological richness with 22 countries with different dialects. Some researchers introduced methods for Arabic annotations, such as AraSenCorpus [19], Sentialg [22], Arasenti-tweet [23] and ZAEBUC [24]. The majority of SA works in the literature review is based on manual annotation by Arabic native speakers [25–29] or by automatic annotation tools [22, 30–34]. Some of them use Google Translator to translate from Arabic to English, and then they apply automatic annotation tools [24]. In such manner, Google deals with modern standard Arabic, not Arabic dialects. Therefore, this still remains a challenge in the Arabic

language due to its complex morphology, dialectal variation, ambiguity and contextual understanding.

Therefore, the purpose of this study is to propose a robust model to detect the users' sentiment to improve business intelligence and to propose methods for automatic Arabic annotation. The proposed model is a hybrid transfer learning model designed to distinguish between the good and bad user comments; it consists of XLM-RoBERTa, AraBERT Ver2 and Arabic BERT models. These models are all based on Bidirectional Encoder Representations from Transformers (BERT) which are Transformers architectures. In addition, this hybrid model comprises two methods which have been used; the first being the voting mechanism, and the second being the fusion feature. In the automatic Arabic annotation method, three annotator tools, i.e. CAMeLBERT, TextBlob and Farasa, are utilised. This method is called the Artificial Intelligence Annotator (AIA). Additionally, a novel dataset that has been collected from mobile apps, which consists of 223,341 user-generated comments, is introduced to validate the performance of the proposed model and to determine how well the AIA method accurately assigns the user comments to the relevant classes. The same dataset is also annotated by three Arabic native speakers; this process is called Human-based Annotation (HA). Then, several NLP methods, such as data cleaning, preprocessing and data annotation, are applied on the dataset to validate the proposed hybrid transfer learning model and the AIA. Several experiments are conducted using ML classifiers, DL models and transformers. The results of these experiments were in terms of the most common metrics, which are, recall, precision, F1-score, Area Under Curve (AUC)–Receiver Operating Characteristic Curve (ROC) and accuracy. Experiment results show that the proposed model is based on the fusion features and voting mechanism which outperformed all the baselines in terms of accuracy, i.e. 97.24% and 98.11%, respectively. Additionally, the experiment results show the AIA method is closely tied to HA in Arabic corpora annotation.

The contributions of this study are multifaceted and can be summarised as follows:

- Proposed a hybrid model of transformers (transfer learning) to detect business SA based on mobile apps that are related to home delivery and taxi. This model consists of XLM-RoBERTa, AraBERT Ver2 and ArabicBERT. In this model, two methods are used: the voting mechanism and the fusion feature.
- Proposed a hybrid automatic Arabic annotation method for the Arabic corpora, which consists of the CAMelBERT, TextBlob and Farasa methods.
- Introduced a novel Arabic dataset which consists of 223,341 user-generated comments collected from reviews of mobile apps. It was reduced after annotation process to 63,313 user comments using AIA method and 54,988 using the HA method.
- Compared the model performance in terms of accuracy with various ML classifiers, DL models, transformers for AIA and HA methods.

The remainder of this paper is organized as follows. Section II provides a review of recent studies that focus on sentiment analysis and business intelligence. Section III explains the methods used to achieve the objectives of this study. The results and experiments are presented in Section IV. Discussion is given in Section V. Finally, the paper concludes in Section VI, summarizing the findings and implications of the study.

II. Related Studies

This section explores the studies that focus on sentiment analysis from a business intelligence perspective and also focuses on the Arabic annotation methods for the Arabic corpora.

In business intelligence and social media, Kurnia & Suharjito [1], developed a business intelligence dashboard to observe the performance of each topic or channel of news posted on social media apps such as Facebook and Twitter. The research tested different ML classifiers like Naive Bayes (NB), Support Vector Machine (SVM), and Decision Tree (DT) to categorize text from social media. The data used was posts from Facebook and Twitter. The research shows that SVM was the most accurate, reaching 78.99% accuracy. Similarly gathering a dataset from twitter, Khan [3]., conducted a case study on the official PlayStation account to illustrate their focus on sentiment analysis from a business intelligence standpoint, notably examining emotions and feelings expressed on Twitter. One thousand tweets from Twitter make up the dataset that was used. User emotions are compared between mentioning the official account and speaking generally as part of the analysis. When people contact their friends and followers instead of merely citing official accounts, the study finds that users are more open about their ideas and discontent. Also revolving on the idea of social media, Sánchez-Núñez et al. [35], provided an approach to model the business intelligence for identifying the users' abnormal emotion from their post through the social media marketing platform. To this end, to detect unusual manifestations in microblogs, the model uses the principles of multivariate Gaussian distribution and joint probability density. Data used in the study was obtained from micro-blogs, aggregated from 100 users over the period of 5 years with a sample size of 10,275. Thus, the accuracy rate that the presented model was achieving was approximately 83%. 87% for identification as an individual user or malicious bot. 84% for monthly identification of the subjects' abnormally endowed emotions.

In business intelligence and sentiment analysis, Swain & Cao [36], focused on utilizing sentiment analysis in order to extract the valuable insights from social media data related to supply chain management. The dataset which was used in the study includes and has over 600 randomly sampled companies from various industries, with data collected from forums, blogs, and microblogs such as Twitter. Tokenization, stemming, and feature selection utilizing filtering strategies such document frequency and mutual information are among the language processing approaches. With F-measures of 0.70 and 0.91 on the test set, respectively, the sentiment analysis approach uses a four-dimension classification algorithm and a positive-negative sentiment classification algorithm. Another

study about sentimental analysis, Aqarwal [37]., aimed to have a better understanding about customer sentiments and goals in order to improve overall business strategy and overall customer satisfaction by applying DL models, namely Recurrent Neural Networks (RNNs) and a Convolutional Neural Networks (CNNs). The study does not specifically disclose the dataset that was utilized for analysis. English was the language used for sentiment analysis. The model uses deep learning techniques, such CNNs and RNNs, to effectively capture the subtleties of customer feedback, resulting in high accuracy rates in sentiment classification. Similarly focusing on customer satisfaction, Prananda & Thalib [38], focused on utilizing sentiment analysis and predictive analytics for business intelligence purposes, specifically in the context of analyzing customer reviews for GO-JEK services. It used predictive analytics and sentiment analysis for business intelligence, particularly when examining customer reviews for GO-JEK services. The dataset is made up of 3,111 tweets from various countries that had keywords associated with GO-JEK that were gathered in January 2019. For sentiment analysis classification, the research uses computer learning techniques like neural networks, SVMs, NB, and DT. The DT method does perform the best, it achieved a score of 0.55 in precision, recall, and f1-score. Almost Identically, Capuano [39], proposed a Hierarchical Attention Networks based approach for sentiment analysis application in customer relationship management. The model was trained on a dataset of more than 30,000 items, 40% of which were collected from an Italian IT company and 60% were collected from public datasets to balance the class distribution. The experimental results show that the proposed approach attains high accuracy rates, with a macro-averaged F1-score of 0.89 for the Italian language and 0.79 for English. The incremental learning mechanism does not affect the overall system performance, improving the model's performance over time.

In the same way, Srinivasan et al. [40], worked on the sentimental analysis of the impact of COVID-19 on social life using ML classifiers. The data which is used for the analysis is gathered from Twitter which contains manual sentiment labels on tweets like "Extremely Positive", "Positive", "Neutral", "Negative" and "Extremely Negative", it also has 41,158 rows of data. The data is collected from Kaggle COVID-19 NLP Text Classification dataset. They worked on the BERT model for the sentimental analysis and check the sentiment of the tweets from various countries and for India. Further focusing on the idea of NLP methods, Sanchez-Nunez et al. [41], concentrated on the opinion mining, sentiment analysis and emotion understanding particularly in the context of advertising. The research draws data from the WoS database and embraces articles published between 2010 and 2019. The study uses NLP and SA methods to analyze latent trends in consumer perceptions of international brands and products. Similarly, Gołębiowska et al. [42], elaborated on cybersecurity aspects in business intelligence analytics via sentiment analysis and big data. It addresses the need for dealing with large volumes of information and the corresponding needs and challenges of latest security demands. The analyzed dataset contains a broad range of information and may originate from different websites, social, scientific, political as well as business topics. The sentiment analysis should help to investigate user knowledge on information technologies, industry 4.0 and the related dangers and risks.

Similarly utilizing big data, Sreesurya et al. [43], centered on employing big data sentiment analysis with Hypex, an improved Long Short-Term Memory (LSTM) technique for retrieving business intelligence from reviews and comments. The dataset used to train is the 5-core Amazon dataset, which comprises of about 18 million reviews. The language processing method utilized entails converting text into word vectors using the GloVe model that has been trained previously. The model presents a testing accuracy rate as an indication of how the model performs when it comes to predicting the sentiment of the reviews. Hypex is the proposed activation function which surpasses general activation function for providing better accuracy in terms of sentiment classification for business intelligence. Further expanding on the idea of using big data, Niu et al. [44]., presented the Optimized Data Management using Big Data Analytics (ODM-BDA) model as a solution of optimizing organizational decision making by adopting big data technology and cloud computing strategies. It is intended to facilitate the Increase in efficiency in processing data, optimization of profits, and upgrade methods of decision. In this particular study, the improvement facilitated by the proposed ODM-BDA framework is achieved through a simulation analysis using 10 to 100 iterations to compare the enhancement in accuracy and duration.

Alike, Saura et al. [45]., employed a Latent Dirichlet Allocation (LDA) and Sentiment Analysis with SVM algorithm and to analyze the User Generated Content through Twitter for arriving at the critical factors that make a startup successful. The studied data set entails 35,401 tweets that include the #Startups hashtag. The analysis is performed through Python LDA 1. 0. 5 software and MonkeyLearn's Sentiment Analysis algorithm. In our model, the accuracy rate was obtained more than 0. 797 for positive sentiment, and >0. 802 for neutrality and above 0 for positive sentiment analysis. Similarly employing an SVM algorithm, Al-Otaibi et al. [46]., discussed about a system designed to measure customer satisfaction using sentiment analysis on Twitter data. The dataset used consists of 5513 hand-classified tweets, with positive and negative sentiments selected for training. The system employs the SVM classifiers for sentiment classification, achieving an accuracy rate of 87% on a 4000-tweet testing dataset.

In Arabic annotation, Guellil et al. [47], developed "ArAutoSenti" which aims to annotate Arabic text automatically based off its understanding on the sentiment behind the Arabic text. "ArAutoSenti" achieves an F1-score of 88%. It is important to mention that this method was applied for the under-resourced Algerian dialect. While Guelil [22], proposed SentiALG in this study, which is a tool intended for automatic annotation for the Algerian dialect in sentiment analysis. The dataset for it consisted of 8,000 messages. Correspondingly, Jarrar et al. [48], proposed an Arabic corpus called "SALMA", which consists of around 34,000 tokens which are all sense-annotated. It is also worthy to mention that a smart web-based annotation tool was developed to support scoring multiple senses against a given word. In the same way,

Al-Laith et al. [19]., presented a way to annotate large prices of texts in Arabic Corpus called "AraSenCorpus". A neural network was used to train a set of models on a manually labeled dataset containing 15,000 tweets.

III. Methods and Materials

This section describes the methods that were used to carry out this study. There are six phases which were conducted namely; data collection, data cleaning, data annotation, data pre-processing, feature engineering, building models and model evaluation, all are as shown in Fig. 1.

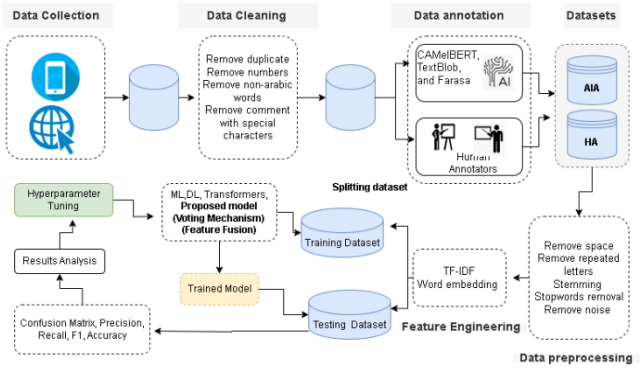


Fig. 1. Phases of the study.

A. Data Collection

In this phase, the data collected by user generated comments from the Play store of Mobile apps was through using the Python programming language. The main criteria are select the user comments and mobile apps are as follows: firstly, the period of collection for the user comments was between Jan 2022 to March 2024. Secondly, the mobile apps that are related to home delivery and transportation in total are eight Mobile apps. Lastly, the text for user comments is related to the Arabic language. All the user comments were downloaded into a separate CSV file for each application, then, all the files were combined together into one CSV file. At the end of this phase the total number of downloaded comments was approximately 223,341.

B. Data Cleaning

This phase had the task of preparing the dataset, in this phase several steps were used to remove duplicates of user comments, remove any user comments that are non- Arabic, remove the spaces from the files if there are any comments which are only empty, remove comments which consisted of special characters due to these types of comments being meaningless in the area of sentiment analysis.

C. Data Annotation

In the data annotation phase, there are two methods which have been utilized, namely; the first being the AIA tools which were used in the annotation process and the second being the HA method. In AIA, three tools have applied which are, TextBlob, Farasa, and CAMeLBERT. TextBlob is a python library built on Natural Language Toolkit (NLTK) and has been used for NLP tasks. Particularly "ar-textBlob" using the Standford API for Arabic techonizer. In this study it is used for sentiment analysis which indicates whether the user comments are positive or negative. Farasa at the Qatar Computing Research Institute has been developed for Arabic NLP. The CAMeLBERT is a pre-trained model based on the BERT architecture and specifically for Arabic NLP. In this type of annotation which is known as "ensemble techniques" is based on a voting mechanism which the decision is based on the majority. Therefore, if two classifiers indicated that the user comments belong to a specific class (either positive or negative) then the decision will be based on that. Thus, by the output of this method, the first dataset was constructed. Table I shows the description of the first dataset.

TABLE I. First Dataset Description (AIA)

Item(s)	No.comments	Min.Length	Max. Length
Positive	28,465	6	88
Negative	34,848	3	76
Total	63,313		

In the HA method, which has been conducted based on human annotation, three Arabic speakers helped in annotating the downloaded user comments. The user comments have been assigned to a class based on the decision of the majority, that means if two annotators agree about a specific decision on the user comments being assigned to the relevant class (either positive or negative) then the decision will be taken. Thus, the output of this method the second dataset has been constructed. Table II shows the description of the second dataset.

TABLE II. Dataset Description (HA)

Item(s)	No.comments	Min.Length	Max. Length
Positive	25,363	4	85
Negative	29,625	2	76
Total	54,988		

D. Feature Engineering

This phase is before feeding the models; each user comment is converted to numerical representation which is known as word representation. Such numerical representation is used as input to train and test the models. There are two types of word representation used in this study; the first being the Term-Frequency Inverse Document Frequency (TF-IDF) and word embedding's. The TF-IDF is represent how the words is important in the dataset, it's calculated as in mathematical Formula (1). While the second type of word embedding discovers the sematic relation between the words in user comments. Each word represented in vector dimension that contains values that measure the closeness of the word syntactic and semantic that happened in the training process.

$$TF - IDF(W, UC) = TF(W, UC) X\ IDF(W, UD) \quad (1)$$

Where, W is the word in the User Comment (UC). UD is represented in the dataset for all user comments. The TF is calculated based on the following mathematical Formula (2).

$$TF = \frac{F_{W,UC}}{N_{UC}} \quad (2)$$

Where $F_{W,UC}$, is the number of times the word (W) appears in the UC, and N represents the total number of W in UC.

While IDF is calculated as in the mathematical Formula (3).

$$IDF = \log \frac{N}{1+n_W} \quad (3)$$

where, n_w is the number of UC that W appears in it.

E. Building Models

This subsection demonstrates the models that were used to evaluate the proposed models' performance compared. These models consisted of ML classifiers, DL models and transformers. The most common classifiers in ML were selected, which are NB, Logistics Regression (LR), Random Forest (RF), K-Nearest Neighbors (KNN), DT and SVM [49]. While in DL experiments a LSTM, Bidirectional (BiLSTM), Gated Recurrent Unit (GRU), and CNN-LSTM were utilized, and in transform learning (transformers) [50], RoBERTa, MARBERT, distilbert, CAMeLBERT-DA, CAMeLBERT-Ca, AraELECTRA, ArabicBERT(Qarib), and AraBERTVer2 were utilized.

The proposed model consists of three transformers, which are built on the concept of transfer learning, as shown in Fig. 2. These models which were used in the proposed model are XLM-RoBERTa, AraBERT Ver2 and Arabic BERT. These models were trained on a large diverse Arabic dataset. These models are called pretrained models which can capture Arabic language patterns.

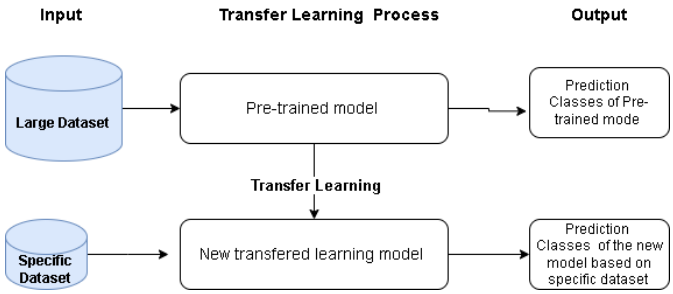


Fig. 2. Transfer learning.

Additionally, these models are built on the BERT architecture. In this manner, the performance of the model improves. These pretrained models, called 'fine tuning models', can be utilised on a specific dataset. Therefore, these three state-of-the-art models can be utilised to benefit each one to improve the model's performance through SA on Arabic business intelligence. XLM-RoBERTa is a multilingual pretrained model which can be trained on a huge opera, including the Arabic language. It can help in discovering the linguistics between Arabic user comments and SA. By contrast, AraBERT Ver2 is an updated version of AraBERT and is essentially trained on Arabic text from different sources. It can handle and deeply understand the relation of Arabic words in the language's unique morphological characteristics. Arabic BERT is trained on a large dataset from different sources, and it is different from the datasets of AraBERT ver2.

The proposed model comprises of two methods, namely, the voting mechanism and feature fusion. In the voting mechanism, the models predict the class of the input comments, and each model gives its prediction separately. Then, the ensemble methods, i.e. the voting mechanism, are applied. In this manner, the decision is based on majority of the output of the pretrained models, as shown in Fig. 3.

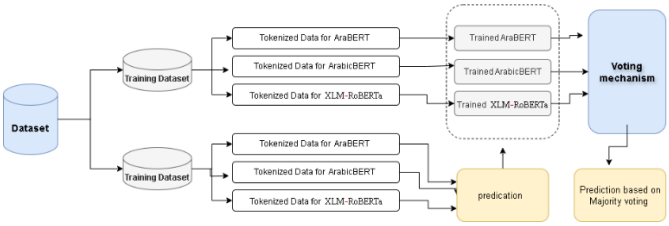


Fig. 3. Voting mechanism method.

The second method is feature fusion, which works by receiving the vectors from three pretrained models. Then, these vectors are combined to a large vector that is subsequently used to feed fully connected layers that are produced in the final prediction of the classes. Fig. 4 illustrates the feature fusion method.

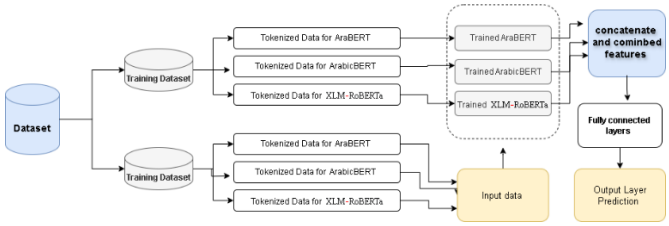


Fig. 4. Feature fusion method.

F. Model Performance Evaluation

In order to validate the proposed method of automatic Arabic annotation and also the proposed model of hybrid transformers models (transfer learning). The models' common measurement matrix has been used which is the precision, recall, F1, Accuracy and AUC-ROC.

Precision is the ratio of correctly predicted positive observations to the total predicted positives. Precision is calculated based on the correctly positive user comments (UCs) predicted to the total number of predicted positive UCs as presented in mathematical Formula (4). In the instance in which the precision is high, that means the model predicted the positive or negative UCs correctly. Recall is the ratio of the True positive UCs predicted to the actual positive UCs as displayed in the mathematical Formula (5). The F1-score or also called the F-measure is the homogenous between the precision and recall. While the accuracy is the total number of correct predicted positive UCs over the total number of UCs in the dataset as portrayed in the mathematical Formula (7).

$$\text{Precision} = \frac{True\ positive\ UCs}{True\ positive\ UCs + false\ positive\ UCs} \quad (4)$$

$$\text{Recall} = \frac{True\ positive\ UCs}{True\ positive\ UCs + false\ negative\ UCs} \quad (5)$$

$$\text{F1} = 2X\ \frac{PrecisionXRecall}{Precsiosn + Recall} \quad (6)$$

$$\text{Accuracy} = \frac{True\ positive\ UCs + True\ negative\ UCs}{Total\ number\ of\ UCs\ is\ dataset} \quad (7)$$

AUC-ROC is the graphical representation that shows the model performance in distinguishing between the positive user comments and the negative UCs. AUC-ROC is calculated based on the True Positive Rate (TPR) as presented in mathematical Formula (8) and False Positive Rate (FPR) as shown in mathematical Formula (9). The final value is between 0 and 1, the higher the values are, indicates how well the model has performed.

$$TPR = \frac{True\ postives\ UCs\ for\ both\ classes}{True\ Psotive + Flase\ Negative} \quad (8)$$

$$FPR = \frac{False\ \ postives\ UCs\ for\ both\ classes}{False\ \ psotive + True\ Negative} \quad (9)$$

IV. Results and Experiments

This section describes the settings of the experiments for the four types of experiments conducted: ML experiments, DL experiments, transformer experiments and the proposed model experiments. The experiment results are then explained.

A. Experimental Settings

All the experiments are conducted using Google Colab, utilising the GPU and other hardware-related matters. The programming language of choice is Python. In the ML experiment, the scikit-learn package is used to split the dataset and to import and use the ML classifiers. In the DL experiments, the TensorFlow framework is used for building DL models. The transformer package is imported to conduct the transformers experiments, utilising the hugging face platform to access the pretrained models. In all the experiments, the dataset is divided into two parts, 70% for training and 30% for testing the models. Hyperparameters for ML, DL, transformers and the proposed model is presented in Tables III, IV and V respectively.

TABLE III. ML Hyperparameters

Classifier	Values
DT	Criterion'gini'
KNN	n_neighbors 5
LR	penalty 'l2'
RF	n_estimators 100
SVM	C 1.0

TABLE IV. DL Hyperparameters

Parameter	Value
LSTM Units	64
Dropout	0.5
Batch Size	32
Optimizer	Adam
Activation function (Hidden Layers)	ReLU
Activation function (output)	Sigmod

TABLE V. Hyperparameters Transformers

Item(s)	Values
Batch Size	16
Number of Epochs	6
weight_decay	0.01
logging_steps	100
learning_rate	2e-5

B. Experimental Results

In this subsection, the results of the four experiments are analysed. In the ML experiments, several experiments are conducted using the AIA and HA methods. Table VI shows the results based on the most commonly used measurements, i.e. precision, recall, F1-score and accuracy.

TABLE VI. Comparison Between the Four Measurements Using the AIA Method

Classifiers	Precision	Recall	F1-Score	Accuracy
DT	81.53%	64.19%	65.64%	78.05%
KNN	73.45%	75.90%	74.28%	77.41%
LR	83.08%	80.27%	81.46%	85.21%
NB	74.91%	77.88%	75.84%	78.59%
RF	81.99%	80.76%	81.33%	84.77%
SVM	82.82%	79.77%	81.04%	84.93%

Table VI presents the performance metrics of various classifiers used for SA in the context of business intelligence, specifically in extracting insights from user comments on mobile apps. LR shows the highest accuracy of 85.21% and an F1-score of 81.46%, indicating its strong ability to classify sentiments from user comments accurately. RF and SVM also perform well, with accuracies of 84.77% and 84.93%, respectively, and F1-scores exceeding 81%. By contrast, the DT classifier has the lowest recall at 64.19% and F1-score at 65.64%, suggesting that it may not be as effective for this application compared with the other models. KNN and NB show moderate performance, with KNN achieving a recall of 75.90% and an F1-score of 74.28%, whereas NB shows a balanced performance with a precision of 74.91% and an F1-score of 75.84%. The confusion matrix is shown in Fig. 5, and the AUC–ROC is presented in Fig. 6. When the HA method is used, the best accuracy is recorded when using the LR classifier and low accuracy is achieved using KNN. Table VII shows the comparison between the four measurements using the HA method for the ML classifiers. The confusion matrix and AUC–ROC are presented in Fig. 7 and Fig. 8, respectively.

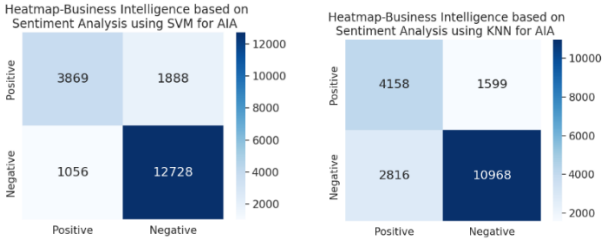


Fig. 5. Confusion matrix SVM and KNN using AIA method.

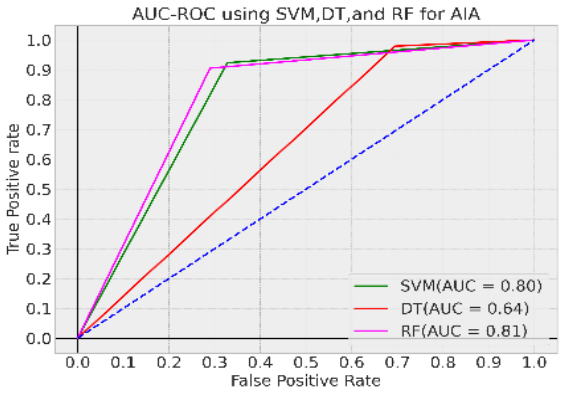


Fig. 6. AUC-ROC- AIA method.

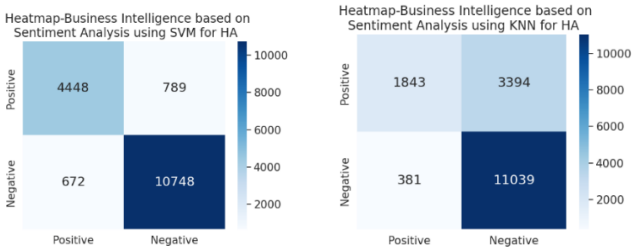


Fig. 7. Confusion matrix SVM and KNN using HA method.

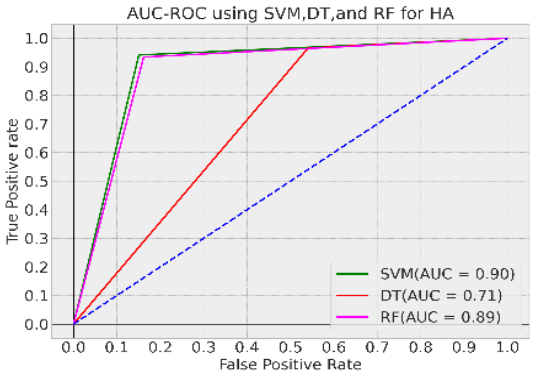


Fig. 8. AUC-ROC –HA method.

TABLE VII. THE COMPARISON BETWEEN THE FOUR MEASUREMENTS USING HA METHOD

Classifiers	Precision	Recall	F1-Score	Accuracy
DT	82.75%	71.25%	73.56%	80.64%
KNN	79.68%	65.93%	67.40%	77.34%
LR	89.96%	89.13%	89.53%	91.06%
NB	83.00%	86.13%	84.07%	85.54%
RF	88.99%	88.61%	88.80%	90.39%
SVM	90.02%	89.52%	89.76%	91.23%

Table VII presents the performance metrics of various classifiers used for SA in the context of business intelligence, focusing on user comments from mobile apps. LR and SVM stand out with the highest performance, achieving accuracy rates of 91.06% and 91.23%, respectively, and F1-scores of 89.53% and 89.76%. R also shows strong performance with an accuracy of 90.39% and an F1-score of 88.80%. NB demonstrates balanced performance with an accuracy of 85.54% and an F1-score of 84.07%. By contrast, DT and KNN perform less effectively, with DT achieving an accuracy of 80.64% and an F1-score of 73.56%, whilst KNN has the lowest accuracy at 77.34% and an F1-score of 67.40%. Overall, LR, SVM, and RF are the most effective classifiers for extracting sentiment-based insights from user comments, highlighting their suitability for enhancing business intelligence through SA.

In the DL experiments, specifically during the utilisation of the AIA method, GRU demonstrates the highest performance with an accuracy of 90.01% and an F1-score of 87.93%, indicating its effectiveness in accurately classifying sentiments from user comments. The accuracy of training and validation is presented in Fig. 9. fNN-LSTM, whilst still effective, shows a slightly lower performance with an accuracy of 89.37% and an F1-score of 87.34%.

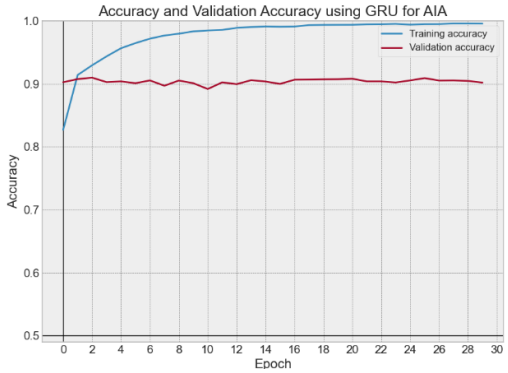


Fig. 9. Accuracy of training and validation using GRU for AIA.

In the HA method, LSTM demonstrates the highest performance with an accuracy of 95.54% and an F1-score of 94.85%, making it the most effective in accurately classifying sentiments from user comments. The accuracy of training and validation is shown in Fig. 10. GRU follows closely with an accuracy of 95.38% and an F1-score of 94.64%, indicating its strong potential as well. CNN-LSTM, whilst showing slightly lower recall, maintains high precision and F1-score with an accuracy of 95.09%. BiLSTM, although slightly behind the others, still performs robustly with an accuracy of 94.94% and consistent precision, recall and F1-score of 94.14%. Fig. 11 shows the precision, recall, F1-score and accuracy using the AIA and HA methods.

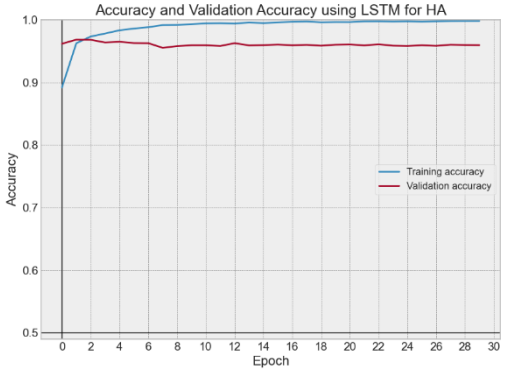


Fig. 10. Accuracy of training and validation using LSTM for HA.

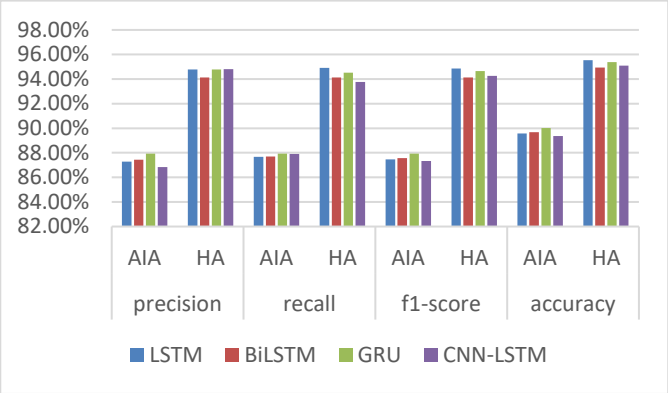


Fig. 11. Model performance between the AIA and HA methods.

In the pretrained models (transformers) and the proposed models, the experiments were conducted using the AIA and HA methods. The experimental results of the AIA and HA methods for the pretrained (transformers) and the proposed models is presented in Table VIII. In the proposed model, two methods, namely, voting mechanism and feature fusion, are used.

Table VIII summarises the performance of the pretrained models (transformers) used for the SA of user comments and reviews on mobile apps for enhancing business intelligence. Each model was evaluated across two approaches: AIA and HA. Amongst the models, ArabBERTVEr2 achieved an accuracy of 92.48% in AIA and 97.06% in AIA. The proposed Model 1 (feature fusion) demonstrated the highest accuracy with 93.96% AIA and 98.11% HA. These accuracies reflect how effectively each model can classify sentiments expressed in user feedback, providing valuable insights for improving mobile app performance and user satisfaction in business contexts. Additionally, the proposed Model 2 (voting mechanism) achieved an accuracy of 92.65% and 97.24%. The confusion matrix and AUC–ROC for the proposed model using the feature fusion method is presented in Fig. 12 and Fig. 13, respectively. While Fig. 14 and Fig. 15 shows confusion matrix and AUC-ROC of the proposed model using the voting mechanism respectively.

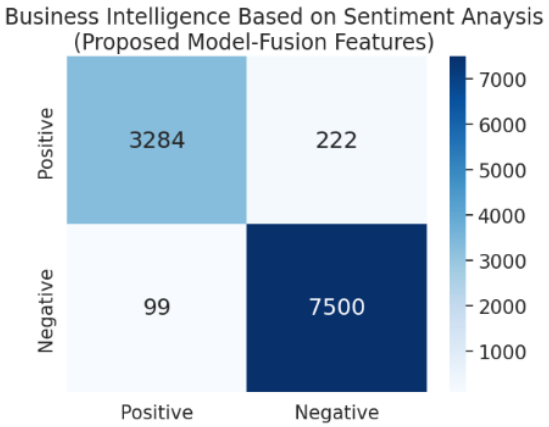


Fig. 12. Confusion matrix of the proposed model using feature fusion.

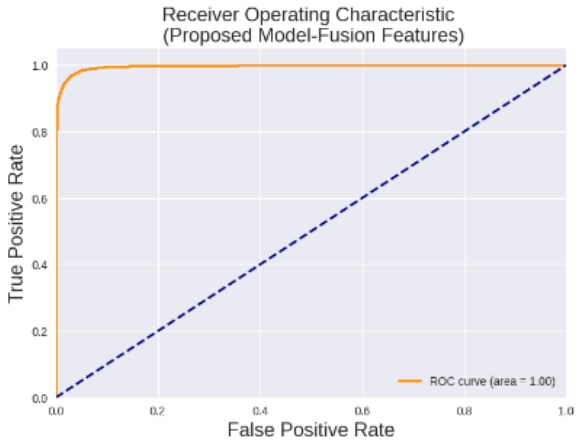


Fig. 13. AUC–ROC of the proposed model using feature fusion.

TABLE VIII. COMPARISON BETWEEN THE PRECISION, RECALL, F1-SCORE AND ACCURACY FOR TRANSFORMERS AND THE PROPOSED MODEL

Model(s)	Precision		Recall		F1-score		Accuracy	
	AIA	HA	AIA	HA	AIA	HA	AIA	HA
ArabBERTVEr2	93.43%	97.50%	96.13%	98.22%	94.76%	97.86%	92.48%	97.06%
ArabicBERT-Qarib	90.37%	97.01%	95.46%	98.26%	92.84%	97.63%	89.58%	96.74%
AraELECTRA	92.23%	97.34%	96.08%	98.34%	94.12%	97.84%	91.50%	97.03%
CAMeLBERT-DA	94.40%	96.86%	94.99%	97.47%	94.69%	97.17%	92.46%	96.11%
CAMeLBERT-Ca	92.57%	97.05%	95.59%	97.78%	94.06%	97.41%	91.45%	96.44%
distilbert	93.75%	97.57%	96.05%	98.22%	94.89%	97.89%	92.68%	97.11%
MARBERT	94.10%	97.27%	95.39%	97.59%	94.74%	97.43%	92.51%	96.48%
MERTICRobBERt	90.78%	97.33%	97.09%	98.20%	93.83%	97.76%	90.96%	96.92%
Proposed Mode 1	93.94%	98.22%	92.78%	97.94%	93.36%	98.07%	**93.96%**	**98.11%**
Proposed Mode 2	92.24%	97.62%	93.12%	96.95%	92.66%	97.28%	**92.65%**	**97.24%**

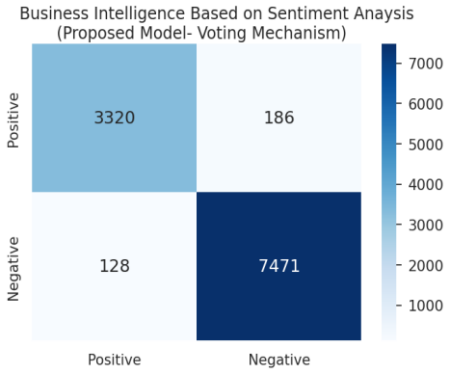


Fig. 14. Confusion matrix of the proposed model using the voting mechanism.

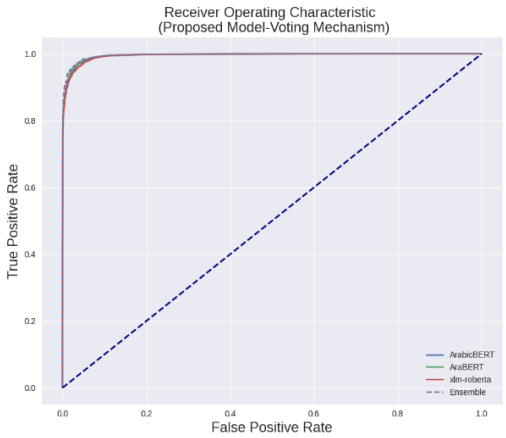


Fig. 15. AUC–ROC of the proposed model using the voting mechanism.

V. Discussion

This study has two findings related to the proposed hybrid transform learning model and the proposed automatic Arabic annotation methods.

Firstly, the experiment results reveal that the performance of the proposed model outperforms the aforementioned experiments, i.e. DL, ML, and transformers experiments. To go into depth of the precise method in which the proposed model achieved the highest results in, is the feature fusion which attained an accuracy score of 98.11%, which proves that it is better equipped for identifying the sentiment behind user comments in the Arabic language on whether the sentiment behind the written comment is positive or negative than the voting mechanism which reached an accuracy if 97.24% respectively. This study demonstrates that utilising transfer learning (pretrained models/transformers) enhances the model's performance in terms of accuracy in identifying and distinguishing positive and negative sentiments in Arabic user comments. The reason is that the pretrained models have been trained on large datasets that help in exploring and discovering the semantic and syntactic relationships between Arabic words.

Secondly, in the proposed Arabic automatic method, AIA saves time and effort in annotating Arabic user comments because it is completely automated as opposed to the HA method, which requires much more effort in annotating Arabic user comments and at least three native Arabic speakers to perform the voting mechanism. Nevertheless, the HA method achieves a somewhat higher accuracy than the AIA method, whilst the difference between the two methods in performance is approximately 5%. This result is observed in all the conducted experiments.

VI. Conclusion

Transformers have been utilised to improve the capabilities in handling natural language processing. Therefore, this study proposes a robust hybrid transfer learning model to enhance business intelligence by accurately detecting users' sentiments. The model combines XLM-RoBERTa, AraBERT Ver2 and Arabic BERT, and the model additionally utilises a voting mechanism and feature fusion to improve the models' performance. Additionally, the study introduces AIA, which integrates CAMeLBERT, TextBlob and Farasa. Furthermore, a novel dataset of user-generated comments from mobile apps is introduced. Results demonstrate that the proposed model leveraging feature fusion and the voting mechanism outperforms all baselines with an accuracy of 97.24% and 98.11%, respectively. Furthermore, the AIA method is closely matched with the HA in the Arabic corpora annotation, confirming its reliability and accuracy. In the future work, enhanced Arabic annotation methods are to be through utilising transformers and to build lexical dictionary, expanding the dataset to include more diverse sources and languages that could help generalise the model's applicability across various domains. Apply the automatic annotation to Arabic dialects across the 22 Arabic-speaking countries, each Arabic speaking country with its own unique variations. This approach can significantly enhance the classification process.

References

[1] Kurnia, P. F. (2018). "Business intelligence model to analyze social media information". Procedia Computer Science, 135, 5-14.

[2] Mehta, P., & Pandya, S. (2020). "A review on sentiment analysis methodologies, practices and applications". International Journal of Sci entific and Technology Research, 9(2), 601-609.

[3] Khan, S. (2022, February). "Business Intelligence Aspect for Emotions and Sentiments Analysis". In 2022 First International Conference on Electrical, Electronics, Information and Communication Technologies (ICEEICT) (pp. 1-5). IEEE.

[4] Wankhade, M., Rao, A. C. S., & Kulkarni, C. (2022). "A survey on sentiment analysis methods, applications, and challenges". Artificial Intelligence Review, 55(7), 5731-5780.

[5] Gohil, S., Vuik, S., & Darzi, A. (2018). "Sentiment analysis of health care tweets: review of the methods used". JMIR public health and surveillance, 4(2), e5789.

[6] Khan, M. T., & Khalid, S. (2016). "Sentiment analysis for health care. In Big data: concepts, methodologies, tools, and applications", (pp. 676-689). IGI Global.

[7] Aattouchi, I., Elmendili, S., & Elmendili, F. (2021). "Sentiment Analysis of Health Care". In E3S Web of Conferences (Vol. 319, p. 01064). EDP Sciences.

[8] Hajrizi, R., & Nuçi, K. P. (2020). "Aspect-based sentiment analysis in education domain". arXiv preprint arXiv:2010.01429.

[9] Shanthi, I. (2022). "Role of educational data mining in student learning processes with sentiment analysis: A survey". In Research Anthology on Interventions in Student Behavior and Misconduct (pp. 412-427). IGI Global.

[10] Alhujaili, R. F., & Yafooz, W. M. (2022, May). "Sentiment analysis for youtube educational videos using machine and deep learning

approaches". In 2022 IEEE 2nd international conference on electronic technology, communication and information (ICETCI) (pp. 238-244). IEEE.

[11] Lin, H. C. K., Wang, T. H., Lin, G. C., Cheng, S. C., Chen, H. R., & Huang, Y. M. (2020). "Applying sentiment analysis to automatically classify consumer comments concerning marketing 4Cs aspects". Applied Soft Computing, 97, 106755.

[12] Reyes-Menendez, A., Saura, J. R., & Filipe, F. (2020). "Marketing challenges in the# MeToo era: Gaining business insights using an exploratory sentiment analysis". Heliyon, 6(3).

[13] Mehraliyev, F., Chan, I. C. C., & Kirilenko, A. P. (2022). "Sentiment analysis in hospitality and tourism: a thematic and methodological review". International Journal of Contemporary Hospitality Management, 34(1), 46-77.X.4

[14] Ahmed, A. A. A., Agarwal, S., Kurniawan, I. G. A., Anantadjaya, S. P., & Krishnan, C. (2022). "Business boosting through sentiment analysis using Artificial Intelligence approach". International Journal of System Assurance Engineering and Management, 13(Suppl 1), 699-709.

[15] Sudirjo, F., Diantoro, K., Al-Gasawneh, J. A., Azzaakiyyah, H. K., & Ausat, A. M. A. (2023). "Application of ChatGPT in Improving Customer Sentiment Analysis for Businesses". Jurnal Teknologi Dan Sistem Informasi Bisnis, 5(3), 283-288.

[16] Yin, J. Y. B., Saad, N. H. M., & Yaacob, Z. (2022). "Exploring Sentiment Analysis on E-Commerce Business: Lazada and Shopee". Tem journal, 11(4), 1508-1519.

[17] Mishev, K., Gjorgjevikj, A., Vodenska, I., Chitkushev, L. T., & Trajanov, D. (2020). "Evaluation of sentiment analysis in finance: from lexicons to transformers". IEEE access, 8, 131662-131682.

[18] Renault, T. (2020). "Sentiment analysis and machine learning in finance: a comparison of methods and models on one million messages". Digital Finance, 2(1), 1-13.

[19] Al-Laith, A., Shahbaz, M., Alaskar, H. F., & Rehmat, A. (2021). "Arasencorpus: A semi-supervised approach for sentiment annotation of a large arabic text corpus". Applied Sciences, 11(5), 2434.

[20] Almuzaini, H. A., & Azmi, A. M. (2022). "An unsupervised annotation of Arabic texts using multi-label topic modeling and genetic algorithm". Expert Systems with Applications, 203.

[21] Almuqren, L., Alzammam, A., Alotaibi, S., Cristea, A., & Alhumoud, S. (2017). "A review on corpus annotation for Arabic sentiment analysis. In Social Computing and Social Media. Applications and Analytics" 9th International Conference, SCSM 2017, Held as Part of HCI International 2017, Vancouver, BC, Canada, July 9-14, 2017, Proceedings, Part II 9 (pp. 215-225). Springer International Publishing. 117384.

[22] Guellil, I., Adeel, A., Azouaou, F., & Hussain, A. (2018). "Sentialg: Automated corpus annotation for algerian sentiment analysis. In Advances in Brain Inspired Cognitive Systems", 9th International Conference, BICS 2018, Xi'an, China, July 7-8, 2018, Proceedings 9 (pp. 557-567). Springer International Publishing.

[23] Al-Laith, Ali, Muhammad Shahbaz, Hind F. Alaskar, and Asim Rehmat. "Arasencorpus: A semi-supervised approach for sentiment annotation of a large arabic text corpus." Applied Sciences 11, no. 5 (2021): 2434.

[24] Habash, N., & Palfreyman, D. (2022, June). "ZAEBUC: An annotated Arabic-English bilingual writer corpus". In Proceedings of the Thirteenth Language Resources and Evaluation Conference (pp. 79-88).

[25] Alahmary, R. M., Al-Dossari, H. Z., & Emam, A. Z. (2019, January). "Sentiment analysis of Saudi dialect using deep learning techniques", In 2019 International Conference on Electronics, Information, and Communication (ICEIC) (pp. 1-6). IEEE.

[26] Rahab, H., Zitouni, A., & Djoudi, M. (2021). "SANA: Sentiment analysis on newspapers comments in Algeria". Journal of King Saud University-Computer and Information Sciences, 33(7), 899-907.

[27] Al-Thubaity, A., Alharbi, M., Alqahtani, S., & Aljandal, A. (2018, April). "A Saudi dialect Twitter Corpus for sentiment and emotion analysis". In 2018 21st Saudi computer society national computer conference (NCC) (pp. 1-6). IEEE.

[28] Oussous, A., Benjelloun, F. Z., Lahcen, A. A., & Belfkih, S. (2020). "ASA: A framework for Arabic sentiment analysis". Journal of Information Science, 46(4), 544-559.

[29] Khalifa, S., Habash, N., Eryani, F., Obeid, O., Abdulrahim, D., & Al Kaabi, M. (2018, May). "A morphologically annotated corpus of Emirati Arabic", In Proceedings of the eleventh international conference on language resources and evaluation (LREC 2018).

[30] Elnagar, A., & Einea, O. (2016, November). "BRAD 1.0: Book reviews in Arabic dataset". In 2016 IEEE/ACS 13th International Conference of Computer Systems and Applications (AICCSA) (pp. 1-8). IEEE.

[31] Elnagar, A., Lulu, L., & Einea, O. (2018). "An annotated huge dataset for standard and colloquial arabic reviews for subjective sentiment analysis". Procedia computer science, 142, 182-189.

[32] Gamal, D., Alfonse, M., El-Horbaty, E. S. M., & Salem, A. B. M. (2019). "Twitter benchmark dataset for Arabic sentiment analysis". Int J Mod Educ Comput Sci, 11(1), 33.

[33] Abdellaoui, H., & Zrigui, M. (2018). "Using tweets and emojis to build tead: an Arabic dataset for sentiment analysis". Computación y Sistemas, 22(3), 777-786.

[34] Abo, M. E. M., Shah, N. A. K., Balakrishnan, V., Kamal, M., Abdelaziz, A., & Haruna, K. (2019, April). "Ssa-sda: subjectivity and sentiment analysis of sudanese dialect Arabic". In 2019 International Conference on Computer and Information Sciences (ICCIS) (pp. 1-5). IEEE.

[35] Sánchez-Núñez, P., Cobo, M. J., De Las Heras-Pedrosa, C., Peláez, J. I., & Herrera-Viedma, E. (2020). "Opinion mining, sentiment analysis and emotion understanding in advertising: a bibliometric analysis". IEEE Access, 8, 134563-134576.

[36] Swain, A. K., & Cao, R. Q. (2019). "Using sentiment analysis to improve supply chain intelligence". Information Systems Frontiers, 21, 469-484.

[37] Agarwal, S. (2022). "Deep learning-based sentiment analysis: Establishing customer dimension as the lifeblood of business management". Global Business Review, 23(1), 119-136.

[38] Prananda, A. R., & Thalib, I. (2020). "Sentiment analysis for customer review: Case study of GO-JEK expansion". Journal of Information Systems Engineering and Business Intelligence, 6(1), 1.

[39] Capuano, N., Greco, L., Ritrovato, P., & Vento, M. (2021). "Sentiment analysis for customer relationship management: an incremental learning approach". Applied intelligence, 51, 3339-3352.

[40] Srinivasan, S. M., Shah, P., & Surendra, S. S. (2021). "An approach to enhance business intelligence and operations by sentimental analysis". Journal of System and Management Sciences, 11(3), 27-40.

[41] Sánchez-Núñez, P., Cobo, M. J., De Las Heras-Pedrosa, C., Peláez, J. I., & Herrera-Viedma, E. (2020). "Opinion mining, sentiment analysis and emotion understanding in advertising: a bibliometric analysis". IEEE Access, 8, 134563-134576.

[42] Gołębiowska, A., Jakubczak, W., Prokopowicz, D., & Jakubczak, R. (2021). "Cybersecurity of business intelligence analytics based on the processing of large sets of information with the use of sentiment analysis and Big Data". European Research Studies Journal, 24(4).

[43] Sreesurya, I., Rathi, H., Jain, P., & Jain, T. K. (2020). "Hypex: A tool for extracting business intelligence from sentiment analysis using enhanced LSTM". Multimedia Tools and Applications, 79, 35641-35663.

[44] Niu, Y., Ying, L., Yang, J., Bao, M., & Sivaparthipan, C. B. (2021). "Organizational business intelligence and decision making using big data analytics". Information Processing & Management, 58(6), 102725.

[45] Saura, J.R.; Palos-Sanchez, P.; Grilo, A. Detecting Indicators for Startup Business Success: Sentiment Analysis Using Text Data Mining. Sustainability 2019, 11, 917.

[46] Al-Otaibi, S., Alnassar, A., Alshahrani, A., Al-Mubarak, A., Albugami, S., Almutiri, N., & Albugami, A. (2018). "Customer satisfaction measurement using sentiment analysis". International Journal of Advanced Computer Science and Applications, 9(2).

[47] Guellil, I., Azouaou, F., & Chiclana, F. (2020)." ArAutoSenti: automatic annotation and new tendencies for sentiment classification of Arabic messages". Social Network Analysis and Mining, 10, 1-20.

[48] Jarrar, M., Malaysha, S., Hammouda, T., & Khalilia, M. (2023). "Salma: Arabic sense-annotated corpus and wsd benchmarks". arXiv preprint arXiv:2310.19029.

[49] Alhejaili, R., Alhazmi, E. S., Alsaeedi, A., & Yafooz, W. M. (2021, September). "Sentiment analysis of the COVID-19 vaccine for Arabic tweets using machine learning". In 2021 9th International conference on reliability, infocom technologies and optimization (Trends and Future Directions)(ICRITO) (pp. 1-5). IEEE.

[50] Yafooz, W. M., Al-Dhaqm, A., & Alsaeedi, A. (2023). "Detecting kids cyberbullying using transfer learning approach: Transformer fine-tuning models". In Kids Cybersecurity Using Computational Intelligence Techniques (pp. 255-267). Cham: Springer International Publishing.

A Method by Utilizing Deep Learning to Identify Malware Within Numerous Industrial Sensors on IoTs

Ronghua MA
Zhengzhou Railway Vocational and Technical College, Teacher Work Department of the Party Committee, Zhengzhou 450052, China

***Abstract*—The industrial sensors of IoT is an emerging model, which combines Internet and the industrial physical smart objects. These objects belong to the broad domains like the smart homes, the smart cities, the processes of the industrial and the military, the agriculture and the business. Due to the substantial advancement in Industrial Internet of Things (IIoT) technologies, numerous IIoT applications have been developed over the past ten years. Recently, there have been multiple reports of malware-based cyber-attacks targeting IIoT systems. Consequently, this research focuses on creating an effective Artificial Intelligence (AI)-powered system for detecting zero-day malware in IIoT environments. In the current article, a combined framework for the detection of the malware basis on the deep learning (DL) is proposed, that uses the dual-density discrete wavelet transform for the extraction of the feature and a combination from the convolutional neural network (CNN) and the long-term short-term memory (LSTM). The method is utilized for malware detection and classification. It has been assessed using the Malimg dataset and the Microsoft BIG 2015 dataset. The results demonstrate that our proposed model can classify malware with remarkable accuracy, surpassing similar methods. When tested on the Microsoft BIG 2015 and Malimg datasets, the accuracy achieved is 95.36% and 98.12%, respectively.**

***Keywords*—Malware; malware detection; industrial sensors; Internet of Things (IoTs); Deep Learning (DL)**

I. INTRODUCTION

The advancement of various technologies such as the sensors, the wireless communication, the embedded computing, the automatic tracking, the widespread access to the Internet and the dispensed services increases the possible of the accretion of the smart sensors in our daily lives via Internet. The convergence of Internet and the smart sensors, which can connect with together, describes IoTs. This novel example has been detected as one from the foremost significant factors on the industries of the data and the communication technology in the coming years [1].

The goal of the IoTs technology is to enable the objects for the connection at any time and any place with anything and anyone, who uses any path or any network as optimally. IoTs is a new evolution from Internet. IoTs is the new technology that pays attention to the pervasive presence of the environment and deals with the diversity of the smart objects with the wireless connections and the wired connections for the communication with together. These objects work together, to create the new applications or the new services and to achieve the common goals. In fact, they are the development challenges for the creation of a smart big world. A world that is the real, the digital and the virtual and is converging towards the formation of the smart environments. This world creates the smarter environments of the energy, the transportation, the cities health and many others [2].

However, the integration of the smart objects in the real world by Internet can bring the threats of the security in the several of our daily behaviors [3]. According to the wide standards of the communication, the limited power of the computing and the great number of the connected sensors, the common actions of the security against the threats cannot work effectively on IoTs. Therefore, the development of the specific solutions of the security for IoTs is necessary, to enable the organizations users, to detect total weaknesses of a network [4]. Several ongoing projects for evolution of the security in IoTs include the methods that provide the data confidentiality, the authentication of the control of the access on IoTs, the privacy, the trust among the users and the implementation of the security policies. [5]. Nevertheless, even with the methods, IoTs are assailable to the several attacks. The attacks which are done, to interrupt and to disrupt the networks. For this reason, the required method of the defense is the creation of the models for the detection of the attackers. The development of the web-based technologies and the cloud computing will mark the future revolution in the digital technologies. Also, it will lead to the increased health, the productivity, the convenience and a wide range of the useful information for the individuals and the organizations. On the other hand, there will be challenges in the field of the personal privacy, the complexity of the intrusion technology and the creation of a digital gap [6].

The security establishment is perhaps the biggest challenge in the IoTs network. The security in the current Internet is also considered as a big challenge, but in the Internet of Things, this issue takes on the greater dimensions. One of the reasons for this issue is the distribution of the network and the more entry points into the system. Also, the objects that are supposed to be connected to the Internet, usually have a simpler architecture than the computers, and this implementation makes the security tools as the difficult. The IoTs technology is much closer to the real life than the current Internet; In fact, the intrusion into this network will be equivalent to the intrusion into the daily life of the users [4]. Due to the security problems on the real world and in the technology of IoTs, and according to the problems of the intrusion into the networks, it is very necessary to present the optimal method, in order to discover the intrusion and to keep the security on the networks [7].

*Corresponding Author.

Following an extensive review of the literature, it has been noted that current methods face several limitations and security challenges, such as low accuracy, insufficient large datasets, limited scalability, and high prediction times for detecting zero-day malware or unknown malicious activities. Therefore, this work proposes an efficient zero-day malware detection framework utilizing a hybrid deep learning model for IIoT systems. The main contributions of this paper are as follows: i) It introduces a novel AI-powered zero-day malware detection system for IIoT, utilizing an image visualization technique by combining CNN and LSTM models. ii) D3WT is employed for deep feature extraction, breaking down malware images into approximate and detailed coefficients. iii) The proposed hybrid model is tested on three major cross-platform malware datasets and compared with advanced models. The method is applied to detect and to classify the malware. The background of research is provided in Section II. Our method is described in Section III. Section IV shows the experiments and the evaluation of the results. Section V also presents the conclusions and the effective suggestions by using the obtained findings.

II. Related Works

In the recent decade, the many models based on the theory of the game on scope of the security in the networks have been done, to model the analysis and to optimize the efficiency of IDSs in the related technology to IoTs, such as the mobile contingency networks [8, 9], WSNs [10], the cloud computing [11] and the physical cyber networks [12]. The research in [11] has presented the various intrusions, that affect the availability of the privacy and the integrity in the cloud computing. They have distributed the methods of the used IDSs in the cloud into three categories: the host-based, the network-based and the hypervisor-based. They are also reviewed the advantages and the disadvantages of every protocol and are recognized the problems, to create the cloud computing as a trustworthy architecture for the providing of IoTs. The research in [9] shows that a malware detection model is capable to handle and to control the several protocols of the communication by combining the rules of the signature and the procedures for the detection of the anomaly. The research in [10] has done a wide review on IDSs in WSNs and has provided a comparative evaluation among IDSs for WSNs, according to the architecture of the network and the method of the detection.

The research in study [13] presents a comprehensive analysis of the security from the several protocols of Internet. Specifically, the authors discuss about the security topics in IEEE 802.15.4 against 6LOWPAN, the Routing Protocol of IPv6 for RPL, the protocols of DTLS and CoAP. The research in [8] has investigated IDSs for the mobile contingency networks, by relying on the detection algorithm. A categorization basis on the tree for IDSs has been introduced according to the character of the used method for the processing in detection model. The research in [14] presents an IDS for LOWPAN-RPL6 that is capable to recognize the Sinkhole attack, the Sybile attack and the Selective attack, by using a hybrid approach that combines the different parameters. The research in study [15] has presented an IDS basis on the features with supervisor, by using forward neural network. In this paper, the feature selection is done on the ISCX-IDS 2012 dataset and the Android CIC dataset. In order to do the feature selection phase, SVM with the incremental learning has been used, which with the ranking of 43 features in the dataset, 20 features with the highest rank have been selected. Then, by using a neural network, the final detection is made with the accuracy equal to 94% and 98.7%.

The research in study [4] has presented an optimal platform, to show the possible application of the practical in the malware propagation suppression for perseverance of the privacy in the smart objects on IoTs, via an IDS by the game theory calculation of Bayesian. The research in study [16] has examined the security of IoTs, the challenges, the solutions and the threats. After checking and evaluating the possible threats and after specifying the security actions in scope of IoTs, they have done the risk analysis of the quantitative and the qualitative that examines the threats of the security on every layer. The research in [7] has investigated a new plan, by using a combination from the classical encryption and the quantum encryption, to improve the security of the Internet network. A research title which is related to the anomaly-based intrusion detection systems [3], by evaluating solutions and the researches and by using role of DL in IDS, discusses the efficiency of the proposed methods, and also, by identifying the challenge from the past researches, it recommends the deep learning-based guidelines.

The research [17], in an article, in addition to the presentation of a model based on the combination from the artificial neural networks for the intrusion detection, it provides an algorithm for extraction of the optimal features on Cup KDD, which is the standard dataset for the testing of the intrusion detection methods in the computer networks. The researches efforts in the field of IDSs for IoTs have begun and speeded up. By taking the research background, it is important to state which the presented approaches have not deeply checked the abilities and the laxities of every detection model and each placement strategy. The many authors have relied on a few kinds of the attacks. Finally, the very easy validation schemes have presented the foundation for the reproduction of the other proposed approaches.

III. The Presented Approach

Here, the details of our proposed approach are provided. This approach is described to detect the zero-day malwares along with its family by the greater accuracy for the industrial sensors of IoT. The presented method is disturbed in $three$ main steps: the first includes the data preprocessing and the image resizing, the second includes the feature extraction by using D3WT and the third includes a hybrid model of LSTM-CNN for the automatic detection of the malwares. Fig. 1 displays the framework of our presented method that is created by integrating LSTM-CNN and D3WT. The coefficients of the approximation and the detail are exploited by D3WT. The exploited features are combined as input of LSTM-CNN, to create the fused images. The proposed approach is evaluated by Microsoft BIG 2015 and Malimg, which contain the different types from the malwares. The details of this approach are provided on the below subsections.

A. Data Pre-Processing

The data preprocessing is a necessary part in every method basis on AI, in order to increase its efficiency. In the current

article, the extraction of the feature has been done from the raw dataset of the malware. Then, the obtained features (like the opcodes, the strings and the bytecodes) are converted to the digits of the binary. Next, a collection from 8 bits are converted to the grayscale image. The converted images have the various sizes in terms of the height, nevertheless, the width of the images is the constant. The conversion of this grayscale images by the bytecodes is provided by using the Python tools. In the next step, the image preprocessing (like the image resizing to a specific size equal to 224×224) is performed on this approach. Then, D3WT is used as a method for the extraction of the feature, to extract the coefficients of the approximation and the detail.

B. Extraction of Features

On our approach, D3WT is used to analyze the inputs by using the banks of the filter. It follows an iterative method. This approach includes 2 wavelets (the high pass) and a scale function (the low pass). A wavelet is the offset by another wavelet. The theoretical flow diagram for 2 filter banks of D3WT is displayed on Fig. 2. D3WT is basis on 3-channel theory for the bank of the filter of the complete reconstruction. A matrix with $three$ columns is applied for the scaling and the function of the wavelet. The filter of the scaling $\theta(\alpha)$ is placed on first column and 2 high-pass filters, which are denoted by $\varphi_1(\alpha)$ and $\varphi_2(\alpha)$, are placed on second column and the third column. The function of the scaling is denoted by $T_0(-v)$ and 2 high-pass filters are represented $T_1(-v)$ and $T_2(-v)$. The input $J(v)$ passes via the model of the filter, and the analysis operation is done by the bank of the filter for the analysis, that creates 3 sub-bands. The sub-bands are down-sampled by 2. The output of this filter is 3 signals $C(v)$, $\mathrm{D}_1(\mathrm{v})$ and $\mathrm{D}_2(\mathrm{v})$. These items are the coefficients with the low frequency (the approximation coefficient) and 2 coefficients with the high frequency (the detail coefficient), respectively [18].

The synthesis filter bank is used to inversely transform the extracted low-pass coefficient $T_0(v)$ and 2 high-pass filters $T_1(v)$ and $T_1(v)$, with the high sampling by 2, and then, in order to receive the output signal, $K(v)$ is fused. D3WT performs the iterative operation with the over-sampled filter bank, to ensure the perfect reconstruction conditions. This work leads to the transformation of the shift constants, namely $T_0(v)$, $T_0(v)$ and $T_0(v)$, which $\mathrm{K}_{\mathrm{output}}(\mathrm{v}) = \mathrm{J}_{\mathrm{Input}}(\mathrm{v})$. Also, D3WT is applied to convert the samples of the malware in the coefficients of the detail on every level from the decomposition and into the coefficients of the approximation on maximum level. 2 wavelets, namely $\varphi_1(\alpha)$ and $\varphi_1(\alpha)$, are generated to be separated by $1/2$, as shown in the following equation [19]:

$$\varphi_1(\alpha) = \varphi_2 \times (\alpha - 0.5) \quad (1)$$

The following equations are given for a multi-resolution framework, that should satisfy θ and φ_i:

$$\theta(\alpha) = \sqrt{2}\sum_v T_0(v)\theta(2\alpha - v) \quad (2)$$

$$\varphi_i(\alpha) = \sqrt{2}\sum_v T_i(v) \times \theta(2\alpha - v) \ \ where\ i = 1.2 \quad (3)$$

C. Classification of Malware Samples

Regarding CNNs, it should be said that they are used in the classification of the image for the object recognition and the classification work. The prominent amicability of CNN is according to its ability for the automatic extraction of the important features from the input samples. CNNs are a set from three important layers: convolution, pooling and fully-connected. In the layer of the convolution, the input features are checked, and the input sample is filtered. This layer does the dot product among 2 matrices, for the creation of the weight matrix and the weighted sum in the layer of the kernel. This filter does the operation of the batch among the pixel values of the input. The proper parameters for the filter size layer, the stride and the zero padding, help to increase in results of the convolution kernels. Also, ReLU is applied, to enhance the non-linearity on the map of the feature. ReLU computes the activation with the setting of the input as 0. ReLU has the values of the negative and the positive. The negatives are displayed by 0 and the positives are indicated by the max value. In the layer of the fully-connected on CNN, a classification is performed, to sequentially determine the given input of the layer of the convolutional and the pooling. In the layer of the pooling, the desired operation is applied, to reduce the dimensionality of the features, by selecting the greatest values by the area for the creation of the matrix. The ultimate layer is the fully-connected, which is applied, to flatten total features in the vectors with the single feature. It presents the communication between the neurons of the previous layer and the next layer. This point causes that feature maps from the input to the output. The output of fully-connected is taken to SoftMax, to predict the samples of the malicious [20].

Regarding LSTMs, it should be said that they include a memory unit and $three$ interactive gates: the input, the forget and the output. The memory unit is applied, to protect the late state against the prior state. The gate of the input is applied, to restrict the amount of the input data for the training in network, which is stored on the state of the unit in the time "t". The gate of the forget determines which whether the data of the input should rouse forward or take away, for the entering to the gate of the input in the time "$t - 1$". The gate of the output describes the data of the output. This gate is applied, to deal the problems of the vanishing gradient, that appear in the implementation of 3 gates [20].

In this paper, a combined model of LSTM-CNN is applied, to detect the malware. CNN is superior for the processing of the large volume from the great-dimensional datasets and for the automatic extraction of the informational features by the images. It is done by using the techniques of the optimization with the great dimensional. The efficiency of CNNs is influenced by the samples size of the training and the testing. In the dataset with the variable size, CNN requires to be fine-tuned. The goal of the use from the feature extraction is the minimization of the data dimensions and the time of the training for CNN, that leads to the improvement in accuracy of the detection. On our approach, first, D3WT is used, to extract the first level feature. The fused features include the coefficients of the approximation and the detail, which serve as the input of CNN. Then, CNN exploits the necessary informational features by the minimal dimensions. On our combined approach, 4 layers of the convolution, 4 layers of the pooling, 4 layers of the dropout, one layer of the fully-connected (the flatten, the dense and the dropout), one layer of LSTM and SoftMax are applied. The fetched features are taken as the input of LSTM (by max-pooling). Then, they are passed

to a layer of the fully-connected that transforms total features to a vector with the single feature (by SoftMax). CNN extracts the features as efficiently and as automatically. On our combined approach, CNN works as the encoder for the encoding of features (by convolution), and LSTM decodes the encoded data. Eventually, a layer of the fully-connected is applied, to classify the malware. Thus, the foremost features of 2 networks are integrated, to model an impressive combined network [21].

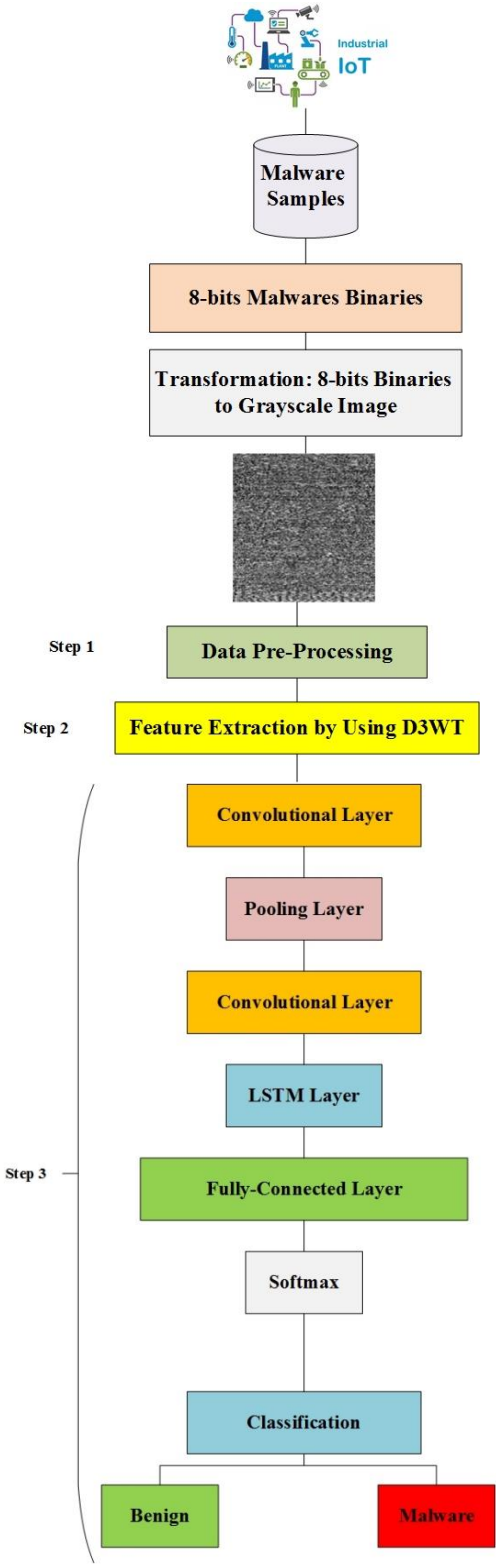


Fig. 1. The general framework of our presented method.

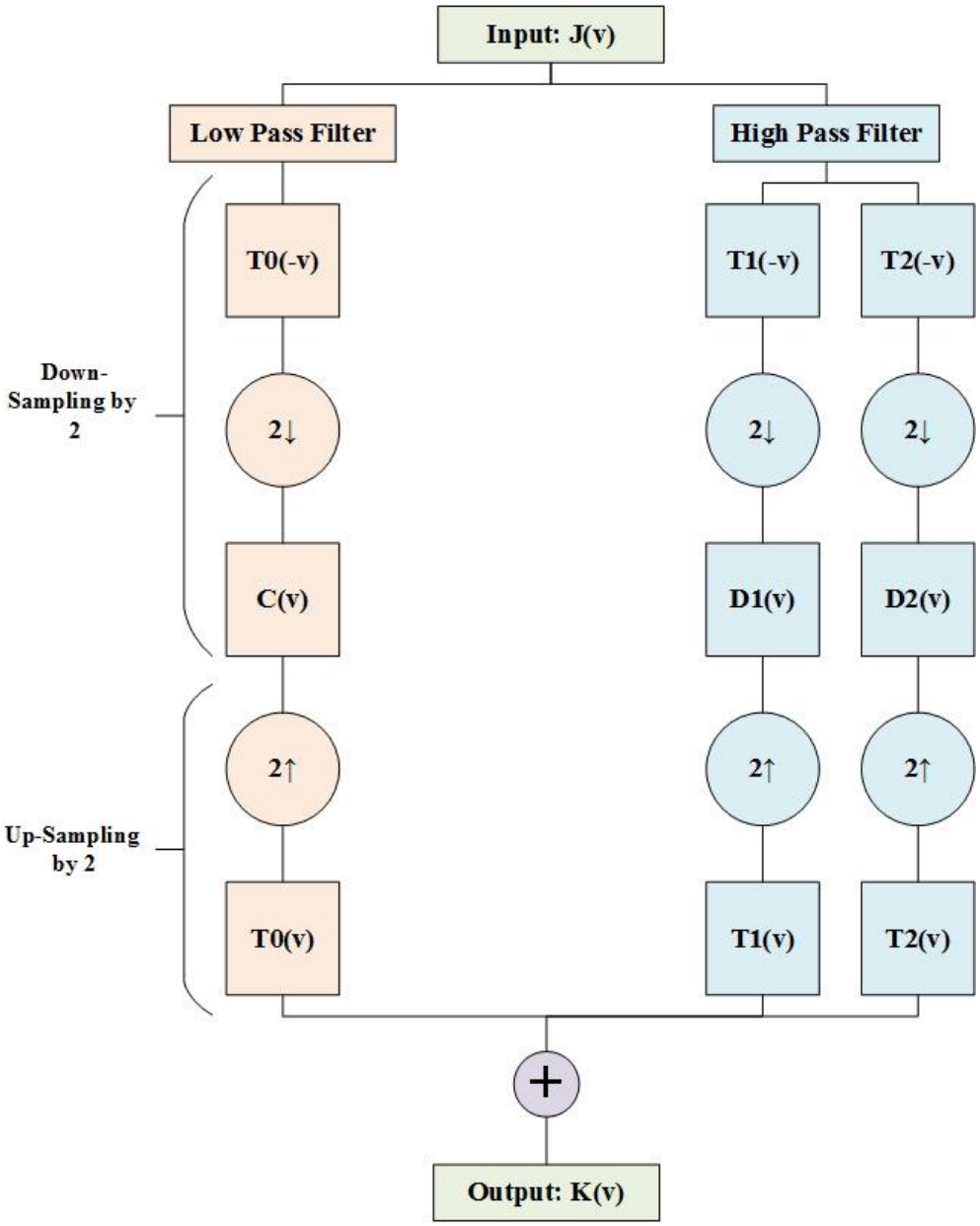


Fig. 2. The D3WT basis on 3-channel filter bank.

IV. Evaluation and Results

Here, the details of datasets and tests and results are provided. Python has been applied, to implement these tests. The proposed approach is implemented in the computer with RAM 8G and Intel(R) CPU Core(TM) i7 3.0 GHz. CNN is implemented on GPU and the graphics card is GEFORCE 840M for NVIDIA. The data of training, testing and validation are randomly selected from each dataset, and the evaluation procedures are done as one by one. In the stages of training, validation and testing, the data selection rates are set at 70%, 10%, and 20%, respectively. CNN includes 4 layers of the convolution, 4 layers of the pooling, 4 layers for the normalization of the batch and ReLU with the various sizes of the filter (32, 64 and 128). The sizes of pooling, stride and kernel are equal to 2×2, 2×2 and 3×3. After the fourth layer of the convolution, the features are passed via LSTM and the layer of the fully-connected. The layer of the fully-connected includes the layers of flatten, dense and dropout. Eventually, SoftMax applied, to classify the malware labels.

A. Datasets and Evaluation Criteria

To demonstrate the efficiency of our approach, the several criteria for the evaluation have been applied. These criteria are: sensitivity, accuracy, F1-score and specificity. These criteria are computed as the below:

$$Accuracy = \frac{TN+TP}{TN+FN+TP+FP} \quad (4)$$

$$Sensitivity = \frac{TP}{FN+TP} \quad (5)$$

$$Specificity = \frac{TN}{TN+FP} \quad (6)$$

$$F1 - Score = \frac{2\times TP}{2\times TP+FP+FN} \quad (7)$$

TP is the true positive and FP displays the false positive, while TN displays the true negative and FN displays the false negative. The mentioned criteria are the first stage for the interpretation of the efficiency for our approach. The comparison procedures are performed for the proposed combined model with 2 DNNs. These two neural networks are: AlexNet and Resnet-50.

The experiments are performed on *two* comprehensive datasets: Malimg and Microsoft BIG 2015. The dataset of Malimg [22] includes 9339 samples from the malwares. Every sample from the malwares on dataset belongs to one of 25 classes. In addition, number of the samples in a class is different. The classes of the malware are: Agent.FYI, Adialer.C, Allaple.L, Allaple.A, Alueron.gen!J, Autorun.K, Benign, C2LOP.P, C2LOP.gen!g, Dialplatform.B, Dontovo. A, Fakerean, Instantaccess, Lolyda.AA1, Lolyda.AA2, Lolyda.AA3, Lolyda.AT, Malex.gen!J, Obfuscator. AD, Rbot!gen, Skintrim.N, Swizzor.gen!E, VB.AT, Yuner.A and Wintrim.BX.

The dataset of Microsoft BIG 2015 [23] includes 21741 samples from the malwares, which are belonging to *nine* classes, and are: Lollipop, Kelihos_ver1, Ramnit, Vundo, Kelihos_ver3, Tracur, Obfuscator, Gatak, .ACY and Simda. The similar to with Malimg, the number of the samples from the malwares in the classes is not uniformly dispensed. Every sample from the malwares is displayed by 2 files: ".asm" and ".byte". In the experiments, ".byte" are only used, to form the malware images.

B. Results

In this section, the outcomes of the performance for our proposed approach and its comparison with the other models are presented. The evaluation criteria define the efficiency of the models. The acute case behind classification is a criterion for the evaluation, which is applied for the understanding of the efficiency of a model [24]. Therefore, the multiple criteria in experimental outcomes and the discussion have been used, to demonstrate the efficiency of our approach. Fig. 3 to Fig. 6 and Fig. 7 to Fig. 10 show the values of the accuracy, the sensitivity, the specificity and the F1-score for AlexNet, Resnet-50 and the presented approach on Microsoft BIG 2015 and Malimg, respectively. According to these results, it can be said which our approach works superior than the DNN architecture. Also, the performance of our approach shows the similar efficiency results on 2 datasets, while the performance of the other DNNs is significantly different on 2 datasets. The above situations show which our approach is stronger, and outperforms than 2 DNNs.

In the next step, the types of the malwares are analyzed along with the confusion matrices. Tables I, II and III show the matrices of the confusion in Microsoft BIG 2015 for *nine* types of malware by using the proposed approach and AlexNet and ResNet-50. Here, the accuracy rate for each type of the malwares is shown by using the confusion matrices. The matrix of the confusion for the proposed approach shows that it provides the better results for the entire malware classification, with the exception of vundo. In addition, the matrix of the confusion for ResNet-50, which is displayed in Table II, provides the better detection for vundo, in compared to the other approaches. In this situation, total models can quickly detect the malwares of simda and tracur.

Finally, a functional comparison with the advanced results is performed. Tables IV and V display the accuracy in Malimg and Microsoft Big 2015 by using our approach and the other models. It should be attended which the efficiency of the presented approach is better than the advanced models, insomuch it creates a greater value for the accuracy.

C. Discussion

The promising results of the proposed framework, which uses an image visualization approach, demonstrate its ability to accurately identify various malware families and new variants. These results indicate that image-based visualization is both effective and efficient for identifying malware samples across different classes. This model has significant potential for detecting advanced malware by analyzing large volumes of input data and classifying them into sub-families. However, the model's performance in training and testing may decline with smaller datasets. Additionally, inaccurate samples of both malware and benign instances can lead to poor malware identification. Despite the effectiveness of our proposed hybrid deep learning architecture in detecting and classifying various malware variants and families, there are still some limitations that need to be addressed.

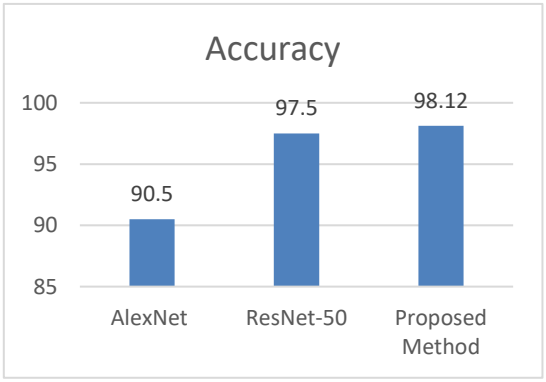


Fig. 3. The comparison of the accuracy of the various models on Malimg.

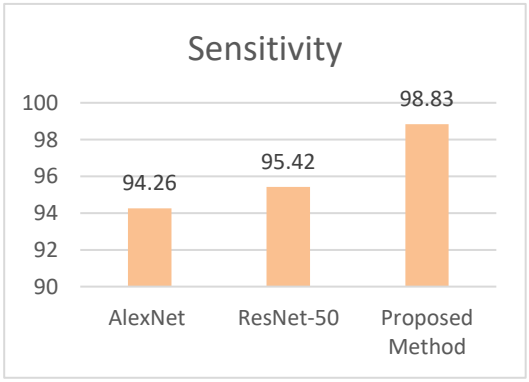


Fig. 4. The comparison of the sensitivity various models on Malimg.

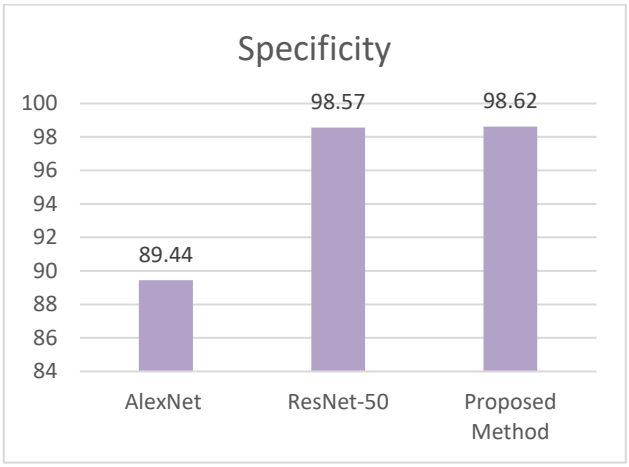


Fig. 5. The comparison of the specificity of the various models on Malimg.

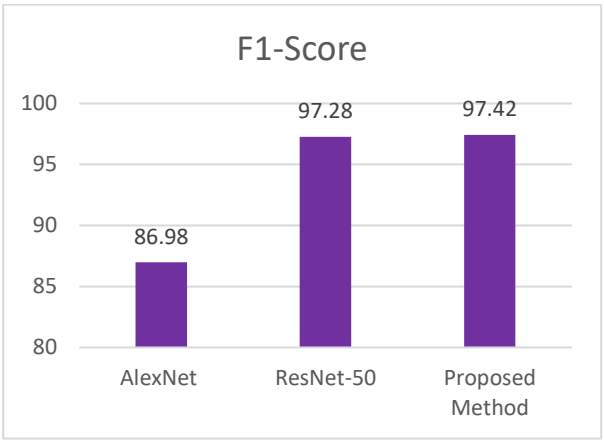


Fig. 6. The comparison of the F1-score of the various models on Malimg.

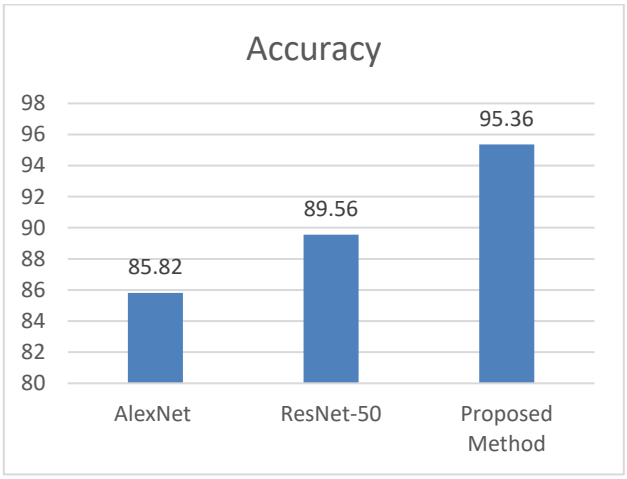


Fig. 7. The comparison of the accuracy of the various models on Microsoft BIG 2015.

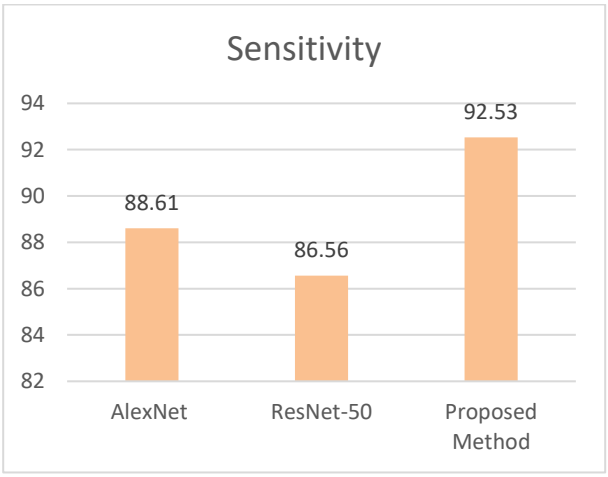


Fig. 8. The comparison of the sensitivity various models on Microsoft BIG 2015.

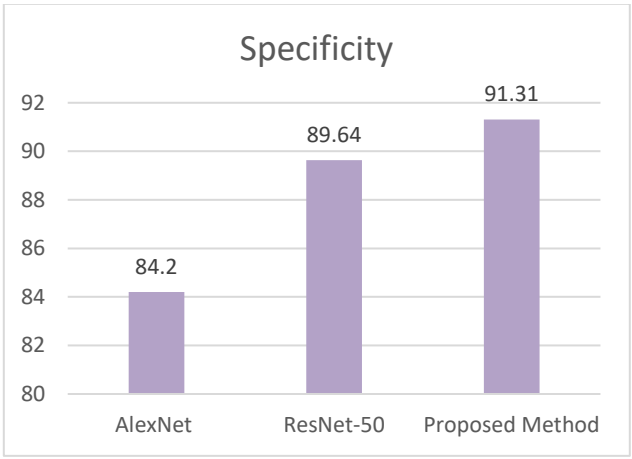


Fig. 9. The comparison of the specificity of the various models on Microsoft BIG 2015.

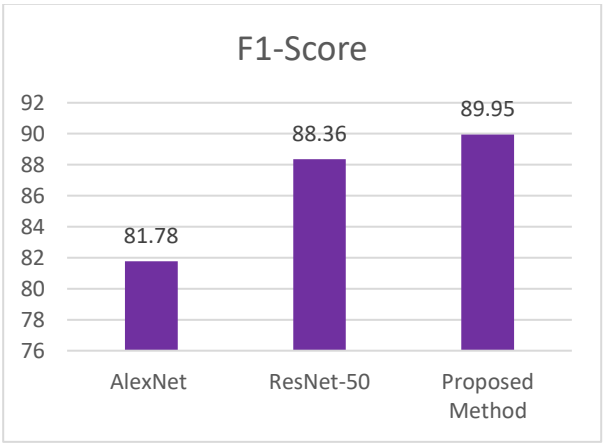


Fig. 10. The comparison of the F1-score of the various models on Microsoft BIG 2015.

TABLE I. THE MATRIX OF THE CONFUSION IN MICROSOFT BIG 2015 FOR NINE TYPES OF THE MALWARE BY USING OUR APPROACH

	Ramnit	Lollipop	Kelihos_ver1	Kelihos_ver3	Vundo	Simda	Tracur	Obfuscator.ACY	Gatak
Ramnit	92.2	2.7	0.2	1.2	0.5	1.3	0.2	0.6	1.1
Lollipop	0.1	95.1	0.8	0.4	0.7	0.3	1.3	1	0.3
Kelihos_ver1	1.1	1.4	92.5	2.5	0.1	0.7	0.4	1.2	0.1
Kelihos_ver3	0.5	0.9	2.1	94.6	0.5	0.1	0.6	0.4	0.3
Vundo	0.4	1.8	0.6	0.3	95.6	0.6	0.4	0.1	0.2
Simda	0.2	0.1	0.4	0.2	0.8	97.2	0.3	0.4	0.4
Tracur	0.1	0.3	0.2	0.3	0.1	0.1	98.4	0.2	0.3
Obfuscator.ACY	1.9	1	0.1	0.1	0.1	0.3	1.7	92.7	2.1
Gatak	0.2	0.5	0.9	0.8	0.2	0.5	1.2	0.1	95.6

TABLE II. THE MATRIX OF THE CONFUSION IN MICROSOFT BIG 2015 FOR NINE TYPES OF THE MALWARE BY USING RESNET-50

	Ramnit	Lollipop	Kelihos_ver1	Kelihos_ver3	Vundo	Simda	Tracur	Obfuscator.ACY	Gatak
Ramnit	83.1	3.6	0.8	2.5	1.5	2.1	1.3	2.7	2.4
Lollipop	1.5	86.3	4.5	2.7	3.5	0.5	0.2	0.2	0.6
Kelihos_ver1	0.2	1.3	79.4	5.6	2.7	4.4	0.6	2.6	3.2
Kelihos_ver3	3.2	2.6	1.8	82.9	1.2	1.3	2.7	1.6	2.7
Vundo	0.3	0.1	0.5	0.2	96.8	0.6	0.1	0.5	0.9
Simda	0.3	0.6	0.7	0.1	0.6	96	0.4	0.7	0.6
Tracur	0.2	0.4	0.3	0.1	0.4	0.1	98.1	0.3	0.1
Obfuscator.ACY	3.1	0.1	0.3	0.7	2.3	1.2	0.7	89.9	1.8
Gatak	1	0.7	2.1	2.7	0.1	0.4	0.1	0.3	93.5

TABLE III. THE MATRIX OF THE CONFUSION IN MICROSOFT BIG 2015 FOR NINE TYPES OF THE MALWARE BY USING ALEXNET

	Ramnit	Lollipop	Kelihos_ver1	Kelihos_ver3	Vundo	Simda	Tracur	Obfuscator.ACY	Gatak
Ramnit	80.6	5.2	1.3	0.9	0.1	4.6	6.7	0.2	0.4
Lollipop	3.6	82.7	3.6	3.1	0.6	1.7	4.3	0.1	0.3
Kelihos_ver1	1.4	2.5	83.5	3.2	1.1	0.1	0.6	6.4	1.2
Kelihos_ver3	4.5	1.6	6.5	78.8	3.8	2.1	0.1	0.8	1.8
Vundo	0.6	0.7	1.2	6.5	80	3.2	5.6	0.7	1.5
Simda	0.5	0.1	0.2	0	1.1	97.3	0.5	0.1	0.2
Tracur	0.4	0.2	0.1	0.2	0.3	1.2	96.6	0.9	0.1
Obfuscator.ACY	1.6	2.6	0.5	1.3	0.2	0.1	2.8	87.5	3.4
Gatak	0.9	1.6	4.5	4.4	0.2	2.6	0.2	0.1	85.5

TABLE IV. THE COMPARISON OF OUR APPROACH WITH THE ADVANCED ALGORITHMS ON MALIMG

Method	Accuracy
Method in [25]	93.72
Method in [26]	94.50
Method in [27]	95.33
Method in [28]	96.08
Method in [29]	96.30
Proposed Method	98.12

TABLE V. THE COMPARISON OF OUR APPROACH WITH THE ADVANCED ALGORITHMS ON MICROSOFT BIG 2015

Method	Accuracy
Method in [25]	93.57
Method in [26]	93.40
Method in [27]	94.64
Method in [28]	94.24
Method in [29]	91.27
Proposed Method	95.36

D. Future Works

The proposed deep learning architecture shows some resistance to obfuscation, as evidenced by satisfactory results on the Microsoft BIG 2015 Dataset, which includes obfuscated malware samples. However, the method has not been tested against adversarial attacks with crafted inputs. Future research aims to evaluate the method's resilience to evasion attacks. Misclassification can occur due to similarities in features among different malware families. The model has only been tested on the Malimg and Microsoft BIG 2015 datasets. Future research could involve evaluating the model on additional datasets. The current architecture was implemented with limited computational power and resources. Plans for future work include deploying the model in a cloud computing environment to leverage greater computational power and resources. Additionally, future studies will use fewer hidden layers to reduce model complexity and will focus on extending the model by incorporating explainable AI and the latest feature optimization techniques to enhance real-time malware detection.

V. CONCLUSIONS AND SUGGESTIONS

The IoTs technology deepens the attendance of the connected sensors to the Internet in our daily behaviors, and brings the many benefits in the quality of the life. Also, it has generated the related problems to the security issues. Accordingly, the security solutions for the Internet of Things should be developed. In the current article, a new impressive cross-platform malware detection method based on artificial intelligence is designed for the industrial sensors of IoTs. In the presented approach, D3WT is used for the extraction of the feature. In addition, a combination from CNN and LSTM is used, to detect the malwares. Our approach is evaluated on Malimg and Microsoft BIG 2015. First, the proposed approach is compared with every network as separately. The obtained results confirm which our approach can impressively classify the malwares with the great values of the accuracy, the sensitivity, the specificity and the F1-score. When tested on the Microsoft BIG 2015 and Malimg datasets, the accuracy achieved is 95.36% and 98.12%, respectively. Then, the presented approach has been analyzed with the advanced models. The obtained outcomes show the benefit and the superiority of our approach against the similar models.

The model has only been tested on the Malimg and Microsoft BIG 2015 datasets. Future research could involve evaluating the model on additional datasets. The current architecture was implemented with limited computational power and resources. Plans for future work include deploying the model in a cloud computing environment to leverage greater computational power and resources. For further study, provision of a detection system that specifically classifies the malwares which uses the obfuscation techniques, can be considered. In addition, the next researches can rely on the deployment of the presented approach, by integrating the model basis on AI by the latest techniques for the optimization of the feature, to enhance the detection of the malware in the real time.

REFERENCES

[1] Miorandi, D., Sicari, S., De Pellegrini, F., Chlamtac, I., 2012. Internet of Things: vision, applications and research challenges. Ad Hoc Netw.10(7).1497-1516.

[2] Atzori, L. Iera, A., Morabiti, G., 2010. The internet of things: A survey, computer Network, V54, 15, 2787-2805.

[3] Borgia, E., 2014. The Internet of Things vision: key features, applications and open issues. Comput Commun.54, 1-31.

[4] Shigen Shen; Longjun Huang; Haiping Zhou; Shui Yu; En Fan; Qiying Cao, Multistage Signaling Game-Based Optimal Detection Strategies for Suppressing Malware Diffusion in Fog-Cloud-Based IoT Networks, IEEE Internet of Things Journal.

[5] Sicari, S., Rizzardi, A., Grieco, L., Coen-Porisini, A., 2015. Security, privacy and trust in Internet of Things: the road ahead. Comput. Netw. 76 (0), 146–164.

[6] Mudgerikar, A., Sharma, p., & Bertino, E.,2019. A system- level Intrusion Detection System for IoT Devices. In Proceedings of the 2019 ACM Asia Conference on Computer and Communications Security, 493-500.

[7] N. Dosti.,2019. New mechanism to enhance IoT network security using quantum and classical cryptography (in Persian), Journal of Electronical & Cyber Defence, Vol 4.

[8] Kumar, S., Dutta, K., 2016. Intrusion detection in mobile ad hoc networks: techniques, systems, and future challenges. Secur. Commun. Netw. 9 (14), 2484–2556.

[9] Midi, S., Krishna, P., Agarwal, H., Saxena, A., Obaidat, M., 2011. A learning automata based solution for preventing Distributed Denial of Service in Internet of Things. In: Internet of Things (iThings/CPSCom), International Conference on and Proceedings of the 4th International Conference on Cyber, Physical and Social Computing, 114–122.

[10] Butun, I., Morgera, S., Sankar, R., 2014. A survey of intrusion detection systems in wireless sensor networks. Commun. Surv. Tutor. IEEE 16 (1), 266–282.

[11] Modi, C., Patel, D., Borisaniya, B., Patel, H., Patel, A., Rajarajan, M., 2013. A survey of intrusion detection techniques in Cloud. J. Netw. Comput. Appl. 36 (1), 42–57.

[12] Mitchell, R., Chen, I.-R., 2014. A survey of intrusion detection techniques for cyberphysical systems. ACM Comput. Surv. (CSUR) 46 (4), 55.

[13] Granjal, J., Monteiro, E., Silva, J.S., 2012. On the effectiveness of end-to-end security for Internet-integrated sensing applications. In: Green Computing and Communications (GreenCom), IEEE, 87–93.

[14] Le, A., Loo, J., Chai, K.K., Aiash, M., 2016. A specification-based IDS for detecting attacks on RPL-based network topology. Information 7 (2), 25.

[15] Arwa Aldweesh, Abdelouahid Derhab., 2020. Deep learning approaches for anomaly-based intrusion detection systems: A survey, taxonomy, and open issues. Knowledge-Based Systems. Vol 189.

[16] Klempous, Ryszard, et al., 2007. Adaptive misbehavior detection in wireless sensors network based on local community agreement. 14th Annual IEEE International Conference and Workshops on the Engineering of Computer- Based System.

[17] A. Marosi, E. Zabah, H. Ataee khabaz.,2020. Network intrusion detection using a combination of artificial neural networks in a hierarchical manner (in Persian), Journal of Electronical & Cyber Defence , Vol 8, pp. 89-99.

[18] Al-Timime ZS. Signal denoising using double density discrete wavelet transform. J Al-Nahrain Univ Sci 2017;20(4):125–9. https://doi.org/10.22401/ jnus.20.4.19.

[19] Qiao YL, Song CY. Double-density dual-tree wavelet transform based texture classifcation. In: IIH-MSP 2009 - 2009 5th Int. Conf. Intell. Inf. Hiding Multimed. Signal Process. 1; 2009. p. 1322–5. https://doi.org/10.1109/IIH-MSP.2009.148.

[20] Islam MZ, Islam MM, Asraf A. A combined deep CNN-LSTM network for the detection of novel coronavirus (COVID-19) using X-ray images. Informatics Med Unlocked 2020;20:100412. https://doi.org/10.1016/j.imu.2020.100412.

[21] Shahzad F, Mannan A, Javed AR, Almadhor AS, Baker T, Al-Jumeily OBE D. Cloud-based multiclass anomaly detection and categorization using ensemble learning. J Cloud Comput 2022;11(1). https://doi.org/10.1186/s13677-022-00329-y.

[22] L. Nataraj, S. Karthikeyan, G. Jacob, and B. S. Manjunath, ‘‘Malware images: Visualization and automatic classification,’’ in Proc. 8th Int. Symp. Visualizat. Cyber Secur. (VizSec), 2011, pp. 1–7.

[23] Microsoft Malware Classification Challenge (Big 2015). Accessed: Apr. 20, 2021. [Online]. Available: https://www. kaggle.com/c/malwareclassification.

[24] T. Saranya, S. Sridevi, C. Deisy, T. D. Chung, and M. K. A. A. Khan, ‘‘Performance analysis of machine learning algorithms in intrusion detection system: A review,’’ Procedia Comput. Sci., vol. 171, pp. 1251–1260, Jan. 2020.

[25] J.-S. Luo and D. C.-T. Lo, ‘‘Binary malware image classification using machine learning with local binary pattern,’’ in Proc. IEEE Int. Conf. Big Data (Big Data), Dec. 2017, pp. 4664–4667.

[26] Z. Cui, F. Xue, X. Cai, Y. Cao, G.-G. Wang, and J. Chen, ‘‘Detection of malicious code variants based on deep learning,’’ IEEE Trans. Ind. Informat., vol. 14, no. 7, pp. 3187–3196, Jul. 2018, doi: 10.1109/tii.2018.2822680.

[27] D. Gibert, ‘‘Convolutional neural networks for malware classification,’’ M.S. thesis, Univ. Rovira Virgili, Tarragona, Spain, Oct. 2016.

[28] A. Singh, A. Handa, N. Kumar, and S. K. Shukla, ‘‘Malware classification using image representation,’’ in Proc. Int. Symp. Cyber Secur. Cryptogr. Mach. Learn. Cham, Switzerland: Springer, Jun. 2019, pp. 75–92.

[29] R. Vinayakumar, M. Alazab, K. P. Soman, P. Poornachandran, and S. Venkatraman, ‘‘Robust intelligent malware detection using deep learning,’’ IEEE Access, vol. 7, pp. 46717–46738, 2019.

Quantifying the Effects of Homogeneous Interference on Coverage Quality in Wireless Sensor Networks

Qingmiao Liu[1], Qiang Liu[2], Minhuan Wang[3]
School of Management Science and Engineering, Shandong University of Finance and Economics, Jinan, China[1, 2]
Tongji University, Shanghai, China[3]

Abstract—This study develops a coverage perception interference model for Wireless Sensor Networks, focusing on the challenges of homogeneous interference within Regions of Interest. Traditional perception models often overlook areas that, while covered, do not meet the required coverage standards for accurate classification. This model addresses both uncovered areas and those inadequately covered, which are susceptible to classification errors. A propositional space for the coverage model is defined to assess the impact of homogeneous interference on sensor nodes, with the aim of quantifying its effects on network coverage quality and stability in complex environments. The study emphasizes the generation of Basic Probability Assignments using Dempster-Shafer theory, a robust framework for managing uncertain information in sensory data. Probability Density Functions derived from historical and real-time data are utilized to facilitate precise BPA calculations by integrating over specific attribute ranges, thereby enhancing the accuracy and reliability of target detection. Algorithms are also developed to calculate the interference effect BPA, which are integrated with perception coverage models to improve the assessment and optimization of coverage quality. The research enhances the methodological understanding of managing interference in WSNs and offers practical strategies for improving sensor network operations in environments affected by significant interference, boosting the reliability and effectiveness of critical surveillance and monitoring applications.

Keywords—Wireless sensor networks; homogeneous interference; basic probability assignment; coverage quality; Dempster-Shafer theory

I. INTRODUCTION

In the current global landscape, security and surveillance of sensitive areas such as borders, critical infrastructures, and urban centers are of paramount importance. The increasing complexity and sophistication of threats, ranging from unauthorized human intrusions to vehicular entries and wildlife disturbances, necessitate advanced technological solutions that can operate under diverse environmental conditions and provide real-time, reliable data. Wireless Sensor Networks (WSNs) have emerged as a pivotal technology [1-9] in this domain, offering the potential to revolutionize the way these areas are monitored. However, despite their significant potential, the deployment of WSNs in security applications faces several critical challenges that hinder their effectiveness.

One of the primary challenges is the high rate of false positives, which can lead to unnecessary alarms and subsequently drain the resources and attention of security personnel. False positives are predominantly caused by the inability of traditional WSNs to accurately classify the nature of the intrusion [10-14]. For instance, the motion sensors in a network might be triggered by non-threatening entities such as small animals or environmental factors like wind. Such inaccuracies not only undermine the reliability of security protocols but also reduce the trust in these systems.

Moreover, the effectiveness of WSNs is often limited by their capacity for real-time processing and analysis of sensor data. Security applications require immediate responses to detected threats, and any delay in the processing can result in a failure to prevent an intrusion [15-20]. Additionally, the diverse range of intruders and the subtleties in their characteristics necessitate sophisticated algorithms capable of making nuanced distinctions. Current systems predominantly employ simplistic threshold-based algorithms [21-23], which are not only prone to errors but also lack the ability to learn and adapt from past data, thus failing to improve over time.

Current WSN implementations primarily focus on detecting the presence of an intruder rather than classifying the type of intrusion accurately [24-26]. This is a significant limitation, as different types of intrusions require different responses. For instance, the approach to dealing with a human trespasser might differ substantially from that for a wild animal entering a restricted area. Most existing systems utilize basic motion sensors that trigger alarms when interrupted, regardless of the cause. Such systems are unable to distinguish between false alarms caused by non-threatening entities and genuine security breaches, leading to high rates of false positives. This limitation is further exacerbated by the lack of integration of advanced classification algorithms within the sensor networks. While there are robust individual sensor technologies capable of complex data processing [27-30], their integration into network-wide systems that perform real-time analysis and classification is not adequately addressed in existing research.

The integration of advanced computational models, such as those based on Dempster-Shafer theory [31-35], presents a promising solution to these challenges. The Dempster-Shafer theory of evidence allows for the combination of evidence from different sources to arrive at a degree of belief (represented by a belief function) that can handle uncertainty more effectively than traditional probabilistic methods.

Previous studies on Wireless Sensor Networks have primarily concentrated on basic detection algorithms and simplistic models [36-40] that often fail to account for the complexities introduced by homogeneous interference. These studies typically emphasize threshold-based detection

mechanisms, which are inadequate in environments where interference affects multiple sensor nodes uniformly. As a result, these approaches struggle to maintain accurate coverage, leading to increased false positives and unreliable surveillance outcomes. Furthermore, existing research has largely overlooked the integration of advanced probabilistic models to address the uncertainty and ambiguity in sensor data caused by interference.

Given the shortcomings of previous research, there is a clear and pressing need for further investigation into how homogeneous interference impacts the coverage quality of Wireless Sensor Networks. This paper addresses these critical gaps by evaluating the effectiveness of WSNs under the influence of such interference, particularly focusing on interference that uniformly affects multiple sensors. Our study proposes a novel coverage perception interference model that leverages the Dempster-Shafer theory to enhance the robustness and accuracy of WSNs, enabling them to maintain effective coverage even in challenging conditions. Through detailed analysis and modeling, we explore how homogeneous interference compromises the network's ability to sustain reliable coverage and identify vulnerable zones. By incorporating Basic Probability Assignments within the Dempster-Shafer evidence framework, this research provides a nuanced understanding of interference effects, offering practical solutions to improve network reliability. This contribution not only advances the current body of knowledge in WSNs but also establishes a foundation for future research aimed at developing more resilient and adaptive sensor networks capable of operating effectively in complex, interference-prone environments.

This paper is structured as follows: Section II provides a detailed overview of the recognition framework and the mathematical models used in the study, including the implementation of the Dempster-Shafer theory for handling uncertainties in Wireless Sensor Networks (WSNs). Section III introduces the S-I perimeter coverage algorithm, which accounts for homogeneous interference, and discusses its operational rules and algorithmic steps. In Section IV, the simulation results are presented and analyzed, highlighting the impact of interference on coverage quality and the effectiveness of the proposed algorithm. Finally, Section V concludes the paper by summarizing the key findings and suggesting potential directions for future research in improving WSN coverage reliability under interference conditions.

II. PRELIMINARIES

A. Recognition Framework

Wireless Sensor Networks (WSNs), strategically deployed within designated Regions of Interest (ROIs), play a crucial role in enhancing the security and surveillance across vulnerable and sensitive areas by monitoring unauthorized entries from a variety of entities such as humans, animals, and vehicles. These networks are comprised of a coordinated array of sensor nodes, each designated as S_1 （ i=1,2,3…, ξ . These nodes are meticulously programmed and equipped to classify potential intruders into several distinct categories, such as C_1 , C_2 , C_3,...,C_φ, facilitating a targeted response to different types of security breaches.

Each sensor node within the network is outfitted with cutting-edge technology capable of capturing a wide array of detailed attributes from each detected entity. These include visual identifiers like shape and size, infrared signatures that reveal body heat, variations in ambient temperature, and precise measurements of movement speeds. This rich dataset enables a comprehensive and multifaceted analysis of each intrusion, substantially increasing the accuracy of both detection and subsequent classification processes. Beyond mere physical detection, the nodes are also equipped to sense more subtle indicators such as acoustic signals and electromagnetic properties. This capability is essential for distinguishing between organic and mechanical intruders, effectively differentiating between humans, animals, and vehicles. This nuanced approach to intrusion detection is crucial for deploying appropriate security measures and for minimizing false alarms, which are common in less sophisticated systems.

Through the use of advanced data analytics in this article, the information captured by the sensors is analyzed in real-time. This not only ensures timely detection but also enables the network to categorize each object based on its unique attributes and behavior patterns. The adaptability and responsiveness of these networks to a range of environmental stimuli and potential threats are vital, enhancing the overall security protocol of the area.

In our investigation, we explore scenarios involving three primary types of targets, illustrating a methodology and conceptual framework that are universally applicable to scenarios involving the classification of implicit targets into φ (φ>3) categories. Utilizing Dempster-Shafer (D-S) theory, we establish a discernment framework denoted as $\Theta = \{C_1,C_2,C_3\}$. The power set of Θ, represented as 2^Θ, encompasses $2^\Theta = \{ \emptyset , A_1 =\{ C_1 \}, A_2=\{C_2\},A_3=\{C_3\},A_4=\{C_1,C_2\},A_5=\{C_1,C_3\},A_6=\{C_2, C_3\}, A_7=\{C_1,C_2,C_3\}\}$, Each element within 2^Θ, $A_\omega(\omega=1,2\ldots,7)$ forms a subset of Θ, symbolizing a specific proposition. Here the propositions$\{C_1\}$、$\{C_2\}$ or $\{C_3\}$ correspond to the sensor classifying the target into categories$\{C_1\}$、$\{C_2\}$ or $\{C_3\}$, respectively. The propositions $\{C_2,C_3\},\{C_1,C_3\}$, $\{C_1,C_2,C_3\}$ indicate ambiguity in the sensing results, suggesting that the target could potentially belong to the combined categories, whereas $\emptyset$ represents the empty set. These propositions are categorized into two types: 1) singleton propositions, where the corresponding subset contains a single element, such as $\{C_1\}$、$\{C_2\}$ and $\{C_3\}$; and 2) multiple subset propositions, where the subset comprises multiple elements, such as $\{C_1,C_2\},\{C_2,C_3\},\{C_1,C_3\}$ and $\{C_1,C_2,C_3\}$.

When a target enters the sensing range of sensor S_i, the sensor perceives each attribute of the target. Based on these attributes, S_i categorizes the target into one of the defined categories, resulting in a classification that is expressed through a basic probability assignment (BPA). The BPA derived from attribute θ by sensor S_i, denoted as m_θ^i, represents a mapping from 2^Θ to the interval [0,1] and adheres to the following conditions:

$$\begin{cases} m_\theta^i(\emptyset) = 0 \\ \sum_{A\subseteq\Theta} m_\theta^i(A) = 1 \end{cases} \quad (1)$$

The sum of all probability masses assigned across the power set must equal one, ensuring a complete and exhaustive representation of all possible classification outcomes.

The probability of the empty set, ∅, is zero, reflecting the premise that every observation can be attributed to at least one category within the framework.

This structured approach not only enhances the precision of target classification within complex environments but also significantly contributes to the development of robust, scalable sensor networks capable of adapting to diverse surveillance and monitoring challenges.

Within the proposition space Θ, A represents any subset, and , $m_\theta^i(A)$ denotes the mass or credibility allocated to proposition A. For instance, upon detecting the attributes θ of an intruding target, sensor node S_i assigns category probability values as follows: $m_\theta^i(C_1) = p_1$, $m_\theta^i(C_2) = p_2$, $m_\theta^i(C_3) = p_3$, $m_\theta^i(\{C_1, C_2\}) = p_4$, $m_\theta^i(\{C_1, C_3\}) = p_5$, $m_\theta^i(\{C_2, C_3\}) = p_6$, $m_\theta^i(\{C_1, C_2, C_3\}) = p_7$, where the sum of p_1 through p_7 equals 1. Given that the invading target A possesses θ attributes, sensor node S_i can measure θ BPA values. For each attribute i=1,2,...,θ, sensor S_i combines these θ BPA values using a special fusion rule denoted as ⊕, thereby deriving a new BPA value m^i, which represents the final detection outcome of the target by sensor S_i.

$$m^i = m_1^i \oplus m_2^i \oplus m_3^i \oplus \cdots m_\theta^i \quad (2)$$

B. Implicitly Targeted BPA Function Generation

The generation of Basic Probability Assignment (BPA) using the Dempster-Shafer (D-S) theory represents a critical step in applying evidence theory to the accurate characterization and identification of targets within sensor networks. The quality of BPA generation crucially influences the precision of perception results and the accuracy of target classification. Currently, the methodology for generating BPA predominantly relies on classification-based approaches, which can be summarized through the following steps: Derivation of classification criteria from historical data; Collection of attribute data from targets awaiting identification, followed by obtaining initial classification results based on the derived criteria; Transformation of these initial classification results into BPA through specific rules. The initial step, forming classification criteria, is pivotal and broadly categorized into three approaches: Expert Systems and Rule-Based Reasoning; Statistical analysis; Machine learning. While each approach offers distinct advantages, they also present challenges such as the maintenance of rule systems, complexity in parameter setting for fuzzy logic, reliance on prior knowledge in Bayesian inference, and the high cost of data annotation in supervised learning. The chosen approach in this research involves using PDFs to fit the attribute values of category targets, subsequently using these fits to generate BPAs for unclassified targets. This method leverages the Gaussian distribution of attribute values to reflect individual variability and measurement errors.

A novel method for determining BPA has been proposed, involving the division of a dataset into a training and a test set. Gaussian models for p attributes are established using the training set and then tested using the test set to determine similarities. Attribute weights are adjusted based on the overlap degree among categories to fine-tune the similarity scores and finalize the BPA, as depicted in the methodology section. This approach provides a structured, objective classification by assessing and integrating new information dynamically, which is essential for robust decision-making in sensor networks.

C. Attribute Modeling Approach

In the discernment framework $\Theta=\{C_1, C_2, C_3 \cdots C_n\}$, each category φ_i (i=1,2,3 $\cdots \tau$ ） is associated with a Gaussian distribution and is characterized by p_j(j=1,2,3⋯P） attributes. For each category φ_i, the mean and standard deviation for each attribute are derived from the training samples as follows:

The mean value for attribute p, represented as $\bar{X}_p$ is calculated using the formula:

$$\bar{X}_p = \frac{1}{N}\sum_{\tau=1}^{N} x_{\tau,p} \quad (p = 1,2,3 \cdots P) \quad (3)$$

where, $x_{\tau,p}$ is the value of attribute p for the τ-th sample in category φ_i.

The standard deviation for attribute p, denoted as σ_p, is determined by:

$$\sigma_p = \sqrt{\frac{1}{N-1}\sum_{\tau=1}^{N}(x_{\tau,p} - \bar{X}_{\tau,p})^2} \quad (p = 1,2,3 \cdots P) \quad (4)$$

These statistical measures establish the parameters for the Gaussian distribution model of each attribute, which is defined as:

$$\mu_\alpha(x) = e^{-\frac{(x-\bar{X}_p)^2}{2\sigma_p^2}} \quad (5)$$

In this model, the Gaussian-type attribute model is considered a singleton proposition where both $\mu_\alpha(x)$ and $\mu_\beta(x)$ represent individual propositions within the framework. Complex or composite subset propositions are formed by the overlapping regions of these Gaussian membership functions. For example, the composite subset proposition $\{\alpha\beta\}$ is defined by:

$$\mu_{\alpha\beta}(x) = \min\{\mu_\alpha(x), \mu_\beta(x)\} \quad (6)$$

This expression effectively captures the lowest membership value between the two propositions, representing the degree of certainty that the value x belongs to both categories α and β simultaneously. This formulation allows for a nuanced understanding of the intersections and relationships between different category attributes in a multi-dimensional attribute space, facilitating a more precise and sophisticated approach to category classification in statistical analysis and machine learning applications.

D. Similarity Measurement

In the realm of multi-category classification, the construction of attribute weights plays a pivotal role in achieving unbiased results through a comprehensive evaluation of each attribute. The efficacy of an attribute in distinguishing between categories is inversely proportional to the degree of similarity

among the categories it connects. Specifically, if a given attribute exhibits substantial overlap across multiple category models, its discriminatory power is diminished, increasing the likelihood of misclassification and reducing its reliability. Consequently, the Basic Probability Assignment (BPA) generated from such attributes contributes minimally in a multi-category classification setting.

Conversely, attributes that demonstrate low similarity among categories possess enhanced discriminatory capability, thereby affirming their reliability and increasing their contribution to the BPA in the classification process. It becomes imperative to amplify the role of attributes with substantial contributions while diminishing the influence of those with minimal impacts, to foster more objective classification outcomes.

This section introduces and discusses the concept of attribute weighting, where each attribute's weight is inversely related to its degree of overlap among categories, reflecting its discriminative strength and reliability. Suppose μ_{ij} (i=1,2,…,k;j=1,2,…,l) represents the membership function for the j-th attribute of the i-th category, and $\mu_{hj}^{\triangle}$(h=1,2,…,$\frac{k^2-k}{2}$;j=1,2,…,l) denotes the generalized triangular fuzzy number model for the composite subset proposition {AB} for the j-th attribute in the h-th composite category. Let S(x) symbolize the integral or total sum over the defined range of the membership function. The weight ω_j (j=1,2,…,k) for the j-th attribute can be formulated as:

$$\omega_j = 1 - \frac{\sum_{h=1}^{\frac{k^2-k}{2}} S(\mu_{hj}^{\triangle})}{\sum_{i=1}^{k} S(\mu_{ij}) - \sum_{h=1}^{\frac{k^2-k}{2}} S(\mu_{hj}^{\triangle})} \quad (7)$$

Here, a higher value of ω_j (j=1,2,…,k) indicates greater overlap and similarity, thereby assigning a lower weight to the attribute. This weighting approach ensures that attributes contributing significantly to the classification accuracy are emphasized, while those with lesser discriminative power are de-emphasized, optimizing the classification framework for better performance and reliability. This strategy is crucial for enhancing model accuracy and robustness in complex classification landscapes, where the correct identification of category boundaries is vital for effective decision-making and analysis.

In the domain of sensor-based classification, establishing robust Basic Probability Assignments (BPA) for the perception of implicit targets requires a systematic application of Gaussian membership functions. This method takes into account the inherent variability and potential measurement inaccuracies associated with each target attribute, adhering closely to statistical norms found in Gaussian distributions.

For a practical application, consider a category, such as C_1, which comprises multiple targets each characterized by a series of attributes. The attribute values for these targets are systematically analyzed to formulate a data matrix, with each row representing a target and columns corresponding to attributes. The statistical distribution of each attribute is characterized by calculating the mean and standard deviation from this ensemble of data points, facilitating the modeling of attribute behaviors within the target population.

The Gaussian membership function for each attribute in category C_1 is defined to encapsulate the likelihood of attribute values deviating within three standard deviations from the mean. Mathematically, this is expressed as:

$$\mu_{\theta}^{C_1}(x) = \begin{cases} e^{-\frac{(x-\bar{X}_\theta)^2}{2\sigma_\theta^2}}, for\ x\ within\ [\overline{X_\theta} - 3\sigma_\theta, \overline{X_\theta} + 3\sigma_\theta] \\ 0, outside\ this\ interval \end{cases} \quad (8)$$

This formulation precisely quantifies the fit of a given data point x to the modeled attribute distribution, thereby assessing its categorical alignment effectively. It's crucial for ensuring that the classifications are both precise and reflective of the actual attribute distributions, thus reducing misclassifications.

Similarly, this method extends to other categories, such as C_2 and C_3, where Gaussian membership functions are established for each respective attribute. By analyzing the intersection or overlap of these functions across different categories, one can discern the level of distinctiveness or similarity between categories. Such intersections can range from no overlap, where category functions are distinctly separate, to various degrees of partial overlap, illustrating complex inter-attribute relationships.

In the analysis of Gaussian membership functions for categorical classification, the spatial relationships between the curves can reveal significant insights into the interaction and distinctiveness of category attributes. These relationships can generally be categorized into four distinct scenarios, each representing different levels of overlap and separation among the category-specific functions:

No Intersection: In the first scenario, the Gaussian functions for each category are entirely distinct, with no overlap. This separation indicates clear demarcation between categories, suggesting that each category possesses unique attribute values that are significantly different from the others. This scenario is ideal for classification tasks, as it implies high discriminative power for the attributes in distinguishing between categories.

In an ideal scenario where these Gaussian functions are distinct and do not intersect—as depicted in the hypothetical Fig. 1—selecting a point on each curve allows us to determine the precise probability that a sensor node categorizes a target into C_1,C_2 and C_3 based on a specific value of attribute θ. Such configurations where $\mu_{\theta}^{C_1}(x)$ 、 $\mu_{\theta}^{C_2}(x)$ and $\mu_{\theta}^{C_3}(x)$ are independent, facilitate straightforward predictions with high confidence levels about the category of the detected targets.

Partial Intersection: The second scenario involves two of the Gaussian curves intersecting while the third remains separate. This configuration implies that while two categories share some similarity in attribute distributions, they both remain distinctly different from the third category. Such a setup can be useful in identifying overlapping characteristics between the two intersecting categories and leveraging this information to enhance classification accuracy for more complex scenarios.

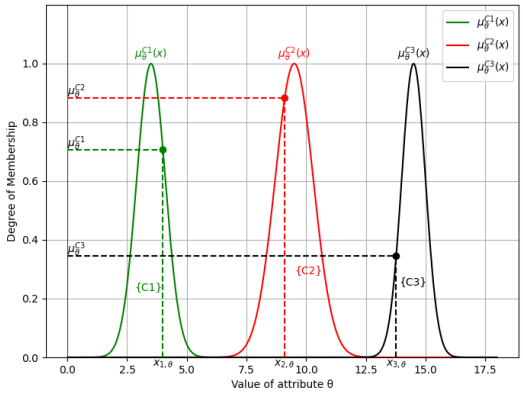


Fig. 1. Scenario of independent gaussian curves.

In contrast, as illustrated in Fig. 2, the Gaussian functions $\mu_\theta^{C_1}(x)$ and $\mu_\theta^{C_2}(x)$ intersect, whereas $\mu_\theta^{C_3}(x)$ remains distinct. The intersection point, noted as $(x_{4,\theta}, \mu_\theta^{C_1,C_2}(x_{4,\theta}))$, marks the area of ambiguity between C_1 and C_2. This area is crucial because it represents the values of θ where the distinction between categories C_1 and C_2 becomes unclear. The upper boundary of this region is defined by the composite subset membership function $\mu_\theta^{C_1,C_2}(x)$, which is the minimum of $\mu_\theta^{C_1,C_2}(x) = min\{\mu_\theta^{C_1}(x), \mu_\theta^{C_2}(x)\}$. This function captures the highest degree of overlap and hence, the maximum uncertainty in classification between these two categories.

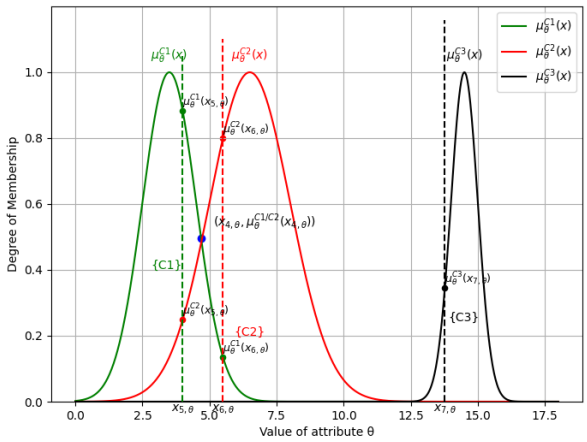


Fig. 2. Two intersecting and one independent GMF curve.

Single Intersection with Others: In the third type, one Gaussian function intersects with each of the other two, but those other two do not intersect with each other. This pattern suggests a central category that shares attributes with the other two categories, which are otherwise distinct from each other. This scenario can be particularly challenging for classification, as it requires careful analysis to ensure accurate category determination.

In Fig. 3, the Gaussian membership functions for the categories C_1,C_2 and C_3 are depicted with specific interactions. The membership functions for C_1 and C_2, $\mu_\theta^{C_1}(x)$ and $\mu_\theta^{C_2}(x)$, intersect at a point defined as $(x_{8,\theta}, \mu_\theta^{C_1,C_2}(x_{8,\theta}))$. Simultaneously, the functions for C_2 and C_3, $\mu_\theta^{C_2}(x)$ and $\mu_\theta^{C_3}(x)$, intersect at $(x_{9,\theta}, \mu_\theta^{C_2,C_3}(x_{9,\theta}))$. The upper limits of these intersections are defined by the composite subset membership functions $\mu_\theta^{C_1,C_2}(x)$ and $\mu_\theta^{C_2,C_3}(x)$, calculated as the minimum values between the respective intersecting functions. This represents a complex but realistic scenario where categories C_1 and C_2 share some common attributes as do categories C_2 and C_3, but C_1 and C_3 remain distinct in this configuration.

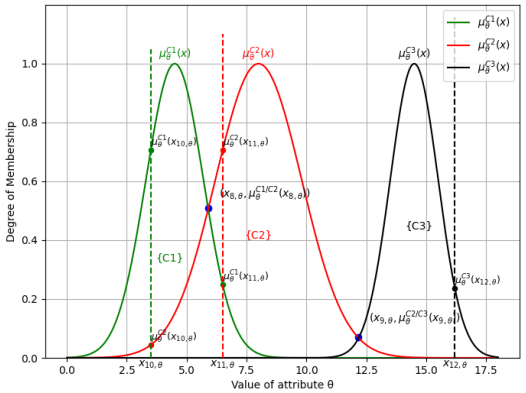


Fig. 3. Two intersecting and separately independent GMF curves.

Mutual Intersection: The final scenario depicts each Gaussian curve intersecting with the other two, indicating a high level of attribute overlap among all three categories. This extensive overlap can lead to higher classification ambiguity and may necessitate more sophisticated analytical techniques or additional data to effectively resolve category assignments.

Fig. 4 illustrates a scenario where all three categories intersect pairwise. The Gaussian curves $\mu_\theta^{C_1}(x)$, $\mu_\theta^{C_2}(x)$ and $\mu_\theta^{C_3}(x)$) each intersect with one another, yielding intersection points at $(x_{13,\theta}, \mu_\theta^{C_1,C_2}(x_{13,\theta}))$、$(x_{14,\theta}, \mu_\theta^{C_2,C_3}(x_{14,\theta}))$ and $(x_{15,\theta}, \mu_\theta^{C_1,C_3}(x_{15,\theta}))$. These points delineate the regions where distinguishing between any two categories becomes challenging due to shared attribute values. The corresponding upper bounds are defined by the composite membership functions $\mu_\theta^{C_1}(x)$,$\mu_\theta^{C_2}(x)$ and $\mu_\theta^{C_3}(x)$), each calculated as the minimum of the intersecting Gaussian functions. This scenario is indicative of a highly intertwined attribute space where each category shares significant overlaps with the others, complicating the classification tasks but also providing rich data for analysis.

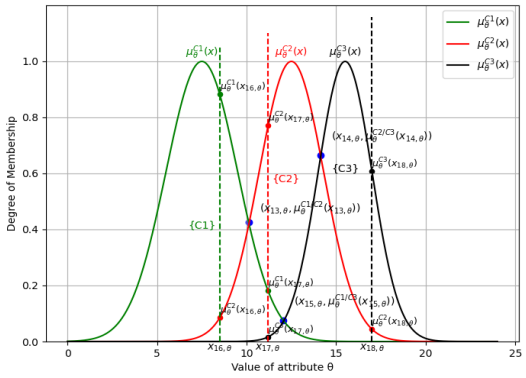


Fig. 4. Pairwise intersecting GMF curves.

E. Interference Effect BPA Generation

In the context of sensor networks, understanding the influence of interference from various sources on sensor nodes is critical, especially how it affects the perception of specific attributes and the subsequent Basic Probability Assignments (BPA) used for decision-making. Interference does not cause sensor failure but introduces biases in the sensors' measurements of attributes, thereby altering the BPA calculations. Here's how to model and calculate the effect of interference on BPA:

Initial BPA Calculation without Interference: First, a sensor S_i measures an attribute θ of a target without any interference. Using the methodology described earlier, the sensor's BPA for the attribute, denoted as m_θ^i, is calculated, representing the sensor's unaltered perception.

BPA Calculation With Interference: The same sensor S_i then measures attribute θ in the presence of a specific interference source G_g. The interference-altered BPA, $m_{\theta,g}^i$, is calculated using the same method as before but adjusted to reflect the combined effect of the original sensor data and the interference. This is done using a specialized operator $\oplus_Z$, where $m_{\theta,g}^i = m_\theta^i \oplus_Z u_{\theta,g}$. This operator blends the original perception effect with the interference effect to produce a new, integrated BPA.

Equation Formulation for Combined BPA: With $m_{\theta,g}^i$ and m_θ^i already derived from the previous steps, the combined BPA equation can be constructed using the operator $\oplus_Z$, finalizing the calculation of the interference-influenced BPA.

Calculating BPA for Other Interference Types: If other interference sources affect sensors designed for different attributes, the above steps (1), (2), and (3) are repeated for each new interference source to calculate its specific impact on BPA.

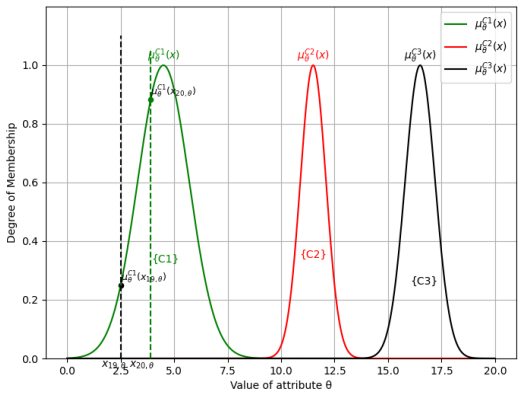


Fig. 5. Membership with interference.

For instance, consider a scenario depicted in a hypothetical Fig. 5, where the original measurement of attribute θ for target 19 is (x_{19}, θ). Under interference, this measurement changes to (x_{20}, θ). Consequently, the BPA generated by the affected sensor node will shift from the original (non-interfered) BPA to a new BPA that is a composite of the original and the interference effects.

In the context of sensor networks, the challenge of accurately interpreting sensor data is compounded when external interference affects the sensor's operation. The concept of Basic Probability Assignment (BPA) is pivotal in quantifying the degree of belief in each possible classification of a target based on sensor data. Consider the scenario where a sensor, without any interference, produces a BPA denoted by mmm. For example, the probabilities that the target belongs to categories C_1, C_2 and C_3 might be given as $m(C_1) = p_1$, $m(C_2) = p_2$, $m(C_3) = p_3$, $m(\{C_1, C_2\}) = p_4$, respectively, with probabilities for combined categories defined similarly.

However, when an interference source affects the sensor, the perception results are altered, introducing deviations in the measured attribute values. The interference effect is itself modeled as a BPA, denoted by g, with its own set of probabilities such as $g(C_1) = p_8$, $g(C_2) = p_9$, $g(C_3) = p_{10}$, $g(\{C_1, C_2\}) = p_{11}$, and so forth, reflecting the impact of the interference on the sensor's ability to classify targets correctly.

The combined effect of the original sensor data and the interference is then calculated using a specialized operator $\oplus_Z$, known as the Dempster combination rule. This rule is employed to integrate the original BPA m and the interference BPA g, resulting in a modified BPA $m^\star$. The probabilities in $m^\star$ are recalculated to reflect this integration, providing new insights into the likely classifications in the presence of interference. For instance, the revised probability that the target belongs to category C_1 in the presence of interference would be updated to $m^\star(C_1) = p_1^\star$, and similarly for the other categories and combinations thereof.

In the context of dealing with sensor interference within wireless networks, the probabilities associated with the Basic Probability Assignment (BPA) for both the non-interfered sensor data and the interference-adjusted data can be methodically calculated using Gaussian membership functions. This calculation treats the probabilities $p_1, p_2, p_3 \cdots\cdots p_7$ associated with the original, undisturbed sensor readings as known quantities. These probabilities reflect the sensor's belief in the target's classification into respective categories without the presence of any distortion.

Similarly, the probabilities $p_1^\star, p_2^\star, p_3^\star \cdots\cdots p_7^\star$ for the interference-affected classifications are derived directly from these Gaussian functions. By applying these well-defined statistical methods, we treat these values as known, calculated based on the sensor's data under the influence of interference.

For the unknown quantities, specifically the interference effect BPA components $p_8, p_9, p_{10} \cdots\cdots p_{14}$, they are not directly observable but can be computed through an established formula that considers the nature of the interference and its impact on the sensor's perception capabilities, refer to the appendix A for detailed computational procedures. This involves leveraging the relationships defined by the Dempster-Shafer theory of evidence, which provides a systematic approach to combine different pieces of evidence, in this case, the original BPA and the interference-induced alterations.

III. S-I PERIMETER COVERAGE ALGORITHM

A. Operational Rules

Sensor Coverage: If a point on the perimeter of sensor node S_i falls within the sensing range of another sensor node S_j, then that point is considered to be covered by S_j.

Segment Coverage: If an entire segment of S_i's perimeter is covered by other sensor nodes excluding S_i itself, it is classified based on the number of covering nodes. For instance, a segment covered by k other nodes is denoted as a k-segment perimeter coverage.

Interference Coverage: If the entire perimeter of a sensor node S_i is within the coverage radius of an interference source I_j, which is the sum of the radius of I_jj and S_i, it is considered to be under the interference perimeter coverage by I_j.

B. Algorithmic Steps

1. Identify Covered Segments: For each sensor node S_j, calculate the segments of other nearby sensor nodes S_i that fall within a double radius distance (2r). These segments are represented by angular intervals $[\alpha_i, L, \alpha_i, R]$.

2. Construct and Sort Points: For all neighboring nodes within a distance less than 2r from S_i, place the points α_i, L and α_i, R on the circular boundary [0, 2π]. These points are sorted in a list L and marked as either the start or end of a covered segment.

3. Determine Coverage Frequency: Traverse the circular boundary [0, 2π] using the sorted list L, from left to right, to determine the coverage status of S_j and count the number of times each segment is covered by other sensor nodes, denoted as $N_{S_{seg}}$.

4. Check Interference Influence: For each interference source I_j, check if S_j is within the interference coverage. If so, place corresponding points α_i, L and α_i, R on [0, 2π].

5. Calculate Interference Coverage Count: Using the sorted list L, determine the number of times each segment is covered by interference sources, represented as $N_{I_{seg}}$.

MSR defined by the arbitrary segmentation of the perimeter of a sensing node S_i are referred to as the MSRs associated with node S_i. For instance, as depicted in Fig. 6, the specific associations between nodes and MSRs are as follows:

MSR 2 and MSR 3 are associated with S_1; MSR 4 and MSR 5 are associated withS_2; MSR 6 and MSR 7、MSR 8 and MSR 9、 MSR 10 and MSR 11are associated with S_3; MSR 9 and MSR 10、MSR 8 and MSR 11are associated with S_4; MSR 6 and MSR 9、MSR 7 and MSR 8、MSR 12 and MSR 13 are associated with S_5; MSR 14 and MSR 15 are associated with S_6; MSR 15 and MSR 16 are associated with S_7; MSR 13 and MSR 14 are associated withS_8.

This structured association allows for a detailed analysis of how each segment of a sensing node's perimeter interacts with the coverage provided by other nodes, facilitating the calculation of coverage frequencies and the influence of interference sources on the network's overall sensing reliability.

MSR1 and MSR2 are within the sensing range ofS_1, MSR2, MSR3, MSR4, MSR21, and MSR22 are within the sensing range of S_2, and MSR4, MSR5, MSR6, MSR9, MSR10, MSR19, MSR20, MSR21, MSR23, and MSR26 fall within the sensing range of S_3. MSR6, MSR7, MSR8, MSR9, MSR25, and MSR26 are covered by S_4, while MSR8, MSR9, MSR10, MSR11, MSR12, MSR18, and MSR19 are within the sensing range of S_5. Similarly, MSRs covered by S_6 to S_{10} can be determined based on their sensing ranges.

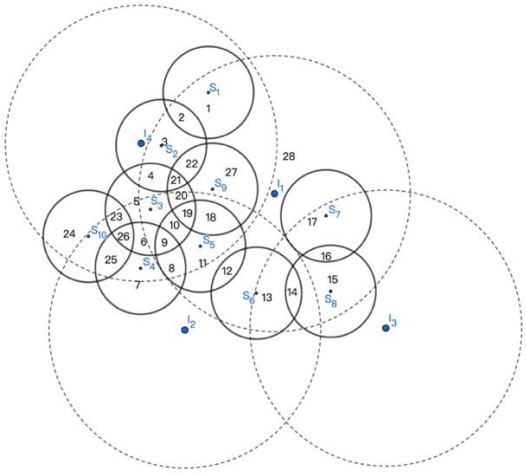


Fig. 6. Perimeter coverage considering homogeneous interference.

The number of times an MSR MSR_p is covered by other sensing nodes is denoted as Q_s, which can be calculated using the function MSR_p $(N_{S_{seg}})$. Similarly, the number of times MSR_q is covered by interference nodes is denoted as G_i, which can be calculated using the function MSR_q $(N_{I_{seg}})$. The function details are as Appendix B.

C. Model Example

After obtaining the sensing coverage count Q_s and the interference coverage count G_i through these functions, the confidence level of the corresponding MSR can be calculated. If a target T_1 enters MSR r, the BPA of the sensing result for T_1 by the sensing node is m, and the interference effect BPA by the interference source is g. Considering the interference effect, the final sensing result in MSR r for the target T_1 is $M_i = (\oplus m)^{Q_s} \oplus (\oplus g)^{G_i}$, where $(\oplus m)^{Q_s}$ represents the combination of Q_s BPAs, such as $(\oplus m)^5$= m ⊕ m⊕ m⊕ m⊕ m. The belief degree bel(MSR_i) of MSR r can then be obtained. According to the reliability metric formula $D^{\delta}_{C_\xi}$, we can compute $D^{\delta}_{C_1}$、$D^{\delta}_{C_2}$、$D^{\delta}_{C_3}$、$D^{\delta}_{\{C_1,C_2\}}$、$D^{\delta}_{\{C_1,C_3\}}$、$D^{\delta}_{\{C_2,C_3\}}$ and $D^{\delta}_{\{C_1,C_2,C_3\}}$.

Based on this analysis, an algorithm to evaluate the reliability of WSN coverage considering the interference is proposed:

For any MSR in the ROI:

Consider the final sensing result $M_i = (\oplus m)^{Q_s} \oplus (\oplus g)^{G_i}$

Compute $D^{\delta}_{C_1}$、 $D^{\delta}_{C_2}$、 $D^{\delta}_{C_3}$、 $D^{\delta}_{\{C_1,C_2\}}$、 $D^{\delta}_{\{C_1,C_3\}}$、 $D^{\delta}_{\{C_2,C_3\}}$ and $D^{\delta}_{\{C_1,C_2,C_3\}}$

To further illustrate the process of this algorithm, we take the sensing node S_3 in Fig. 6 as an example. Besides being perimeter-covered by other sensing nodes S_2、 S_4、 S_5、 S_9, and S_{10}, node S_3 is also influenced by interference sources I_1、 I_2, and I_4. The coverage of various segments of node S_3's perimeter by other sensing nodes is illustrated in Fig. 7.

In a detailed examination of the sensing perimeter associated with sensor node S_3, we observe that interference sources I_1、 I_2, and I_4 impact its operational efficacy. As demonstrated in Fig. 7, the sequence list provides insights into the perimeter segments of S_3 that are influenced by these interference sources. Table I (refer to Appendix C) summarizes the count of coverage occurrences for each segment of S_3's perimeter, reflecting the interplay between sensory and interference coverages.

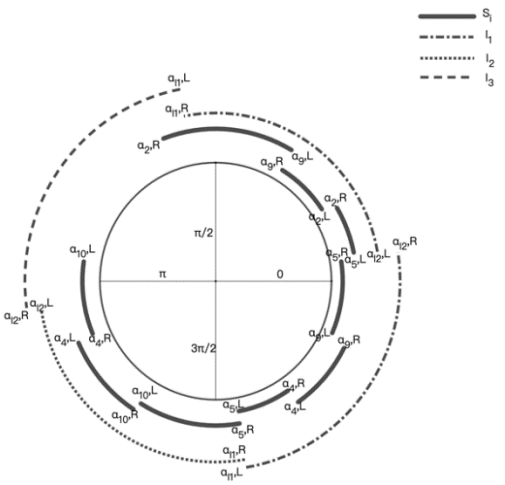


Fig. 7. Perimeter coverage sorted list of S_3.

In the detailed examination of the coverage dynamics within a sensor network, specific attention is directed toward Sensor Node S_3. This node serves as a focal point for evaluating the MSRs related to its periphery, delineated in accordance with the established network protocols and environmental interactions. The process involves a meticulous traversal and assessment of all related MSRs, denoted as MSR_p, to ascertain their sensory coverage status, which is captured in the ensuing coverage information table.

Additionally, the examination extends to the MSRs associated with Sensor Node S_2, including MSR 1 and MSR 2. These regions, lying outside the sensory radius of S_2, experience coverage counts of 1 and 2 respectively, reflecting varying levels of sensor influence. Furthermore, an expansive list of MSRs associated with Sensor Node S_5 includes regions from MSR 5 to MSR 28. Among these, MSR 12, located within S_5's sensory radius, exhibits a coverage count of 2, signifying robust sensor activity, whereas MSR 13, outside this radius, is covered once, indicating lesser sensory influence.

The remaining MSRs, specifically MSR 14 to MSR 17, relate to Sensor Node S_6. MSR 14, found within the sensory radius, is covered twice, affirming its significant sensory engagement. In contrast, MSR 15, outside the sensory radius, demonstrates a reduced coverage count of 1. Similar patterns are observed with MSR 16 related to Sensor Node S_8, covered twice within the sensory radius, and MSR 17, with a coverage count of 1 outside the radius.

Post compilation of sensory coverage counts for all pertinent MSRs, the interference coverage counts, denoted G_i, for these regions are calculated using the function MSR_q. This leads to a nuanced understanding of both the sensory and interference dynamics impacting each MSR.

Upon obtaining the sensory (Q_s) and interference (G_i) coverage counts for the smallest sensing regions, the ultimate sensory results considering interference effects are derived through the formula $M_i = (\bigoplus m)^{Q_s} \bigoplus (\bigoplus g)^{G_i}$. Subsequently, the reliability indices $D^{\delta}_{C_\xi}$ for various configurations are computed, providing a quantitative measure of the network's coverage reliability across diverse environmental and operational scenarios.

IV. Model Simulation

In the ongoing research to enhance the comprehension and reliability of belief coverage in sensory networks, the integration of Monte Carlo simulation offers a robust methodology to analyze the impact of varying parameters on system robustness. This advanced approach not only facilitates a detailed assessment across multiple scenarios but also augments the analytical capabilities concerning sensory and interference node deployments within a defined Region of Interest (ROI).

In Fig. 8, thirteen sensory nodes and four interference nodes are randomly positioned within a 100 by 100 ROI. The sensory range, depicted by blue circles, extends a radius of 10 units, while the interference range, illustrated with red dashed circles, extends a radius of 15 units. Using this setup as a basis, we evaluate the reliability metrics $D^{\delta}_{C_1}$、 $D^{\delta}_{C_2}$、 $D^{\delta}_{C_3}$、 $D^{\delta}_{\{C_1,C_2\}}$、 $D^{\delta}_{\{C_1,C_3\}}$、 $D^{\delta}_{\{C_2,C_3\}}$ and $D^{\delta}_{\{C_1,C_2,C_3\}}$. These assessments employ a Monte Carlo method to explore how various parameters influence system reliability, increasing the resolution of our point matrix to 500 by 500 to enhance simulation accuracy. The Gaussian Membership Functions (GMFs) for categories C_1, C_2, and C_3 are defined with respective means and standard deviations of 7.5 and 2, 12.5 and 2.5, and 15.5 and 2.

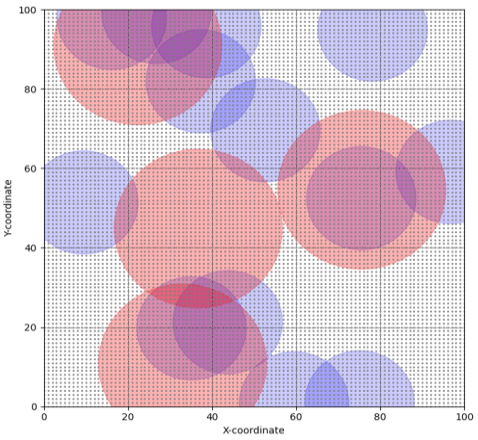


Fig. 8. Random deployment of sensor nodes and interference nodes.

These functions are graphically represented in Fig. 9. Notable intersections of these functions occur at (9.71, 0.54) between $\mu^{C_1}(x)$ and $\mu^{C_2}(x)$, and at (11.48, 0.14) between $\mu^{C_1}(x)$ and $\mu^{C_3}(x)$, and at (14.15, 0.8) between $\mu^2(x)$ and $\mu^3(x)$. Fig. 10 and Fig. 11 illustrate the computed reliability indices. Setting the confidence threshold δ at 0 and the interference factor I at 1, where I=1 implies no consideration of interference effects, the comprehensive ROI coverage rate is 0.9739.

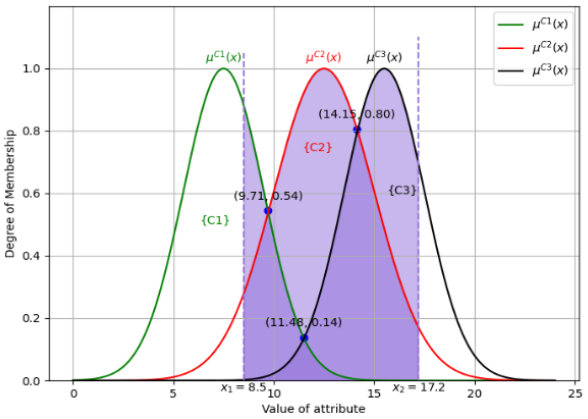


Fig. 9. Gaussian membership function simulation.

Assuming classification is based solely on a single attribute that varies from 8.5 to 17.2, For values of x ranging from 0 to 4.85, $D^0_{C_1}$ dominates at 0.9739, with other classifications scoring zero, indicating that targets within the full sensory coverage are classified exclusively as C_1. At critical values of x such as 9.71 and 14.15, peak values are observed for $D^0_{\{C_1,C_2\}}$ and $D^0_{\{C_2,C_3\}}$, respectively.

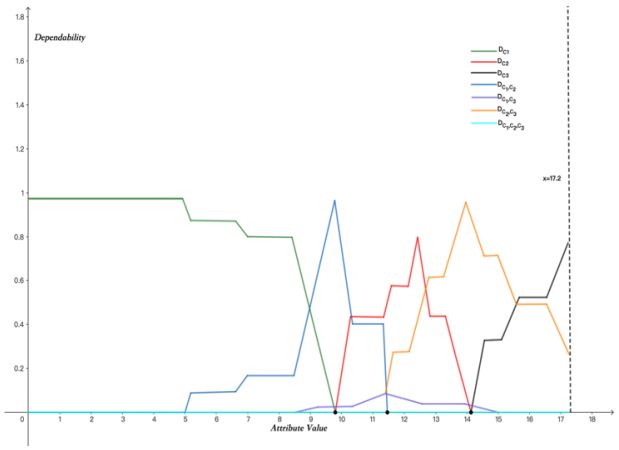


Fig. 10. Reliability trend (**I**=1).

As x increases, the category shifts: from 4.85 to 9.71, the proportion of regions classifying the target as C_1 decreases, while $D^0_{\{C_1,C_2\}}$ grows, indicating ambiguity in classification between C_1 and C_2. Between 9.71 and 12.5, $D^0_{C_2}$ increases directly with x, peaking at x=12.5. Beyond 14.15, more regions start classifying the target as C_3, although some areas remain uncertain between C_2 and C_3, until a definitive classification as C_3 becomes predominant as x approaches 17.2.

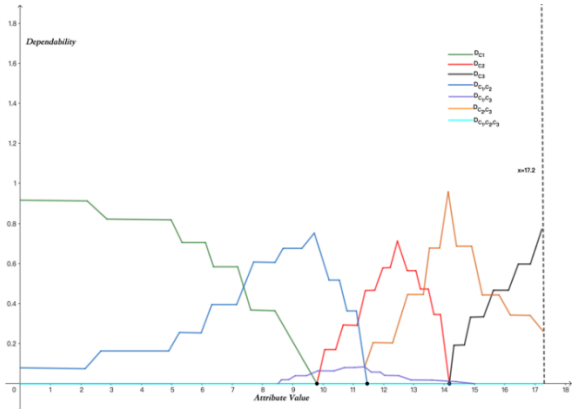


Fig. 11. Reliability trend (**I**=0.7).

In Fig. 11, with δ set at 0 and I at 0.7, significant shifts occur in the reliability metrics compared to when I=1, underscoring how interference significantly impacts sensory coverage reliability. Despite these variations, the overall coverage rate remains 0.9739, affirming robust system performance under varied conditions. As x ranges from 0 to 7.25, most of the ROI classifies targets as C_1; between 11.48 and 12.67, C_2 becomes more likely; and from 15.53 to 17.2, the classification increasingly favors C_3.

Setting the confidence threshold δ to zero allows for a more detailed observation of the reliability metrics for different classifications: $D^\delta_{C_1}$、$D^\delta_{C_2}$、$D^\delta_{C_3}$、$D^\delta_{\{C_1,C_2\}}$、$D^\delta_{\{C_1,C_3\}}$、$D^\delta_{\{C_2,C_3\}}$ and $D^\delta_{\{C_1,C_2,C_3\}}$, with attribute values ranging from 0 to 17.2 and interference factors between 0.7 and 1. The overall trends for these functions are illustrated in Fig. 12 to Fig. 17.

In Fig. 12, within the attribute value range of 9.71 to 17.2, changes in the interference factor do not affect the reliability outcome $D^\delta_{C_1}$. However, for values from 0 to 9.71, as the interference factor decreases, the reliability for category C_1 similarly declines, and the lower the interference factor, the greater the reduction in reliability, indicating that the disruptive effects of interference sources are particularly significant within specific attribute value ranges.

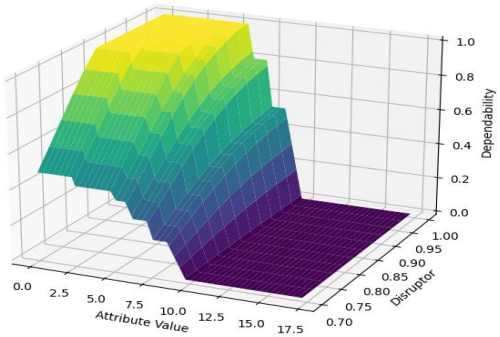


Fig. 12. Overall trend of the function $D^\delta_{C_1}$.

This figure clearly demonstrates the relationship between external intrusion attribute values and the reliability of the sensing coverage system. Across the higher attribute value range of 9.71 to 17.2, regardless of changes in the interference factor,

the system's sensing classification reliability remains stable, indicating that the sensor network's monitoring of attribute values has become stable and robust. When attribute values fall below 9.71, the reliability of the coverage system significantly decreases with increasing interference factor; thus, when designing sensing coverage systems, particular attention needs to be paid to the effects of interference in scenarios with low attribute values, as smaller interference factors mean greater disruption from external interference sources, leading to faster declines in system reliability. This may be due to the low intensity of the target's relevant attributes under these conditions, making them more susceptible to environmental noise and electromagnetic interference, thus reducing the sensing nodes' detection accuracy.

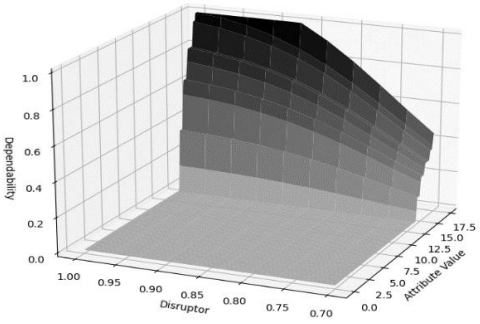


Fig. 13. Overall trend of the function $\boldsymbol{D^{\delta}_{C_2}}$.

In Fig. 13, for attribute values within the ranges of 0 to 9.71 and 14.15 to 17.2, there are no MSRs that classify the sensing results as category C_2, indicating that the attribute values of the intrusion targets are either too low to elicit an adequate system response or too high, exceeding the optimal operational range of the sensing coverage system. From 9.71 to 14.15, the reliability results of sensing classification for $D^{\delta}_{C_2}$ initially increase and then decrease, particularly as the attribute value reaches 12.5, where the characteristics of the intrusion target highly align with the features of category C_2, resulting in maximum reliability for this classification—a peak symbolizing the sensing coverage system's highest confidence level in assigning targets to category C_2. Thereafter, the degree of membership for classifying targets as C_2 gradually diminishes until it reaches zero, with the interference factor I and the reliability of classification $D^{\delta}_{C_2}$ being directly correlated—the larger the interference factor, the higher the classification reliability.

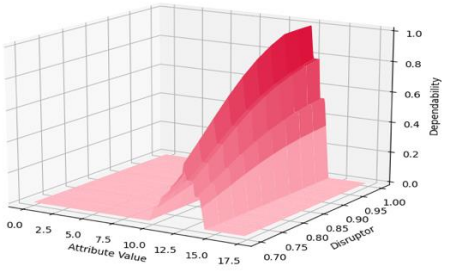


Fig. 14. Overall trend of the function $\boldsymbol{D^{\delta}_{C_3}}$.

In Fig. 14, within the range of 0 to 14.15, the reliability outcome of sensing classification $D^{\delta}_{C_3}$ does not fluctuate with changes in the interference factor, but from 14.15 to 17.2, the degree of membership for classifying targets as C_3, $D^{\delta}_{C_3}$, steadily increases with rising attribute values. Similarly, $D^{\delta}_{C_3}$ shows a positive correlation with the interference factor I, with an increase in the factor enhancing the classification result's reliability, which, in turn, results in a decrease in $D^{\delta}_{\{C_2,C_3\}}$ as shown in Fig. 17.

Fig. 15 illustrates that from 0 to 8.5, the changes in $D^{\delta}_{C_1}$ and $D^{\delta}_{\{C_1,C_2\}}$ display a symmetrical trend due to the intensified effect of interference, causing uncertainty in the sensing classification results right from the start ($D^{\delta}_{\{C_1,C_2\}} \neq 0$). From 8.5 to 9.71, the number of MSRs unable to determine the category of the intrusion target gradually increases, and from 9.71 to 11.48, the reliability of classification results $D^{\delta}_{\{C_1,C_2\}}$ consistently decreases, due to the attribute values increasing the affiliation of the intrusion targets to category C_2more frequently.

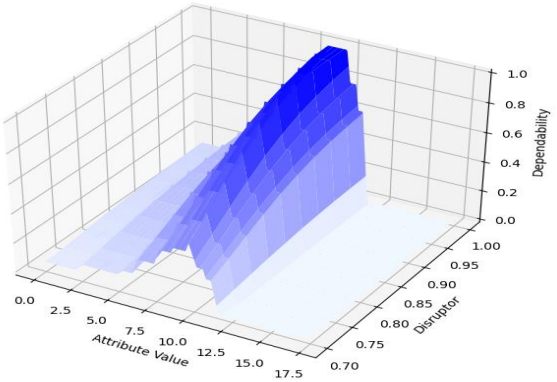


Fig. 15. Overall trend of the function $\boldsymbol{D^{\delta}_{\{C_1,C_2\}}}$.

Fig. 16 indicates that within the range of 8.5 to 15, there are portions of the sensing coverage area unable to correctly classify intrusion targets as either C_1 or C_3, reaching a peak number of problematic MSRs at 11.48, although this has limited impact on the overall reliability of classification results, indirectly showing that these parts of the sensing coverage due to negative effects from interference sources have insufficient detection accuracy.

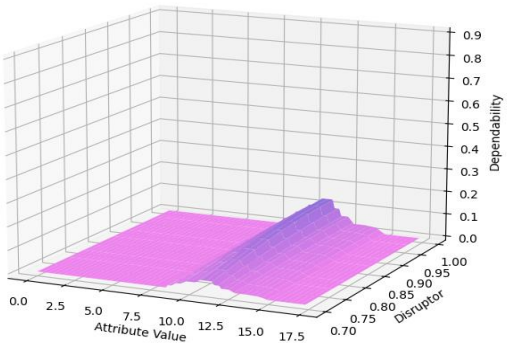


Fig. 16. Overall trend of the function $\boldsymbol{D^{\delta}_{\{C_1,C_3\}}}$.

Fig. 17 shows that the interference factor impacts the reliability of sensing coverage $D^{\delta}_{\{C_2,C_3\}}$ only within the specific attribute value range of 11.48 to 17.2, not significantly in all cases, allowing for an assessment of the dynamic nature and sensitivity of the sensing coverage system to external condition changes. From 11.48 to 14.15, as the interference factor increases, the reliability of coverage $D^{\delta}_{\{C_2,C_3\}}$ also gradually improves, with the number of MSRs unable to correctly classify targets within the ROI incrementally increasing. From 14.15 to 17.2, the weaker the interference effect, the stronger the reliability of sensing coverage $D^{\delta}_{\{C_2,C_3\}}$.

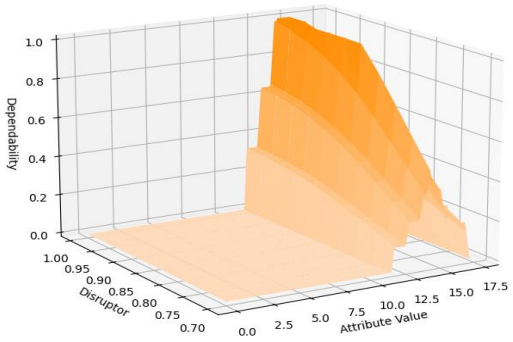


Fig. 17. Overall trend of the function $D^{\delta}_{\{C_2,C_3\}}$.

V. Conclusion

The research presented in this study contributes a novel and robust approach to enhancing the security and reliability of Wireless Sensor Networks (WSNs) by addressing the critical issue of false positives and improving intruder classification accuracy. The use of Dempster-Shafer theory for evidence combination is a key innovation, offering a powerful means to manage the uncertainty inherent in sensor data, particularly in environments where interference and overlapping signals are prevalent. This methodology stands out for its ability to integrate multiple sources of evidence, thereby refining the overall decision-making process and enhancing the reliability of intrusion detection systems.

Despite these strengths, there are several areas where the model could be further refined and extended. One significant limitation of the current approach is its static nature. While the model effectively reduces false positives by distinguishing between different types of intruders, it does so based on predefined Gaussian membership functions and belief structures that may not adapt well to rapidly changing conditions. In dynamic environments, where the nature of threats can evolve quickly, this rigidity could lead to reduced effectiveness over time. Therefore, integrating machine learning algorithms into the framework could be a promising direction for future research. Such integration would enable the model to learn from real-time data, adapt its parameters dynamically, and improve its accuracy in response to changing environmental factors and threat landscapes.

Moreover, while the model has been demonstrated to work effectively in the context of high-security areas, its application to larger and more complex WSNs raises questions about scalability. As the network size increases, the computational demands associated with processing and integrating data from numerous sensors could become significant. This potential bottleneck suggests a need for optimization techniques that can maintain the model's efficiency in larger deployments. For instance, hierarchical or distributed processing methods could be explored to manage the computational load more effectively, ensuring that the system remains responsive and reliable even as the scale of the network grows.

Another important consideration is the model's applicability beyond the immediate context of high-security monitoring. The principles underlying the proposed framework—such as the use of evidence theory and the focus on managing uncertainty—could be valuable in a range of other domains. For example, in wildlife tracking or traffic management, where sensor networks must operate under varying and often unpredictable conditions, the ability to accurately classify and respond to different types of entities is crucial. Expanding the framework to address these broader applications could provide substantial societal benefits, making WSNs more versatile and effective across diverse fields.

In addition to these technical considerations, it is also worth reflecting on the broader implications of this research in the context of WSN development. The growing reliance on sensor networks in critical infrastructure and security applications means that the robustness and reliability of these systems are of paramount importance. By providing a framework that can better manage the inherent uncertainties and complexities of these environments, this research contributes to the advancement of WSN technology as a whole. However, as with any emerging technology, continuous improvement and adaptation are necessary to keep pace with evolving challenges. Future research should not only focus on technical enhancements but also consider the ethical and societal implications of deploying increasingly autonomous and intelligent sensor networks in sensitive areas.

In conclusion, while the current study represents a significant step forward in the development of more reliable and adaptable WSNs, there is ample room for further exploration and refinement. By addressing the limitations of the current model and expanding its applicability, future work can build on this foundation to create even more effective and versatile sensor networks, capable of meeting the demands of a wide range of modern applications.

Acknowledgment

We extend our deepest gratitude to the Shandong Province Social Science Planning Research Project (Grant No. 23CXWJ04), the National Natural Science Foundation of China (Grant No. 61403230), and the Natural Science Foundation of Shandong Province (Grant No. ZR2020MG011) for their generous support. Their contributions were invaluable in the realization of our research. The insights and outcomes presented in this work are a testament to their commitment to advancing knowledge in our field.

References

[1] MOLKA-DANIELSEN J, ENGELSETH P, OLEšNANíKOVá V, et al. Big data analytics for air quality monitoring at a logistics shipping base via autonomous wireless sensor network technologies; proceedings of the

2017 5th international conference on enterprise systems (ES), F, 2017 [C]. IEEE.

[2] SHAKOOR N, NORTHRUP D, MURRAY S, et al. Big data driven agriculture: big data analytics in plant breeding, genomics, and the use of remote sensing technologies to advance crop productivity [J]. The Plant Phenome Journal, 2019, 2(1): 1-8.

[3] KAUSHIK S. Big Medical Data Analytics Using Sensor Technology [M]. Efficient Data Handling for Massive Internet of Medical Things: Healthcare Data Analytics. Springer. 2021: 45-70.

[4] UDDIN M, SYED-ABDUL S. Data analytics and applications of the wearable sensors in healthcare: an overview [J]. Sensors, 2020, 20(5): 1379.

[5] BLAKE R, MICHALIKOVA K F. Deep learning-based sensing technologies, artificial intelligence-based decision-making algorithms, and big geospatial data analytics in cognitive internet of things [J]. Analysis and Metaphysics, 2021, 20: 159-73.

[6] ALEXAKIS T, PEPPES N, DEMESTICHAS K, et al. A distributed big data analytics architecture for vehicle sensor data [J]. Sensors, 2022, 23(1): 357.

[7] BATOOL S, SAQIB N A, KHATTACK M K, et al. Identification of remote IoT users using sensor data analytics; proceedings of the Advances in Information and Communication: Proceedings of the 2019 Future of Information and Communication Conference (FICC), Volume 1, F, 2020 [C]. Springer.

[8] HAILE M A, HAILE D T, ZERIHUN D. Real-time sensor data analytics and visualization in cloud-based systems for forest environment monitoring [J]. International Journal of Advances in Signal and Image Sciences, 2023, 9(1): 29-39.

[9] HARB H, MANSOUR A, NASSER A, et al. A sensor-based data analytics for patient monitoring in connected healthcare applications [J]. IEEE Sensors Journal, 2020, 21(2): 974-84.

[10] LIAO H-J, LIN C-H R, LIN Y-C, et al. Intrusion detection system: A comprehensive review [J]. Journal of Network and Computer Applications, 2013, 36(1): 16-24.

[11] OZKAN-OKAY M, SAMET R, ASLAN Ö, et al. A comprehensive systematic literature review on intrusion detection systems [J]. IEEE Access, 2021, 9: 157727-60.

[12] PANIGRAHI R, BORAH S. A detailed analysis of CICIDS2017 dataset for designing Intrusion Detection Systems [J]. International Journal of Engineering & Technology, 2018, 7(3.24): 479-82.

[13] SHENFIELD A, DAY D, AYESH A. Intelligent intrusion detection systems using artificial neural networks [J]. Ict Express, 2018, 4(2): 95-9.

[14] WU X, HONG D, CHANUSSOT J. Convolutional neural networks for multimodal remote sensing data classification [J]. IEEE Transactions on Geoscience and Remote Sensing, 2021, 60: 1-10.

[15] BUTUN I, MORGERA S D, SANKAR R. A survey of intrusion detection systems in wireless sensor networks [J]. IEEE communications surveys & tutorials, 2013, 16(1): 266-82.

[16] DREWEK-OSSOWICKA A, PIETROŁAJ M, RUMIńSKI J. A survey of neural networks usage for intrusion detection systems [J]. Journal of Ambient Intelligence and Humanized Computing, 2021, 12(1): 497-514.

[17] KABIR E, HU J, WANG H, et al. A novel statistical technique for intrusion detection systems [J]. Future Generation Computer Systems, 2018, 79: 303-18.

[18] KARATAS G, DEMIR O, SAHINGOZ O K. Deep learning in intrusion detection systems; proceedings of the 2018 international congress on big data, deep learning and fighting cyber terrorism (IBIGDELFT), F, 2018 [C]. IEEE.

[19] KHRAISAT A, GONDAL I, VAMPLEW P, et al. Survey of intrusion detection systems: techniques, datasets and challenges [J]. Cybersecurity, 2019, 2(1): 1-22.

[20] LANSKY J, ALI S, MOHAMMADI M, et al. Deep learning-based intrusion detection systems: a systematic review [J]. IEEE Access, 2021, 9: 101574-99.

[21] KHAN B A, SHARIF M, RAZA M, et al. An approach for surveillance using wireless sensor networks (WSN) [J]. Journal of Information & Communication Technology, 2007, 1(2): 35-42.

[22] MOSTAFAEI H, CHOWDHURY M U, OBAIDAT M S. Border surveillance with WSN systems in a distributed manner [J]. IEEE Systems Journal, 2018, 12(4): 3703-12.

[23] LIAO Y, MOLLINEAUX M, HSU R, et al. Snowfort: An open source wireless sensor network for data analytics in infrastructure and environmental monitoring [J]. IEEE Sensors Journal, 2014, 14(12): 4253-63.

[24] TIDJON L N, FRAPPIER M, MAMMAR A. Intrusion detection systems: A cross-domain overview [J]. IEEE Communications Surveys & Tutorials, 2019, 21(4): 3639-81.

[25] BHATI N S, KHARI M, GARCíA-DíAZ V, et al. A review on intrusion detection systems and techniques [J]. International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, 2020, 28(Supp02): 65-91.

[26] CASAS P, MAZEL J, OWEZARSKI P. Unsupervised network intrusion detection systems: Detecting the unknown without knowledge [J]. Computer Communications, 2012, 35(7): 772-83.

[27] LAOUIRA M L, ABDELLI A, OTHMAN J B, et al. An efficient WSN based solution for border surveillance [J]. IEEE Transactions on Sustainable Computing, 2019, 6(1): 54-65.

[28] AL GHAMDI A, ASEERI M, AHMED M R. A novel trust and reputation model based WSN technology to secure border surveillance [J]. International Journal of Future Computer and Communication, 2013, 2(3): 263.

[29] SERT S A, ONUR E, YAZICI A. Security attacks and countermeasures in surveillance wireless sensor networks; proceedings of the 2015 9th International Conference on Application of Information and Communication Technologies (AICT), F, 2015 [C]. IEEE.

[30] VIANI F, OLIVERI G, DONELLI M, et al. WSN-based solutions for security and surveillance; proceedings of the The 40th European Microwave Conference, F, 2010 [C]. IEEE.

[31] YAGER R R, KACPRZYK J, FEDRIZZI M. Advances in the Dempster-Shafer theory of evidence [M]. John Wiley & Sons, Inc., 1994.

[32] YANG B-S, KIM K J. Application of Dempster–Shafer theory in fault diagnosis of induction motors using vibration and current signals [J]. Mechanical Systems and Signal Processing, 2006, 20(2): 403-20.

[33] DENOEUX T. A neural network classifier based on Dempster-Shafer theory [J]. IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans, 2000, 30(2): 131-50.

[34] ZADEH L A. A simple view of the Dempster-Shafer theory of evidence and its implication for the rule of combination [J]. AI magazine, 1986, 7(2): 85-.

[35] KLIR G J, RAMER A. Uncertainty in the Dempster-Shafer theory: a critical re-examination [J]. International Journal of General System, 1990, 18(2): 155-66.

[36] BENELHOURI A, IDRISSI-SABA H, ANTARI J. An evolutionary routing protocol for load balancing and QoS enhancement in IoT enabled heterogeneous WSNs [J]. Simulation Modelling Practice and Theory, 2023, 124: 102729.

[37] HOSSEINZADEH M, YOO J, ALI S, et al. A cluster-based trusted routing method using fire hawk optimizer (FHO) in wireless sensor networks (WSNs) [J]. Scientific Reports, 2023, 13(1): 13046.

[38] JARADAT Y, MASOUD M, JANNOUD I, et al. Analysis of the optimal number of clusters and probability in homogeneous unreliable WSNs [J]. Multimedia Tools and Applications, 2023, 82(25): 39633-52.

[39] LOHAR L, AGRAWAL N K, GUPTA P, et al. A novel approach based on bio - inspired efficient clustering algorithm for large - scale heterogeneous wireless sensor networks [J]. International Journal of Communication Systems, 2023, 36(8): e5472.

[40] WANG N, ZHANG S, ZHANG Z, et al. Lightweight and Secure Data Transmission Scheme Against Malicious Nodes in Heterogeneous Wireless Sensor Networks [J]. IEEE Transactions on Information Forensics and Security, 2023.

APPENDIX A

$$K=m(C_1)\,g(C_2)+m(C_1)\,g(C_3)+m(C_1)\,g(\{C_2,C_3\})+m(C_2)\,g(C_1)+m(C_2)\,g(C_3)+m(C_2)\,g(\{C_1,C_3\})+m(C_3)\,g(C_1)+m(C_3)\,g(C_2)+m(C_3)\,g(\{C_1,C_2\})=p_1p_9+p_1p_{10}+p_1p_{13}+p_2p_8+p_2p_{10}+p_2p_{12}+p_3p_8+p_3p_9+p_3p_{11}$$

$$p_1^{\star}=m^{\star}(C_1)=\frac{1}{1-K}[m(C_1)\,g(C_1)+m(C_1)\,g(\{C_1,C_2\})+m(C_1)\,g(\{C_1,C_3\})+m(C_1)\,g(\{C_1,C_2,C_3\})+\ m(\{C_1,C_2\})g(C_1)+\ m(\{C_1,C_3\})g(C_1)+m(\{C_1,C_2,C_3\})g(C_1)]=\frac{1}{1-K}(p_1p_8+p_1p_{11}+p_1p_{12}+p_1p_{14}+p_4p_8+p_5p_8+p_7p_8)$$

$$p_2^{\star}=m^{\star}(C_2)=\frac{1}{1-K}[m(C_2)\,g(C_2)+m(C_2)\,g(\{C_1,C_2\})+m(C_2)\,g(\{C_2,C_3\})+m(C_2)\,g(\{C_1,C_2,C_3\})+\ m(\{C_1,C_2\})g(C_2)+\ m(\{C_2,C_3\})g(C_2)+m(\{C_1,C_2,C_3\})g(C_2)]=\frac{1}{1-K}(p_2p_9+p_2p_{11}+p_2p_{13}+p_2p_{14}+p_4p_9+p_6p_9+p_7p_9)$$

$$p_3^{\star}=m^{\star}(C_3)=\frac{1}{1-K}[m(C_3)\,g(C_3)+m(C_3)\,g(\{C_1,C_3\})+m(C_3)\,g(\{C_2,C_3\})+m(C_3)\,g(\{C_1,C_2,C_3\})+\ m(\{C_1,C_3\})g(C_3)+\ m(\{C_2,C_3\})g(C_3)+m(\{C_1,C_2,C_3\})g(C_3)]=\frac{1}{1-K}(p_3p_{10}+p_3p_{12}+p_3p_{13}+p_3p_{14}+p_5p_{10}+p_6p_{10}+p_7p_{10})$$

$$p_4^{\star}=m^{\star}(\{C_1,C_2\})=\frac{1}{1-K}[\,m(\{C_1,C_2\})\,g(C_1)+m(\{C_1,C_2\})\,g(C_2)+m(\{C_1,C_2\})\,g(\{C_1,C_2\})+m(\{C_1,C_2\})\,g(\{C_1,C_2,C_3\})+\ m(C_1)g(\{C_1,C_2\})+m(C_2)g(\{C_1,C_2\})+m(\{C_1,C_2,C_3\})g(\{C_1,C_2\})]=\frac{1}{1-K}(p_4p_8+p_4p_9+p_4p_{11}+p_4p_{14}+p_1p_{11}+p_2p_{11}+p_7p_{11})$$

$$p_5^{\star}=m^{\star}(\{C_1,C_3\})=\frac{1}{1-K}[\,m(\{C_1,C_3\})\,g(C_1)+m(\{C_1,C_3\})\,g(C_3)+m(\{C_1,C_3\})\,g(\{C_1,C_3\})+m(\{C_1,C_3\})\,g(\{C_1,C_2,C_3\})+\ m(C_1)g(\{C_1,C_3\})+m(C_3)g(\{C_1,C_3\})+m(\{C_1,C_2,C_3\})g(\{C_1,C_3\})]=\frac{1}{1-K}(p_5p_8+p_5p_{10}+p_5p_{12}+p_5p_{14}+p_1p_{12}+p_3p_{12}+p_7p_{12})$$

$$p_6^{\star}=m^{\star}(\{C_2,C_3\})=\frac{1}{1-K}[\,m(\{C_2,C_3\})\,g(C_2)+m(\{C_2,C_3\})\,g(C_3)+m(\{C_2,C_3\})\,g(\{C_2,C_3\})+m(\{C_2,C_3\})\,g(\{C_1,C_2,C_3\})+\ m(C_2)g(\{C_2,C_3\})+m(C_3)g(\{C_2,C_3\})+m(\{C_1,C_2,C_3\})g(\{C_2,C_3\})]=\frac{1}{1-K}(p_6p_9+p_6p_{10}+p_6p_{13}+p_6p_{14}+p_2p_{13}+p_3p_{13}+p_7p_{13})$$

$$p_7^{\star}=m^{\star}(\{C_1,C_2,C_3\})=\frac{1}{1-K}[\,m(\{C_1,C_2,C_3\})\,g(C_1)+m(\{C_1,C_2,C_3\})\,g(C_2)+m(\{C_1,C_2,C_3\})\,g(C_3)+m(\{C_1,C_2,C_3\})\,g(\{C_1,C_2\})+m(\{C_1,C_2,C_3\})g(\{C_1,C_3\})+\ m(\{C_1,C_2,C_3\})g(\{C_2,C_3\})+m(\{C_1,C_2,C_3\})g(\{C_1,C_2,C_3\})]=\frac{1}{1-K}(p_7p_8+p_7p_9+p_7p_{10}+p_7p_{11}+p_7p_{12}+p_7p_{13}+p_7p_{14})$$

APPENDIX B

Algorithm 1: S-I Perimeter Coverage

Initialize:
S = {S_1,S_2, ..., S_i}: Set of sensor nodes
I = {I_1, I_2, ..., I_j}: Set of interference sources
r_s: Sensing radius of sensor nodes
r_i: Interference radius of interference sources
For each sensor node S_i in S do
Initialize Coverage Status [S_i] to 0
Covered Segments is initially empty
For each sensor node S_j in S, j ≠ i do
If distance (S_i, S_j) ≤ 2 * r_s then
Calculate the segment of Si covered by S_j
Add this segment to Covered Segments
End If
End For
Sort the segments in Covered Segments
Calculate the coverage frequency for S_i based on the sorted segments
End For
For each sensor node S_i in S do
Interference Segments is initially empty
For each interference source I_j in I do
If distance (S_i, I_j) ≤ r_s + r_i then
Calculate the interference segment on S_i due to I_j
Add this segment to Interference Segments
End If
End For
Combine the interference segments with the original covered segments
Recalculate the final coverage status for S_i considering the interference
End for
Return:
return the final coverage status of all sensor nodes

Algorithm 2: Calculate MSR_p Coverage Count

Initialize:
Let S_i be the sensor node under consideration.
Let MSR_p be a segment related to the boundary of node S_i.
Compute:
For each MSR_p associated with node S_i:
If MSR_p is on the inner side of the boundary:
$Q_s = N_{S_{seg}} + 1$ // $N_{S_{seg}}$ is the count of sensor nodes covering the segment S_{seg}.
Else if MSR_p is on the outer side of the boundary:
$Q_s = N_{S_{seg}}$.
End If.
End For.
Return:
Q_s // Return the coverage count for the segment MSR_p.

Algorithm 3: Calculate MSR_q Interference Count

Initialize:
Let S_i be the sensor node under consideration.
Let MSR_q be a segment related to the boundary of node S_i.
Compute:
For each MSR_q associated with node S_i:
If MSR_q is on the inner side of the boundary:
$G_i = N_{I_{seg}} + 1$ // $N_{I_{seg}}$ is the count of sensor nodes covering the segment I_{seg}
Else if MSR_p is on the outer side of the boundary:
$G_i = N_{I_{seg}}$.
End If
End For.
Return:
G_i // Return the coverage count for the segment MSR_p.

APPENDIX C

TABLE I. COVERAGE COUNT STATISTICS FOR S_3 PERIMETER

Perimeter Segment	Sensory Node Coverage Count	Interference Source Coverage Count
$[\alpha_{5,L}, \alpha_{2,R}]$	1	3
$[\alpha_{2,L}, \alpha_{9,R}]$	2	2
$[\alpha_{9,L}, \alpha_{2,R}]$	1	2
$[\alpha_{2,L}, \alpha_{10,R}]$	0	1
$[\alpha_{10,L}, \alpha_{4,R}]$	1	1
$[\alpha_{4,L}, \alpha_{10,R}]$	2	2
$[\alpha_{10,L}, \alpha_{5,R}]$	1	2
$[\alpha_{5,L}, \alpha_{4,R}]$	2	3
$[\alpha_{4,L}, \alpha_{9,R}]$	1	3
$[\alpha_{9,L}, \alpha_{5,R}]$	2	3

Lightweight and Efficient High-Resolution Network for Human Pose Estimation

Jiarui Liu[1], Xiugang Gong[2]*, Qun Guo[3]
School of Computer Science and Technology, Shandong University of Technology, Zibo, Shandong, 255000, China

***Abstract*—To address the challenges of high parameter quantities and elevated computational demands in high-resolution network, which limit their application on devices with constrained computational resources, we propose a lightweight and efficient high-resolution network, LE-HRNet. Firstly, we designs a lightweight module, LEblock, to extract feature information. LEblock leverages the Ghost module to substantially decrease the number of model parameters. Based on this, to effectively recognize human keypoints, we designed a Multi-Scale Coordinate Attention Mechanism (MCAM). MCAM enhances the model's perception of details and contextual information by integrating multi-scale features and coordinate information, improving the detection capability for human keypoints. Additionally, we designs a Cross-Resolution Multi-Scale Feature Fusion Module (CMFFM). By optimizing the upsampling and downsampling processes, CMFFM further reduces the number of model parameters while enhancing the extraction of cross-branch channel features and spatial features to ensure the model's performance. The proposed model's experimental results demonstrate accuracies of 69.3% on the COCO dataset and 88.7% on the MPII dataset, with a parameter count of only 5.4M, substantially decreasing the number of model parameters while preserving its performance.**

***Keywords*—Human pose estimation; model lightweighting; Ghost module; attention mechanism; multi-scale feature fusion**

I. INTRODUCTION

Human pose estimation, as a core topic in the field of computer vision, aims to recognize and locate keypoints of the human body from images or videos. The key to this task lies in accurately understanding and analyzing human posture and movement, which is crucial for computer vision to comprehend and process complex scenes. Human pose estimation plays an important role in numerous application areas, such as sports analysis, human-computer interaction, and security monitoring [1] [2].

The research on human pose estimation has evolved from early model-based and traditional learning algorithm-based methods, such as graphical models and handcrafted feature extraction [3], to recent deep learning-based methods. Deep learning methods, particularly Convolutional Neural Network (CNN) [4], have significantly improved the accuracy and robustness of techniques for recognizing and locating key-points of the body. This heatmap-based approach effectively handles complex scenes and multi-person pose estimation tasks. Since heatmaps can intuitively represent the positional probability of each keypoint, the model can accurately recognize key points even in cases of partial occlusion or overlap of the human body.

In recent years, numerous classic human pose estimation algorithms have emerged [5][10], achieving significant advancements in recognizing and locating human keypoints in images or videos. Particularly, High-resolution network (HRNet) [11], with their unique network structure and high-resolution feature representation capabilities, can achieve effective human pose estimation while maintaining high accuracy, making them widely applicable in various scenarios. However, due to their complex network structure and large number of parameters and high computational demands, high-resolution networks face difficulties when deployed on resource-constrained devices. Lite-HRNet [12] effectively reduces the model's parameter count by incorporating a Conditional Channel Weighting module. Dite-HRNet [13] introduces dynamic lightweight processing, multi-scale context information extraction, and long-range spatial dependency modeling in high-resolution networks, ensuring model performance with lower parameters. X-HRNet [14] incorporates Spatially Unidimensional Self-Attention (SUSA) for lightweight processing, significantly reducing model parameters without compromising accuracy. These methods have made significant progress in model lightweighting. However, human pose estimation is a task highly sensitive to positional information, and lightweighting high-resolution networks can lead to the loss of critical human keypoint positional information. During multi-scale feature fusion, frequent upsampling and downsampling operations introduce a computational burden. Furthermore, downsampling reduces the spatial detail in feature maps, which is difficult to recover during upsampling.

In response to the issues mentioned above, we propose a lightweight and efficient high-resolution network, LE-HRNet. We utilize the Ghost module to reduce the model's parameter count and introduce a novel attention mechanism in LE-HRNet to enhance the detection of keypoint positional information. This approach ensures model performance while lowering both the parameter count and computational load. Additionally, we optimize the multi-scale feature fusion stage to further decrease computational demands and enhance the extraction of channel and spatial dimensional feature information. The main contributions of this paper are summarized as follows:

- We designed a lightweight module, LEblock, for extracting feature information. We used the Ghost module instead of standard convolution to reduce the model's parameter count, and designed a Multi-Scale Coordinate Attention Mechanism to enhance the

*Corresponding Author.

detection capability of human key points, ensuring the model's performance.

- We optimized the multi-scale feature fusion stage and proposed a Cross-Resolution Multi-Scale Feature Fusion Module. This module optimizes the upsampling and downsampling processes, and by learning cross-branch channel information and spatial features, it ensures the model's performance while further reducing the model's parameter count.
- We conducted experimental validation on the COCO dataset and MPII dataset to demonstrate the effectiveness of the proposed method.

The structure of this paper is organized as follows: Section II introduces the main methods proposed in this paper. Section III conducts experimental verification on the COCO and MPII datasets and analyzes the experimental results. Section E summarize the research results and discuss the future work of LE-HRNet.

II. PROPOSED METHOD

HRNet is widely used for visual tasks that require detailed features, such as human pose estimation and semantic segmentation. Unlike most existing methods that recover high-resolution features from low-resolution features, HRNet connects high-resolution to low-resolution subnets in parallel, maintaining high-resolution feature representation throughout the entire network. It extracts feature information using residual block [15] and achieves multi-scale information exchange through multi-scale feature fusion. This design allows HRNet to effectively retain and utilize high-resolution detailed feature information, enabling more precise capture of image details during the multi-scale feature fusion process. HRNet processes multiple resolution feature maps in parallel within its structure and facilitates information interaction and fusion between feature maps of different resolutions, allowing it to simultaneously acquire local detailed information and high-level semantic information.

Although HRNet has made significant progress in terms of performance, its high parameter count and computational demands make it challenging to apply to devices with constrained computational resources, hindering its practical application value. To address this issue, we propose a lightweight high-resolution network, LE-HRNet, based on HRNet, aimed at human pose estimation. Its structure is shown in Fig. 1.

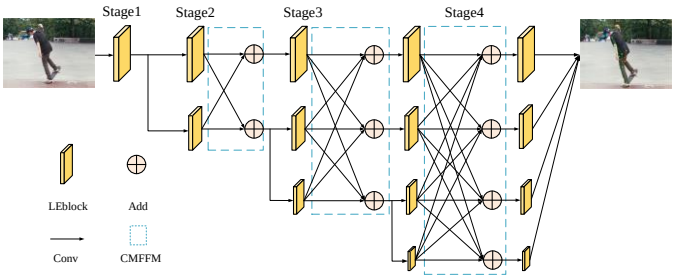


Fig. 1. LE-HRNet structure diagram.

As shown in Fig. 1, we use LEblock instead of residual blocks for feature extraction. This block effectively lowers the number of model parameters and computational demands while minimizing performance degradation, ensuring the model's ability to detect keypoints. Between different resolution feature maps, we optimize the sampling process and use CMFFM for multi-scale feature fusion. This process further reduces the parameter count and computational load, and enhances performance by learning cross-resolution channel and spatial information.

A. ELblock

HRNet uses residual block as the feature extraction module. While residual block effectively enhance the model's feature representation capability, they also bring a large number of parameters and computational load. Therefore, this paper proposes a lightweight block, LEblock, based on the residual block. The structure of LEblock is shown in Fig. 2. We substitute the conventional 3×3 convolution in the residual block with the Ghost module [16] to reduce the number of model parameters and computational overhead. To minimize performance loss and enhance the detection capability of human keypoints while lightweighting the model, we designed and added a Multi-Scale Coordinate Attention Mechanism, which improves the detection of human keypoints with a smaller computational load.

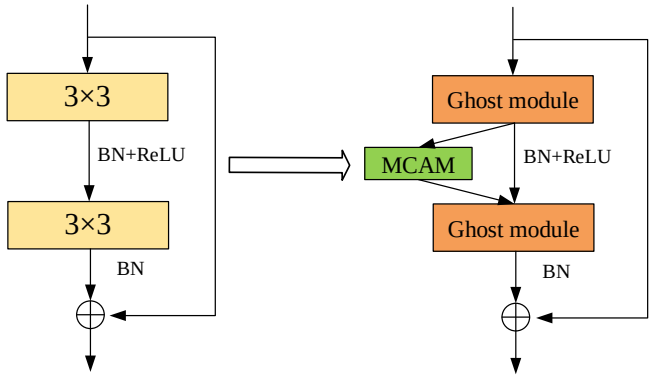


Fig. 2. The left is the residual block and the right is the ELblock.

B. Ghost Module

In traditional convolution operations, a large number of parameters and computations are generated, many of which are redundant. Ghost module optimizes the convolution process to obtain more image features with fewer parameters, thereby achieving model lightweighting. Ghost module decomposes the standard convolution process into three main steps: first, the number of feature map channels is reduced to generate the initial feature map; second, the initial feature map undergoes linear transformations to generate Ghost feature maps; finally, the initial feature map and the generated Ghost feature maps are concatenated to form the final output feature map. The specific transformation process of the Ghost module is shown in Fig. 3.

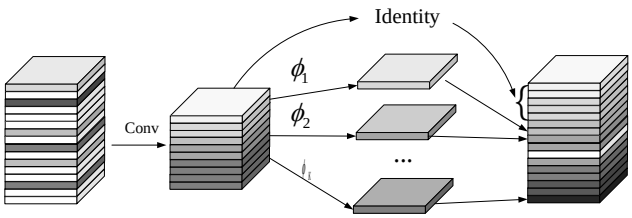


Fig. 3. The Ghost module.

Assuming the input feature map size is $C\times H\times W$, the number of output channels is N, and the convolution kernel size is $k\times k$, the number of parameters for standard convolution is:

$$Param_{conv} = k\times k\times C\times N \tag{1}$$

Among them, C and N are usually quite large, which results in a large number of parameters.

To address this, Ghost Convolution compresses the number of channels in the first step, reducing the channels to $m = N/s$. Next, linear transformations are used to generate Ghost feature maps with the number of channels $m\times(s-1) = N/s$. Finally, the feature maps obtained from the first two steps are concatenated along the channel dimension using an identity operation, resulting in an output feature map with N channels. Assuming $k = d$ and s is much smaller than C, the number of parameters for Ghost module can be calculated as follows:

$$Param_{Ghost} = k\times k\times C\times m + (s-1)\times d\times d\times m \tag{2}$$

Compared to standard convolution, the parameter compression ratio r_{param} for Ghost Convolution is:

$$r_{param} = \frac{Param_{conv}}{Param_{Ghost}} \approx \frac{s\times C}{s+C-1} \approx s \tag{3}$$

Through the Ghost module, the number of parameters can be reduced by a factor of s, achieving a significant reduction in parameter count.

C. Multi-Scale Coordinate Attention Mechanism

Although using the Ghost module to replace 3×3 convolution can decrease the parameter count and enhance computational efficiency, it also weakens the model's feature representation capability, leading to performance degradation and affecting the final prediction results. To enhance the detection capability of the model, attention mechanisms are commonly employed. SE (Squeeze-and-Excitation) [17] and ECA (Efficient Channel Attention) [18] enhance feature representation by re-weighting the channels of feature maps. The SE block integrates information between channels through global average pooling and fully connected layers, while the ECA enhances features through local cross-channel interactions. CBAM (Convolutional Block Attention Module) [19] combines channel attention and spatial attention, extracting global feature information through global average pooling and max pooling, capturing inter-channel dependencies and important spatial information, further enhancing feature representation and model performance. Human pose estimation is a task highly sensitive to positional information, making this information crucial. Coordinate Attention Mechanism[20] encodes spatial information by performing global average pooling in horizontal and vertical directions on the input feature map, and then fuses channel information to generate coordinate attention weights, re-weighting the input feature map. This not only enhances channel feature representation but also captures critical spatial information, thereby improving the model's feature expression capability and overall performance. However, Coordinate Attention Mechanism promotes channel fusion by using channel dimension reduction and expansion, which, although reducing the number of parameters, results in the loss of feature information during the reduction process. Additionally, 1×1 convolution are limited in their ability to extract local feature information and overlook the positional dependencies between different keypoints. Therefore, we propose a Multi-Scale Coordinate Attention Mechanism (MCAM) rove the model's ability to detect human keypoints. The structure of MCAM is shown in Fig. 4.

MCAM enhances the semantic and spatial information of feature maps by using feature grouping, parallel sub-branches, and multi-scale feature learning, producing better pixel-level attention without losing channel dimension information. For the input feature map, to avoid performance loss caused by channel dimension reduction, the input feature map is divided

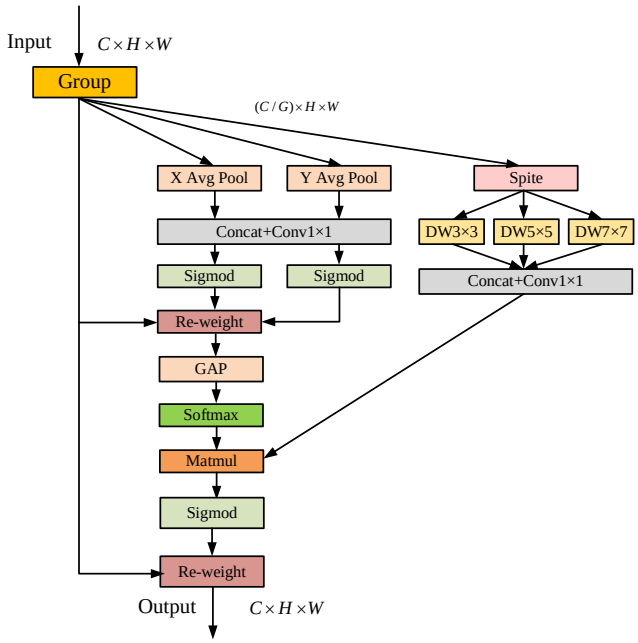


Fig. 4. The structure of multi-scale coordinate attention mechanism.

into multiple sub-feature maps to extract different semantic information. Assuming the input feature map is F_{input} and the output feature map is F_{output}, the generated multiple sub-feature maps are as follows:

$$F_{submap1}, F_{submap2}, ..., F_{submapg} = Group(F_{input}) \tag{4}$$

For each sub-feature map, three parallel sub-branches are used to extract coordinate position information and multi-scale

feature information. The first two sub-branches use operations similar to the coordinate attention mechanism, performing global average pooling in the vertical and horizontal directions to generate direction-aware attention maps. Then, using concatenation and 1×1 convolution, channel fusion is promoted without channel dimension reduction. Finally, the fused coordinate information feature maps are output through the sigmoid activation function. The formula is as follows:

$$\begin{cases} z_x = f_x^{GAP}(F_{submapi}) \\ z_y = f_y^{GAP}(F_{submapi}) \end{cases} \tag{5}$$

$$F_{mid} = f_{conv1\times1}([z_x, z_y]) \tag{6}$$

$$\begin{cases} g_x = \sigma(F_{mid}) \\ g_y = \sigma(F_{mid}) \end{cases} \tag{7}$$

$$F_{Coord_i} = F_{submapi} \times g_x \times g_y \tag{8}$$

In another sub-branch, we split into three branches and apply depthwise convolution with different kernel sizes of 3×3, 5×5, and 7×7. Smaller kernels can extract local feature information, while larger kernels, due to their larger receptive fields, can more easily extract relevant features between different keypoints. By integrating multi-scale feature information from local to global, we can enhance the model's ability to detect human keypoints in complex scenes, further improving the overall performance and robustness of the model. Finally, multi-scale feature maps are generated through concatenation and 1×1 convolution. The formula is as follows:

$$\begin{cases} F_3, F_5, F_7 = f_{splite}(F_{submapi}) \\ F_{multi} = f_{1\times1}([f_{DW3\times3}(F_3), f_{DW5\times5}(F_5), f_{DW7\times7}(F_7)]) \end{cases} \tag{9}$$

The feature map with coordinate information is modeled in the channel dimension through GAP and Softmax, and fused with the multi-scale feature map to ultimately output the feature map that integrates multi-scale feature information and positional information. The formula is as follows:

$$\begin{cases} \omega_i = \sigma(f_{GAP}(f_{soft\max}(F_{Coord_i})) \otimes F_{mulit}) \\ F_{outputi} = \omega_i \times F_{submapi} \\ F_{output} = [F_{output1}, F_{output2}, ..., F_{outputg}] \end{cases} \tag{10}$$

Compared to existing attention mechanisms, MCAM not only offers higher computational efficiency but also avoids the compression and expansion in the channel dimension, which reduces the loss of feature information. MCAM effectively enhances the detection capabilities for human keypoints by integrating coordinate positional information with multi-scale feature information.

D. Cross-Resolution Multi-Scale Feature Fusion Module

HRNet enhances the network's understanding of feature maps from local to global by integrating multi-scale feature information through a process of multi-scale feature fusion, improving recognition accuracy and adaptability. However, frequent upsampling and downsampling operations increase the computational burden. Additionally, these sampling processes can lead to the loss of spatial feature information, adversely affecting the model's performance.

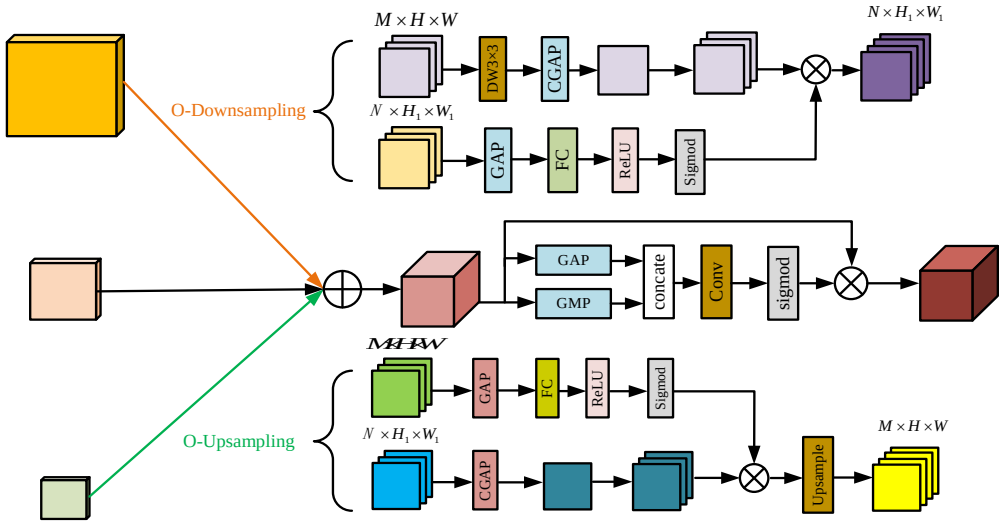


Fig. 5. The structure of cross-resolution multi-scale feature fusion module.

To address this, this paper optimizes the sampling process and proposes a cross-resolution multi-scale feature fusion module, as shown in Fig. 5. Taking the downsampling process as an example, a depthwise convolution with a stride of 2 is applied to the high-resolution branch feature map to reduce its computational load. Global average pooling is then performed along the channel dimension, reducing the number of channels to 1, ensuring that the information of each channel is uniformly preserved in a single feature map, thereby maintaining the essential information of the channels. The pooled feature map is then duplicated *N* times, where *N* is the number of channels of the low-resolution feature map.

For the low-resolution branch feature map, global average pooling is performed along the spatial dimension, and the channel weights are generated through a fully connected layer,

ReLU, and Sigmoid functions. The channel weight information is multiplied with the newly generated low-resolution feature map to achieve fusion, resulting in the optimized downsampled feature map. This process reduces the computational load during downsampling and significantly enhances the response to important feature channels while suppressing the response to unimportant channels by learning cross-resolution channel weight information. The optimized downsampling computation formula is as follows:

$$\begin{cases} \omega = \sigma(\mathrm{Re}\,LU(f_{FC}(f_{GAP}(F_{low-R})))) \\ F_{down} = \omega \otimes f_{copy}(f_{CGAP}(f_{DW3\times3}(F_{high-R}))) \end{cases} \tag{11}$$

Assuming the size of the high-resolution feature map is $M \times H \times W$ and the size of the low-resolution feature map is $N \times H_1 \times W_1$, the compression ratio r_{flops} of the computational load in the downsampling process is:

$$r_{flops} = \frac{Flops_{new}}{Flops_{original}} = \frac{H_1 \times W_1 \times M \times 3 \times 3}{H_1 \times W_1 \times N \times M \times 3 \times 3} = \frac{1}{N} \tag{12}$$

From the formula, it can be seen that this optimization effectively reduces the computational load in the downsampling process. The upsampling process is similar to the downsampling process. The processing flow for upsampling is as follows:

$$\begin{cases} \omega = \sigma(\mathrm{Re}\,LU(f_{FC}(f_{GAP}(F_{high-R})))) \\ F_{up} = f_{upsamp}(\omega \otimes f_{copy}(f_{CGAP}(F_{low-R}))) \end{cases} \tag{13}$$

We optimize the upsampling and downsampling processes in two main steps. The first step involves channel-wise aggregation compression, where information from different channels is merged into one channel. This representation encapsulates key information from multiple channels, resulting in a comprehensive feature representation. The second step focuses on learning cross-resolution channel weight information and using these weights to model the to-be-sampled feature maps along the channel dimension. This optimizes feature selection and reorganization, enhancing the model's representational ability and processing efficiency. These steps not only effectively reduce the computational load but also enhance the model's adaptability and sensitivity to features of different scales. By adjusting channel weights, we can provide varying degrees of emphasis on features at different levels, thus balancing detail and global information better during the upsampling or downsampling processes.

To retain more spatial information and enhance the ability to extract spatial features, we draw on the ideas of CBAM and use a spatial attention mechanism to extract spatial information. First, global average pooling and global max pooling are used to capture the average feature information and salient features of the feature map, respectively. These pooled features are then combined to form a more comprehensive feature representation. A 7×7 convolution is applied to further extract richer spatial feature information, and a Sigmoid activation function is used to generate spatial weights. These spatial weights are fused with the original feature map to produce a weighted and enhanced feature map, effectively preserving and highlighting the spatial details in the feature map.

III. EXPERIMENT

A. Datasets and Evaluation Metric

COCO (Common Objects in Context) [21] is a large-scale dataset widely used in computer vision, particularly suitable for human pose estimation, object detection, and image segmentation. This dataset offers rich scene complexity and extensive category coverage, including over 200,000 images and 250,000 human-annotated object in-stances. For human pose estimation, COCO meticulously annotates 17 keypoints cov-ering the major joints and parts of the body, making it an essential resource for re-searching and developing advanced human pose recognition algorithms.

MPII [22] dataset is a large-scale dataset focused on human pose estimation, containing over 25,000 images spanning 410 types of activities. Each image is detailed and annotated with 16 human body keypoints, including the head, neck, shoulders, elbows, hands, hips, knees, and feet. These images are derived from everyday life scenes, encompassing both individual and multi-person interactions, making MPII not only extensively used in academic research but also crucial for developing practical application algorithms in pose recognition.

In the COCO dataset, the performance of human pose estimation is primarily assessed using Object Keypoint Similarity (OKS). OKS is an evaluation metric that compares the similarity between predicted keypoints and true keypoints. The formula for calculating OKS is as follows:

$$OKS = \frac{\sum_i \exp\left[\frac{-d_i^2}{2s^2k_i^2}\delta(v_i > 0)\right]}{\sum_i \delta(v_i > 0)} \tag{14}$$

Where d_i is the Euclidean distance between the ground truth and predicted keypoint i; k is the constant for keypoint i; s is the scale of the ground truth object; v_i is the is the ground truth visibility flag for keypoint i; $\delta(v_i > 0)$ is the Dirac-delta function which computes as 1 if the keypoint i is labeled, otherwise 0.

OKS can be understood as a normalized measure of the error between the predicted and true annotations for each keypoint, which takes into account the size of the human body and the specific sensitivity of each keypoint. Based on OKS, the COCO dataset also calculates Average Precision (AP) at multiple thresholds, ranging from OKS=0.50 (looser matching) to OKS=0.95 (very strict matching).

MPII uses PCK (Percentage of Correct Keypoints) as the main metric to evaluate model performance. PCK measures the percentage of predicted keypoints that match the true keypoints within a certain distance threshold. The specific calculation formula is as follows:

$$PCK = \frac{1}{n}(\sum_{i=1}^{n} \delta(dist(p_i, q_i) \le \alpha \cdot \max(h, w)) \tag{15}$$

Where n is the number of keypoints; p_i is the predicted position of the i-th keypoint; q_i is the actual position of the i-th keypoint; $dist(p_i, q_i)$ is the distance between the pre- dicted keypoint p_i and the actual keypoint q_i ; α is a predefined threshold, and ; $dist(p_i, q_i) < \alpha$ means the prediction is considered accurate.

B. Experimental Setup

The experimental setup for this paper is as follows: Intel(R) Xeon(R) Silver 4310 CPU @ 2.10GHz, 64GB RAM, two RTX A5000 GPU with 24GB VRAM each, Ubuntu 22.04.3 LTS, Python 3.8. The deep learning framework used is Pytorch 3.9, with CUDA 11.5 for accelerated computing.

When training on the COCO training set, images from the COCO training set are cropped and scaled to a fixed size of 256×192. Adam is used as the optimizer during network training, with an initial learning rate of 0.001. The learning rate is reduced to 0.0001 at the 170th epoch, and then to 0.00001 at the 210th epoch, with a total of 230 epochs of training. During training, random image rotation and horizontal flipping are also used for data augmentation. When training on the MPII dataset, cropped images are uniformly scaled to a fixed size of 256×256. Other training details are the same as those for the COCO dataset, using the same parameter settings and experimental environment.

C. Result and Analysis

The performance comparison of LE-HRNet with other human pose estimation algorithms on the COCO validation set, with an input size of 256×192, is shown in Table I. As seen from the table, compared to HRNet, LE-HRNet reduces the number of parameters and computational load by 81.1% and 78.7% respectively, while the AP only decreases by 4.1%. LE-HRNet achieves a significant reduction in model parameters and computational load with minimal performance loss, maintaining a balance between model performance and parameter size. Compared to Hourglass and CPN, LE-HRNet has lower parameters and higher performance. Compared to SimpleBaseline, LE-HRNet's AP is only 1.1% lower, but its number of parameters is much lower than SimpleBaseline. Compared to lightweight models like MobileNetV2[23] and ShuffleNetV2[24], LE-HRNet's AP is higher by 4.7% and 9.4% respectively, and LE-HRNet also has an advantage in terms of parameters and computational load. Compared to even smaller lightweight models like Lite-HRNet, Dite-HRNet, and X-HRNet, LE-HRNet has more parameters, but its AP is higher by 2.1%, 1.0%, and 1.9% respectively. Unlike these models which aggressively pursue lightweight design, LE-HRNet focuses more on balancing parameter size and performance, ensuring model performance while reducing the number of parameters.

With an input size of 384 × 288, the performance comparison on the COCO test set is shown in Fig. 6. LE-HRNet achieved an AP score of 72.1, outperforming other networks while maintaining a balance between performance and computational load.

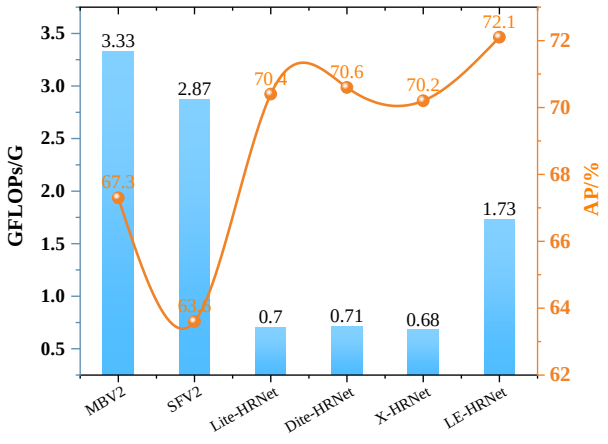


Fig. 6. Performance comparison of different algorithms on the COCO test set.

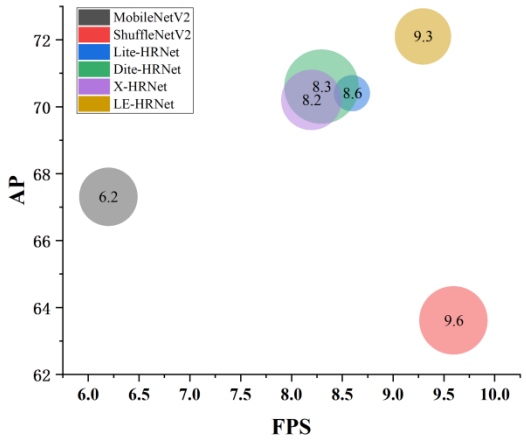


Fig. 7. Performance comparison of inference speed of different models.

The comparison of inference speed between some methods and LE-HNet was conducted in this paper. The testing of lightweight models emphasizes performance under limited resources, so the tests in this paper were conducted using only the CPU, specifically an Intel(R) Xeon(R) Silver 4310. The tests were carried out under consistent experimental conditions, and the results are shown in Fig. 7. Compared to other models, LE-HRNet achieves an inference speed of 9.6 FPS while maintaining high performance, making it faster than Lite-HRNet, Dite-HRNet, and X-HRNet. Although other lightweight models, such as ShuffleNetV2, have slightly faster inference speeds, their accuracy is lower and they fail to accurately detect human key points. LE-HRNet offers a better balance, and the verification of its inference speed demonstrates that LE-HRNet is more suitable for edge computing platforms.

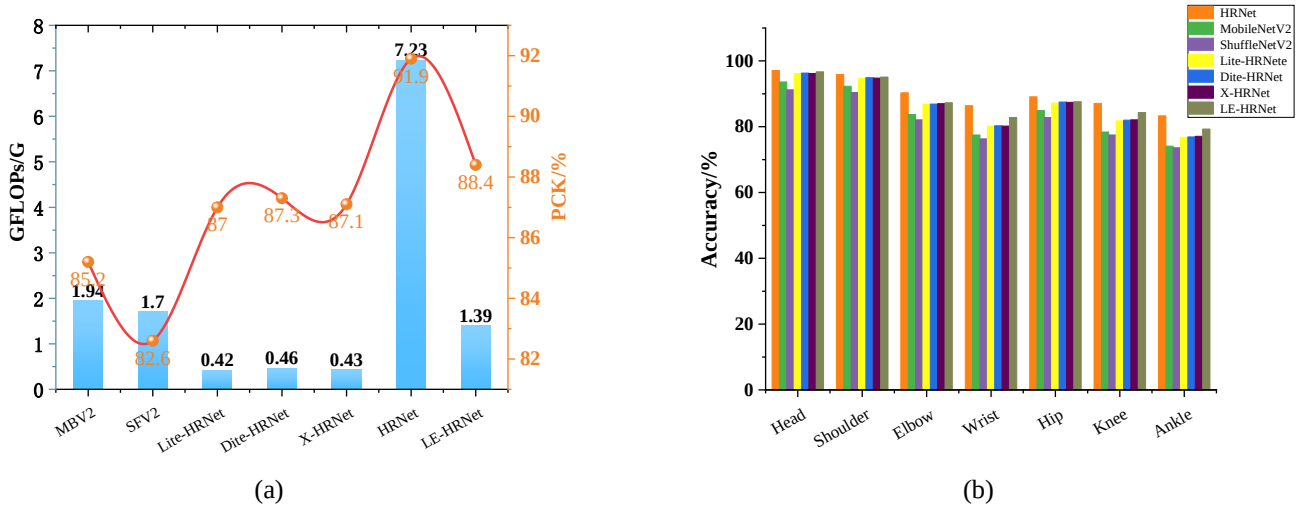
TABLE I. PERFORMANCE COMPARISON OF DIFFERENT ALGORITHMS ON THE COCO VALIDATION SET

Model	Params /10^6	GFlops/G	*AP*/%	AP^{50}/%	AP^{75}/%	AP^{M}/%	AP^{L}/%	*AR*/%
Hourglass	25.1	14.3	66.9	-	-	-	-	-
CPN	27.0	6.20	68.6	-	-	-	-	-
SimpleBaseline	34.0	8.90	70.4	88.6	78.3	67.1	77.2	76.3
HRNet	28.5	7.10	73.4	89.5	80.7	70.2	80.1	78.9
MobileNetV2	9.6	1.48	64.6	87.4	72.3	61.1	61.1	70.7
ShuffleNetV2	7.6	1.28	59.9	85.4	66.3	56.6	66.2	66.4
Lite-HRNet	1.8	0.31	67.2	88.0	75.0	64.3	73.1	73.3
Dite-HRNet	1.8	0.3	68.3	88.2	76.2	65.5	74.1	74.2
X-HRNet	2.1	0.3	67.4	87.5	75.4	64.5	73.3	73.5
LE-HRNet	5.4	1.51	69.3	88.6	77.2	66.1	74.3	74.6

Table II shows the comparison results with different human pose estimation algorithms on the MPII validation set. Compared to HRNet, LE-HRNet significantly reduces the number of parameters and computational load, with only a 3.6% decrease in accuracy. Compared to MobileNetV2 and ShuffleNetV2, LE-HRNet has lower parameters and an accuracy improvement of 3.0% and 5.6%, respectively. Compared to Lite-HRNet, Dite-HRNet, and X-HRNet, LE-HRNet improves accuracy by 1.7%, 1.1%, and 1.4%, respectively, with a slight increase in parameters and computational load. This demonstrates that LE-HRNet maintains model performance while ensuring low parameter count.

TABLE II. PERFORMANCE COMPARISON OF DIFFERENT ALGORITHMS ON THE MPII VALIDATION SET

Model	Params/10^6	GFLOPs/G	PCKh/%
MobileNetV2	9.6	1.9	85.4
ShuffleNetV2	7.6	1.7	82.8
Lite-HRNet	1.8	0.4	87.0
Dite-HRNet	1.8	0.4	87.6
X-HRNet	2.1	0.4	87.3
HRNet	28.5	7.6	92.3
LE-HRNet	5.4	1.44	88.7



(a) (b)

Fig. 8. Performance comparison of different algorithms on the MPII test set. (a) Comparison of different algorithms on GFLOPs and PCK; (b) Recognition accuracy of different algorithms at each keypoint.

Fig. 8 shows the performance comparison of different algorithms on the MPII test set. From the results in the figure, LE-HRNet has higher keypoint recognition accuracy compared to other lightweight algorithms. Particularly for some challenging keypoints such as Wrist, Knee, and Ankle, the accuracy improvement is more significant than for other keypoints. This is mainly due to the MCAM, which significantly enhances the detection capability for human keypoints.

We randomly selected a set of images from the COCO dataset for visual analysis. This set includes various scenarios such as single-person and multi-person scenes, as shown in Fig. 9. From the figure, it can be seen that LE-HRNet can accurately identify human keypoints in single-person scenarios.



Fig. 9. Visualization results on the COCO dataset.

In multi-person scenes, especially when there is overlap between body parts, LE-HRNet, with its strong feature extraction capabilities, can accurately infer the positions of the occluded keypoints by extracting multi-scale feature information and utilizing other visible keypoints. The visual analysis of different scenarios demonstrates that LE-HRNet maintains excellent detection performance in various complex scenes.

D. Ablation Experiment

To validate the effectiveness of the proposed method, we conducted ablation experiments on the COCO dataset for LEblock and CMFFM. The experimental results are shown in Table III. By replacing the residual block with LEblock, the model's parameter count is reduced by 76.4%, while the AP score only decreases by 4.5%, demonstrating that the lightweight module LEblock can significantly reduce the model's parameter count with minimal performance loss. Building on this, we inserted CMFFM for multi-scale feature fusion. The model's parameter count was further reduced, and the AP increased by 0.4%. This improvement is due to the optimization of the sampling process and the decomposition of the convolution process, which reduced the parameter count. Additionally, learning cross-resolution channel weight information effectively models channel features, and the spatial attention mechanism preserves more spatial detail features.

TABLE III. ABLATION EXPERIMENTS ON LEBLOCK AND CMFFM ON THE COCO DATASET

Model	LEblock	CMFFM	Params	AP/%
HRNet	×	×	28.5M	73.4
LE-HRNet	√	×	6.7M	68.9
	√	√	5.4M	69.3

To further validate the effectiveness of MCAM, we conducted ablation experiments on MCAM, with the results shown in Table IV. With the addition of MCAM, the parameter count increased by only 1.4M, while performance improved by 1.5%, demonstrating that MCAM effectively enhances the model's ability to detect human keypoints.

TABLE IV. ABLATION EXPERIMENTS ON MCAM

Model	Params	AP/%
+ELblock	6.7M	68.9
+ELblock (No MCAM)	5.3M	67.4

E. Discussion

To enable human pose estimation on mobile devices or edge computing devices, we propose a series of methods to streamline the high-resolution network. High-resolution networks are widely used in scenarios such as human pose estimation and semantic segmentation due to their high recognition accuracy. However, the high parameter count and computational complexity of these models make it difficult to deploy them on devices with limited computational resources. To address this, we propose replacing the standard 3×3 convolution with a Ghost module to reduce computational load, and we further optimize the upsampling and downsampling processes to improve computational efficiency. Additionally, to maintain model performance while reducing computation, we introduce a multi-scale coordinate attention mechanism that effectively minimizes performance loss due to lightweighting. Through this series of methods, we have successfully streamlined the high-resolution network and achieved favorable inference speed on low-power devices.

IV. Conclusion and Future Work

To address the issues of large parameter count and high computational complexity in high-resolution network models, we propose a lightweight and efficient high-resolution network module. We use the Ghost module to replace 3×3 convolution to reduce the parameter count and computational load of the model. Simultaneously, to minimize the loss of feature information during the lightweight process and ensure model performance, we designed a Multi-Scale Coordinate Attention Mechanism. This mechanism effectively enhances the detection of human keypoints by integrating multi-scale feature information and coordinate positional information without compromising performance. Finally, we optimized the multi-scale feature fusion stage, modeling both channel and spatial features while reducing the parameter count, further enhancing the model's performance. Experiments on multiple datasets validated the effectiveness of our proposed method.

In future work, we will deploy LE-HRNet on mobile devices and apply it in physical education, such as high jump, long jump, and swimming. By using LE-HRNet to identify key points of students' movements and calculate similarity with standard actions, we will be able to score students' movements and provide improvements for non-standard actions, which will aid in students' sports training.

Acknowledgment

This study was supported by Shandong Provincial Undergraduate Teaching Reform Project (Grant Number: Z2021450), Shandong Provincial Natural Science Foundation of P.R. China (Grant Number: ZR2020QF069), National College Students' Innovation and Entrepreneurship Training Program (Grant Number: 202310433069), and Shandong University of Technology Postgraduate Teaching Reform Project (Grant Number: 4053222063).

References

[1] L. Song, G. Yu, J. Yuan, and Z. Liu, “Human pose estimation and its application to action recognition: A survey,” J. Vis. Commun. Image Represent., vol. 76, p. 103055, 2021.

[2] C. Zheng, W. Wu, C. Chen, T. Yang, S. Zhu, J. Shen, N. Kehtarnavaz, and M. Shah, “Deep learning-based human pose estimation: A survey,” ACM Comput. Surv., vol. 56, no. 1, pp. 1–37, 2023.

[3] P. F. Felzenszwalb and D. P. Huttenlocher, “Pictorial structures for object recognition,” Int. J. Comput. Vis., vol. 61, pp. 55–79, 2005.

[4] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, “Gradient-based learning applied to document recognition,” Proc. IEEE, vol. 86, no. 11, pp. 2278–2324, 1998.

[5] A. Toshev and C. Szegedy, “Deeppose: Human pose estimation via deep neural networks,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2014, pp. 1653–1660.

[6] A. Newell, K. Yang, and J. Deng, “Stacked hourglass networks for human pose estimation,” in Proc. Comput. Vis. ECCV 2016: 14th Eur. Conf., Amsterdam, The Netherlands, Oct. 11-14, 2016, Part VIII, Springer, 2016, pp. 483–499.

[7] S.-E. Wei, V. Ramakrishna, T. Kanade, and Y. Sheikh, “Convolutional pose machines,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2016, pp. 4724–4732.

[8] Z. Cao, T. Simon, S.-E. Wei, and Y. Sheikh, “Realtime multi-person 2D pose estimation using part affinity fields,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2017, pp. 7291–7299.

[9] Y. Chen, Z. Wang, Y. Peng, Z. Zhang, G. Yu, and J. Sun, “Cascaded pyramid network for multi-person pose estimation,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2018, pp. 7103–7112.

[10] B. Xiao, H. Wu, and Y. Wei, “Simple baselines for human pose estimation and tracking,” in Proc. Eur. Conf. Comput. Vis. (ECCV), 2018, pp. 466–481.

[11] K. Sun, B. Xiao, D. Liu, and J. Wang, “Deep high-resolution representation learning for human pose estimation,” in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., 2019, pp. 5693–5703.

[12] C. Yu, B. Xiao, C. Gao, L. Yuan, L. Zhang, N. Sang, and J. Wang, “Lite-HRNet: A lightweight high-resolution network,” in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 10440–10450, 2021.

[13] Q. Li, Z. Zhang, F. Xiao, F. Zhang, and B. Bhanu, “Dite-HRNet: Dynamic lightweight high-resolution network for human pose estimation,” arXiv preprint arXiv:2204.10762, 2022.

[14] Y. Zhou, X. Wang, X. Xu, L. Zhao, and J. Song, “X-HRNet: Towards lightweight human pose estimation with spatially unidimensional self-attention,” in 2022 IEEE International Conference on Multimedia and Expo (ICME), pp. 01–06, 2022, IEEE.

[15] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 770–778, 2016.

[16] K. Han, Y. Wang, Q. Tian, J. Guo, C. Xu, and C. Xu, “Ghostnet: More features from cheap operations,” in Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp. 1580–1589, 2020.

[17] J. Hu, L. Shen, and G. Sun, “Squeeze-and-excitation networks,” in Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 7132–7141, 2018.

[18] Q. Wang, B. Wu, P. Zhu, P. Li, W. Zuo, and Q. Hu, “ECA-Net: Efficient channel attention for deep convolutional neural networks,” in Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp. 11534–11542, 2020.

[19] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, “CBAM: Convolutional block attention module,” in Proceedings of the European conference on computer vision (ECCV), pp. 3–19, 2018.

[20] Q. Hou, D. Zhou, and J. Feng, “Coordinate attention for efficient mobile network design,” in Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp. 13713–13722, 2021.

[21] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick, “Microsoft COCO: Common objects in context,” in Computer Vision--ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part V 13, pp. 740–755, 2014, Springer.

[22] M. Andriluka, L. Pishchulin, P. Gehler, and B. Schiele, “2D human pose estimation: New benchmark and state of the art analysis,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3686–3693, 2014.

[23] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, “MobileNetV2: Inverted residuals and linear bottlenecks,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 4510–4520, 2018.

[24] N. Ma, X. Zhang, H.-T. Zheng, and J. Sun, “ShuffleNet V2: Practical guidelines for efficient CNN architecture design,” in Proceedings of the European conference on computer vision (ECCV), pp. 116–131, 2018.

Enhanced Resume Screening for Smart Hiring Using Sentence-Bidirectional Encoder Representations from Transformers (S-BERT)

Asmita Deshmukh, Anjali Raut
Hanuman Vyayam Prasarak Mandal's College of Engineering and Technology, Amravati, 444605, Maharashtra, India

***Abstract*—In a world inundated with resumes, the hiring process is often challenging, particularly for large organizations. HR professionals face the daunting task of manually sifting through numerous applications. This paper presents 'Enhanced Resume Screening for Smart Hiring using Sentence-Bidirectional Encoder Representations from Transformers (S-BERT)' to revolutionize this process. For HR professionals dealing with overwhelming numbers of resumes, the manual screening process is time consuming and error-prone. To address this, here the proposed solution is developed for an automated solution leveraging NLP techniques and a cosine distance matrix. Our approach involves pre-processing, embed- ding generation using S-BERT, cosine similarity calculation, and ranking based on scores. In our evaluation on a dataset of 223 resumes, our automated screening mechanism demonstrated remarkable efficiency with a screening speed of 0.233 seconds per resume. The system's accuracy was 90%, showcasing its ability to effectively identify relevant resumes. This work presents a powerful tool for HR professionals, significantly reducing the manual workload and enhancing the accuracy of identifying suitable candidates. The societal impact lies in streamlining hiring processes, making them more efficient and accessible, ultimately contributing to a more productive and equitable job market.**

***Keywords*—S-BERT; resume; automated screening; job; CV**

I. Introduction

In the contemporary landscape of recruitment, the influx of numerous resumes for a single job opening poses a considerable challenge for Human Resources (HR) professionals [1]. The traditional method of manual resume screening, while (being) essential, is not without its shortcomings [2]. This process, laden with time-consuming intricacies, demands meticulous attention in detail to ensure the identification of the most qualified candidates [3]. However, the reliance on conventional keyword matching methods in automated screening introduces its own set of challenges, often resulting in false positives and negatives [4].

To address these challenges, this research paper introduces a pioneering approach to automated resume screening [5]. Leveraging the capabilities of Sentence-Bidirectional Encoder Representations from Transformers (S-BERT), a cutting-edge natural language processing (NLP) model, our methodology offers a novel perspective to the intricate task of identifying the most suitable candidates for a given role. S-BERT's unique ability to generate contextualized representations of text enables a nuanced understanding of resumes, allowing for the identification of relevant skills and experiences even when not explicitly articulated.

A. Bert

BERT, or Bidirectional Encoder Representations from Transformers, is a groundbreaking language model developed by Google AI, significantly impacting natural language processing (NLP). Operating on the Transformer architecture, it excels in learning intricate relationships between words and phrases, crucial for understanding textual meaning. BERT demonstrates state-of-the-art performance across NLP tasks, including natural language understanding (NLU), natural language generation (NLG), and natural language inference (NLI). Its functionality involves tokenizing input text, embedding tokens into meaningful vectors, adding positional embeddings, passing tokens through Transformer encoders to understand relationships, and generating final embeddings for the desired NLP task. BERT finds applications in enhancing search engine results, improving machine translation accuracy, developing context-aware chatbots, and generating concise text summaries. As an evolving tool, BERT holds immense potential to transform human-computer interactions.

B. S-Bert

S-BERT, short for Sentence-BERT, is a BERT model adaptation designed for sentence embedding computation. These embeddings, representing sentence meaning, are valuable for tasks like semantic similarity, clustering, and information retrieval. In contrast to BERT, which undergoes masked language modeling and next sentence prediction training, S-BERT trained on natural language inference. This task involves predicting if pairs of sentences entail, contradict, or are neutral. S-BERT utilizes a triplet loss function, minimizing distances for similar sentence pairs and maximizing them for dissimilar ones. During application, S-BERT processes each sentence independently, generating vector outputs for the first token. This single-pass approach is computationally more efficient than BERT's pairwise sentence comparison. S-BERT excels in producing semantically meaningful embeddings due to its emphasis on understanding sentence relationships. Demonstrating superiority over BERT, S-BERT excels in downstream tasks, including semantic textual similarity, paraphrase identification, and clustering.

Fig. 1 shows the process flow of S-BERT (Sentence-BERT), a machine learning algorithm that uses a shared encoder and a distance metric to train a model. The shared encoder is

used to train the model, and the distance metric is used to measure the distance between two points.

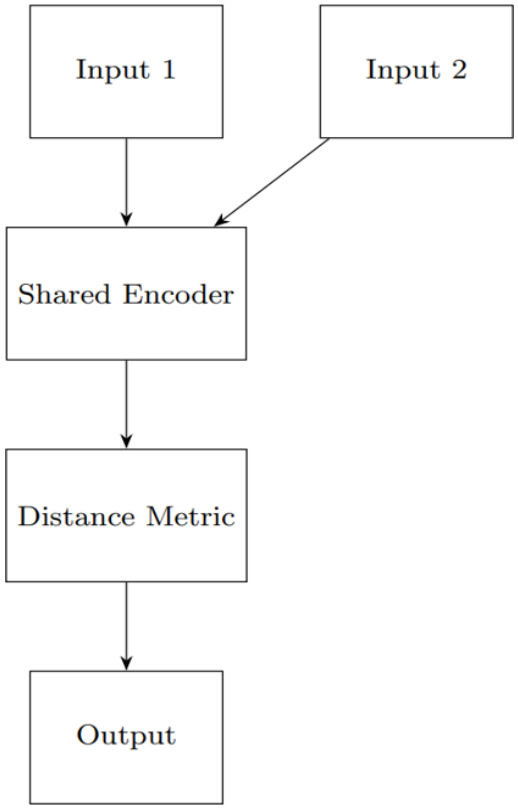


Fig. 1. Process flow diagram of sentence BERT (S-BERT).

Process flow:

Input: S-BERT takes two input sentences as input.

Shared encoder: The shared encoder is a neural network that learns to represent each sentence as a dense vector. The shared encoder is trained using a natural language inference (NLI) task, where the model is given a pair of sentences and must predict their relationship: entailment, contradiction, or neutral. This NLI training enables S-BERT to capture the semantic differences between sentences, leading to more meaningful sentence embeddings. The shared encoder generates embeddings for the two input sentences. The embeddings are dependent on the inputs, meaning that the embedding for a sentence is different depending on the other sentence in the pair.

Distance metric: The embeddings are then fed to a distance metric to calculate the distance between the two sentences. A common distance metric used for S-BERT is the cosine similarity.

Output: Based on the distance, a decision is made about whether the sentences are similar or dissimilar. If the distance is small, then the sentences are considered to be similar. If the distance is large, then the sentences are considered to be dissimilar.

This paper outlines the proposed methodology, which involves generating embedding from both resumes and job descriptions using S-BERT and subsequently measuring their alignment through cosine similarity. The ranking of resumes based on these scores facilitates an efficient and accurate screening process.

The effectiveness of our approach is validated through a comprehensive evaluation on a dataset of 223 resumes, showing an impressive accuracy of 90%. Beyond these quantitative metrics, our method's resilience to common pitfalls such as keyword stuffing and its efficiency, with a screening speed of 0.233 seconds per resume, mark a significant advancement in the realm of automated resume screening.

As the need for streamlined and unbiased hiring processes intensifies, our research stands as a beacon for HR professionals, offering a solution that not only enhances efficiency but also contributes to the broader goals of diversity, equity, and inclusivity in the workforce. The ensuing sections delve into the intricacies of our proposed methodology, the experimental results, and the potential implications of our work on the future landscape of smart hiring practices.

Fig. 2 illustrates the conceptual framework of the proposed Resume Screening System for Smart Hiring using Sentence-Bidirectional Encoder Representations from Trans- formers (S-BERT). The model processes 200 resumes in PDF format, initially converting them to Excel format. Subsequently, text normalization techniques such as lemmatization and stemming are applied. Keywords are then extracted, forming sentences for each resume. S-BERT generates embeddings for these sentences. A parallel process is conducted on the job description, creating job description embeddings. Cosine similarity is computed between the sentence and job description embeddings, determining the ranking of resumes based on these similarity scores. While previous studies have made significant strides in automated resume screening, several gaps remain that underscore the urgency of our research. First, many existing systems rely heavily on keyword matching, which can be easily gamed and may miss candidates with relevant skills expressed in different terms. Second, the contextual understanding of resumes has been limited, often failing to capture the nuanced relationships between skills, experiences, and job requirements. Third, there has been insufficient focus on mitigating biases in automated screening processes, potentially perpetuating unfair hiring practices. To address these gaps, our research proposes the use of Sentence-BERT (S-BERT), a state-of-the-art natural language processing model that offers deeper contextual understanding and semantic analysis of resume content. This approach allows for more nuanced matching between resumes and job descriptions, reducing reliance on exact keyword matches. Additionally, by incorporating cosine similarity measures, our method provides a more holistic evaluation of candidate suitability. To tackle bias concerns, proposed solution suggest rigorous testing and continuous refinement of the model with diverse datasets. Our research not only aims to enhance the efficiency of resume screening but also to improve its fairness and accuracy, thus addressing critical shortcomings in existing automated hiring systems.

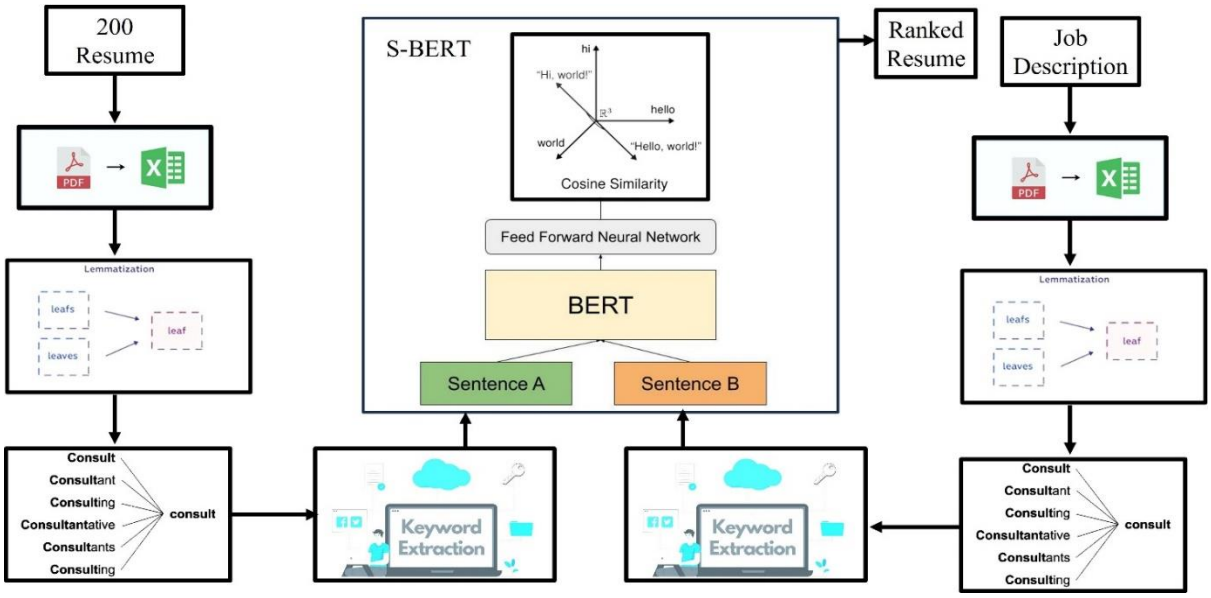


Fig. 2. Concept diagram of the proposed resume screening system for smart hiring using Sentence-Bidirectional Encoder Representations from Transformers (S-BERT).

II. Literature Review

Automated resume screening has evolved as a critical domain within the recruitment process, driven by the need for efficiency, accuracy, and mitigation of biases in hiring [6]. Traditional methods, reliant on manual screening, have proven to be time-consuming, error-prone, and susceptible to human biases [7]. This literature review explores key studies and methodologies in the realm of automated resume screening, culminating in the proposal of an advanced system utilizing Sentence-Bidirectional Encoder Representations from Transformers (S-BERT).

Early endeavors in automated resume screening focused on basic keyword matching and rule-based systems. Kopparapu introduced a system for information extraction from unstructured resumes using natural language processing (NLP) techniques [4]. This marked an initial attempt to streamline the screening process. However, these early systems struggled with nuanced contextual understanding.

Recognizing the limitations of keyword-based approaches, recent years have witnessed a surge in the application of advanced NLP techniques to resume screening. Gundlapalli et al. demonstrated the effective use of NLP tools to screen medical records for evidence of homelessness [8].

Feller et al. [9] explored NLP for predictive modeling of HIV diagnoses, showcasing the potential for contextual understanding beyond explicit mentions [10].

The application of NLP in diverse domains highlights its versatility. Nayor et al. [11] employed NLP for accurate calculation of adenoma detection rates in the context of screening colonoscopies [12]. Trivedi et al. [13] used NLP for large-scale labeling of clinical records, emphasizing the potential for automating data extraction from existing records [14].

The integration of machine learning into resume screening systems has been pivotal. Roy et al. [15] introduced a machine learning approach for automating a resume recommendation system, illustrating the intersection of NLP and machine learning in enhancing screening mechanisms [16]. Recent breakthroughs in NLP, particularly with models like S-BERT, have brought contextualized representations to the forefront. Delimayanti et al. [17] presented a content-based suggestion system using cosine similarity and KNN algorithms [18]. This highlighted the importance of contextual understanding, which is a hallmark of S-BERT. While the literature showcases promising advancements, challenges remain. Ndukwe et al. [19] discussed the need for careful development and evaluation of NLP models to ensure fairness and mitigate biases [20].

Choi et al. [21] emphasized the efficiency of resume screening through NLP but acknowledged the challenges posed by computational expenses. Recent research on automated resume screening and ranking has explored various approaches using natural language processing and deep learning techniques. Several studies have investigated the use of transformer-based models like BERT and its variants for this task. James et al. [22] and Mukherjee employed DistilBERT and XLM [23] for resume shortlisting and ranking. Kinger et al. [24] combined YOLOv5 for resume parsing with DistilBERT for ranking. Sentence-BERT (S-BERT), introduced by Reimers and Gurevych [25], has gained traction for generating semantically meaningful sentence embeddings. Subsequent work has evaluated and refined S-BERT, including TA-SBERT. Seo et al., [26] and Chu et al., [27], these approaches aim to capture nuanced semantic relationships between resume content and job requirements, moving beyond simple keyword matching. Additionally, researchers have explored combining embeddings with other techniques, such as named entity recognition and domain-specific knowledge Vanetik and Kogan, [28]; Yu et al., [29], to further improve matching accuracy [30]. While progress has been made, challenges remain in mitigating biases and ensuring fair evaluation across diverse candidate pools [31].

III. METHODOLOGY

The objective of this study is to develop an automated resume screening mechanism to assist the Human Resources (HR) department in the initial filtering of resumes, ensuring the provision of the most pertinent candidates for further evaluation, such as interviews. The dataset used in this study consisted of 223 resumes in PDF format. These resumes were collected from various job applicants across different fields and experience levels. To facilitate processing, the PDF files were converted to CSV (Comma-Separated Values) format. This conversion preserved the textual content of the resumes while organizing it into a structured tabular format. The CSV structure included columns for different resume sections such as personal information, education, work experience, skills, and additional qualifications. This standardized format allowed for easier extraction of relevant information and application of natural language processing techniques. The conversion from PDF to CSV was performed using a custom Python script that utilized PDF parsing libraries to extract text and organize it into appropriate CSV fields. This structured dataset provided a consistent foundation for the subsequent steps in our automated resume screening process.

The proposed automated screening mechanism leverages Natural Language Processing (NLP) techniques and a cosine distance matrix to evaluate the alignment of resumes with the corresponding job description. The methodology employs Sentence-Bidirectional Encoder Representations from Transformers (S-BERT), a sentence-level model, to extract embeddings that capture the contextual information from the resumes. These embeddings are then compared to those generated from the job description, with the aim of ranking the resumes based on their relevance.

Stop words are common words that do not carry much meaning and can cause noise in text analysis. Removing stop words helps to improve the efficiency and accuracy of the NLP process. Pre-defined list of stop words in English was used to remove stop words from the resumes and job description.

Lemmatization is a NLP technique that groups words with the same meaning together. This is done by reducing words to their root form. For example, the words “cats” and “kittens” would both be lemmatized to the root word “cat”. Lemmatization helps to improve the accuracy of the NLP process by ensuring that words with the same meaning are treated similarly.

Stemming is a NLP technique that reduces words to their common stem. This is done by removing suffixes and prefixes. For example, the words “running” and “ran” would both be stemmed to the common stem “run”. Stemming when used in combination with lemmatization, produces better results.

The resumes and job description were preprocessed using the following steps:

- Stop words were removed.
- Lemmatization was performed.
- Stemming was performed.

The S-BERT model was employed to generate embeddings from the preprocessed text of both resumes and the job description. Embeddings are numerical representations of words that capture their meaning and context. S-BERT is a sentence-level model that generates embeddings that capture the contextual information from the sentences.

Cosine similarity is a metric used to quantify the similarity between two vectors. In this study, cosine similarity was used to measure the similarity between the embeddings of the resumes and the job description. Resumes with higher cosine similarity scores are considered to be more relevant to the job description.

The resumes were ranked based on their cosine similarity scores. The resumes with the highest cosine similarity scores were ranked at the top of the list.

This methodology aims to reduce the reliance on subjective referral-based hiring by introducing an automated screening process. This approach ensures a more transparent and standardized selection mechanism, promoting fairness in the company’s hiring process.

The keywords from each resume are concatenated to form a sentence. S-BERT is then applied to these sentences to generate embeddings of a specific length, e.g., 4x96 bytes.

Cosine similarity is then used to calculate the matching score between the job description embedding and each resume embedding. The matching score is a value between 0 and 1, with 1 being the best match and 0 being no match.

Finally, the cosine scores are ranked in descending order. The resume with the highest cosine score is ranked first.

A step-by-step explanation of the process is as given below:

- Concatenate keywords to form a sentence: The keywords from each resume are concatenated to form a sentence. This sentence captures the essence of the resume and highlights the applicant’s key skills and experience.
- Generate S-BERT embeddings: S-BERT is applied to the sentences to generate embeddings of a specific length. Embeddings are vector representations of text that capture the semantic meaning and context of the words.
- Calculate cosine similarity: Cosine similarity is used to calculate the matching score between the job description embedding and each resume embedding. Cosine similarity is a measure of similarity between two vectors. The higher the cosine similarity score, the more similar the two vectors are.
- Rank cosine scores: The cosine scores are ranked in descending order. The resume with the highest cosine score is ranked first.

Fig. 3 illustrates the resume matching mechanism utilizing S-BERT and cosine similarity in our proposed automated screening system. The process begins with extracting key sentences from both the resume and the job description. These sentences are then fed into the S-BERT model, which generates embeddings - dense vector representations of the text with dimensions of 4 x 96 bytes for each input. The embeddings capture the semantic meaning of the sentences, allowing for a

nuanced comparison beyond simple keyword matching. Once the embeddings are generated, a cosine matching algorithm computes the similarity between the resume and job description embeddings. This similarity score quantifies how well the content of the resume aligns with the requirements outlined in the job description. Finally, a ranking algorithm uses these similarity scores to order the resumes, with higher scores indicating better matches for the position. This approach enables a more contextual and meaningful comparison between candidates and job requirements, addressing limitations of traditional keyword-based screening methods.

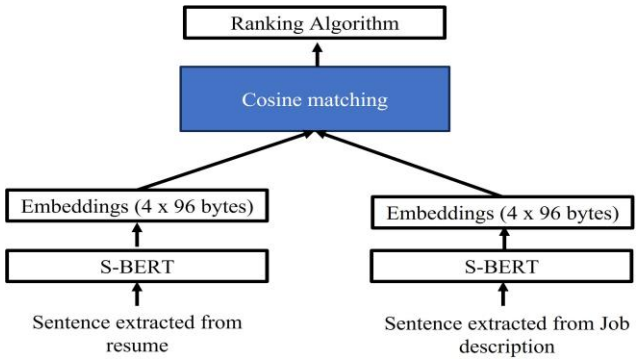


Fig. 3. Resume matching mechanism using S-BERT and cosine similarity.

The proposed resume matching mechanism has a number of advantages:

- It is able to accurately match resumes to job descriptions, even when the resumes are in different formats.
- It is able to identify resumes that are relevant to the job description, even if the resumes do not contain all of the keywords that are listed in the job description.
- It is able to rank resumes based on how well they match the job description, making it easy for recruiters to identify the most qualified candidates.

IV. Results

The automated resume screening mechanism was rigorously (applied) evaluated on a dataset consisting of 223 resumes in PDF format. The results underscore the efficacy of the proposed methodology in identifying and ranking relevant candidates based on alignment with the job description.

1) Screening speed: The screening process demonstrated re- markable efficiency, achieving a speed of 0.233 seconds per resume. This rapid processing speed ensures the practical applicability of the automated mechanism to scenarios involving substantial resume inflow.

2) Accuracy metrics: The evaluation metrics utilized for gauging the performance of the screening mechanism encompassed.

3) Accuracy: The mechanism exhibited an accuracy rate of 90%, indicating a high precision in identifying resumes that align with job requirements.

4) Precision: Precision, representing the percentage of correctly identified relevant resumes out of the total identified as relevant, reached 85%.

5) Recall: The recall rate, measuring the percentage of relevant resumes correctly identified out of the total relevant resumes, achieved a commendable 75%.

6) Ranking consistency: The ranking mechanism displayed consistent performance, ensuring that resumes were consistently and accurately prioritized based on their alignment with the job description.

7) Efficiency in large-scale processing: Experimental outcomes suggested that the proposed solution maintains efficiency even as the dataset scales. This scalability aspect is crucial for handling real-world scenarios involving a substantial volume of incoming resumes.

8) Impact on workload: The implementation of the automated screening mechanism resulted in a substantial reduction in the workload of the initial screening team. This points to its potential in enhancing the efficiency of the early stages of the hiring process.

9) Robustness to updates: The generated embeddings, once created, demonstrated robustness to updates in resumes. Unless there were significant changes in the content, the same embeddings could be reused for subsequent screenings, contributing to processing efficiency.

Fig. 4 shows the output of an automated resume screening system that uses S-BERT to calculate the similarity between each resume and a job description. The system first extracts words from the resumes and forms sentences from them. The top six rectangular brackets contain words extracted from different resumes and the sentences formed by them in rectangular brackets. Then, it uses S-BERT to calculate the similarity between each sentence and the job description. The second-last line shows the similarity scores between the job description and each of the 30 resumes, and the last line shows the execution time in seconds. The system can be used to quickly identify resumes that are most relevant to a job opening. This can save recruiters time and help them find the best candidates for the job.

The graph illustrates the correlation between the number of resumes and the screening time in an S-BERT-based automated resume screening system. As the number of resumes increases, the screening time also rises, but not in a linear fashion. For instance, screening five resumes takes about 0.3 seconds, while screening 10 resumes takes approximately 0.9 seconds, and screening 30 resumes extends to about 4.9 seconds. This non-linear trend implies that the screening time increases at a varying rate.

Several explanations could account for this phenomenon. One possibility is that the efficiency of the automated screening algorithm improves with experience, allowing it to swiftly identify and discard unsuitable resumes by learning patterns. Conversely, in traditional mechanisms, screeners might experience fatigue with increased resume volume, resulting in slower screening times. In summary, the data indicates that the number of resumes significantly influences screening time. HR managers adopting this solution should consider this relationship when planning their workflow.

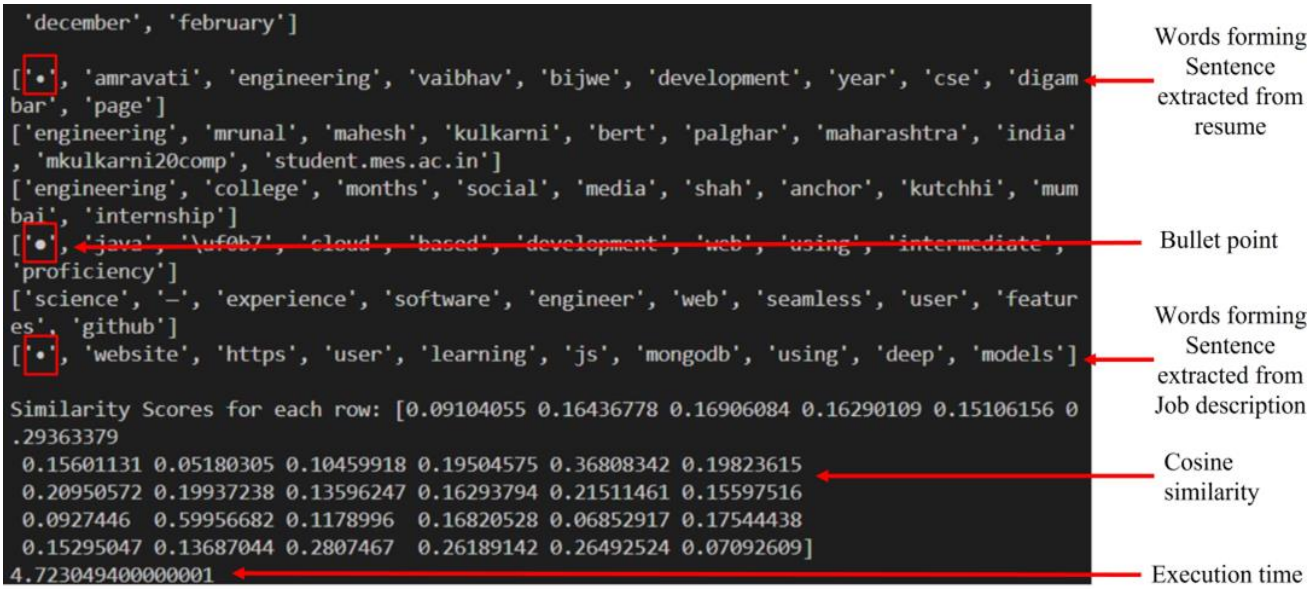


Fig. 4. Output screenshot of the proposed automated resume screening system using S-BERT.

Error computation and calculations were conducted based on feedback from the HR, who serves as the final end user of the developed system. This proposed work curated seven job profiles from the HR manager and selected 10 resumes for each of the seven job descriptions. Some resumes were common, given that candidates had multiple eligibilities. Throughout the system, this approach ranked the top three candidates and asked HR to rank the candidates based on their experience. To assess the system, the proposed work assigned numerical labels (1 to 10) to all resumes in each case, and HR provided rankings that precisely matched our numbering system. Table I presents the results obtained through HR rankings versus the results obtained with the proposed Automated Screening System (AS).

TABLE I. VALIDATION TABLE FOR SEVEN DISTINCT JOB DESCRIPTIONS (JD) COMPARING RANKINGS FROM HR (HUMAN RESOURCES MANAGER) AND AS (AUTOMATED SYSTEM)

Type of Screening	Rank 1	Rank 2	Rank 3
HR (JD1)	5	7	8
AS (JD1)	5	8	7
HR (JD2)	6	7	9
AS (JD2)	6	7	9
HR (JD3)	6	3	2
AS (JD3)	3	6	2
HR (JD4)	5	7	3
AS (JD4)	5	7	8
HR (JD5)	4	3	2
AS (JD5)	4	1	7
HR (JD6)	3	4	9
AS (JD6)	3	4	9
HR (JD7)	3	8	9
AS (JD7)	3	8	4

The error calculations follow the proposed rule base outlined as follows:

If a completely new resume appears on the list, not present in the best three resumes suggested by the HR manager, proposed solution assigns values based on the rank of the resume. If the 1st ranked resume is replaced, a value of -1 is assigned, indicating a 100% error, aligning with our acceptable 3 resumes policy. If a totally new resume appears at the 2nd rank, the error is set to -0.9, where the minus sign indicates an issue, and the value between 0 to 1 indicates the extent of the error in percentage; in this case, 90% unacceptable is denoted by -0.9. For the 3rd rank, the value is reduced to -0.8 as it is the last resume on the list that was missed.

Our primary objective is to prioritize bringing all 3 recommended resumes to the output and ensuring they are in the correct order: 1, 2, and 3.

In another case, when resumes match in the top 3 but are not in the correct order, the proposed solution provides a table to validate this scenario. For Case 1, where rank 1 by HR corresponds to ranks 1, 2, and 3 by AS, the resulting errors are 0, -0.4, and -0.7. This implies that if the rank 1 resume in HR matches the rank 1 in AS, there is no error (0). If the rank 1 HR resume is found at Rank 2 in AS, the error is -0.4, indicating a 40% error. Finally, if the rank 1 HR resume is found at Rank 3 in AS, the error is -0.7.

Similarly, for rank 2 HR, the values are -0.3 (Rank 1 AS), 0 (Rank 2 AS), and -0.3 (Rank 3 AS). The magnitude of error is reduced from 0.4 to 0.3 as the position is lowered to rank 2.

For rank 3 HR, the following errors were computed: -0.4 for Rank 1 AS, -0.2 for Rank 2 AS, and 0 for Rank 3 as this approach is utilized for error computation to validate the effectiveness of the automated screening technique.

The scatter plot in Fig. 5 illustrates the effectiveness of the proposed S-BERT-based automated resume screening system across seven distinct job descriptions: Software Engineer, Data

Scientist, Product Manager, Marketing Manager, Sales Representative, Customer Support Representative, and Human Resources Manager. Accuracy is quantified as the percentage of resumes correctly classified as relevant or not for each job, based on HR evaluations. The proposed mechanism attains an overall accuracy of 90% across all job descriptions, with the Sales Representative role having the lowest accuracy at 43.33%, and the highest accuracy observed for Data Scientist, Product Manager, and Customer Support Representative. The overall error rate of the system is 21.42%, indicating S-BERT's efficacy in automating the resume screening process.

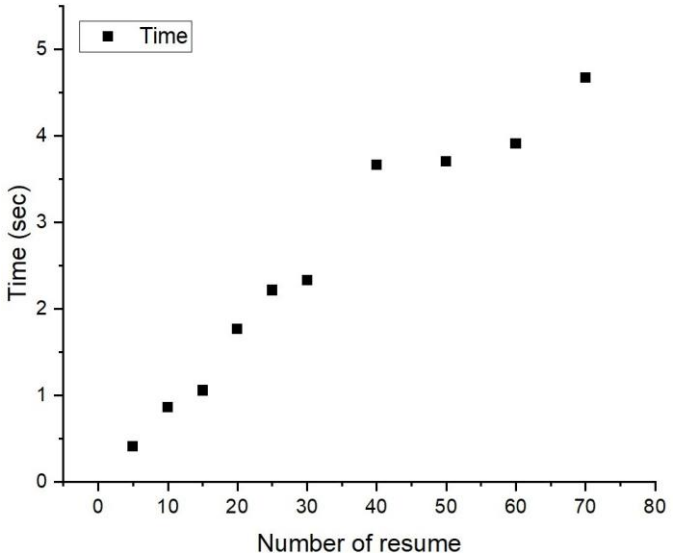


Fig. 5. Plot of number of resumes vs. time required for screening.

This S-BERT-based automated screening tool offers HR managers a means to efficiently handle large volumes of resumes while maintaining high accuracy. This efficiency allows HR managers to redirect their focus toward critical tasks like candidate interviews and hiring decisions. The presented bar graph represents the average accuracy of the system. Future opportunities include testing the mechanism on a larger resume dataset to affirm accuracy and generalizability, extending its application to screen for diverse job types, and developing a user-friendly web or mobile app to enhance accessibility for HR managers.

V. Discussions

The proposed automated resume screening mechanism presents a paradigm shift in the recruitment landscape, offering distinct advantages over traditional methods. Firstly, its reliance on Natural Language Processing (NLP) techniques empowers the system to extract nuanced information from resumes, ensuring resilience against tactics like keyword stuffing [10]. This not only enhances the accuracy of candidate evaluation but also reduces the potential for falsification in the initial screening stages [4]. Secondly, the incorporation of a cosine distance matrix for ranking resumes based on alignment with the job description adds a layer of sophistication [22]. By prioritizing relevance over specific keywords, the system ensures that the most suitable candidates rise to the top [18]. This is a crucial departure from conventional keyword-based screening, aligning the system with the broader goal of identifying candidates based on their actual qualifications and experience [16]. Despite these advantages, it's essential to acknowledge the system's developmental stage and the inherent biases that can be present in NLP models [25]. Rigorous development and evaluation processes are vital to mitigate biases and ensure fairness. Moreover, the system's performance hinges on the quality of training data, emphasizing the need for diverse and comprehensive datasets [8] [13]. Importantly, while the automated screening mechanism streamlines the initial filtering process, it doesn't replace human judgment. HR professionals must review top-ranked resumes for final decisions, emphasizing the collaborative nature of technology and human expertise in the hiring process [2].

The proposed automated resume screening mechanism opens avenues for future research in the realm of human resources. Firstly, it can be a valuable tool for studying factors contributing to resume success [19]. Analyzing top-ranked resumes can provide insights into the skills and experiences highly valued by employers, guiding both candidates and educators in aligning with industry expectations [3]. Secondly, the system acts as a catalyst for developing advanced methods to enhance the accuracy and fairness of automated resume screening. Future research could focus on refining NLP techniques, making them more robust to biases and capable of capturing richer contextual information from resumes [24]. Thirdly, the proposed mechanism sets the stage for the development of new tools and resources for HR professionals. Dashboards visualizing automated screening results could empower HR teams with quick insights, making the hiring process more transparent and efficient [20]. The proposed automated resume screening mechanism not only addresses current challenges in hiring but also paves the way for future innovations and research in the dynamic field of human resources. Future endeavors will likely focus on refining the system's performance, addressing identified limitations, and exploring new frontiers in the evolving intersection of technology and recruitment practices.

VI. Conclusions

Proposed method offers a transformative solution to the challenges faced in contemporary hiring processes. The conventional method of manual screening, while crucial, is marred by inefficiencies, biases, and the overwhelming influx of resumes. This research introduces an automated screening mechanism leveraging state-of-the-art natural language processing (NLP) techniques, particularly S-BERT, to revolutionize the initial phases of candidate evaluation.

The results of our evaluation on a dataset of 223 resumes reveal the remarkable efficiency of the proposed methodology. With a screening speed of 0.233 seconds per resume, the system showcased practical applicability in scenarios with substantial resume inflow. The accuracy metrics demonstrated a high precision in identifying relevant resumes, with an accuracy of 90%. The ranking mechanism exhibited consistency, ensuring resumes were prioritized accurately based on their alignment with job descriptions.

Beyond quantitative metrics, our automated screening mechanism significantly reduces the workload on initial screening teams, presenting a scalable solution for handling

large volumes of incoming resumes. The robustness of generated embeddings to updates in resumes enhances processing efficiency, allowing for reuse unless there are substantial content changes. This work not only contributes to the efficiency of the hiring process but also aligns with broader societal goals. By automating and streamlining the screening process, this manuscript contribute to making hiring practices more efficient, transparent, and accessible. Moreover, the adoption of advanced NLP techniques like S-BERT helps mitigate biases and promotes diversity and inclusivity in candidate selection.

As one move forward, the implications of this research extend beyond the immediate context. The automated screening mechanism presented here not only serves as a tool for HR professionals but also as a beacon for future developments in smart hiring practices. The integration of cutting-edge NLP models signifies a step toward a future where technology enhances, rather than hinders, the human aspect of hiring.

DECLARATION OF STATEMENTS

Author contribution: The conceptualization was jointly undertaken by Asmita Deshmukh (AD) and Anjali Raut (AR). AD was responsible for data collection, coding, and experimentation. Additionally, AD took the lead in preparing the initial draft of the manuscript, while AR handled corrections. Furthermore, data analysis and graphic design were conducted by AD.

Data and Code availability: Upon a reasonable request, both the data and code utilized in this research will be provided.

Funding: This study is not linked to any funding sources.

Conflict of interest and Competing interests: The authors affirm that they have no conflicts of interest or competing interests to disclose.

REFERENCES

[1] V. Sinha and P. Thaly, “A review on changing trend of recruitment practice to enhance the quality of hiring in global organizations,” Management: journal of contemporary management issues, vol. 18, no. 2, pp. 141–156, 2013.

[2] D. S. Chapman and J. Webster, “The use of technologies in the recruiting, screening, and selection processes for job candidates,” International journal of selection and assessment, vol. 11, no. 2-3, pp. 113–120, 2003.

[3] M. Travis, Mastering the art of recruiting: how to hire the right candidate for the job. Bloomsbury Publishing USA, 2015.

[4] S. K. Kopparapu, “Automatic extraction of usable information from unstructured resumes to aid search,” in 2010 IEEE International Conference on Progress in Informatics and Computing, vol. 1, 2010, pp. 99–103.

[5] R. T. Hassan and N. S. Ahmed, “Evaluating of efficacy semantic similarity methods for comparison of academic thesis and dissertation texts,” Science Journal of University of Zakho, vol. 11, no. 3, pp. 396–402, 2023.

[6] E. Derous and A. M. Ryan, “When your resume is (not) turning you down: Modelling ethnic bias in resume screening,” Human Resource Management Journal, vol. 29, no. 2, pp. 113–130, 2019.

[7] S. Nabi, “Comparative analysis of AI Vs Human based hiring process: A Survey,” in 2023 International Conference on Computational Intelligence and Knowledge Economy (ICCIKE), 2023, pp. 432–437.

[8] A. V. Gundlapalli, M. E. Carter, M. Palmer, T. Ginter, A. Redd, S. Pickard, S. Shen, B. South, G. Divita, and S. Duvall, “Using natural language processing on the free text of clinical documents to screen for evidence of homelessness among US veterans,” in AMIA Annual Symposium Proceedings, vol. 2013, 2013, p. 537.

[9] D. J. Feller, J. Zucker, M. T. Yin, P. Gordon, and N. Elhadad, “Using clinical notes and natural language processing for automated HIV risk assessment,” Journal of acquired immune deficiency syndromes (1999), vol. 77, no. 2, p. 160, 2018.

[10] R. Devika, S. V. Vairavasundaram, C. S. J. Mahenthar, V. Varadarajan, and K. Kotecha, "A Deep Learning Model Based on BERT and Sentence Transformer for Semantic Keyphrase Extraction on Big Social Data," IEEE Access, vol. 9, pp. 165252–165261, 2021.

[11] J. Nayor, L. F. Borges, S. Goryachev, V. S. Gainer, and J. R. Saltzman, “Natural language processing accurately calculates adenoma and sessile serrated polyp detection rates,” Digestive diseases and sciences, vol. 63, pp. 1794–1800, 2018.

[12] A. Mukherjee, "Resume Ranking and Shortlisting with DistilBERT and XLM," 2024 IEEE International Conference for Women in Innovation, Technology & Entrepreneurship (ICWITE), IEEE, pp. 301-304, 2024.

[13] H. M. Trivedi, M. Panahiazar, A. Liang, D. Lituiev, P. Chang, J. H. Sohn, Y.-Y. Chen, B. L. Franc, B. Joe, and D. Hadley, “Large scale semi-automated labeling of routine free-text clinical records for deep learning,” Journal of digital imaging, vol. 32, pp. 30–37, 2019.

[14] B. A. Sherazi, S. Laeer, S. Krutisch, A. Dabidian, S. Schlottau, and E. Obarcanin, “Functions of mHealth Diabetes Apps That Enable the Provision of Pharmaceutical Care: Criteria Development and Evaluation of Popular Apps,” International Journal of Environmental Research and Public Health, vol. 20, no. 1, p. 64, 2022.

[15] P. K. Roy, S. S. Chowdhary, and R. Bhatia, “A Machine Learning approach for automation of Resume Recommendation system,” Procedia Computer Science, vol. 167, pp. 2318–2327, 2020.

[16] Y. S. Swarupa and S. Aruna, "Natural Language Processing for Resume Screening," zkginternational.com, 2024.

[17] M. K. Delimayanti, M. Laya, B. Warsuta, M. B. Faydhur-rahman, M. A. Khairuddin, H. Ghoyati, A. Mardiyono, and R. F. Naryanto, “Web-Based Movie Recommendation System using Content-Based Filtering and KNN Algorithm,” in 2022 9th International Conference on Information Technology, Computer, and Electrical Engineering (ICITACEE), 2022, pp. 314–318.

[18] D. Pawade, T. Joshi, and S. Parkhe, "Survey on Resume and Job Profile Matching System," 2023 6th International Conference on Computing, Communication and Automation (ICCCA), IEEE, 2023.

[19] I. G. Ndukwe, C. E. Amadi, L. M. Nkomo, and B. K. Daniel, "Automatic Grading System Using Sentence-BERT Network," Artificial Intelligence in Education: 21st International Conference, AIED 2020, Springer, pp. 224–227, 2020.

[20] G. M. GR, S. Abhi, and R. Agarwal, "A Hybrid Resume Parser and Matcher using RegEx and NER," 2023 International Conference on Computer Communication and Informatics (ICCCI), IEEE, 2023.

[21] H. Choi, J. Kim, S. Joe, and Y. Gwon, "Evaluation of BERT and ALBERT sentence embedding performance on downstream NLP tasks," 2020 25th International Conference on Pattern Recognition (ICPR), IEEE, pp. 5482–5487, 2021.

[22] V. James, A. Kulkarni, and R. Agarwal, "Resume Shortlisting and Ranking with Transformers," International Conference on Intelligent Systems and Machine Learning, Springer, pp. 99–108, 2022.

[23] B. Wang and C. C. J. Kuo, "SBERT-WK: A sentence embedding method by dissecting BERT-based word models," IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 28, pp. 2146–2157, 2020.

[24] S. Kinger, D. Kinger, S. Thakkar, and D. Bhake, "Towards smarter hiring: resume parsing and ranking with YOLOv5 and DistilBERT," Multimedia Tools and Applications, Springer, 2024.

[25] N. Reimers and I. Gurevych, "Sentence-BERT: Sentence embeddings using siamese BERT-networks," arXiv preprint arXiv:1908.10084, 2019.

[26] J. Seo, S. Lee, L. Liu, and W. Choi, "TA-SBERT: Token Attention sentence-BERT for improving sentence representation," IEEE Access, vol. 10, pp. 39119–39128, 2022.

[27] Y. Chu, H. Cao, Y. Diao, and H. Lin, "Refined SBERT: Representing sentence BERT in manifold space," Neurocomputing, vol. 555, p. 126453, 2023.

[28] Vanetik and G. Kogan, "Job vacancy ranking with sentence embeddings, keywords, and named entities," Information, vol. 14, no. 8, p. 468, 2023.

[29] S. Yu, J. Su, and D. Luo, "Improving BERT-Based Text Classification with Auxiliary Sentence and Domain Knowledge," IEEE Access, vol. 7, pp. 176600–176612, 2019.

[30] C. Liu, W. Zhu, X. Zhang, and Q. Zhai, "Sentence Part-Enhanced BERT with Respect to Downstream Tasks," Complex & Intelligent Systems, vol. 9, no. 1, pp. 463–474, 2023.

[31] M. Alsuhaibani, "Deep Learning-based Sentence Embeddings using BERT for Textual Entailment," International Journal of Advanced Computer Science and Applications, vol. 14, no. 8, 2023.

Machine Learning Techniques for Protecting Intelligent Vehicles in Intelligent Transport Systems

Yuan Chen
College of Intelligent Transportation, Hunan Communication Polytechnic, Changsha, 410132, China

***Abstract*—Intelligent transport system (ITS) is the development direction of future transport systems, in which intelligent vehicles are the key components. In order to protect the safety of intelligent vehicles, machine learning techniques are widely used in ITS. For intelligent protection in ITS, the study introduces an improved driving behaviour modelling method based on Bagging Gaussian Process Regression. Meanwhile, to further promote the accuracy of driving behaviour modelling and prediction, Convolutional Neural Network-Long and Short-term Memory Network-Gaussian Process Regression are used for effective feature extraction. The results show that in the straight overtaking scenario, the mean absolute error, root mean square error and maximum absolute error of the improved Bagging Gaussian process regression method are 0.5241, 0.9547 and 10.7705, respectively. In the corner obstacle avoidance scenario, the improved Bagging Gaussian process regression method is only 0.6527, 0.9436 and 14.7531. Besides, the mean absolute error of the Convolutional Neural Network-Long and Short-term Memory Network-Gaussian process regression algorithm is only 0.0387 in the case of the input temporal image frame number of 5. This denoted that the method put forward in the study can provide a more accurate and robust modeling and prediction of driving behaviours in complex traffic environments, and it has a high application potential in the field of safety and protection of intelligent vehicles.**

Keywords*—Intelligent vehicle protection; machine learning techniques; Gaussian Process Regression; convolutional neural networks; long and short-term memory networks

I. Introduction

With the rapid development of science and technology, intelligent transport systems have become a hot topic in today's society. In the intelligent transport system, intelligent vehicles play a crucial role [1]. However, with the popularity of intelligent vehicles, their safety problems are becoming more and more prominent. Driving behaviour modelling methods can help intelligent vehicles achieve safer and more efficient driving by predicting and simulating vehicle and driver behaviours. Traditional driving behaviour modelling methods include rule-based methods, probabilistic model-based methods, and so on. However, rule-based methods are difficult to adapt to the complex and changing traffic environment, and the formulation of rules often requires a lot of manual intervention and lacks adaptivity. Meanwhile, probabilistic model-based methods need to learn from a large amount of historical data, and model building requires high computational resources and time costs [2, 3]. Aiming at such problems, the study will explore the machine learning techniques for protecting intelligent vehicles in ITS, and introduce an improved driving behaviour modelling method based on Bagging Gaussian Process Regression (Bagging GPR), in order to achieve more accurate driving behaviour modelling and prediction In order to achieve more accurate modelling and prediction of driving behaviour, Convolutional Neural Network-Long Short-Term Memory-Gaussian Process Regression (CNN-LSTM-GPR) is used for effective feature extraction, to achieve good results in the field of intelligent vehicle security protection field to achieve good results. The study is composed of six main sections. The introduction is given in Section I. Section II gives details about the previous research work. Section III introduces the advanced driving behaviour modelling methods, the improved Bagging GPR driving behaviour modelling method is introduced in the first section, and the CNN-LSTM-GPR-based feature extraction method is presented in the second section. Section IV focuses on the experimental validation of the studied proposed method. Section V and Section VI summarize and discuss the experimental results and propose future directions. The contribution of the research is the introduction of advanced driving behaviour modelling methods, which help to improve the accuracy of driving behaviour modelling and prediction. These methods have important application value in the safety protection of intelligent vehicles and can effectively protect the safe driving of intelligent vehicles.

II. Literature Review

GPR is a nonparametric model for regression analysis of data with a Gaussian process prior, which is widely used. Deringer V L et al. put forward the GPR machine learning method to investigate the nature of atoms in chemistry and materials science. The method constructs an interatomic potential or force field using a Gaussian approximation of the potential framework and can be fitted with arbitrary scalars, vectors and tensors. The outcomes denoted that the method can effectively and accurately predict atomic properties [4]. Liu and other scholars developed two related data-driven models for the prediction of the effective capacity of lithium-ion batteries with a systematic understanding of the covariance function in GPR. The results illustrated that the raised models can accurately predict the battery capacity under different cycling modes [5]. Band et al. In order to accurately predict the groundwater level in arid areas, the researchers proposed to use the Support Vector Regression, GPR and its combination with Wavelet Transform for the prediction. The results show that the wavelet transform method combined with GPR has a strong performance advantage in GWL prediction compared to GPR [6]. Hewing L et al. researchers raised a model prediction control method with GPR for modelling and control of nonlinear dynamical systems. The method integrates the nominal system with the additional nonlinear part of the dynamics modelled as GPR. The findings

indicated that the method can effectively assess the residual uncertainty [7]. Zhang Y's team developed a GPR-based predictive model for the optimisation of the magneto-thermal effect and relative cooling power of ferromagnetic lanthanum manganites. The model mainly searches for the correlation between RCP and lattice parameters by statistical learning. The findings denoted that the GPR model can predict RCP values efficiently and cost-effectively [8].

Driver behaviour recognition can detect and correct driver violations in time, reduce the accident rate and play an important role in vehicle protection. Xing and other scholars proposed a deep convolutional neural network-based driver activity recognition system for driver behaviour recognition and safe driving. The system utilizes a Gaussian mixture model to segment the original image to extract the driver's body from the background. The outcomes denoted that the raised system can accurately recognize seven driver activities with an average accuracy of 81.6% [9]. McDonald et al. found that an advanced driver assistance system needs to have a proper understanding of the driver's state, and therefore proposed a method for inference of self-vehicle driver's intention. The results show that the interaction between these modules has a significant impact on the lane change intention inference system [10]. Kabzan et al. researchers put forward a learning-based method to the problem of self-driving racing car control. The method was improved using a simple nominal vehicle model and GPR was used to account for model uncertainty. Test results show that on a full-size AMZ driverless race car, the method can improve the model and cut down the lap time by 10% [11]. Researchers such as Mozaffari have proposed a deep learning-based method for the problem of behavioural prediction of autonomous vehicles, which predicts the future state of a nearby vehicle by observing the surrounding environment. This approach provides better performance in more complex environments than traditional methods [12]. Hoel C J's team introduced a generalised framework that combines planning and learning in order to better address the tactical decision-making challenges of autonomous driving. The approach, based on the AlphaGo Zero algorithm and extended to continuous state spaces, shows better performance than the baseline approach [13].

In summary, numerous researchers and scholars have carried out extensive studies on GPR methods as well as driver behaviour recognition methods, but few scholars have applied improved GPR methods to driver behaviour recognition. Therefore, the study introduces the advanced GPR method for driver behaviour modelling, which is utilised with a view to achieving intelligent vehicle prediction and avoiding potential hazards.

III. Results

To protect the safety of intelligent vehicles, the study proposes an improved Bagging GPR driving behaviour modelling method, which uses GPR to design the base regressor and Bagging method to improve the whole effectiveness of the algorithm, as well as self-sampling method for the generation of new datasets. In order to achieve more accurate driving behaviour modelling and prediction, the study adopts CNN-LSTM-GPR for effective feature extraction, which mainly uses CNN to extract features from the input time-series image data, and inputs the extracted features into LSTM for processing Finally, the processed features are used as inputs for model fitting and prediction using GPR.

A. GPR-Based Modelling of Intelligent Vehicle Driving Behaviour

In intelligent transportation systems, to deal with different complex driving environments and promote adaptability and safety, driving behaviour modelling techniques play a key role. Based on this, the study introduces the GPR algorithm, which is a non-parametric Bayesian learning method that can be used for regression problems and is suitable for dealing with problems with uncertainty [14, 15]. In driving behaviour modelling, the perceived state quantities usually include the vehicle's attitude, speed, acceleration and other sensor data, and the GPR algorithm can be used to learn these perceived state quantities to build a model of the driving behaviour and perform real-time prediction and control. Therefore, the GPR algorithm is an effective method for modelling driving behaviour based on perceptual state quantities. However, the basic GPR algorithm has the issue of imitating a large amount of data and uneven distribution, to improve the algorithm's effectiveness, the study further proposes an improved Bagging GPR approach for driving behaviour modelling. The method firstly uses GPR to design the base regressor, and then uses the Bagging method to further improve the algorithm's whole effectiveness. The main process is to first randomly sample the driving behaviour in the training set by self-sampling method with playback, then train N GPR base learners using GPR, and lastly integrate the outputs of each base learner by averaging method to obtain the predicted values [16]. The driving behaviour modelling framework based on Bagging GPR is shown in Fig. 1.

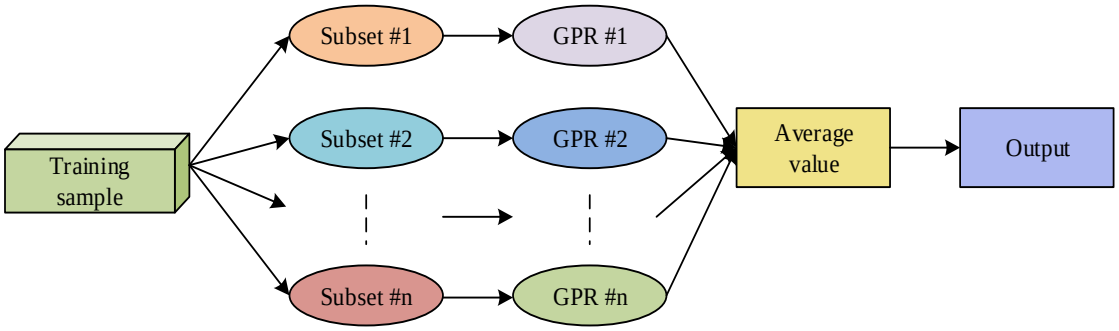


Fig. 1. A driving behaviour modeling framework based on Bagging GPR.

In the Bagging GPR-based driving behaviour modelling method, the self-sampling method is a method to generate a new dataset from an initial dataset. The specific steps of the method are, firstly, given a dataset containing m samples. Then, m random sampling operations are performed. In each sampling, a sample is randomly selected and placed into the new dataset, which is then placed back into the initial dataset. According to the limiting formula for probability estimation of the self-sampling method, it can be deduced that about 36.82% of the samples in the initial dataset never appeared in the new dataset, while about 63.2% of the samples appeared in the new dataset, which is calculated as shown in Eq. (1).

$$\lim_{m\to\infty}(1-\frac{1}{m})^m=\frac{1}{e} \tag{1}$$

A Gaussian process is a set of random variables, where any finite amount of random variables satisfy the joint Gaussian distribution.GPR mainly uses kernel methods for nonlinear mapping, and is a nonparametric model with good generalisation performance and global mapping ability. Given a training sample set D, where each element is a binary group containing an input vector s_i and an output vector a_i . The GPR model's output is worked out as shown in Eq. (2).

$$a=f(s)+\varepsilon \tag{2}$$

In Eq. (2), ε denotes the error, which satisfies the Gaussian distribution, and its representation is shown in Eq. (3).

$$\varepsilon\sim N(0,\delta_n^2) \tag{3}$$

In Eq. (3), δ_n^2 denotes the variance of the output error. a The prior distribution of is indicated as expressed in Eq. (4).

$$a\sim N(0,\delta_n^2+K) \tag{4}$$

In Eq. (4), K represents the covariance matrix. Based on the prediction samples as well as the training samples a joint Gaussian prior distribution can be obtained, which is represented as shown in Eq. (5).

$$\begin{bmatrix}a\\a*\end{bmatrix}\sim N\left(0,\begin{bmatrix}K(s,s)+\delta_n^2 & K(s,s*)\\ K(s,s*)^T & K(s*,s*)\end{bmatrix}\right) \tag{5}$$

In Eq. (5), $s*$ and $a*$ denote the input vector and output vector of prediction, respectively. $K(s,s*)$ denotes the prediction and the training samples' the covariance matrix, and $K(s*,s*)$ denotes the prediction sample's self-covariance matrix. With the kernel function, the construction of the covariance matrix is mainly carried out. The kernel function chosen for the study is the radial basis function kernel, which is represented as shown in Eq. (6).

$$K(s,s*)=\alpha^2\exp(-\frac{(s-s*)^2}{2l^2}) \tag{6}$$

In Eq. (6), l is the kernel width of the radial basis function kernel, and α represents the hyperparameters, and the optimal hyperparameters are mainly obtained by the great likelihood method. After obtaining the optimal α , the the predicted value $a*$'s posteriori probability will be got based on the new $s*$, which is calculated as shown in Eq. (7).

$$p(a*|a)=\frac{p(a,a*)}{p(a)} \tag{7}$$

In Eq. (8), $a*$'s distribution is calculated.

$$a*\sim N(\hat{y}(s*),\hat{\delta}(s*)) \tag{8}$$

In Eq. (8), $\hat{y}(s*)$ represents the mean value, which is calculated as denoted in Eq. (9).

$$\hat{y}(s*)=K^T(s*)\cdot(K+\delta_n^2)^{-1}\cdot a \tag{9}$$

$\hat{\delta}(s*)$ denotes the variance, which is calculated as shown in Eq. (10).

$$\hat{\delta}(s*)=K(s*,s*)-K^T(s*)\cdot(K+\delta_n^2)^{-1}\cdot K(s*) \tag{10}$$

The steps of GPR application are shown in Fig. 2.

Aiming at the characteristics of large volume and uneven distribution of driving behaviour data, the study adopts the Bagging algorithm to promote of GPR algorithm's effectiveness. Bagging is an integrated learning method used to improve the accuracy of learning algorithms, and integrating and combining it can cut down the variance of the output, and promote the accuracy and stability of the algorithm. The Bagging method contains three parts, namely, sampling, training the base learner, and combining the output of three parts, and averaging method is used to obtain the output of the strong regressor, which is calculated as expressed in Eq. (11).

$$\hat{y}(s)=\frac{1}{N}\sum_{k=1}^{N}GPR_k(s) \tag{11}$$

The study proposes a Bagging GPR driving behaviour modelling method with improved sampling, due to the Bagging GPR algorithm's insensitivity to samples with large learning errors. The method raises the comprehensive effectiveness of the integrated regressor by increasing the sampling probability of samples, increasing the fluctuation of training samples, and increasing the attention of the base learner to samples with large training errors, which further reduces the maximum prediction error of the proposed algorithm [17]. The flow of the improved Bagging GPR approach for driving behaviour modelling is shown in Fig. 3.

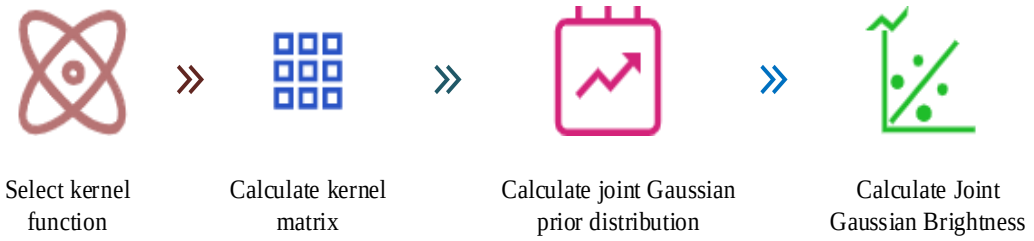


Fig. 2. Application steps of the GPR.

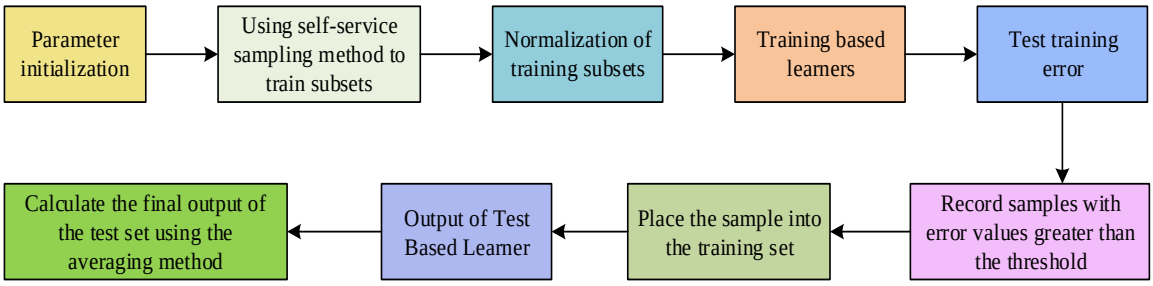


Fig. 3. Flow chart of improved Bagging GPR driving behaviour modelling method.

B. *Intelligent Vehicle Feature Extraction Based on CNN-LSTM-GPR*

To achieve driving behaviour modelling and prediction, effective feature extraction is required as input. Considering the characteristics of human drivers, driving behaviour learning should focus on the correlation between before and after states and actions. The study focuses on CNN-LSTM for time-series image feature extraction and fusion. Among them, CNN is a dedicated algorithm for image processing. CNN is not only able to efficiently downsize a large amount of image data into a small amount of feature data, but is also able to effectively extract the features related to a specific task through training. Convolutional, pooling, and fully connected layers are composed of CNN networks [18]. The output of the l convolutional layer is calculated as shown in Eq. (12).

$$C^l = \sigma(z^l) \tag{12}$$

In Eq. (12), σ represents the activation function, and z^l denotes the variables of the activation function of the l layer. The output of the pooling layer at l is calculated as shown in Eq. (13).

$$D^l = pool(C^l) \tag{13}$$

The output of the fully connected layer is calculated as shown in Eq. (14).

$$F^n = \sigma(F^{n-1} * W^n + b^n) \tag{14}$$

In Eq. (14), F^{n-1}, W^n, and b^n denote the output, weight matrix, and bias of the nth fully connected layer, respectively. The LSTM network is mainly applied to deal with the temporal prediction problem, whose input is the features extracted by the CNN at the time t, and the output is the predicted output at the time t. The dimensionality of the output is mainly related to the amount of output nodes of the LSTM network. The LSTM model can be viewed as a stack of cell units, each of which controls the transfer of information through a specially designed "gate" structure. The output of each cell consists of state and implicit layers, while the output of each implicit layer is jointly determined by three gates, including the inputting gate, forgetting gate and outputting gate. Each Cell unit selectively remembers and forgets the information through these three gates, and then passes it to the next Cell unit [19]. The LSTM network's structure is shown in Fig. 4.

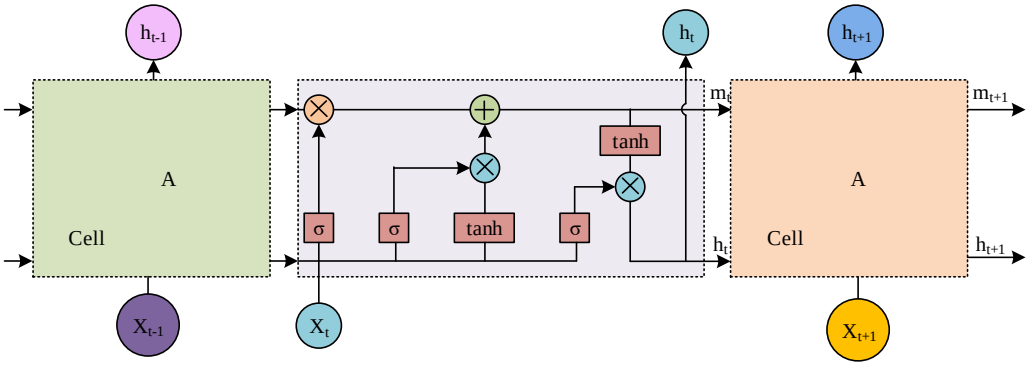


Fig. 4. Structure diagram of LSTM.

The CNN-LSTM's error loss function is mainly used to measure the degree of deviation between the labels and the output of the network. This loss function defines a criterion for evaluating the performance of the network and aims to cut down the difference between the actual and the desired output. During the training, the optimisation algorithm continuously adjusts the network parameters to minimise the loss function, thus making the network's prediction more accurate. Its calculation is shown in Eq. (15).

$$E = \frac{1}{N}\sum_{i=1}^{N}\left\|y_i - a_i\right\|_2^2 \tag{15}$$

In Eq. (15), y_i represents the CNN-LSTM network's output, a_i represents the labels of the expert demonstration teaching, and N is the size of the Batch size. The CNN-LSTM method considers the temporal correlation between image sequences while considering the image feature extraction. This method can promote the simulation accuracy of driver behaviour. The neural network's layer is a combination of convolutional and pooling layers. The specific configuration of the convolutional layers is as follows: the first, second, third, fourth, and fifth convolutional layers have a convolutional kernel size of 5 × 5, 5 × 5, 5 × 5, 3 × 3, and 3 × 3, respectively, and a feature map number of 24, 36, 48, 64 and 64, respectively. The first, second, fourth, and fifth convolutional layers have a pooling downsampling window size of 2 × 2 steps with a step size of 2, while the third layer has a pooling downsampling window size of 1 × 2 steps with a step size of 2. Subsequently these feature maps are fed into the fully-connected layer, which has a node count of 512. The fully connected layer is followed by two LSTM layers. Finally, the output of the LSTM layer is connected to the output layer, which outputs the normalized value of the steering wheel angle [20]. A schematic diagram of the CNN-LSTM feature extraction network structure is shown in Fig. 5.

The research focuses on combining CNN-LSTM with GPR to enhance the understanding of the mapping relationship between temporal features and driving behaviour. The core idea of the method is to utilize GPR to further optimize the features of the CNN-LSTM to promote the fully connected layer's structure. The CNN-LSTM accumulates errors layer-by-layer during the training process, which limits its generalisation ability. Since the fully connected layer is similar to that of CNN, its generalisation ability is limited. By using the GPR method, the perfect temporal feature extraction effect of CNN-LSTM is fully utilized, while the fitting mapping ability between temporal features and driving behaviour is further improved by GPR, to improve the comprehensive performance of the algorithm. The detailed process of the CNN-LSTM-GPR method is denoted in Fig. 6.

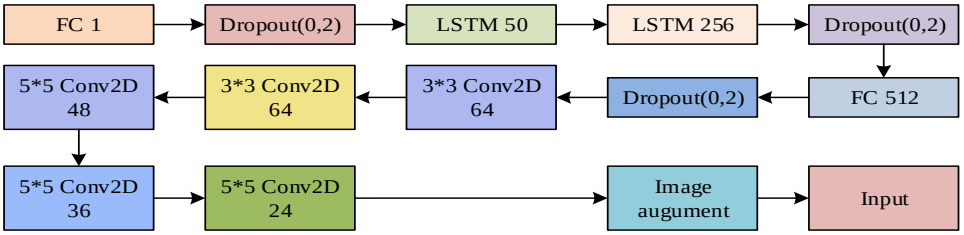


Fig. 5. Schematic diagram of CNN-LSTM feature extraction network structure.

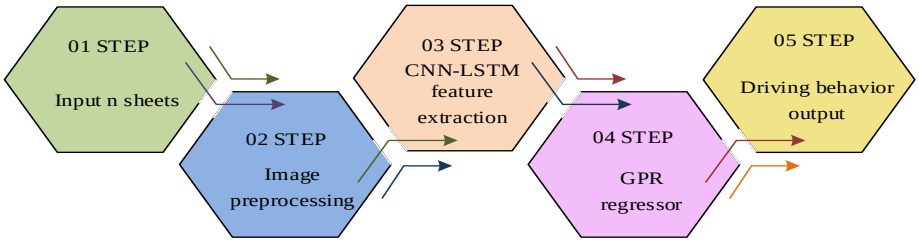


Fig. 6. The process of CNN-LSTM-GPR method.

IV. Intelligent Vehicle Protection Analysis Based on GPR Methodology

This chapter focuses on the experimental analysis of the improved Bagging GPR and the CNN-LSTM-GPR methods proposed by the study. Among them, for the improved Bagging GPR method, the study analyses the effect of driving behaviour modelling and verifies the performance from two scenarios, namely the straight overtaking scenario and the corner obstacle avoidance scenario. The study verifies the effectiveness of each algorithm from different input image frames, and compares the steering wheel corner prediction results of different algorithms to evidence the performance of the raised method.

A. Driving Behaviour Modelling Analysis Based on Improved Bagging GPR Methodology

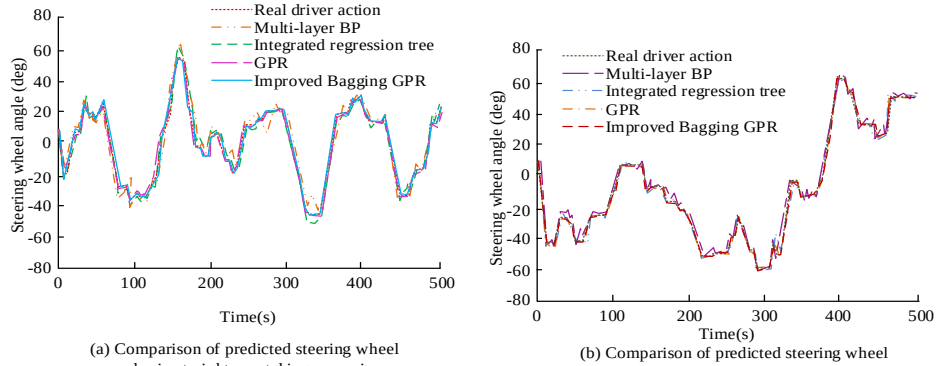
To verify the effectiveness of the improved Bagging GPR method, firstly, the driving behaviours of the straight overtaking scenario as well as the corner obstacle avoidance scenario are modelled and learned. At the same time, three driver behaviour modelling methods, namely, multi-layer Back Propagation Algorithm (BP), Integrated Regression Tree and GPR, are selected for performance comparison with them. The parameter settings of each method are expressed in Table I.

TABLE I. PARAMETER SETTINGS FOR EACH METHOD

Method	Project	Parameter
Multi-layer BP network	Number of nodes of the two hidden layers	75, 15
Integrated regression tree	Integrated learning cycle	100
	Learner	Regression tree
GPR	Kernel function	Gaussian kernel function
	Nuclear width	1
	Noise parameters	0.1
Improved Bagging GPR	Number of iterations of the Bagging	20
	The size of the Bagging	2000
	Error threshold	3

The steering wheel angle prediction results of different driving modelling methods for different scenarios are indicated in Fig. 7. From Fig. 7, in contrast with the remaining three modelling methods, the driving behaviour modelling method of the improved Bagging GPR has a better fitting performance with the actual steering wheel angle and a higher matching accuracy. Whereas, the multi-layer BP algorithm has the largest deviation from the actual steering wheel angle, representing its worst modelling performance. It indicates that the proposed algorithms in the study have high prediction accuracies in modelling driving behaviours in both straight overtaking scenarios as well as cornering obstacle avoidance scenarios.

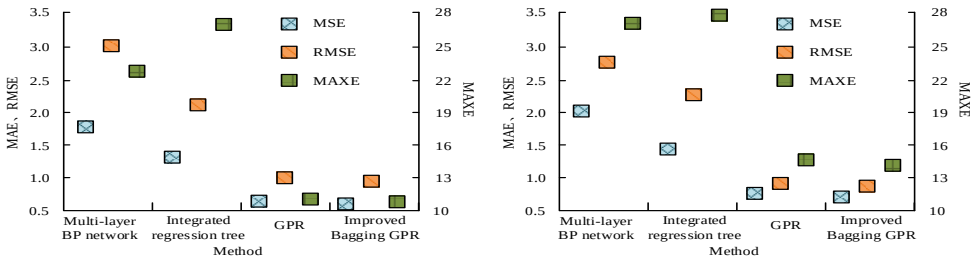
The experiments continue to use Mean Square Error (MSE), Root Mean Square Error (RMSE) and Maximum Absolute Error (MAXE) to experimentally validate the different driving behaviour modelling methods. The comparative results of the steering wheel corner prediction performance of each method in different scenarios are expressed in Fig. 8. From Fig. 8, the improved Bagging GPR method outperforms the remaining three methods for steering wheel angle prediction in the straight overtaking scenario as well as in the corner obstacle avoidance scenario. Among them, in the straight overtaking scenario, the MAE, RMSE, and MAXE of the improved Bagging GPR method are 0.5241, 0.9547, and 10.7705, respectively, whereas those of the multilayer BP algorithm are as high as 1.7763, 3.0334, and 23.0549, respectively. The integrated regression tree is as high as 1.2863, 2.1538, 27.349, and 1.2863, respectively, 2.1538, and 27.3626, respectively. The indexes of GPR are 0.5569, 0.9638, and 10.9934, respectively, which are improved by 0.0328, 0.0091, and 0.2184 compared with the improved Bagging GPR method. At the same time, in the corner obstacle avoidance scenario, the MAE, RMSE, and MAXE of the improved Bagging GPR method are 0.660, 0.660, and 0.660, respectively. MAE, RMSE, and MAXE are 0.6527, 0.9436, and 14.7531, respectively, which are 1.383, 1.8274, and 12.8996 less than the multi-layer BP algorithm, suggesting that the improved Bagging GPR method has better performance and stability when dealing with complex driving behaviour modelling problems.



(a) Comparison of predicted steering wheel angles in straight overtaking scenarios

(b) Comparison of predicted steering wheel angles in obstacle avoidance scenarios on curves

Fig. 7. Prediction results of steering wheel angles in different scenarios.



(a) The results of various indicators in the scenario of overtaking on the straight track

(b) The results of various indicators in the obstacle avoidance scene of bends

Fig. 8. Comparison of steering wheel angle prediction performance of various methods in different scenarios.

B. Feature Extraction Analysis Based on CNN-LSTM-GPR Algorithm

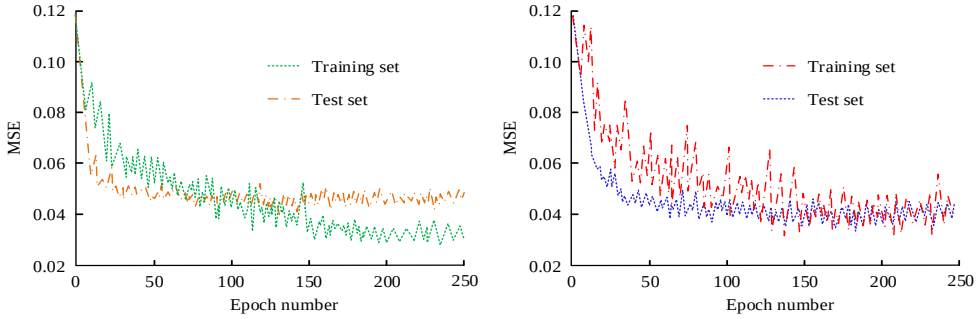
The study first uses a CNN-LSTM network to extract the features of various input sequence images. In terms of training parameter settings, the specific settings are denoted in Table II.

The experiments have been conducted on Apollo dataset using CNN-LSTM network with an input image sequence of 3 and an input image sequence of 5. The variation of the training error loss function is denoted in Fig. 9. The MSE of the training and test set for the case of the input image sequence of 3 is lower than the case of the input image sequence of 5. In particular, the MSE of the test set with an input image sequence of 3 converges to 0.0415 at about 25 iterations, which is 0.014 higher than the case with an input image sequence of 5. This indicates that, for the task of learning driving behaviours, the use of the CNN-LSTM network with longer input image sequences can extract the image features in a better way and help to improve the model's accuracy and generalization ability.

The experiments continued with GPR to fit the features for mapping driving behaviour. The CNN-LSTM-GPR algorithm's parameters were set as below: the kernel function was RBF, the kernel width parameter was 0.5, and the noise parameter was 0.3. The experiments were conducted using 50-dimensional features extracted by the CNN-LSTM on the Apollo dataset for driving behaviour Learning. To ensure the accuracy of the experiments, the study conducted a total of 10 experiments and took the effective average as the final experimental results. Meanwhile, MSE is chosen as the evaluation index of the experiment. The test outcomes of various driving behaviour learning methods under different input timing image frame numbers are shown in Fig. 10. The CNN-LSTM-GPR algorithm can obtain lower MSE under different input time-series image frame numbers, among which, under the input time-series image frame number of 3, the MSE of the CNN-LSTM-GPR algorithm is only 0.0405, which is 0.010 less than that of the CNN-LSTM algorithm under the input time-series image frame number of 5, the MSE of the CNN-LSTM-GPR algorithm is only 0.0405, which is 0.010 less than that of the CNN-LSTM algorithm. The MSE of the CNN-LSTM-GPR algorithm is 0.0387, which is reduced by 0.0023 in contrast with the CNN-LSTM algorithm. The MSE values of the individual algorithms for the input temporal image frame number of 5 are lower compared to the case where the input temporal image frame number is 3. It shows that the CNN-LSTM-GPR algorithm has higher accuracy in learning to mimic the driving behaviour of the temporal images and the performance of the algorithms is better at a higher number of input temporal image frames.

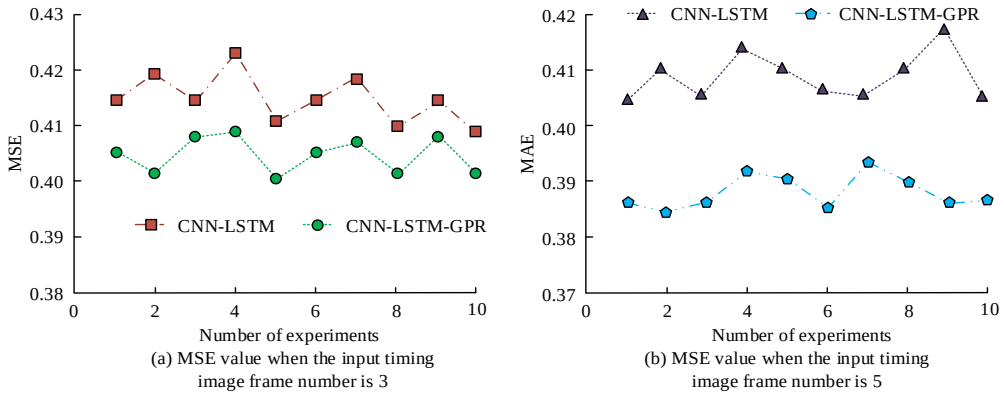
TABLE II. TRAINING PARAMETER SETTINGS

Project	Parameter
Learning rate	0.0001
Optimizer	Adam
Dropout	0.2
Batch size	20
Number of training samples	5000
Number of training rounds	200
Training time	9h



(a) Change in loss function of input timing image frame number 3 (b) Change in loss function of input timing image frame number 5

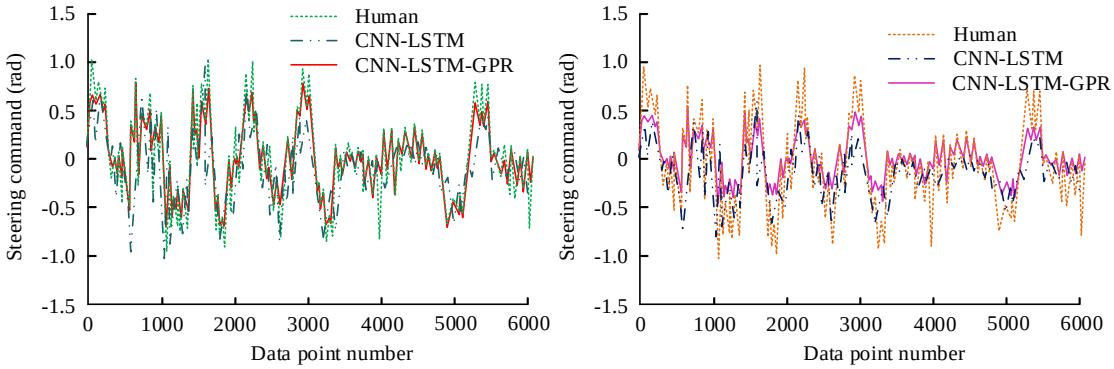
Fig. 9. Changes in training error loss function for different input image sequences.



(a) MSE value when the input timing image frame number is 3

(b) MSE value when the input timing image frame number is 5

Fig. 10. MSE values under different input timing image frames.



(a) Comparison of steering wheel angle prediction when the input image frame number is 3

(b) Comparison of steering wheel angle prediction when the input image frame number is 5

Fig. 11. Steering wheel angle prediction results of different driving behaviour learning methods based on time series information.

To evidence the function of the CNN-LSTM-GPR algorithm, the study further compares the steering wheel corner prediction outcomes of various driving behaviour learning methods with temporal information on the test set, and the research findings are indicated in Fig. 11. From Fig. 11, the CNN-LSTM-GPR based driving behaviour learning model can simulate human driving behaviour better compared to the CNN-LSTM model. It's learned driving actions are smoother, the coherence between actions is more solid, and the learning error is smaller, which outperforms the CNN-LSTM model. It can also be seen that when five consecutive frames of images are input, the driving behaviours simulated by the CNN-LSTM-GPR-based driving behaviour learning model fluctuate less and the movements are more coherent. This further confirms that introducing more temporal information helps to improve the performance of driving behaviour learning.

Further research was conducted to verify the computational efficiency of the CNN-LSTM-GPR algorithm, using CNN-LSTM algorithm, Extreme Gradient Boosting (XGBoost) algorithm, and Seasonal Autoregressive Integrated Moving Average (SARIMA) model for performance comparison. The calculation time for different algorithms in predicting driving behavior is shown in the table. According to the table, the computation time of the CNN-LSTM-GPR algorithm is only 0.5213 seconds, which is lower compared to the XGBoost and SARIMA algorithms. The computation time of the CNN-LSTM algorithm is slightly lower than that of the CNN-LSTM-GPR algorithm, because the algorithm integrates Gaussian Process Regression to handle uncertainty, thereby increasing computation time. However, overall, the CNN-LSTM-GPR algorithm has better predictive performance.

TABLE III. CALCULATION TIME FOR DIFFERENT ALGORITHMS

Algorithm	Runtime (s)
CNN-LSTM	0.4926
XGBoost	1.6397
SARIMA	15.3969
CNN-LSTM-GPR	0.5213

V. DISCUSSION

An improved Bagging GPR method has been proposed for intelligent protection in intelligent transportation systems. The results showed that in the scenario of overtaking on a straight line, the MAE, RMSE, and MAXE of the improved Bagging GPR method were 0.5241, 0.9547, and 10.7705, respectively. Meanwhile, in the scenario of obstacle avoidance on curves, the MAE, RMSE, and MAXE of the improved Bagging GPR method are 0.6527, 0.9436, and 14.7531, respectively. Compared to the multi-layer BP algorithm, its various indicators have decreased by 1.383, 1.8274, and 12.8996, respectively. ZHONG Q et al. proposed a tool wear prediction method based on maximum information coefficient and improved Bagging GPR. The results show that this method has significant advantages in predictive performance [21]. The Bagging GPR method demonstrates high prediction accuracy in all aspects. The reason is that the improved Bagging GPR method integrates multiple GPR models, each trained with a different subset of data, and then averages or weights their prediction results, thereby reducing the variance of the model. At the same time, this method increases the attention of the base learner to samples with large training errors, further reducing the maximum prediction error and improving the overall performance of the ensemble regressor.

In the effectiveness verification experiment of the CNN-LSTM-GPR algorithm, the driving behavior learning model based on CNN-LSTM-GPR can better simulate human driving behavior compared to the CNN-LSTM model. The driving actions it learns are smoother, with stronger coherence between actions and smaller learning errors, and its performance is better than that of the CNN-LSTM model. When five consecutive frames of images are input, the driving behavior learning model based on CNN-LSTM-GPR simulates less fluctuation and more coherent actions. Chen H and other researchers proposed a CNN-GPR method for driving behavior learning, which addresses the problems of low learning accuracy and poor generalization performance in traditional driving behavior learning methods. They also introduced LSTM and proposed a CNN-LSTM-GPR method for driving behavior learning using time-series images. The results show that the proposed CNN-LSTM-GPR method can fully utilize the temporal information

of the image, resulting in smaller simulation errors [22]. Compared with the CNN-LSTM method, this method can further improve learning accuracy and exhibit better generalization performance. This is similar to the research findings. The reason is that the CNN-LSTM-GPR method can effectively utilize temporal image information. Temporal information includes the temporal sequence of consecutive frame images, which helps capture dynamic changes and coherence in driving behavior. By introducing the GPR (Gaussian Process Regression) model, it is possible to more accurately model the spatiotemporal dynamics of driving behavior, thereby making the learned driving actions smoother and more coherent.

VI. Conclusion

Intelligent vehicle safety protection in intelligent transport systems is of great significance. For intelligent protection in intelligent transportation systems, the study successively introduces an improved Bagging GPR=driving behaviour modelling method and a feature extraction method with CNN-LSTM-GPR algorithm. The findings denoted that compared with the remaining three modelling methods, the driving behaviour modelling method of the improved Bagging GPR has better fitting performance with the actual steering wheel angle and higher matching accuracy. Whereas, the multi-layer BP algorithm has the largest deviation from the actual steering wheel angle, which represents its worst modelling performance. Meanwhile, when the input image sequence of the CNN-LSTM network is 3, the MSE of the test set converges to 0.0415 at about 25 iterations, which is an improvement of 0.014 compared to the case when the input image sequence is 5. In addition, compared to the CNN-LSTM model, the driving behaviour learning model with CNN-LSTM-GPR can more accurately simulate human driving behaviour. Its learned driving actions are smoother, the articulation between actions is more natural, and its learning error is relatively smaller, so the whole effect is better than that of the CNN-LSTM model. In addition, when five consecutive frames of images are input, the driving behaviours simulated by the CNN-LSTM-GPR-based driving behaviour learning model show less fluctuation and more coherent movements. It shows that the improved Bagging GPR method and CNN-LSTM-GPR feature extraction method can provide more accurate and smooth driving behaviour modelling and learning schemes for intelligent vehicles in ITS. However, the drawback of this study is that it only focuses on specific scenarios and environments, which may limit the universality and scalability of the proposed technology in a wider range of driving conditions and challenges. Future research needs to expand its scope to cover more diverse scenarios, road conditions, and driving behaviours to ensure the effectiveness and robustness of the developed models in real-world applications.

References

[1] Sheibani M, Ou G. The development of Gaussian process regression for effective regional post-earthquake building damage inference. Computer-Aided Civil and Infrastructure Engineering, 2021, 36(3): 264-288.

[2] Haydari A, Yılmaz Y. Deep reinforcement learning for intelligent transportation systems: A survey. IEEE Transactions on Intelligent Transportation Systems, 2020, 23(1): 11-32.

[3] Guo Y, Mustafaoglu Z, & Koundal D. Spam Detection Using Bidirectional Transformers and Machine Learning Classifier Algorithms. Journal of Computational and Cognitive Engineering, 2022, 2(1), 5-9.

[4] Deringer V L, Bartók A P, Bernstein N, Wilkins, D. M., Ceriotti, M., & Csányi, G. Gaussian process regression for materials and molecules. Chemical Reviews, 2021, 121(16): 10073-10141.

[5] Liu K, Hu X, Wei Z, Liu K, Hu X, Wei Z, et al. Modified Gaussian process regression models for cyclic capacity prediction of lithium-ion batteries. IEEE Transactions on Transportation Electrification, 2019, 5(4): 1225-1236.

[6] Band S S, Heggy E, Bateni S M, Karami, H., Rabiee, M., Samadianfard, S.& Mosavi, A. Groundwater level prediction in arid areas using wavelet analysis and Gaussian process regression. Engineering Applications of Computational Fluid Mechanics, 2021, 15(1): 1147-1158.

[7] Hewing L, Kabzan J, Zeilinger M N. Cautious model predictive control using gaussian process regression. IEEE Transactions on Control Systems Technology, 2019, 28(6): 2736-2743.

[8] Zhang Y, Xu X. Relative cooling power modeling of lanthanum manganites using Gaussian process regression. RSC advances, 2020, 10(35): 20646-20653.

[9] Xing Y, Lv C, Wang H, Cao, D., Velenis, E., & Wang, F. Y. Driver activity recognition for intelligent vehicles: A deep learning approach. IEEE transactions on Vehicular Technology, 2019, 68(6): 5379-5390.

[10] McDonald A D, Alambeigi H, Engström J, Markkula, G., Vogelpohl, T., Dunne, J., & Yuma, N. Toward computational simulations of behavior during automated driving takeovers: a review of the empirical and modeling literatures. Human factors, 2019, 61(4): 642-688.

[11] Kabzan J, Hewing L, Liniger A, & Zeilinger, M. N. Learning-based model predictive control for autonomous racing. IEEE Robotics and Automation Letters, 2019, 4(4): 3363-3370.

[12] Mozaffari S, Al-Jarrah O Y, Dianati M, et al. Deep learning-based vehicle behavior prediction for autonomous driving applications: A review. IEEE Transactions on Intelligent Transportation Systems, 2020, 23(1): 33-47.

[13] Hoel C J, Driggs-Campbell K, Wolff K, Laine, L., & Kochenderfer, M. J. Combining planning and deep reinforcement learning in tactical decision making for autonomous driving. IEEE transactions on intelligent vehicles, 2019, 5(2): 294-305.

[14] Schwarting W, Pierson A, Alonso-Mora J, Karaman, S., & Rus, D. Social behavior for autonomous vehicles. Proceedings of the National Academy of Sciences, 2019, 116(50): 24972-24978.

[15] Claussmann L, Revilloud M, Gruyer D, & Glaser, S. A review of motion planning for highway autonomous driving. IEEE Transactions on Intelligent Transportation Systems, 2019, 21(5): 1826-1848.

[16] Pensoneault A, Yang X, Zhu X. Nonnegativity-enforced Gaussian process regression. Theoretical and Applied Mechanics Letters, 2020, 10(3): 182-187.

[17] Wang Z, Yuan C, Li X. Lithium battery state-of-health estimation via differential thermal voltammetry with Gaussian process regression. IEEE Transactions on Transportation Electrification, 2020, 7(1): 16-25.

[18] Sun, Y., Xue, B., Zhang, M., & Yen, G. G. Completely automated CNN architecture design based on blocks. IEEE transactions on neural networks and learning systems, 2019, 31(4): 1242-1254.

[19] Yu Y, Si X, Hu C, & Zhang, J. A review of recurrent neural networks: LSTM cells and network architectures. Neural computation, 2019, 31(7): 1235-1270.

[20] Ma M, Mao Z. Deep-convolution-based LSTM network for remaining useful life prediction. IEEE Transactions on Industrial Informatics, 2020, 17(3): 1658-1667.

[21] ZHONG Q, LI Y, CHEN Y, WU, Z., LIAO, X., MA, J., & LU, J. Tool wear prediction based on MIC and improved Bagging-GPR. Computer Integrated Manufacturing System, 2023, 29(5): 1471.

[22] Chen H, Zeng Y, Huang J, Zhang, Y. Deep Neural Network Gaussian Process Regression Method for End-to-end Driving Behavior Learning. 2020 3rd International Conference on Mechatronics, Robotics and Automation (ICMRA). IEEE, 2020: 64-73.

Automation of Book Categorisation Based on Network Centric Quality Management System

Tingting Liu[1]*, Qiyuan Liu[2], Linya Fu[3]
Library, Harbin University, Harbin 150000, China[1]
Changchun Sixth High School, Changchun 130000, China[2]
Fenghua Middle School, Harbin 150000, China[3]

***Abstract*—In order to improve the efficiency of automatic book classification, the study uses a crawler to crawl book data from regular websites and perform data cleaning and fusion to build a structured knowledge graph. Meanwhile, the processed data is applied to a pre-trained model to improve it, and migration learning is used to improve the results. Fusion of Multiple Attention Mechanisms, Recurrent Neural Network, and Convolutional Neural Network modules into the classification model and feature fusion is used to enhance feature extraction. In addition, the study designed a pre-trained model architecture to help automatically categorise and manage book resources. The results of this study show a significant improvement in the classification of Chinese books on the Chinese Book L2 Subject Classification, iFlytek, and THUCNews datasets with significant performance improvement. The fusion of long and short-term memory and convolutional network Transformer-based bi-directional encoding models improved the accuracy by 0.19%, 1.54% and 0.42% on the two datasets, respectively, while the weighted average F1 scores improved accordingly. Through wireless technology, the automatic classification efficiency of books is realized and the management ability of the library is improved.**

***Keywords*—Crawler; books; automated; classification; Recurrent Neural Network; Multiple Attention Mechanism; knowledge graphs**

I. Introduction

In the context of the big data era, the huge amount of complex text data makes libraries lack detailed subject classification labels when purchasing books, which leads to the inefficiency of acquiring book resources. Currently, this task mainly relies on library managers, which is costly and time-consuming. The introduction of an intelligent model greatly improves the accuracy of book categorisation, which helps to reduce labour costs and improve work efficiency. The robots not only automatically identify and classify books, but also operate continuously, significantly optimising library resource management and user services [1-2]. Existing studies rarely consider subject-specific library classification work, and the use of the latest natural language processing techniques, including large-scale pre-trained models such as the Bidirectional Encoder Representations from Transformers (BERT) model, can significantly improve the efficiency and accuracy of automatic classification of library resources in subject domains [3]. Such models can effectively reduce the dependence on a large amount of manually labelled data by virtue of their powerful semantic capture and feature extraction capabilities. Meanwhile, Knowledge Graph, as an intelligent semantic tool, greatly facilitates information sorting and management by interconnecting dispersed data through constructed entities, relationships and attributes [4]. Although the approach limits the application of the system across domains, it tends to show better performance and efficiency within established domains.

Network Centric Quality Management (NCQM) systems often use web-based technologies to collect, analyse and manage quality-related data to improve the efficiency and effectiveness of the overall system or process [5]. In the context of intelligent model book categorisation, NCQM systems can be used to monitor and optimise robot performance, accuracy and efficiency. Therefore, in order to improve the efficiency and accuracy of book management, reduce the dependence on paper records and manual operations, and improve the convenience of users to obtain information, the following work have been made in this study. First of all, the data is collected by efficient crawler technology, and then sorted and cleaned as the basis for the construction of a knowledge map. Based on Neo4j graph database, the book information knowledge graph is quickly constructed according to the pre-defined book knowledge graph Schema. Secondly, the pre-trained data set is divided into training set, verification set and test set, and the transfer learning strategy is used to fine-tune the pre-trained model to achieve the optimal state. By comparing with Text Convolutional Neural Network (Text-CNN) and other baseline models, the proposed model is fully tested on the Chinese book dataset specially constructed for this study. In addition, a feature enhancement method is proposed. This approach combines the multi-head attention mechanism of pre-trained models with the feature extraction and learning advantages of Convolutional Neural Networks (CNN) and Recurrent Neural Network (RNN). The pre-training model of fusion feature extraction technology has an effective effect on the classification of Chinese books. On the basis of using the pre-training model to classify Chinese books, according to the actual needs of university libraries, an automatic classification system of book resources based on the pre-training model is completed.

The study is divided into six sections. Section I discusses the challenges libraries face in the era of big data and proposes the use of intelligent models and natural language processing techniques to improve classification accuracy and efficiency. Section II reviews the research progress of automatic book

*Corresponding Author.

classification, and mentions the application of BERT pre-training model and knowledge graph in information classification. Section III introduces the research methods, including the collection and processing of book data and the construction of a knowledge graph, the collection and storage of data in Neo4j graph database by web crawler technology, and the application of BERT model for transfer learning and feature fusion. Section IV tests the models and compares the classification effects of different models. Section V and Section VI summarize the research results, confirm the effectiveness of the intelligent model in improving classification efficiency and accuracy, and propose future work directions, such as strengthening the subject classification module and improving user experience.

II. Related Work

The application of intelligent model in book categorisation not only improves the efficiency of library management but also provides better reading services to the users. An innovative feature selection optimisation algorithm has been designed by Janani and Vijayarani. To validate the effectiveness of the algorithm, the researchers further proposed a machine learning-based automatic text categorisation algorithm. The results demonstrated the efficiency and accuracy of the algorithm in extracting key features and classifying text documents by content [6]. Zhang designed a classification system based on support vector machines (SVM) for intelligent text classification, which is dedicated to public security information. The results proved the efficiency and practicality of SVM in dealing with specific information classification tasks [7]. For the field of automated text categorisation, Rezaeian and Novikova have compared the efficacy of plain Bayes and support vector machine algorithms, as well as Gaussian kernel function, polynomial kernel function, and Bernoulli's model in plain Bayes in order to enhance the accuracy of Persian textual materials in categorisation. The results of the study showed the better performance of these methods in Persian text classification [8]. Dizaji and other researchers proposed to combine the imperialist competitive algorithm with support vector machines for text classification. After experimental analyses, the results showed that this method showed higher efficiency and accuracy in text classification tasks [9].

The best performing language models for various tasks in the natural language processing direction are pre-trained language models. However, focusing on the topology of knowledge graphs often ignores the potential differences between knowledge graph embeddings and natural language embeddings, which limits the ability to reason effectively using both implicit and explicit knowledge. To address this problem, Cao and Liu designed an innovative model, ReLMKG, which combines pre-trained language models and relevant knowledge graphs. The efficiency and applicability of the model is demonstrated by testing it on the complex WebQuestions and WebQuestionsSP datasets [10]. According to the characteristics of Chinese library classification, Yuhui Z and other scholars used innovative methods to adapt to the problems of extreme multi-label (XMC) and hierarchical text classification (HTC). This paper extracts semantic features by a lightweight deep learning model and combines hierarchical information and other features with a learning ranking (LTR) framework to improve accuracy and classification depth. This model not only understands deep semantics but also has interpretability, is easy to expand and customize, and is suitable for processing tens of thousands of category labels, providing a solid foundation for comprehensive deep classification [11]. Jiang Y et al. combined domain-specific and general text enhancement strategies, such as category mapping and bilingual theme-based method, to solve the problem of small data volume and unbalanced categories in English books with Chinese picture classification numbers, and added punctuation and conjunction to the English text. Experimental results show that this hybrid strategy can improve the model performance. The visualization of BERT word vectors and the analysis of word information entropy can be applied to the classification and recognition of books [12]. To improve the accuracy of Chinese text Classification, Liu X and other researchers designed Chinese Library Classification based on an adaptive feature selection algorithm and tested a variety of Chinese text types. In this paper, an improved mutual information chi-square algorithm is introduced, which combines word frequency and term adjustment, term frequency-inverse document frequency method, and uses the limit gradient enhancement algorithm to improve the word filtering effect. Experiments show that the proposed algorithm can effectively improve the classification performance of different news texts, but the optimal algorithm selection varies according to the text type [13].

In summary, existing models perform well in specific domains, but are less adaptable across domains. For example, some Chinese book classification models have limited ability to generalize on other languages or topics. Despite the introduction of methods such as Multiple Attention Mechanisms, RNNS, and CNNS, feature extraction and fusion are still inadequate, especially when dealing with long text, which may miss key information or be inefficient. In addition, the lack of a large-scale open-source book knowledge graph limits the model's knowledge reasoning and classification accuracy. Training deep learning models requires a lot of resources and time, which small organizations can't afford. Model tuning and parameter adjustment also require a lot of experiments, and the number of books in different categories is unbalanced, resulting in poor classification performance, especially in small categories. Most methods do not perform well in cross-domain book classification, the model depends on the domain, and the application scope is limited. In order to achieve effective text representation, feature extraction and accurate subject categorisation of book resources.

III. Automated Research on Categorisation of Books by Intelligent Model Based on NCQM System

Due to the lack of available open source book knowledge graph and the data requirement of deep learning model, the research first designed an asynchronous anti-anti-crawler based on aiohttp to collect website book information. The collected data was cleaned and used to build the knowledge graph, and the book information knowledge graph was built automatically by using Neo4j. In order to deal with the low efficiency of traditional book classification and the need for professional knowledge, the research adopts the pre-trained model in deep

learning for text representation, and optimizes the model structure through transfer learning. The pre-processed data set was divided into training, verification and test sets, and the model was fine-tuned to compare with the existing baseline models such as TextCNN and TextRNN. RNN module and CNN module are added to the feature extraction layer of the book classification model, and a feature fusion method is proposed to enhance the features of the pre-trained model. To improve the accuracy of Chinese book classification. According to the actual demand of university library, the automatic classification system of book resources based on pre-training model is finally completed.

A. Construction of Knowledge Graph of Book Information

The core advantage of Knowledge Graph is that it can connect related data elements, and its construction is closer to the natural logic of human processing information [14]. Moreover, knowledge graph relies on the use of graphs to store data, which makes it possible to achieve good data visualisation. There are two main methods for storing knowledge graphs, the first is the resource description framework introduced by W3C, and the second is to use of graph databases for data preservation. Graph databases are popular among developers and users because of their intuitive data representation and visualisation features, and they are not inferior to RDF in terms of data query performance. Therefore, the vast majority of developers tend to use graph databases to build knowledge graphs. Fig. 1 shows the construction steps of the book knowledge graph.

In Fig. 1, this construction step first collects data from multiple online resources through web crawling techniques and then pre-processes them to remove redundancies and error messages. Next, the processed data is subjected to knowledge extraction using a predefined schema in the study, i.e., converting the textual information into a structured ternary form. Subsequently, a knowledge fusion step is carried out to solve the problems of word polysemy and textual ambiguity by technical means. Ultimately, using Neo4j, a graph database platform, the collated ternary data is constructed into an exhaustive knowledge graph of book information and its visualisation is achieved to facilitate a more intuitive understanding and analysis of the data. In view of the characteristics of the website where the data comes from, it is necessary to design a crawler programme to obtain the data quickly. In terms of language platform selection, Python language is chosen as the platform for building the crawler programme [15]. After choosing the platform and tools, the design of the crawler programme begins, and the running flow of the crawler programme is shown in Fig. 2.

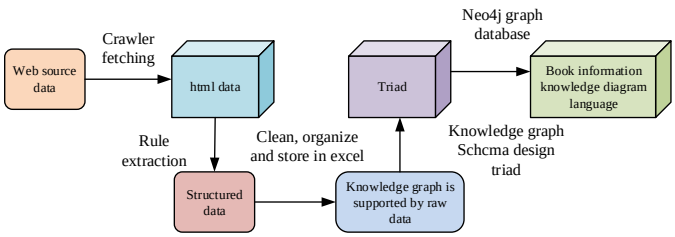


Fig. 1. Flowchart of the construction of book knowledge map.

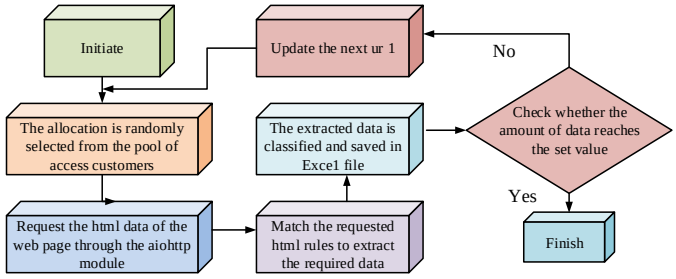


Fig. 2. Crawler operation flow.

According to the flow in Fig. 2, first, the crawler makes use of the asynchronous HTTP web module aiohttp to initiate HTTP requests against the specified book information website URL and obtain the HTML content. The reason for choosing aiohttp as the request tool is that it supports asynchronous processing of HTTP requests, which significantly improves the efficiency of the crawler. Given the large amount of data required to build a knowledge graph of book information, traditional synchronous crawling methods are less efficient, while aiohttp can accelerate the data crawling process. Once the HTML data is successfully retrieved, the program will analyse its content and extract the required textual information. In addition, the study employs the BeautifulSoup tool to crawl the required content from web pages quickly and efficiently. Relying on HTML's Selector for precise positioning, the extraction of target information is achieved. Initially, the acquired data may contain non-targeted components, so it is decided whether to use regular expressions to further remove depending on the situation [16]. Afterwards, it is processed and saved in xlsx format using Python's Xlsxwriter library, a process that converts semi-structured information in HTML into structured data. In constructing the knowledge graph of book information, the study uses Python's py2neo library in combination with the Neo4j graph database. Firstly, pandas are used to extract information from book data in xlsx format, and subsequently entities and relationships are created by py2neo. In order to cope with the polysemy and translation problems in Chinese, translation software is used to translate and staging the foreign language content. When merging entities, if a translated entity is found to have the same primary key as an existing entity, it will be manually reviewed to determine if it should be merged.

B. Automatic Book Classification Method Based on Pre-Trained Model

After initial processing, book title, synopsis and keyword texts are selected for classification. These data were transformed into vectors by a pre-trained model to be adapted for machine processing. The text features are refined through a multi-layer Transformer structured encoder. Subsequently, the learnt text features are fed into the classification system to compute probability values for different classification labels [17]. The final probability distribution is determined by a normalisation method and the label corresponding to the highest probability is assigned as the subject category of the book. The process of Chinese book classification is shown in Fig. 3.

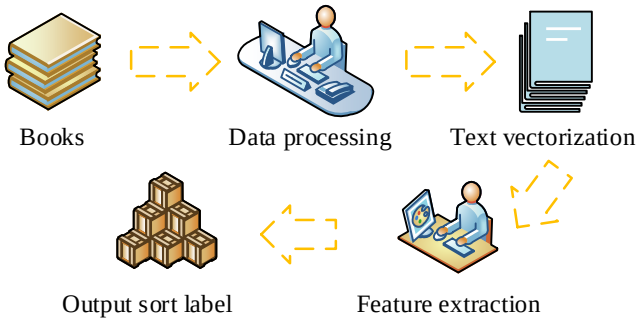


Fig. 3. Process of Chinese book classification.

BERT is a deep learning model based on Transformer, which mainly consists of bi-directional encoders. It is available in two versions, the basic version with 12 layers of encoders and the advanced version with 24 layers [18]. These layers include a multi-head self-attention mechanism and a feed-forward network, as well as residual connectivity and layer normalisation. In pre-training, BERT processes two pieces of text and learns two tasks, next-sentence prediction and masked language modelling. These tasks help BERT to better understand the language structure. Among them, the mask language model is suitable for single or double-paragraph text, while the next sentence prediction is specifically used for two paragraphs. The input representation of BERT consists of the sum of word vectors, block vectors, and position vectors [19]. In order to facilitate the computation, the dimensions of these three vectors are set to e, and the representation of the text is shown in Eq. (1).

$$v = v_t + v_s + v_p \tag{1}$$

In Eq. (1), v_t, v_s, v_p denotes word vector, block vector and position vector respectively, and the length of all three vectors is N. The word vector representation is very similar to the word vector in neural networks, both of which convert a text sequence into a fixed dimensional vector of values. Consider a dictionary with $|\Upsilon|$ words and an e-dimensional word vector space, where the sequence representation x is uniquely encoded into $e^t \in \mathbb{R}^{N\times|\Upsilon|}$ and $W^t \in \mathbb{R}^{e\times|\Upsilon|}$ represents the learnable word embedding matrix. Then, the word vector formula corresponding to x is given in Eq. (2).

$$v_t = e^t W^t \tag{2}$$

The role of the block vector is to tell the model which block the current word belongs to. And the role of the position vector allows the model to obtain the absolute position of each word in the sequence, which enhances the memory ability of the model. Applying BERT to the task of single-sentence text classification, the model is mainly composed of an input layer, a BERT feature extraction layer and a classification output layer. For the input text sequence $x_1 x_2 x_3 \cdots x_n$, the specific markers of the BERT model are added to both ends of the sequence to get the original input of the model X as shown in Eq. (3).

$$X = [CLS]\, x_1 x_2 x_3 \cdots x_n\, [SEP] \tag{3}$$

Next, the mapping of X is performed according to the above processing of word vectors, block vectors, and position vectors, and let the length of the sentence be n, then the input representation of the model is obtained as v, see Eq. (4).

$$v = \text{InputRepresentation}(X) \tag{4}$$

BERT utilises Transformers encoders with a different number of layers (12 layers for the basic version and 24 layers for the large version) to process the input v and learn the implicit associations between lexical items within the text. Let the implied dimension be d and the textual contextual representation be denoted as $h \in \mathbb{R}^{N\times d}$, corresponding to Eq. (5).

$$h = \text{BERT}(v) \tag{5}$$

The BERT model represents the whole sentence vector for next sentence prediction via [CLS] labelling in the pre-training phase. For task alignment, the implicit layer representation of [CLS] is similarly utilised as the vector representation of the whole sentence for single-sentence classification, denoted as h_0. To obtain the corresponding categorical labels of the input sentences are simply the [CLS] implicit layer representation h_0 extracted by BERT into the fully connected layer. Setting K as the number of categories, $W^0 \in \mathbb{R}^{d\times K}$ as the weights, and $b^0 \in \mathbb{R}^K$ as the bias, the label probability distribution $p \in \mathbb{R}^K$ is obtained as shown in Eq. (6).

$$p = Soft\max\left(h_0 W^0 + b^0\right) \tag{6}$$

After obtaining the probability distribution of the classification, it is compared with the real classification labels to calculate the loss, and then backpropagation is performed to update the parameters of the model. Conditional Random Field (CRF) is mainly applied to sequence labelling task to ensure the accuracy of the labelling [20]. The correct sequence is identified by scoring each labelled sequence and calculating the proportion of the total score that a particular labelled sequence scores. The length of the text sequence is l and the length of its true annotation sequence is also l. The total score S_{i,y_i} for position i in the sequence annotation consists of the sum of the current launch score R_{i,y_i}, the total score $S_{i-1,y_{i-1}}$ for the previous position, and the transfer score T_{y_{i-1},y_i}. The total score of the end point represents the score of the whole sequence, which is calculated in Eq. (7).

$$S(X,y) = \sum_{i=1}^{n}\left(T_{y_i,y_{i+1}} + R_{i,y_i}\right) \tag{7}$$

After obtaining the score of a single sequence annotation, in order to confirm whether a sequence annotation is optimal or

not, it is necessary to calculate the proportion of its score to the sum of the scores of all possible sequences, and the expression of this proportion is shown in Eq. (8).

$$P\left(y_{now}\left|X\right.\right)=\frac{e^{S\left(X,y_{now}\right)}}{\sum_{y\in Y_x}e^{S\left(X,y\right)}} \tag{8}$$

In Eq. (8), the larger P is, the closer the current sequence prediction is to the real value. Therefore, Eq. (8) can be used as the loss calculation of CRF model, and its calculation procedure is shown in Eq. (9).

$$\log\left(P\left(y_{now}\left|X\right.\right)\right)=S\left(X,y_{now}\right)-\log\left(\sum_{y\in Y_x}e^{S\left(X,y\right)}\right) \tag{9}$$

Typically, the CRF uses the Viterbi algorithm to find the optimal solution, labelled as the optimal sequence of X. This formula is shown in Eq. (10).

$$\overset{*}{y}=\arg\max_{\tilde{y}\in Y_X}S\left(X,\tilde{y}\right) \tag{10}$$

C. Fine-Grained Book Classification Based on Pre-Trained Model and Feature Fusion

Combining the feature extraction advantages of pre-trained models such as BERT, the newly designed text categorisation model integrates a feed-forward network, using the AdamW optimiser and cyclic unit, convolution and pooling techniques to enhance the feature representation. Based on these techniques, this paper proposes a feature enhancement method of Pre-trained model, PLM-LCN, which combines Long short term memory and Convolution Networks (LCN) and pre-trained language models (PLM). Fig. 4 shows the PLM-LCN model structure.

In Fig. 4, the main flow of the PLM-LCN method model is as follows, firstly, the original input text is encoded into word vectors, block vectors and position vectors, which are combined according to the rules to form the input representation of the Multihead Self-Attention Layer. The text features are extracted after processing in this layer, and the sentence representation with [CLS] characters enters into the convolution and loop unit module. The output features from the convolutional module and the bidirectional long and short term memory network (Bi-LSTM) are combined and spliced to form the ultimate textual representation. These features are used for probabilistic prediction by the classifier and normalised by Softmax, with the highest probability category being the book classification result. The counting formula for the convolution operation is given in Eq. (11).

$$O\left(i,j\right)=\sum_{m=0}^{M-1}\sum_{n=0}^{N-1}\left(I\left(i+m,j+n\right)K\left(m,n\right)\right) \tag{11}$$

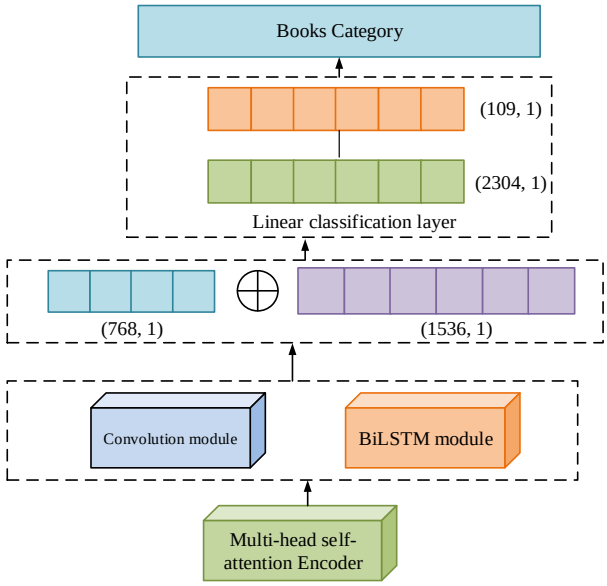


Fig. 4. PLM-LCN model structure.

In Eq. (11), (i,j) represents the output position. I represent the input data into the convolutional layer. K is the convolution kernel and (m,n) denote its size. Attention mechanism, derived from the human characteristic of focusing attention, is introduced into deep learning to highlight key features and downplay secondary information. In this mechanism, Q(Query) locates the target task, while K(Key) and V(Value) form matching pairs. Q It is used to find the corresponding V value in K . The operation of the mechanism and the computation process is shown in Eq. (12).

$$D_V=Attention\left(Q_t,K,V\right)=Soft\max\left(\frac{Q_t,K_s}{\sqrt{d_k}}\right)V_s=\sum_{s=1}^{m}\frac{1}{z}\exp\left(\frac{Q_t,K_s}{\sqrt{d_k}}\right)V_s \tag{12}$$

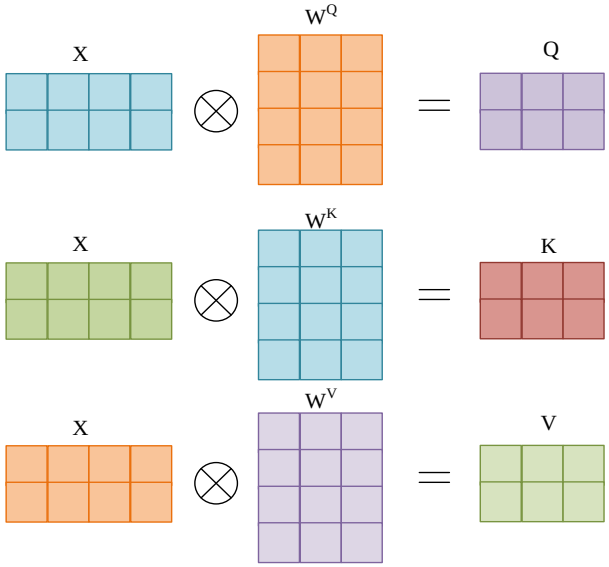


Fig. 5. Feature transformation process in self-attention calculation.

In Eq. (12), z represents the normalisation factor. Q_t represents the query value. The feature transformation process in Self-Attention computation is shown in Fig. 5.

Fig. 5 shows the matrix of word vectors X in the input text, representing a complete sentence. X Multiplying different weight matrices produces the Q,K,V matrix in the attention mechanism, where each row corresponds to a vector of words q,k,v. The parameters of these weight matrices W^Q,W^K,W^V are continuously updated during model training. In order to evaluate the model's learning of a -set of parameters, it is necessary to introduce a loss function, whose computational expression is shown in Eq. (13).

$$CE = -\frac{1}{m}\sum_{i=1}^{m}\sum_{j=1}^{c} y_j^{(i)} \log \hat{y}_j^{(i)} \tag{13}$$

In Eq. (13), $y_j^{(i)}$ denotes the true output result on the j th class for the i st sample, where only the correct class outputs 1 and the rest of the classes output 0. $\hat{y}_j^{(i)}$ denotes the probability that the model belongs to the j th class for the i th sample. Alternatively, Eq. (13) can be transformed into Eq. (14).

$$CE = -\frac{1}{m}\sum_{i=1}^{m} \log \hat{y}_t^{(i)} \tag{14}$$

In Eq. (14), $\hat{y}_j^{(i)}$ denotes the model's probability of predicting the first i sample on the correct category t. To realize the book information query and classification function of the automatic book classification robot system, the system is constructed into four levels. The overall framework of the book's automatic sorting robot is shown in Fig. 6.

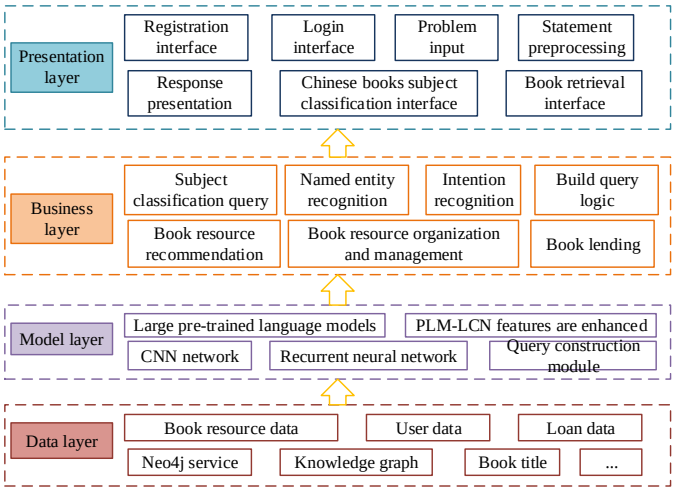


Fig. 6. Overall framework of the book automatic classification robot.

In Fig. 6, the presentation layer mainly realizes the communication between the system and the user. The user enters questions through the visualization window of this layer. The user interaction layer processes the natural language questions provided by the user to some extent, eliminates certain invalid interference content in the questions, and makes the text to be processed more standardized. The model layer calculates the data based on user commands and pushes the results to the presentation layer. The business layer is the bridge of user interaction scenario, covering various functional modules. The data interaction layer mainly relies on the constructed book information knowledge graph. After obtaining the query statement from the logical layer, the Neo4j server queries the book information knowledge graph. The limited number of candidate answers is returned to the user interaction layer after the candidate answers are sorted. The data interaction layer mainly relies on the book information knowledge graph constructed. After obtaining the query statement from the logical layer, the Neo4j server queries the book information knowledge graph. The limited number of candidate answers is returned to the user interaction layer after the candidate answers are sorted.

IV. Analysis of Experimental Results

Model training in deep learning research relies on a large number of matrix operations that place high demands on hardware. Graphics Processing Unit (GPU) is significantly more efficient in handling such tasks with its parallel processing capability and fast dedicated memory. The initial input length of all neural network models is set to 128 and the training period is 10 epochs. BERT-Base-Chinese and its improved version with AdamW optimiser and cross-entropy loss function are selected. Pytorch was chosen as the development framework, and its supporting hardware configuration and RSBC model parameter settings are shown in Table I.

TABLE I. Experimental Environment and Model Parameter Settings

Experimental Environment	Disposition	Model	Argument
Operating system	Windows 10	RoBERTa version	RoBERTa_zh_L12
CPU	Interl I7-11800H	learning_rate	3e-5
GPUs	NVIDIA GeForce RTX 3060	hidden_dim	768
RAM	16G	dropout_rate	0.5
Hard disk	10G disposable space	batch_size	64

The experiments use seven traditional neural network models as baseline models to participate in the comparison experiments. Accuracy (acc) represents the proportion of correctly classified samples in the total samples, and is a basic index for evaluating classification models. Macro Precision (mc_p) calculates and averages the accuracy of each class separately, reflecting the average accuracy of the model in each class, regardless of class imbalance. Macro Recall (mc_r) also calculates and averages the recall rate for each category separately, and measures the average recall rate of the model across all categories, ignoring the class imbalance problem. Macro F1-Score (mc_f) is the harmonic average of macro average accuracy and macro average recall, taking into account the performance of both accuracy and recall. Weighted

Precision (w_p) takes into account the number of samples in each category and makes a weighted average to reflect the comprehensive accuracy rate of the model in different categories, which is applicable in the case of unbalanced categories. Weighted F1-Score (w_fl) is also a harmonic average of weighted accuracy and weighted recall to comprehensively evaluate the model's performance in the case of class imbalance. The experimental results on different neural network models on the secondary subject classification dataset of Chinese books are shown in Fig. 7.

As can be seen from Fig. 6, each benchmark neural network model achieves more than 75% accuracy and average F1 score, indicating that the models show good performance after a number of training cycles. TextRCNN fails to outperform TextRNN. The Fasttext model performs better, with high speed and second only to the TextRNN-Att model. The loss values for the Fasttext model and the TextRNN-Att model has a loss value of 0.62 and 0.59 respectively, while the TextRNN model with embedded attention mechanism has the best result in all comparisons, which highlights the feature extraction advantage of the attention algorithm, and verifies the feasibility of studying the classification research with the pre-trained model based on the attention mechanism. The LERT model retains the original architecture of the BERT, but optimises the training tasks and methods, which provides a good opportunity to study the BERT series of models. The LERT model retains the original BERT architecture but optimises the training tasks and methods, providing guidance for the study of the BERT series of models. The learning rate is set to 1e-5, the batch size is 8, and 10 rounds of training are performed, taking the maximum sequence length of the input text as 16, 32, 64, 128, 192, 256, and 512 for the experiments, respectively. The experimental results corresponding to Fig. 8 are obtained.

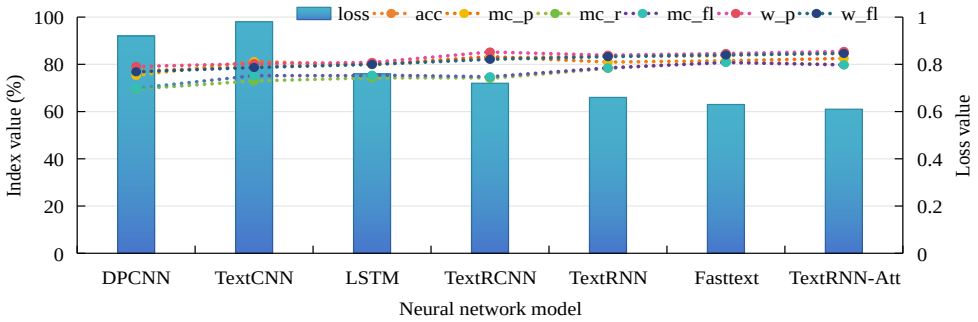
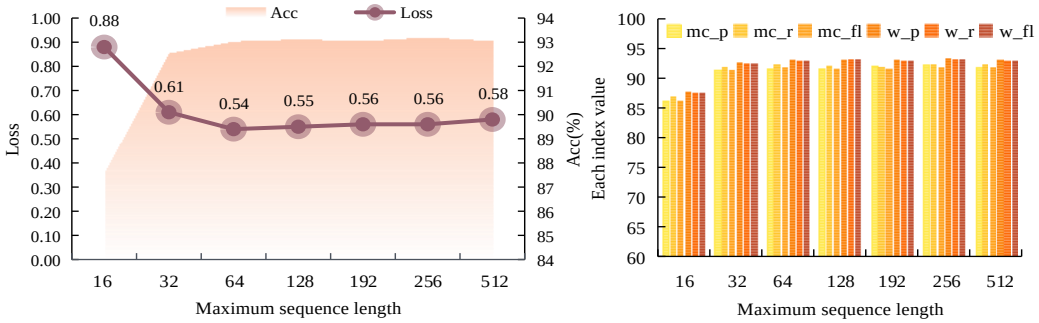


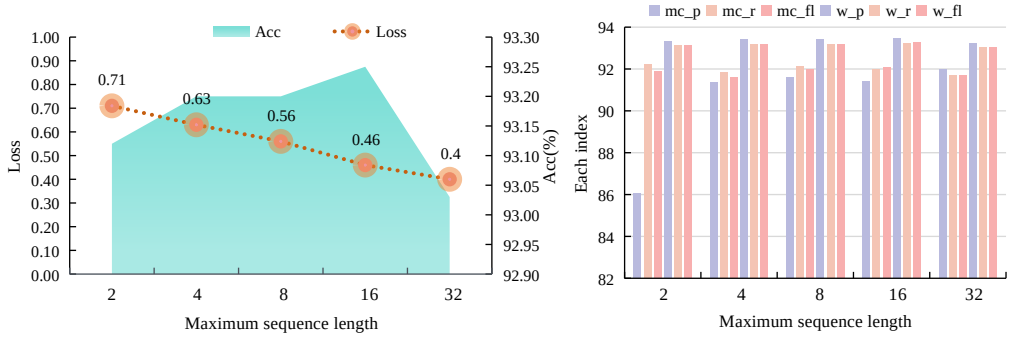
Fig. 7. Comparison of indicators of each model on the Chinese books' secondary subject classification data set.



(a) Loss and Acc of LERT model under different maximum sequence lengths

(b) Each index of LERT model under different maximum sequence lengths

Fig. 8. Experimental results of the LERT model under different maximum sequence lengths.



(a) Loss and Acc of LERT model under different batch sizes

(b) Indicators of LERT model under different batch sizes

Fig. 9. Experimental results of the LERT model under different batch sizes.

As can be seen in Fig. 7, the performance of the LERT model when dealing with different maximum sequence lengths, where the best performance is achieved when 256 is the sequence length. A sequence that is too short (e.g., 16) will lose more information because the important content is truncated, while setting it too long results in filling too many invalid 0-values, which affects feature extraction and computational efficiency. Therefore, setting 256 as the most suitable text length for the model to handle on the task of secondary subject classification of Chinese books is a significantly preferred solution. Fig. 9 shows the experimental results of the LERT model under different batch sizes.

Fig. 9 shows that the performance of the model typically improves with increasing batch size, although the effect decreases after a certain point. The minimum batch is theoretically good for optimisation, but too low may cause unstable convergence and prolong training time. On the contrary, too large batches tend to trigger memory overruns and impair accuracy. Considering the experimental results and computational efficiency, 16 is chosen as the preferred batch size setting for the BERT pre-training model on the Chinese book secondary subject classification dataset. Fig. 10 shows the experimental results of the LERT model at different learning rates.

As can be seen from Fig. 10, the LERT model performs similarly when the learning rate is set to 1e-5 and 5e-5, both outperforming other higher learning rate settings. A learning rate that is too small may slow down the training speed and cause the model to fall into a local optimum, while a learning rate that is too high may cause oscillations in the loss function and prevent the model from converging. Therefore, based on the trade-off between model performance and training efficiency, 1e-5 or 5e-5 is a more appropriate choice of learning rate to promote the model to achieve good training results on the Chinese book secondary subject classification dataset. Experiments are conducted using the Chinese book secondary subject classification dataset, the iFlytek dataset and the THUCNews dataset for ablation experiments on the five models, and the results are shown in Fig. 11.

As can be seen from Fig. 11, the BERT-LCN model exhibits better performance compared to the original BERT when dealing with the Chinese book secondary subject classification, iFlytek and THUCNews datasets. Its accuracy is improved by 0.19%, 1.54% and 0.42%, while the weighted average F1 score achieves an increase of 0.19%, 2.73% and 0.47%, respectively. Meanwhile, due to the original expressive power of the multiple self-attention mechanism, the BERT-LCN model demonstrated significant results on all three different attribute datasets reflecting its highly generalised nature. The ablation experiments were conducted using three datasets on six models and the results are shown in Fig. 12.

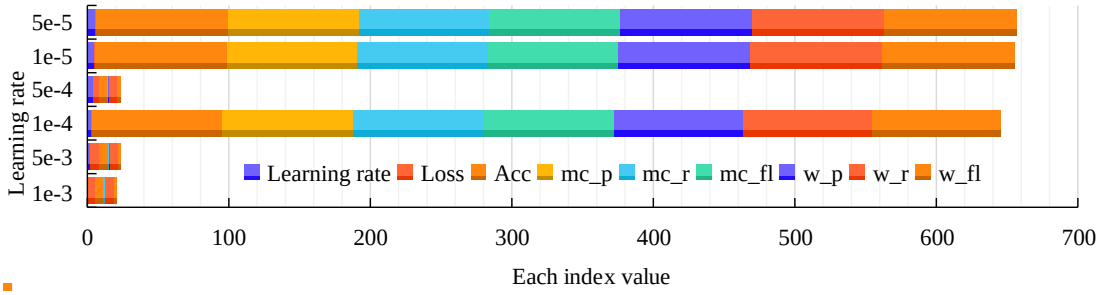


Fig. 10. Experimental results of the LERT model under different learning rates.

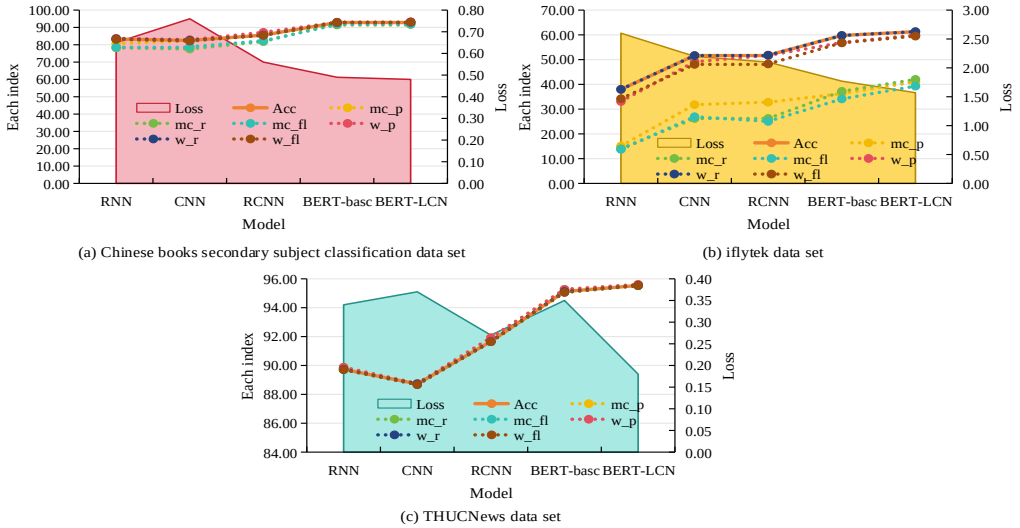
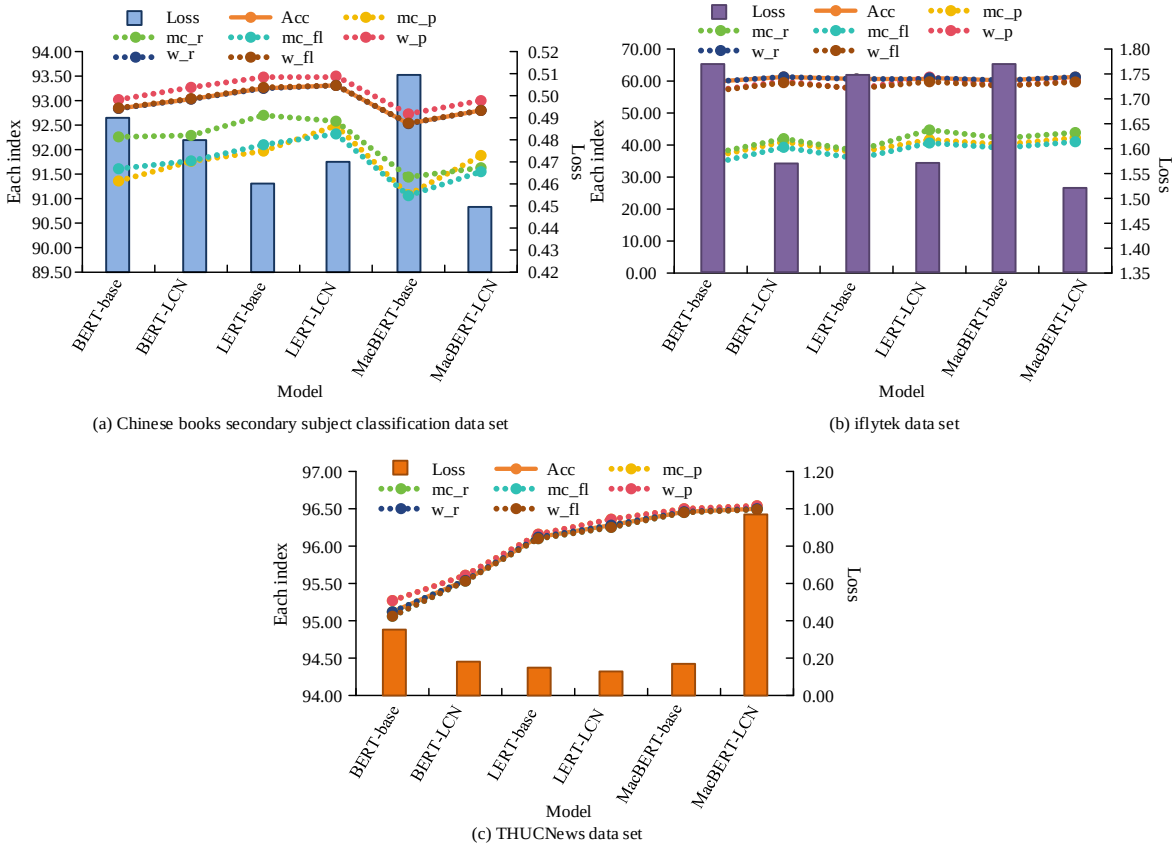


Fig. 11. Ablation results of the BERT-LCN method model on three datasets.



(a) Chinese books secondary subject classification data set

(b) iflytek data set

(c) THUCNews data set

Fig. 12. Experimental results of the PLM-LCN method model on three datasets.

As can be seen from Fig. 12(a), in the BERT-LCN, LERT-LCN and MacBERT-LCN models, the accuracy is increased by 0.18%, 0.07% and 0.28%, the weighted average F1 score is increased by 0.18%, 0.05% and 0.27%, respectively, and the macro-averaged F1 scores are increased compared to the original model. As can be seen from Fig. 12(b), the accuracy of these models increased by 1.55%, 0.32% and 1.04%, with weighted average F1 score enhancements of 2.74%, 2.13% and 1.28%, respectively, and the macro-averaged F1 scores likewise showed significant improvements. As can be seen in Fig. 12(c), these models showed increases in accuracy of 0.43%, 0.17% and 0.04%, and weighted average F1 score boosts of 0.47%, 0.15% and 0.04%, respectively. This difference in effect enhancement may be related to the characteristics of different datasets. In THUCNews and Chinese book subject classification, the baseline model has already achieved a high accuracy of more than 95%, so the room for improvement is relatively small. In contrast, the highest accuracy of the iFlytek baseline model is only 60.64%, so the performance enhancement of the LCN-enhanced model is more significant here.

To further verify the automatic method of book classification based on Network-centric quality management system proposed in this paper, it is compared with existing text classification methods, as shown in Table II.

In Table II, the feature extraction efficiency (96.49±0.74%), flexibility (94.23±0.37%) and accuracy (98.45±0.47%) of this method are superior to other literatures. The corresponding data in the study [6] were 87.37±0.64%, 88.47±0.68% and 86.46±0.42%. Study [8] were 87.97±0.47%, 87.26±0.59% and 89.68±0.73%. Study [10] were 90.67±0.14%, 90.63±0.62% and 92.67±0.36%. All P values were less than 0.05, indicating that the difference between different methods was statistically significant. In summary, the research method has the best performance in each index.

TABLE II. COMPARATIVE RESULTS OF VARIOUS BOOK CLASSIFICATION METHODS

Method	Feature Extraction Efficiency (%)	Flexibility (%)	Accuracy Rate (%)
Research method	96.49±0.74	94.23±0.37	98.45±0.47
Study [6]	87.37±0.64	88.47±0.68	86.46±0.42
Study [8]	87.97±0.47	87.26±0.59	89.68±0.73
Study [10]	90.67±0.14	90.63±0.62	92.67±0.36
P	<0.05	<0.05	<0.05

V. DISCUSSION

Compared with the feature selection optimization algorithm of Janani and Vijayarani et al. [6], this study can not only obtain a large amount of data quickly, but also ensure the high quality and comprehensiveness of the knowledge graph. In terms of feature fusion technology, the method proposed in this study is more flexible and scalable than the Bayesian algorithm and SVM used by Rezaeian and Novikova et al. [8] in Persian text classification, and is suitable for processing large-scale and

complex data sets. The pre-trained model BERT was also used for optimization in combination with transfer learning. Although Cao and Liu's ReLMKG et al. [10] model combined the pre-trained language model and knowledge graph, this study further integrated RNN and CNN to make the model perform better in the task of Chinese book classification. Overall, this study shows significant advantages in data acquisition, feature extraction and model optimization, especially in the Chinese book classification task showing higher accuracy and efficiency.

VI. Conclusion

In the face of the challenge that the Chinese map classification no longer meets the needs of the emerging disciplines, the research proposes to adopt the existing natural language processing technology to solve the problem of classifying Chinese resources in libraries. The research team constructed a Chinese book database, used neural networks and pre-trained models for text analysis, feature extraction and category classification on this dataset, and created an NCQM-based solution to automate the archiving process with the help of an intelligent model. The results show that various benchmark neural network models achieved more than 75% accuracy and average F1 scores on the Chinese book classification dataset, indicating that the dataset is moderately difficult and the model training performs well. Specifically, the single-layer LSTM model outperforms the CNN in text sequence processing, while the TextRCNN does not exceed the performance of the TextRNN. The Fasttext model stands out with its high speed and excellent performance, second only to the TextRNN-Att model with embedded attention mechanism. For models using BERT-LCN, LERT-LCN, and MacBERT-LCN, the results show that they achieved improvements in accuracy and weighted average F1 scores compared to the original models on the Chinese book secondary subject classification, iFlytek, and THUCNews datasets. These enhancements reflect the generalisation ability and feature extraction advantages of the models when dealing with different attribute datasets. In particular, the BERT-LCN model, by combining the CNN and RNN modules, enhances the feature extraction capability and therefore shows significant performance gains on multiple datasets. In addition, there is still room for improvement in the research. In future work, it is planned that the subject classification module will be enhanced to provide diverse model choices and a top ten subject classification display, and the library seat reservation function will be added to improve user experience and convenience.

References

[1] Zeng L. Classification and English translation of book titles in Dunhuang documents. China Terminology, 2022, 24(3): 41-48.

[2] Wang J, Yue K, Duan L. Models and techniques for domain relation extraction: A survey. Journal of Data Science and Intelligent Systems, 2023, 3(1): 16-25.

[3] Li Y, Yang Y, Ma Y, Yu D, Chen Y. Text adversarial example generation method based on BERT model. Computer Application, 2023, 43(10): 3093-3098.

[4] Wang Y, Zhang X, Dang Y, Ye P. Knowledge Graph Representation of Typhoon Disaster Events based on Spatiotemporal Processes. Journal of Geo-Information Science, 2023, 25(6): 1228-1239.

[5] Biswash S K. Device and network driven cellular networks architecture and mobility management technique for fog computing-based mobile communication. Journal of Network and Computer Applications, 2022, 200(4): 1-16.

[6] Balakumar J, Vijayarani S. Automatic text classification using machine learning and optimisation algorithms. Soft Computing, 2021, 25(2): 1129-1145.

[7] Zhang L. Implementation of classification and recognition algorithm for text information based on support vector machine. International Journal of Pattern Recognition and Artificial Intelligence, 2020, 34(8): 1-16.

[8] Rezaeian N, Novikova G. Persian text classification using naive bayes algorithms and support vector machine algorithm. IAES Indonesia Section, 2020, 8(1): 178-188.

[9] Dizaji Z A, Asghari S, Gharehchopogh F S. An improvement in support vector machines algorithm with imperialism competitive algorithm for text documents classification. Signal and Data Processing, 2020, 17(1): 117-130.

[10] Xing C, Yun L. ReLMKG: reasoning with pre-trained language models and knowledge graphs for complex question answering. Applied Intelligence: The International Journal of Artificial Intelligence, Neural Networks, and Complex Problem-Solving Technologies, 2023, 53(10):12032-12046.

[11] Yuhui Z. A study of automated deep classification of literature based on Chinese library classification. Libraly Journal, 2024, 43(395): 61-74.

[12] Jiang Y. English books automatic classification according to CLC. Beijing Da Xue Xue Bao, 2023, 59(1): 11-20.

[13] Liu X, Wang S, Lu S, Yin Z, Li X, Yin L, Tian H W, heng, W. Adapting feature selection algorithms for the classification of Chinese texts. Systems, 2023, 11(9): 483-483.

[14] Suganya G, Mariappan P, Dubey P, Drolia A R, Srihari S. Subjective areas of improvement: A personalised recommendation. Procedia Computer Science, 2020, 172: 235-239.

[15] Sun J, Zhu M, Jiang Y, Liu Y, Wu L. Hierarchical attention model for personalised tag recommendation. Journal of the Association for Information Science and Technology, 2021, 72(3): 173-189.

[16] Zhou Y. Design and implementation of book recommendation management system based on improved apriori algorithm. Intelligent Information Management, 2020, 12(3): 75-87.

[17] Li G, Zhuo J, Li C, Hua J, Yuan T, Niu Z, Ji D, Wu R, Zhang H. Multi-modal visual adversarial Bayesian personalized ranking model for recommendation. Information Sciences, 2021, 572(1): 378-403.

[18] Zhang X, Liu X, Guo J, Bai W, Gan D. Matrix factorization based recommendation algorithm for sharing patent resource. IEICE Transactions on Information and Systems, 2021, E104. D(8): 1250-1257.

[19] Aljunid M F, Manjaiah D H. Multi-model deep learning approach for collaborative filtering recommendation system. CAAI Transactions on Intelligence Technology, 2020, 5(4): 268-275.

[20] Wang X, Ma W, Guo L, Jiang, H, Liu F, Xu C. HGNN: Hyperedge-based graph neural network for MOOC Course Recommendation. Information Processing & Management, 2022, 59(3): 1-18.

Optimization of Distribution Routes in Agricultural Product Supply Chain Decision Management Based on Improved ALNS Algorithm

Liling Liu[1], Yang Chen[2]*, Ao Li[3]
School of Economics and Management, Ji'an Vocational and Technical College, Ji'an 343000, China[1, 2]
Wentworth Graduate College, University of York, Heslington, York, YO10 5DD, UK[3]

Abstract—The transportation of fresh agricultural products is not conducted along a sufficiently precise route, resulting in an extended transportation time for vehicles and a consequent deterioration in product freshness. Therefore, the study proposes an agricultural product transportation path optimization model based on an optimized adaptive large neighborhood search algorithm. The Solomon standard test case is used for the experiment, and the algorithm before and after optimization is compared. From the results, the optimized method was effective for the distribution model C201, R201, and CR201 sets after conducting case analysis. The total cost of the R201 transportation set was the lowest, while C101 had the highest total cost. The lowest vehicle cost consumption was R201 at 600, and the highest was C101 at 2220. The C101 algorithm took 145 s to calculate, and R201 took 199 s. All values of CR201 were average, with high fault tolerance. The proposed method was used to address the optimal operator solution. The C201 example took 244 s to calculate 2350 objective function values. The R201 example took 239 s to obtain 657 objective function values. The CR201 example took 233 s to obtain 764 objective function values. This indicates that the designed method has a significant effect on optimizing the distribution path of agricultural products. Compared with the unimproved algorithm, it has more accurate search ability and lower transportation costs. This algorithm provides path optimization ideas for the agricultural product transportation industry.

Keywords—ALNS; agricultural products; path optimization; cold chain transportation; supply chain

I. INTRODUCTION

In a rapidly developing society, the demand for fresh agricultural and related products among urban residents is increasing day by day. In recent years, the e-commerce industry has developed rapidly. Online ordering of fresh agricultural and related products has become one of the main consumption channels that is widely popular among consumers [1]. The distribution task of agricultural products in urban areas is also increasing due to low distribution efficiency, which greatly affects the industrial development. As fresh agricultural and related products themselves have short freshness and shelf life, they are prone to deterioration over time. Once the freshness is too low, they lose their nutritional value and appearance as products for sale. Therefore, how to ensure product freshness during transportation is the main issue that needs to be urgently addressed in the logistics of the entire supply chain. This poses a challenge to the timeliness of logistics transportation and the cold chain level of transport vehicles [2]. The competition for Fresh Agricultural Products (FAP) to stand out is the entire supply chain. The circulation mode of FAP refers to the transfer mode from the place of origin to the dining table, including various elements involved in the circulation of agricultural products [3]. For the supply chain, it is crucial to increase agricultural product enterprises, production and suppliers of raw materials in the middle and upper reaches, such as vegetables, seedlings and pigs. In addition to being responsible for sowing, picking, breeding, slaughtering and packaging FAP, enterprises must also directly supply raw processed agricultural products to wholesale or retail companies [4, 5]. The supply chain has drawbacks such as high loss, untimely delivery and lack of trust. The reason for this is that certain fresh ingredients that require strict time and storage conditions have increased cold chain transportation pressures and transportation costs.

More and more scholars have noticed that the transformation of agricultural supply chains requires strong technical support. Yu and Rehman proposed an evolutionary game model on the basis of the relationship between agricultural product suppliers and urban residents. This model applied evolutionary game theory to analyze the financing game model. The results indicated that the model could effectively improve the operational capability of agricultural product platforms [6]. Fu et al. introduced contract and trust mechanisms to control the uncertainty. Therefore, a digital system coupling relationship between blockchain and FAP supply chain was proposed. The results indicated that the blockchain-based digital system could help the agricultural supply chain achieve significant industrial transformation [7]. Syofya et al. proposed a value-added approach through transparent methods and supply chain management among commercial actors to address the impact of the Clincy coffee agricultural supply chain on the agricultural economic added value in Chambe Province. The results showed that this method effectively increased the yield of coffee agricultural products [8]. Mukherjee et al. established a decentralized, data-immutable, smart contract supply chain, transparency, and shared database for blockchain technology in complex multi-electronic supply chain. The results indicated that the supply chain provided deep significance for potential practitioners [9]. Luckstead et al. discussed the impact of the pandemic on workers in the food supply chain accepting important job decisions. The study analyzed the attitudes of low-skilled workers towards the processing plant industry during the epidemic. The results showed that gender, current agricultural

workers, and information about COVID-19 and agricultural workers affected respondents' answers [10].

The optimization of supply chain transportation paths cannot be achieved without search algorithms. Prymachenko et al. proposed a method for evaluating multi-modal transportation in transportation enterprises based on the multi-modal transportation route network model. The results indicated that this method could minimize the supply cost [11]. Chang et al. found that there were problems with the route planning of freight buses in urban distribution systems. Therefore, a mixed integer linear programming model was established, and an Adaptive Large Neighborhood Search (ALNS) was developed. The results showed that the correlation of the mathematical model and the model effectiveness was demonstrated through numerical experiments [12]. Hu et al. found that fast online route decisions must be made to fulfill offline retail service commitments. Therefore, a vehicle path optimization method combining an open architecture ALNS algorithm was proposed. The results indicated that this method could achieve offline training of neural network models to generate almost immediate solutions online [13]. Relying on the two levels and multiple centers in the urban logistics joint distribution system, Li et al. analyzed the two-level joint delivery path. An ALNS algorithm was proposed to solve models with multiple deletion and insertion operators. The ALNS algorithm was faster and more effective [14]. Nikzad et al. established a two-stage stochastic mathematical model for asset protection routing under wildfires. This model used the ALNS algorithm to determine routing decisions. The results showed that numerical analysis confirmed the effectiveness [15].

In summary, domestic and foreign researchers have also introduced the ALNS algorithm for the optimization of transportation paths in agricultural supply chains, but few scholars have improved and applied the ALNS algorithm. In response to this issue, a FAP distribution path optimization model is constructed for supply chain decision-making, aiming to improve transportation efficiency. The innovation lies in the ALNS algorithm, which is adapted from the Solomon standard test case for experimental testing. This algorithm fully meets the characteristics of fresh time limit requirements during agricultural product transportation, which benefits to optimize the delivery efficiency of fresh agricultural and sideline product distribution enterprises.

II. Methods and Materials

Aiming at optimizing the distribution path of agricultural product supply chain, an improved ALNS is designed. Firstly, the transportation vehicle routing problem is introduced, and the freshness calculation of agricultural products at each stage of transportation is explained. Secondly, the operational framework of the ALSN algorithm and the transportation process and cost of cold chain vehicles are introduced. Finally, an improved ALSN algorithm model framework is proposed.

A. Distribution Cost of Agricultural Product Supply Chain Ground on Improved ALNS Algorithm

The urban transportation stage of the fresh agricultural and sideline product supply chain, which is the transportation stage of delivering goods from the supply location to the consumer's ordering location [16]. During transportation, agricultural products have the characteristics of high storage difficulty, high distribution cost, freshness requirements, and irreversibility, as well as high timeliness requirements [17]. In accordance with the features of agricultural products, the supply chain distribution path problem is reasonably optimized. Vehicle path refers to the transportation path optimized by the logistics distribution center that meets the delivery requirements under certain dispatching conditions [18]. When planning the vehicle routing problem, it is necessary to cover constraints such as customer needs, location selection of logistics centers, number of transportation vehicles dispatched on different routes, and characteristics of agricultural products. The vehicle routing problem is displayed in Fig. 1.

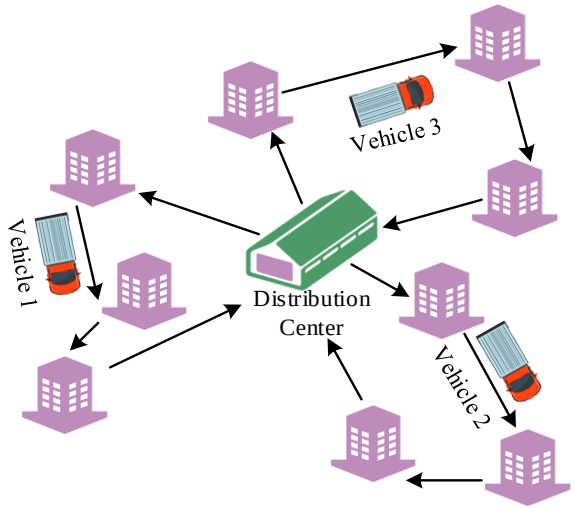


Fig. 1. Sketch map of vehicle routing problem.

Due to their strict freshness requirements, FAP lose their selling and purchasing value once they exceed the optimal freshness period. However, any fresh or similar product has the freshness loss during transportation, so freshness requirements must be an important constraint in planning and optimizing delivery routes. At present, many transportation enterprises and scholars in related fields have attached great importance to the freshness changes during the transportation of agricultural and sideline products. Various prediction algorithms have been proposed, among which linear decreasing functions can be used to represent the freshness reduction. Combining the decreasing functions, transport vehicles depart at time A, the vehicle arrived at time B, and the freshness at random time $t\left(A \le t \le B\right)$ is shown in Eq. (1).

$$\theta(t) = 1 - \frac{t - A}{B - A} \tag{1}$$

In Eq. (1), θ signifies the freshness at time t. The maximum freshness time limit T of the product after transportation time $t\left(0 \le t \le T\right)$ is shown in Eq. (2).

$$\theta(t) = 1 - \frac{t^2}{T^2} \tag{2}$$

In Eq. (2), $1-\frac{t^2}{T^2}$ represents the freshness factor of a monotonic continuous decreasing function. The freshness changes during the transportation after time t are shown in Eq. (3).

$$\theta(t)=\theta_0 e^{-\mu t} \tag{3}$$

In Eq. (3), θ_0 is the product freshness just picked. μt signifies the decreasing freshness index of FAP. To ensure that agricultural products can exhibit clear changes under the common constraints of time and preservation costs, a three parameter Weil function is used for prediction. The freshness variation of agricultural products constructed by the three parameter Weil function is shown in Eq. (4).

$$\theta(t)=\theta_0^{(\mu-f(r))\cdot t} \tag{4}$$

In Eq. (4), μ is the decay rate during transportation calculated by the three parameter Weil function. $f(r)$ represents the cost of preservation investment. FAP is also divided into different categories. To predict the freshness changes of different products, the Arrhenius function is used to construct the freshness changes, as shown in Eq. (5).

$$\theta(t)=\begin{cases}\theta_0-\mu t & if\ \gamma=0\\ \theta_0\cdot\exp(-\mu t) & if\ \gamma=1\end{cases} \tag{5}$$

In Eq. (5), γ represents the reaction level of FAP. In addition to the inherent freshness characteristics, transportation efficiency is also affected by the traffic congestion in different regions, the number of residents traveling at different time points, and license plate restrictions. Search models can effectively improve uncertainty factors. ALNS is an algorithm based on large-scale neighborhood search. The solution process of this algorithm is to first calculate the global optimal solution, then move and insert this optimal solution to iteratively calculate and obtain more domain optimal solutions near the optimal solution range [19]. The optimal solution for insertion and removal in this algorithm can represent the planned consumer ordering location in the transportation path. When different transportation routes pass through this location, the nearby better route is searched again and closely associated with the nearest transportation point. The removal and insertion processes of the ALNS algorithm are shown in Fig. 2.

The ALNS algorithm is affected by the weight values of the insertion and removal operators during the iterative calculation process, resulting in the inability to select the optimal path reasonably when there are too many paths to select. Therefore, it is necessary to determine in advance the effectiveness and necessity of the optimal solution calculation for the operators to be removed and inserted, and make adjustments when the optimization conditions are met. The ALNS algorithm must continuously eliminate bad paths and paths with constant distances by adaptively adjusting the adjustable values, which can continuously improve the accuracy of the algorithm's prediction. The ALNS algorithm has been discovered and used by logistics companies for transportation path optimization problems due to its advantages such as wide applicability and large macro search range. The flowchart of the ALNS algorithm is shown in Fig. 3.

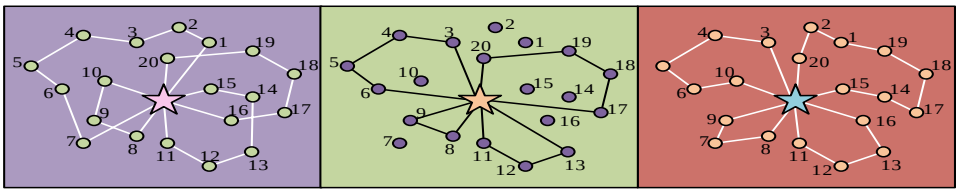


Fig. 2. Process diagram of ALNS algorithm's removal and insertion operations.

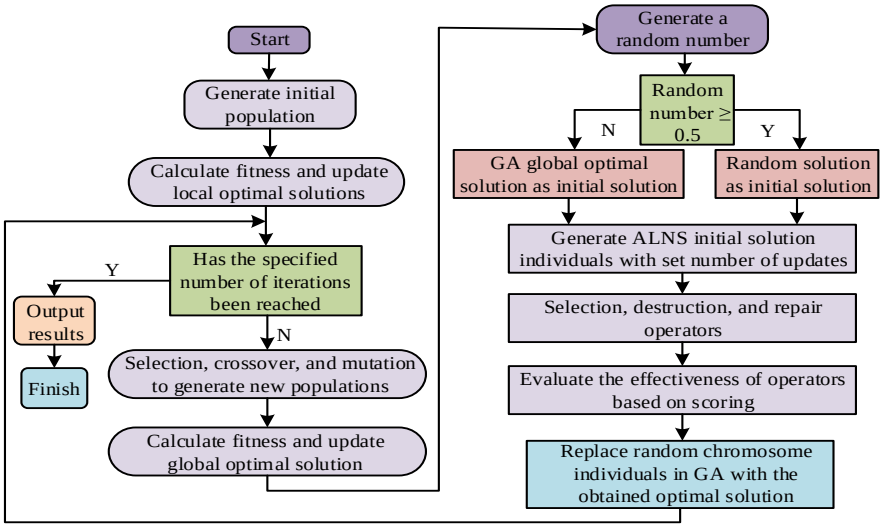


Fig. 3. ALNS algorithm flowchart.

B. Construction of Distribution Route Optimization Model for Agricultural Product Supply Chain Decision Management

The ALNS algorithm optimizes the transportation path, but the transportation mode also has an important impact on the quality of agricultural products. Cold chain transport vehicles can maintain freshness through refrigeration, which is consistent with the principle of refrigeration in refrigerators, ensuring the freshness of agricultural products to the greatest extent possible and reducing the spoilage rate. However, cold chain transportation requires a large amount of energy to cool, resulting in high costs and carbon emissions, which leads to high-cost consumption for logistics enterprises. Therefore, the unit time fuel consumption of the refrigeration unit of the cold chain truck is predicted, as shown in Eq. (6).

$$f_c(g) = R_0 + \frac{R_* - R_0}{Q_{\max}} g \tag{6}$$

In Eq. (6), f_c represents the fuel consumption rate. R_0 is the fuel consumption per unit time when the vehicle is unloaded, and R_* is the fuel consumption at full load. g is the maximum load of the cold chain truck. The fuel consumption of the refrigeration unit generator is displayed in Eq. (7).

$$F(g_{ij}) = \left(R_0 + \frac{R_* - R_0}{Q_{\max}} g_{ij} \right) T_{ijk} \tag{7}$$

In Equation (7), T_{ijk} represents the total time traveled. (i, j) represents the path traveled. $F(g_{ij})$ signifies the fuel consumption of the refrigeration unit generator in the cold chain vehicle. In light of the considerable variation in the loads of different cold chain vehicles and the marked differences in the fuel consumption of refrigeration units at different times of year, the calculation method of fuel consumption is subjected to rigorous and comprehensive analysis. Numbers 1 to 4 represent prefabricated cold, in delivery, loading and unloading, and returning after completion. Cold chain truck transportation is shown in Fig. 4.

Cold chain vehicles need to be placed in the logistics center's cold storage for full refrigeration before the delivery task departs. It can effectively avoid the aggravation of agricultural product spoilage caused by filling effects [20]. The research assumes that the time required for a single cold chain vehicle to enter the warehouse for refrigeration is T_p. Moreover, the fuel consumption calculation of the unloaded cold chain vehicle refrigeration unit at this time is shown in Eq. (8).

$$F_1 = \sum_{k=1}^{K} R_0 Z_k T_p \tag{8}$$

In Eq. (8), F_1 is the fuel consumption of the refrigeration unit. After the final stage of service, the delivery unit needs to be shut down to save energy. At this time, the cold chain truck is in an unloaded state. The fuel consumption calculation for the journey back to the logistics center is shown in Eq. (9).

$$F_2 = \sum i, j \in N, i \neq j \sum_{k=1}^{K} X_{ijk} f_c(g_{ij}) \left(T_{ijk} - T_{j0k} \right) \tag{9}$$

In Eq. (9), F_2 represents the fuel consumption of the cold chain vehicle when the refrigeration unit is turned off and unloaded. Therefore, it can be inferred that the economic cost of refrigeration fuel is calculated, as shown in Eq. (10).

$$C_{21} = P_2 (F_1 + F_2) \tag{10}$$

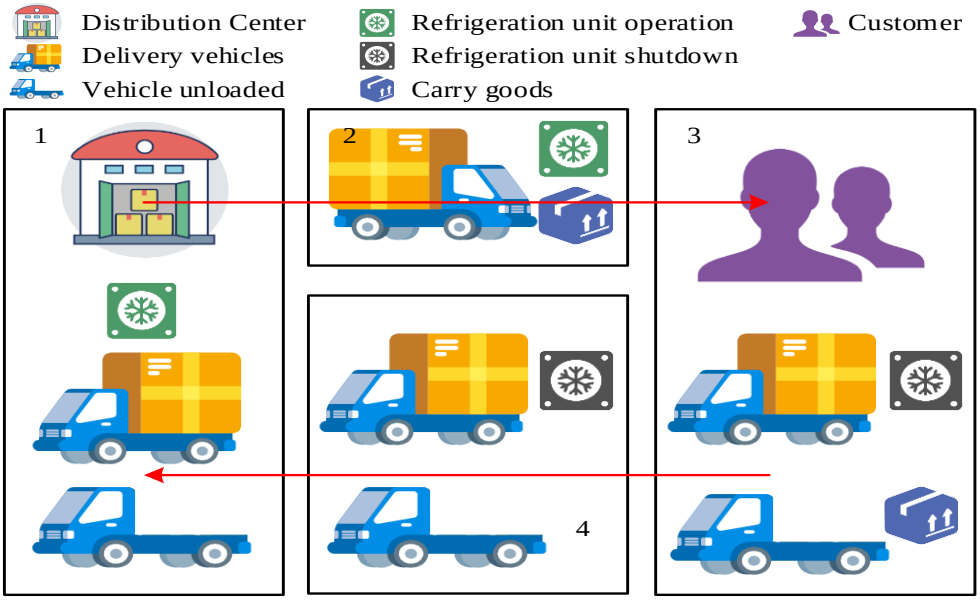


Fig. 4. Cold chain vehicle delivery process diagram.

In Eq. (10), C_{21} represents the economic cost of refrigeration fuel. The fuel consumption and carbon dioxide emissions of delivery vehicles during the delivery process are calculated, as shown in Eq. (11).

$$fuel = \chi \begin{pmatrix} FeNeVe\dfrac{d}{v} + \eta\beta d\left(v\right)^2 + \\ \eta\alpha d\left(G_d + G_i\right) \end{pmatrix} \tag{11}$$

In Eq. (11), $fuel$ represents the fuel consumption of the delivery vehicle. v represents the return speed. d represents the distance between the location of the last delivery task and the logistics center. Fe represents the friction index. Ne represents the engine speed of the cold chain vehicle. Ve represents the carbon emissions of cold chain vehicles. G_d represents the weight of the cold chain vehicle during the return journey. G_i represents the cold chain vehicle load. The research takes into account the distribution costs generated during the distribution process. The final constructed agricultural product distribution path optimization model is shown in Eq. (12).

$$\begin{aligned} \min Z = {} & c_p \sum_{k\in K}\sum_{j\in V'} x_{0jk} + \\ & c_t\left(\sum_{i\in V'} a_i + \sum_{i\in V'} w_i + \sum_{k\in K}\sum_{(i,j)\in A} x_{ijk}t_{ij}\right) \\ & + c_f \sum_{k\in K}\sum_{(i,j)\in A} x_{ijk} f_{ij} + c_e \sum_{k\in K}\sum_{(i,j)\in A} x_{ijk} e_{ij} \end{aligned} \tag{12}$$

In Eq. (12), c_p represents the cost of dispatching. c_t represents the cost of transportation labor. a_i signifies the time when the delivery vehicle arrives at customer i. w_i signifies the waiting time before the vehicle starts transportation. x_{ijk} represents the transportation vehicle k traveling from node i to customer j. c_f represents the fuel consumption cost. c_e represents the carbon emissions cost. To ensure that the freshness of agricultural products received by consumers exceeds the expected requirements, the calculation is shown in Eq. (13).

$$\theta_i \geq \theta_r, \quad \forall i \in V' \tag{13}$$

In Eq. (13), θ_i represents that the agricultural products are within the freshness expected by consumer i. θ_r represents the minimum freshness that consumers can accept. The waiting time for the delivery vehicle of agricultural products to consumers is shown in Eq. (14).

$$w_i = b_i - a_i, \quad \forall i \in V' \tag{14}$$

In Eq. (14), w_i represents the waiting time before the delivery vehicle starts transportation. b_i represents the time when consumer i started being served. The time for the consumer to confirm receipt, the delivery vehicle to leave the delivery point, and proceed to the next service point is displayed in Eq. (15).

$$\tau_i = b_i + s_i, \quad \forall i \in V' \tag{15}$$

In Eq. (15), τ_i signifies the time when the delivery vehicle leaves after completing the task. s_i represents the time when consumers accept agricultural products. The calculation results of the model constructed represent the set of distribution paths for cold chain vehicles. Meanwhile, the study enhances the ALNS algorithm by incorporating insertion and removal operators, specifically the ordering consumer. The improved ALNS algorithm is shown in Fig. 5.

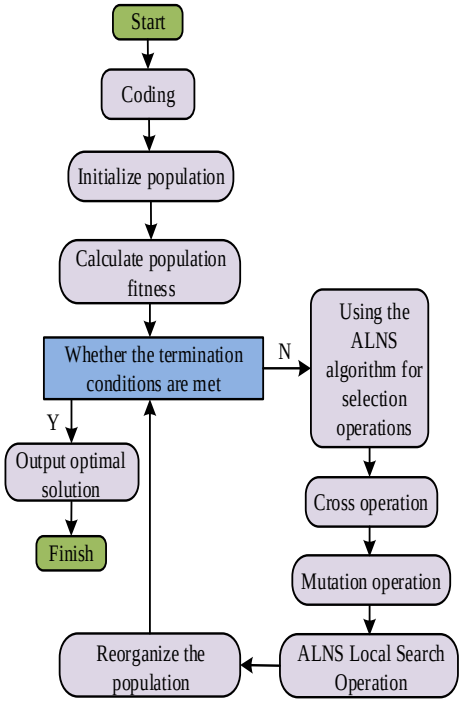


Fig. 5. Optimized ALNS algorithm.

The distribution of agricultural products exhibits regional characteristics, with a greater concentration observed in urban residential areas. The service distance for consumers in close proximity to one another is approximately equivalent [21,22]. The transportation path needs to meet various delivery conditions of the waybill, ensuring smooth driving, less congested road sections, and less freshness loss. Therefore, based on the density of distribution tasks, the supply chain hub is established. The distribution hub combined with the distribution path optimized by ALNS can ensure the freshness of agricultural products reaching consumers. Cold chain vehicles can also minimize consumption. The red dots represent the points at which consumers are required to complete delivery tasks when placing orders. The black five-pointed stars represent the points at which supply chain hubs are constructed. The overview of distribution tasks and hub construction is shown in Fig. 6.

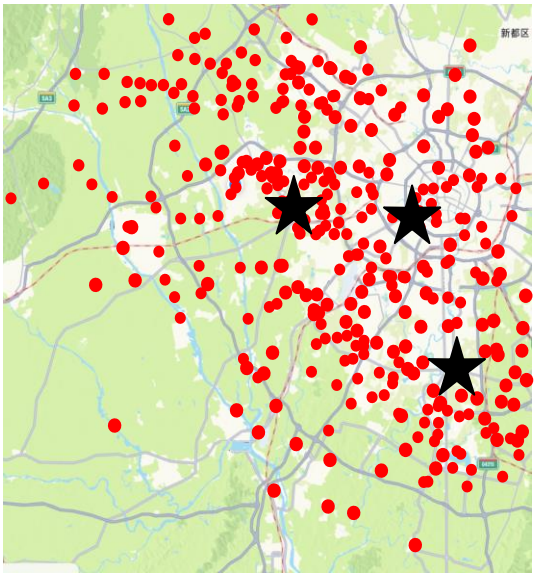


Fig. 6. Overview of distribution tasks and hub construction.

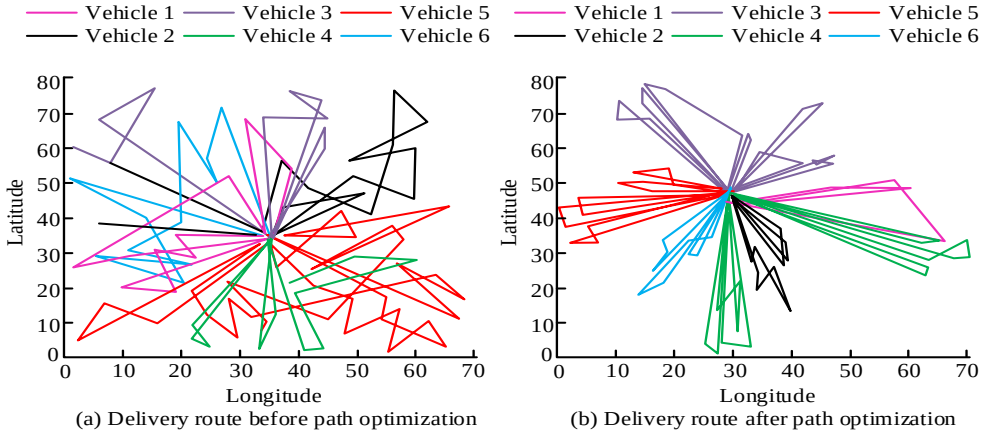
III. RESULTS

To display the effectiveness of the improved ALSN in optimizing the transportation path of agricultural products, a set of case studies were conducted. Firstly, a standard test case was constructed to compare the driving paths of cold chain vehicles. Next, C201, R201, and CR201 were used to conduct case studies to further validate the freshness and total cost of agricultural products. Finally, the path results before and after ALSN algorithm optimization were compared.

A. Effectiveness of Distribution Route Optimization in Agricultural Product Supply Chain Decision Management

Due to the high demand for freshness in agricultural products, a cold chain distribution route model for agricultural products was constructed. Relevant experimental data required for the model was supplemented. The Solomon standard test case was used to conduct numerical experiments on the adapted ALNS algorithm. The experiment adopted Windows 10, 64 bit operating system, and the processor uses Intel® Xeon® Platinum 8124 M, with 64 CB memory. The experiment was conducted using Solomon standard test cases adapted and downloaded from the website neo.lcc.uma.exe/vrp/solution methods/. An example of a set of 100 consumers was analyzed. Class C refers to densely distributed consumption points, Class R refers to dispersed consumption points, and CR refers to consumption points with cross distribution. The experiment mainly focused on C201, R201 and CR201 sets for example analysis. Therefore, the distribution path scheme of the six cold chain vehicles presents two states, as shown in Fig. 7.

From Fig. 7(a), before the optimization of the distribution path, the path was relatively chaotic and cumbersome. Six cold chain vehicles crossed the central hub significantly, resulting in high transportation costs and low efficiency. Fig. 7(b) shows the optimized distribution route. The transportation of each cold chain truck was in an orderly manner. Among them, vehicles 3 and 4 had a wider service range due to their larger capacity, while vehicle 5 had a narrower time window constraint and presented a narrow and short driving path due to its urgent service demand at consumer points. The cost of six vehicles and the freshness delivered to consumers are displayed in Table I.



(a) Delivery route before path optimization (b) Delivery route after path optimization

Fig. 7. Delivery path plan for cold chain vehicles.

TABLE I. THE CORRELATION EVALUATION COEFFICIENTS OF THE EXPERIMENT AND THE CALCULATED RESULTS

Example number	Total cost	Vehicle cost	Labor costs	Fuel cost	Low carbon cost	Average freshness	Medium speed driving ratio	Algorithm running time
CR101	1457	1540	23	35	8	90.96%	19.26%	153 s
CR201	764	700	26	31	7	82.86%	21.17%	195 s
R101	1676	1600	25	33	8	92.55%	22.19%	149 s
R201	657	600	24	27	6	80.81%	20.72%	199 s
C101	2533	2220	106	40	9	80.86%	21.66%	145 s
C201	2350	2200	103	39	8	81.17%	19.31%	194 s

From Table I, in order to evaluate more objective and accurate path optimization examples, the agricultural products delivered to consumers were all delivered with average freshness and moderate transportation speed. The lowest total cost for R201 transport set was 657. The highest total cost of C101 was 2533, with the lowest vehicle cost consumption of R201 at 600 and the highest consumption of C101 at 2220. The minimum running time of the algorithm was C101, taking 145 s, and the maximum time was R201, with a total time of 199 s. All values of CR201 were average values, and the optimized delivery path was within this average value, with high fault tolerance, which could meet consumers' requirements for freshness of agricultural products, and the cost was also within a reasonable range. The ALNS algorithm used the following parameters when calculating the substitution example: maximum number of customers removed 15 (N), weight response coefficient 0.9 (ρ), weight score σ 1 (50), weight score σ 2 (20), weight score σ 3 (5), weight score σ 4 (0), and initial annealing temperature 5000 (Te). The numerical variation of the global optimal solution with specific values is shown in Fig. 8.

In Fig. 8, C201 showed a downward trend before 200 iterations, dropping from 3500 to around 2500. After 200 iterations, the value remained constant at 2400, with small fluctuations. The overall trend of R201 values was roughly consistent with C201, with a decrease from the highest value of 2300 before 250 iterations to 1600. After 250 iterations, the values fluctuated around 1600. CR201 showed significant fluctuations before 450 iterations, with values dropping from 1200 to 700, but the overall change was flat. This indicated that as the iteration increases, the optimal solution converged, and the weights of the three sets of examples decreased. Operators were quickly stacked in the early stage, which maximized the probability of obtaining the global optimal solution.

B. *Analysis of Factors Influencing Delivery Routes based on Improved ALNS Algorithm*

In order to further analyze the ALNS algorithm, it was necessary to compare the weights of the insertion and removal operators in the actual optimization effects during solving. The study mainly analyzed the Worst Removal (WOR), Wait-time Related Removal operator (WRR), Regret Insertion operator (REI). The calculation example was validated to obtain the updated weight values and their adaptability as the number of iterations increased. The iteration is shown in Fig. 9.

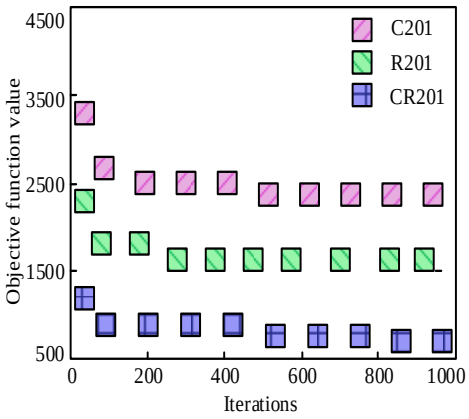


Fig. 8. Objective function iteration diagram.

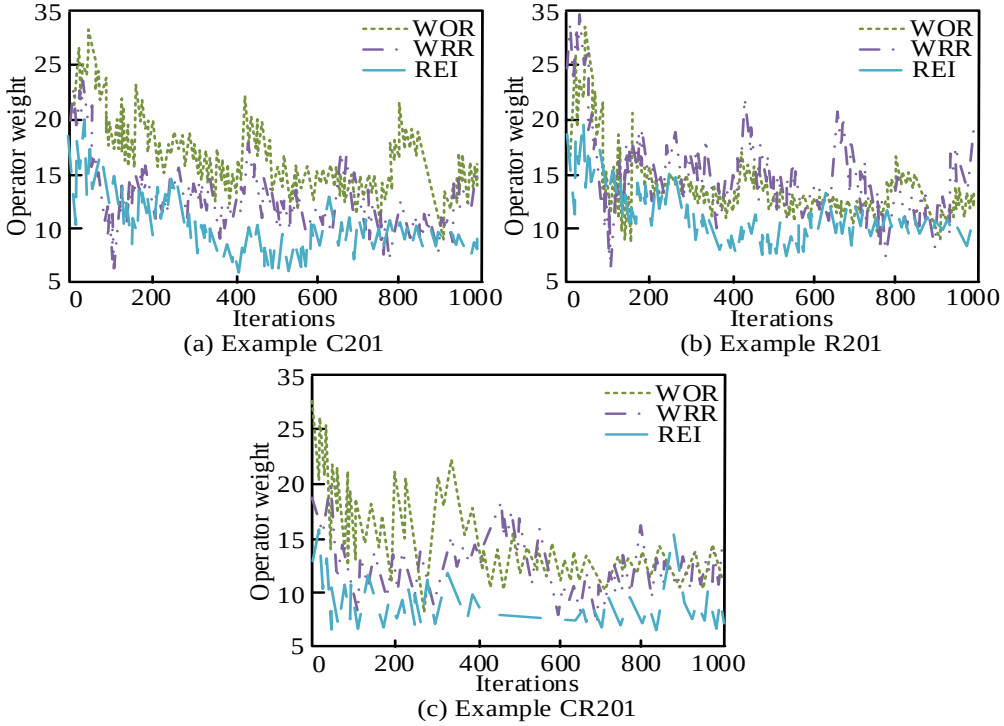


Fig. 9. Operator weights and usage iterations graph.

As shown in Fig. 9(a), the WOR operator had a higher iteration weight in example C201, which was much higher than WRR and REI. It had significant changes in the overall curve, with the highest weight value of 33 and the lowest weight value of 11. The WRR operator was in the middle, with a minimum weight value of 6 and a maximum weight value of 24. The weight values of the REI operator were the highest at 17 and the lowest at 7. In Fig. 9(b), the WRR operator had the highest weight of 35 and the lowest weight of 7. The overall curve position and most of the values were higher than the other two operators, which indicated that WRR had the highest proportion of weights. The WOR operator was the median curve, with a maximum weight of 34 and a minimum weight of 8. The maximum weight value of the REI operator curve was 19, and the minimum was 7. In Fig. 9(c), in the CR201 example, the weight values of the WOR operator were relatively stable in the later stage, with a maximum of 26 and a minimum of 9. The overall fluctuation of the WRR operator curve was uniform, with a maximum of 19 and a minimum of 7. The weight value curve of the REI operator had the smallest variation, with a maximum value of 17 and a minimum value of 6. This demonstrated that the improved algorithm yielded more accurate results. The running results before and after improvement is displayed in Fig. 10.

In Fig. 10, the purple color represented the unimproved ALNS algorithm. Among them, C201 took 194 s to calculate 2495 objective function values, R201 took 199 s to calculate 699 objective function values, and CR201 took 195 s to calculate 768 results. The green color represented the improved ALNS algorithm. Among them, C201 took 244 s to calculate 2350 objective function values, and R201 took 239 s to calculate 657 objective function values. The CR201 example took 233 s to attain 764 objective function values. This indicated that the improved ALNS algorithm had higher efficiency and less time consumption in the same number of iterations. The decay rate also affected the delivery quality of agricultural products. Based on a decay rate of 0.01, the study incorporated the rates of each stage of agricultural products into the improved ALNS algorithm for optimal solution calculation, as shown in Fig. 11.

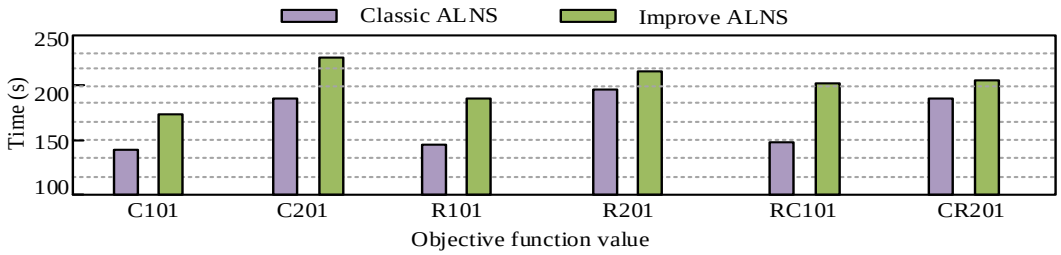


Fig. 10. Comparison of running results before and after algorithm improvement.

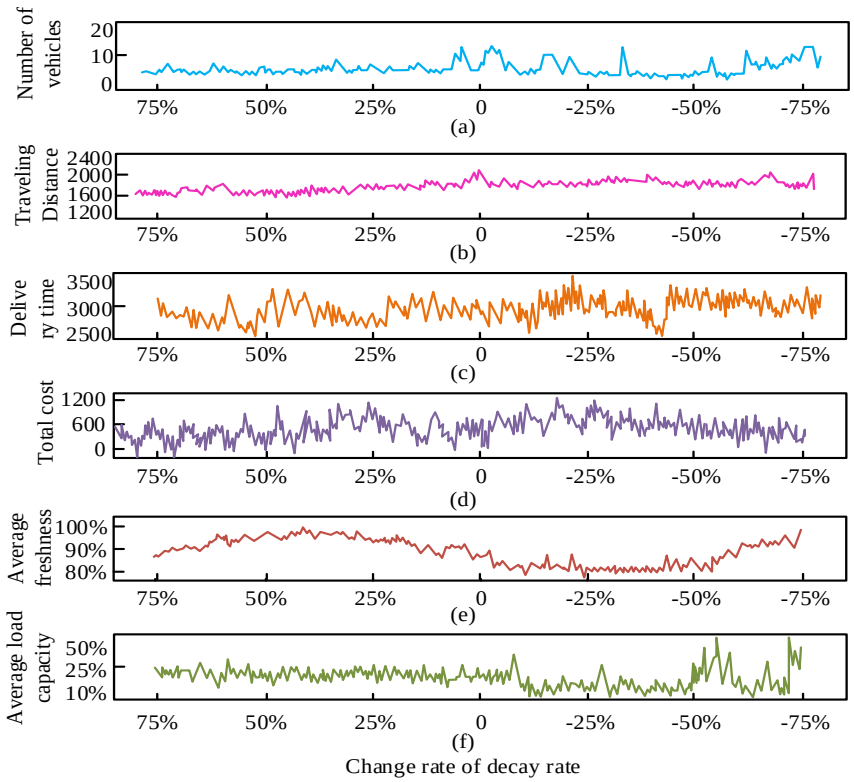


Fig. 11. Analysis of the calculation results of decay rate.

Fig. 11(a) shows the number of transportation vehicles for agricultural products. At a decay rate of 75%, the maximum number of transport vehicles reached 11. In Fig. 11(b), the driving distance was the lowest at a speed of -75%, only 1510 km. In Fig. 11(c), the delivery time was the highest at a 25% decay rate, reaching 3043 minutes, and the lowest at a 50% rate, reaching 2478 minutes. The total cost in Fig. 11(d) was directly proportional to the change in decay rate. In Fig. 11(e), when the decay rate was -75%, the freshness was as high as 91.89%. When the rate was 25%, the freshness remained the lowest at 81.55%. In Fig. 11(f), when the decay rate was -75%, the freshness was as high as 91.89%. When the rate was 25%, the freshness remained the lowest at 81.55%. From this, in practical application, the optimized path for delivering agricultural products was faster and more efficient, with the lowest cost and the best freshness.

IV. Discussion

As the demand for fresh produce delivery increases, consumers have demands for delivery times and product freshness. Therefore, the study proposes an improved ALNS algorithm for optimizing the cold chain distribution path of agricultural products, taking into account product characteristics comprehensively. The results showed that after analyzing the optimized distribution models C201, R201, and CR201, the total cost of the R201 transportation set was the lowest at 657 and the highest at 2533. The lowest vehicle cost consumption was 600 for R201 and 2220 for C101. The minimum runtime of the algorithm was C101, taking 145 s. Moreover, the maximum runtime was R201, with a total runtime of 199 s. All values of CR201 were average, with high fault tolerance. The improved ALNS algorithm was used to solve the operator optimal solution, and the C201 case took 244 s to calculate 2350 objective function values. The R201 example took 239 s to obtain 657 objective function values. The CR201 example took 233 s to obtain 764 objective function values. The study optimized the delivery path by combining the ALNS algorithm to ensure that the average freshness of agricultural products was above 80%. The lowest loss cost was achieved when the spoilage rate was between -25% and -75%. The higher the decay rate, the more cold chain vehicles were used, and the lower the decay rate, the shorter the driving distance. When the delivery time was at a decay rate of 25%, it took the most time, reaching 3043 minutes. Moreover, the lowest consumption was at a rate of 50%, 2478 minutes. The improved ALNS algorithm proposed in the study has significant advantages in optimizing the cold chain distribution path of agricultural products and can provide a reference for path optimization in the agricultural product distribution industry. Nevertheless, research is predominantly grounded in historical empirical data, which limits its practical applicability. In the future, there is the potential for greater use of real-time data in research.

V. Conclusion

The research proposes an improved ALNS algorithm for optimizing the cold chain distribution path of agricultural products. By combining the ALNS algorithm to optimize the distribution path, the average freshness of agricultural products was ensured to be above 80%, and the lowest loss cost was achieved when the decay rate was between -25% and -75%. Ni C et al. also obtained similar data for the verification calculation of the freshness of cold chain vehicles, which proved that the higher the decay rate, the more cold chain vehicles were used, the lower the decay rate, and the shorter the driving distance [23]. When the decay rate was -75%, the freshness reached 91.89%. Moreover, when the rate was 25%, the freshness remained the lowest at 81.55%. Wofuru Nyenke O et al. also obtained similar data on freshness preservation under different decay rates, proving the effectiveness of the research experiment [24]. When the delivery time was at a decay rate of 25%, it took the most time, reaching 3043 minutes. Moreover, the lowest consumption was at a rate of 50%, 2478 minutes. Bao H et al. also obtained similar data in their experiment on the effect of cold chain truck delivery time on the spoilage rate of agricultural products [25]. This indicates that the improved ALNS algorithm proposed in the study has significant advantages in optimizing the cold chain distribution path of agricultural products and can provide a reference for path optimization in the agricultural product distribution industry.

Acknowledgment

This work was supported by “the Network Marketing Research Center for Characteristic Agricultural Product of Ji’an City” of Ji’an City Science and Technology Platform.

References

[1] A. Dwivedi, A. Jha, D. Prajapati, N. Sreenu, S. Pratap, “Meta-heuristic algorithms for solving the sustainable agro-food grain supply chain network design problem,” Modern Supply Chain Research and Applications, Vol.2, pp. 161–177.

[2] F. T. S. Chan, Z. X. Wang, A. Goswami, A. Singhania, M. K. Tiwari, “Multi-objective particle swarm optimisation based integrated production inventory routing planning for efficient perishable food logistics operations,” INT J PROD RES, Vol. 58, pp. 5155–5174.

[3] Y. Zhang, X. Kou, Z. Song, Y. Fan, M. Usman, V. Jagota, “Research on logistics management layout optimization and real-time application based on nonlinear programming,” Nonlinear Engineering, Vol. 10, pp. 526–534.

[4] T. Vaiyapuri, V.S. Parvathy, V. Manikandan, N. Krishnaraj, D. Gupta, K. Shankar, “A novel hybrid optimization for cluste-based routing protocol in information-centric wireless sensor networks for IoT based mobile edge computing,” WIRELESS PERS COMMUN, Vol. 127, pp. 39–62.

[5] S. Kumar, R. Agrawal, “A hybrid C-GSA optimization routing algorithm for energy-efficient wireless sensor network,” WIREL NETW, Vol. 29, pp. 2279–2292.

[6] Z. Yu, A. Rehman Khan S, “Evolutionary game analysis of green agricultural product supply chain financing system: COVID-19 pandemic,” INT J LOGIST-RES APP, Vol. 25, pp. 1115–1135.

[7] H. Fu, C. Zhao, C. Cheng, H. Ma, “Blockchain-based agri-food supply chain management: case study in China,” INT FOOD AGRIBUS MAN, VOL. 23, pp. 667–679.

[8] H. Syofya, A. Chatra, “The influence of traceability of kerinci coffee agricultural products on agricultural value added in Jambi Province,” International Journal of Entrepreneurship and Business Development, Vol. 5, pp. 246–252.

[9] A. A. Mukherjee, R. K. Singh, R. Mishra, S. Bag, “Application of blockchain technology for sustainability development in agricultural supply chain: Justification framework,” OPER MANAGE RES, Vol. 15, pp. 46–61.

[10] J. Luckstead, Jr. R. M. Nayga, H. A. Snell, “Labor issues in the food supply chain amid the COVID-19 pandemic,” APPL ECON PERSPECT P, Vol. 43, pp. 382–400.

[11] H. O. Prymachenko, O. Shapatina, “Pestremenko-Skrypka O S, Shevchenko, A. V., Halkevych, M. V. Improving the technology of product supply chain management in the context of the development of

multimodal transportation systems in the European union countries," International Journal of Agricultural Extension, Vol. 10, pp. 77–89.

[12] Z. Chang, H. Chen, F. Yalaoui, B. Dai, "Adaptive large neighborhood search Algorithm for route planning of freight buses with pickup and delivery," J IND MANAG OPTIM, Vol. 17, pp. 1771–1793.

[13] H. Hu, Y. Zhang, J. Wei, Y. Zhan, X. Zhang, S. H uang, S. Jiang, "Alibaba vehicle routing algorithms enable rapid pick and delivery," INFORMS J APPL ANAL, Vol. 52, pp. 27–41.

[14] Z. Li, Y. Zhao, Y. Zhang, R. Teng, "Joint distribution Location-routing problem and large neighborhood search algorithm," Journal of System Simulation, Vol. 33, pp. 2518–2531.

[15] E. Nikzad, M. Bashiri, "A two-stage stochastic programming model for collaborative asset protection routing problem enhanced with machine learning: a learning-based matheuristic algorithm," INT J PROD RES, Vol. 61, pp. 81–113.

[16] Z. Yi, Y. Wang, Y. J. Chen, "Financing an agricultural supply chain with a capital-constrained smallholder farmer in developing economies," PROD OPER MANAG, Vol. 30, pp. 2102–2121.

[17] Z. Wu, Y. Zhao, N. Zhang, "A literature survey of green and Low-Carbon economics using natural experiment approaches in top field journal," Green and Low-Carbon Economy, Vol. 1, pp. 2–14.

[18] S. K. Dewi, D. M. Utama, "A new hybrid whale optimization algorithm for green vehicle routing problem," SYST SCI CONTROL ENG, Vol. 9, pp. 61–72.

[19] C. Pfeiffer, A. Schulz, "An ALNS algorithm for the static dial-a-ride problem with ride and waiting time minimization," Or Spectrum, Vol. 44, pp. 87–119.

[20] X. Xu, Z. Lin, X. Li, C. Shang, Q. Shen, "Multi-objective robust optimisation model for MDVRPLS in refined oil distribution," INT J PROD RES, Vol. 60, pp. 6772–6792.

[21] Ni C, Dohn K. Research on Optimization of Agricultural Products Cold Chain Logistics Distribution System Based on Low Carbon Perspective. International Journal of Information Systems and Supply Chain Management (IJISSCM), 2024, 17(1): 1-14.

[22] Wofuru-Nyenke O. Routing and facility location optimization in a dairy products supply chain. Future Technology, 2024, 3(2): 44-49.

[23] Liu Q. Logistics Distribution Route Optimization in Artificial Intelligence and Internet of Things Environment. Decision Making: Applications in Management and Engineering, 2024, 7(2): 221-239.

[24] Bao H, Fang J, Zhang J, Wang, C. Optimization on cold chain distribution routes considering carbon emissions based on improved ant colony algorithm. Journal of System Simulation, 2024, 36(1): 183-194.

[25] Tang Q, Qiu Y, Xu L. Forecasting the demand for cold chain logistics of agricultural products with Markov-optimised mean GM (1, 1) model−a case study of Guangxi Province, China. Kybernetes, 2024, 53(1): 314-336.

Harnessing Technology to Achieve the Highest Quality in the Academic Program of University Studies

The Quality Based on ABET and NCAAA Accreditations

Rania Aboalela
Information Systems Department, King Abdulaziz University, Jeddah, Saudi Arabia

***Abstract*—This research aims to utilize information technology to improve education quality, particularly in higher education. A key contribution of this research is the application of generative artificial intelligence, specifically ChatGPT, to validate test questions that meet both international (ABET) and local (NCAAA) academic accreditation standards. The study was conducted within the Information Systems Department's bachelor's program at King Abdulaziz University in the Kingdom of Saudi Arabia, focusing on a website development course. The custom ChatGPT application, named Question Checker, was developed to validate questions generated by instructors. These validation criteria were aligned with the accreditation requirements for technology and computer science programs, ensuring compliance with both ABET and NCAAA standards. The application was tested by validating nine questions related to Student Outcomes, demonstrating its effectiveness in supporting the educational objectives of the program.**

Keywords*—ChatGPT; academic accreditations; technology programs; computer science; Kingdom of Saudi Arabia; website development; NCAAA; ABET

I. Introduction

ChatGPT is a Generative Pre-trained Transformer (GPT) language model designed to generate text in response to natural language inputs. It received widespread media coverage following the launch of a free preview by OpenAI in December 2022 [1][2]. The goal of ChatGPT is to mimic human communication not only in language translation systems but also in other applications such as chatbots and virtual assistants. ChatGPT employs highly effective machine learning techniques and has been trained on extensive textual datasets, creating a robust intelligence that enables it to provide nearly flawless responses to user input. It represents a significant advancement capable of revolutionizing the way humans interact with technology, fostering more conversational and intuitive communication. ChatGPT is applied in various domains, including customer service chatbots, language translation tools, and virtual assistants. Research into its potential applications in education aims to enhance student learning and engagement [3].

In this research, we propose a cutting-edge solution to assist faculty in improving the quality of the questions they generate by incorporating ChatGPT into the question-generation process. Additionally, this study focuses on validating the correctness of questions generated by instructors. This modern method allows teachers to easily request questions relevant to specific academic units while ensuring compliance with rigorous accreditation criteria. It benefits the education system by enabling the creation of questions adaptable to various credits and accreditations, thereby increasing productivity. Our intention is to use Bloom's Taxonomy as a roadmap to methodically construct a robust structure aligned with NCAAA [4] and ABET standards [5]. Establishing the link between Bloom's Taxonomy and the curriculum has received considerable attention and contributes positively to the creation of high-quality questions that assess student understanding [6][7][8].

This study utilized Bloom's Taxonomy to align with various academic accreditations. Subsequently, a novel approach employing ChatGPT was devised to generate questions based on shared criteria. The mapping established correlations between the requirements and contexts of diverse academic accreditations for evaluating students' knowledge through questions. This mapping, which uses Bloom's Taxonomy verbs to gauge different knowledge levels, facilitated the selection of appropriate verbs for questions based on accreditation needs. The integration of this technology streamlines question development, ensuring accurate and swift alignment with academic accreditations, thereby saving instructors' time and aiding examiners in certifying program eligibility. The study also explored faculty members' readiness to adopt this technology for evaluating their questions in terms of compliance with ABET outcomes and NCAAA three-level domain zones. A questionnaire administered to faculty members revealed that 100% of respondents endorsed the technology's role in assisting question development, while 11% refused to accept its role in correcting question composition.

A. The Research Objective

The research aims to leverage generative artificial intelligence (AI), particularly ChatGPT, to improve the quality of test questions in an academic setting, aligning with international ABET and local NCAAA accreditations.

B. Research Questions

Is there a method to use generative artificial intelligence to correct instructors' questions to a high-quality level??

Answer:

Yes, by utilizing customized ChatGPT.

C. Methodology

To This study aims to develop and assess the effectiveness of a customized ChatGPT application designed to enhance the quality of questions in educational assessments. The focus is on ensuring that these questions meet the standards required by international and national accreditation bodies, specifically ABET (Accreditation Board for Engineering and Technology) and NCAAA (National Commission for Academic Accreditation & Assessment).

1) Selection of course and program: The study was conducted within the Information Systems bachelor's degree program at King Abdulaziz University, a leading institution in the Kingdom of Saudi Arabia. This program is globally recognized with ABET accreditation and is in the process of obtaining national certification from NCAAA. For this study, a course on Web Design was selected as the testing ground due to its relevance to both computing and information systems education.

2) Development of the ChatGPT application: To align the questions with accreditation standards, a customized ChatGPT model was developed. The model was tailored to generate and evaluate questions based on the criteria set by ABET and NCAAA. The customization process involved fine-tuning the ChatGPT application to understand and apply the specific requirements of both accreditation bodies.

3) Criteria for question evaluation: The evaluation criteria were derived from the Student Outcomes (SOs) defined by ABET and the domains outlined by NCAAA. The focus was on ensuring that the questions generated by instructors are compatible with these criteria, promoting the development of competencies that are essential for program accreditation.

4) Application of bloom's taxonomy: To further enhance the quality of the questions, Bloom's Taxonomy was employed as a framework. Bloom's Taxonomy classifies educational learning objectives into levels of complexity and specificity. By mapping the accreditation criteria of ABET and NCAAA onto Bloom's Taxonomy, the study aimed to establish clear links between these criteria and the cognitive levels required by the taxonomy. This approach ensured that the questions not only met accreditation standards but also targeted appropriate levels of cognitive learning.

5) Testing and validation: The customized ChatGPT application was tested on the selected Web Design course. The questions generated were evaluated to ensure they met the dual requirements of ABET and NCAAA accreditation. The validation process involved a detailed comparison of the questions against the established criteria, with a particular focus on the alignment with Bloom's Taxonomy.

6) Data collection and analysis: The data collected from the application testing was analyzed to assess the effectiveness of the ChatGPT model in generating questions that satisfy both accreditation standards. The analysis also explored the extent to which the use of Bloom's Taxonomy contributed to the improvement of question quality. The study utilizes the links between the two accreditation criteria to provide material helping ChatGPT evaluate instructors' questions in terms of compatibility with ABET's Student Outcomes (SOs) and the domains outlined by NCAAA [27].

The remainder of this paper is organized as follows:

Section II discusses related works; Section III illustrates the accreditation NCAAA & ABET; Section IV illustrate the Bloom Taxonomy. Section V covers the six students' outcomes of information Systems bachelor Program. Section VI illustrates the mapping Bloom's Taxonomy with NCAAA and ABET. Section VII discusses the design of the proposed custom ChatGPT. Section VIII covers the test of the proposed custom ChatGPT. Section IX covers a questionnaire to assess the acceptance of academic teachers regarding the use of artificial intelligence (AI) and the efficiency of the proposed application. Sections X and XI present the discussion, conclusions and future work, respectively.

II. Literature Review

A. GPT Technology

Number GPT (Generative Pre-trained Transformer) technology is a kind of AI language model created by OpenAI. The main objective of this model is to generate syntactically correct, human-like prose by predicting the next word in a sentence based on contextual information provided by previous words. GPT utilizes deep neural networks to process extensive text data and learn text patterns, enabling the system to generalize contextual and linguistic phrases. As a result, GPT-3, one of the latest versions of GPT, boasts over 175 billion parameters and was trained on a vast volume of internet text data, making it one of the most powerful natural language models available today [9].

B. GPT Technology in Education

There are several ways in which the GPT language model can be used in education [10]:

- GPT technology can be utilized to develop chatbots and virtual language coaches that serve as practice tools for students as they focus on their language skills.
- GPT can serve as a tool for teachers to assist students in improving the quality of their written work.
- GPT can be exploited for grading essays and other types of written assignments without human intervention, saving time and providing students with instant assessment of their progress.
- GPT technology can be applied for students' personalization of interactive learning activities. By analyzing a student's learning processes and preferences, GPT can provide recommendations concerning the type of learning materials that best suit the user, such as articles, videos, and textbooks.

C. ChatGPT

In recent years, Natural Language Processing (NLP) has undergone tremendous development. Nevertheless, the emergence of ChatGPT (Chat Generative Pre-Trained Transformer) has reignited conversations and optimism

surrounding the technology. Developed by OpenAI, ChatGPT was introduced to the public in November 2022 [11]. It quickly gained popularity, reaching over 1 million users in just five days, a stark contrast to Facebook's 300 days, Twitter's 720 days, and Instagram's 75 days [12].

ChatGPT is a large language model with extraordinary comprehension and generation capabilities, closely resembling human speech. Its unparalleled ability to answer questions, engage in conversations, and provide logical and relevant responses within the context of the conversation marks a revolutionary advancement in the development of conversational AI. The diverse applications of ChatGPT and its capacity to enhance across various sectors has brought about new discussions about this cutting-edge AI technology [13]. However, ChatGPT is merely a complicated chatbot at the early stage of Long Short-Term Memory (LSTM) research [14] and cannot be compared to developments in language processing and cognitive sciences. Nevertheless, it is widely used in many industries, including customer service assistance, e-commerce, healthcare, and education. Machine learning, a subfield of AI, enables computers to automatically learn from data, surpassing human-coded instructions. Deep learning has become a powerful predictive tool due to improvements in hardware processing power, data availability, and algorithmic innovations [15], [16], [17], [18]. Furthermore, ChatGPT needs to be fine-tuned for exam purposes as well [19]. ChatGPT is a large language model with remarkable comprehension and speech production capacities akin to those of humans. Its outstanding performance in comprehending questions, dialogue processing, and delivering contextual and coherent responses represents a significant achievement in conversational AI [14]. The first GPT model, GPT-1, was released in 2018, then a successor called GPT-2 in 2019, and later the GPT-3 model in 2020. The model's size, along with the training data and language test scores, have significantly improved since its first version. On November 30, 2022, Open AI released a free behind-the-scenes look at ChatGPT, their AI-powered chatbot expected to be worth $29 billion [20]. A chatbot is a software system that employs artificial intelligence techniques to converse with humans, simulating human communication. Users pose questions, and the system responds promptly. Within five days of its release, ChatGPT had garnered 1 million users [21].

D. ChatGPT for Exam Correction

The use of ChatGPT for generating and correcting exam questions was studied by Aboalela et al. [22]. The study found that faculty members accepted the use of ChatGPT for both producing and correcting questions. Also, the study by Weng et al. [23] evaluated ChatGPT for Taiwan's Family Medicine Board exams, which included English and Chinese. Despite its popularity and extensive database, ChatGPT's accuracy was found to be 41.6%, indicating limited performance in the medical domain. Notably, it performed better on negative-phrase, mutually exclusive, and case scenario questions. Challenges such as the exam's difficulty level and the shortage of traditional Chinese language resources probably contributed to its lower accuracy. While ChatGPT may be useful for learning and exam preparation, improvements are needed for specialized exams.

E. Benefit of ChatGPT in Education

According to Cribben and Zeinali [24] the benefits of ChatGPT in education are as follows: ChatGPT can generate course materials for professors and produce assignments, test questions, and solutions across different courses. Professors can also utilize ChatGPT to instruct a chatbot to answer students' inquiries over the internet when they are not available during office hours. As an illustration, students have the option to submit their queries to an internet-based discussion platform like eClass or Blackboard, utilizing ChatGPT.

In the same context [25] ChatGPT is an educational accessibility website that assists people with disabilities and non-English speakers by providing spoken responses, topic summarization, and translation services. It enables homework with tailored explanations and examples and builds academic skills. It supports teachers in lesson planning, test generation, grading, analysis, and resource planning in higher education. In addition, ChatGPT personalizes learning adapting to individual styles and performance and assists in exam preparation by reviewing notes, formulating answers, and identifying strengths and weaknesses.

F. Challenges of Using ChatGPT in Higher Education

The obstacles associated with using ChatGPT in higher education, as identified by EU Business School [26], are as follows: ChatGPT's extensive consumption of internet material may lead to unintentional acquisition of preconceived notions and prejudices, potentially resulting in discrimination against various demographic groups. Both students and teachers should acknowledge its inherent subjectivity of it, thoroughly scrutinizing its output for any prejudice.

III. NCAAA AND THE ABET

Discussed first is the approach that would be most effective in developing a common application to both organizations. These two accrediting authorities have different purposes. The process of NCAAA accreditation process focuses on specialized scientific institutions or academic programs. Certain minimum standards and quality requirements must be met by an institution or program to obtain accreditation from the NCAAA. The reputation of an academic institution or its program depends on whether it is accredited or not, as it provides a global reference point for students. Academic excellence, reflected in both local and global reputation, attracts high caliber researchers and practitioners seeking for assurance of quality education. Accreditation aims to ensure that the outputs of educational institutions and academic programs meet societal needs; it also seeks to foster cooperation between the education system and the professional labor market [28]. This fosters trust and belief in academic programs among the community and helps them attain a stable financial standing.

The NCAAA grants accreditation to the university, as well as to each individual program offered by the institution. The NCAAA accredits both the university as a whole and each individual program it offers. On the other hand, ABET serves as a quality assurance mechanism specifically for programs in applied and natural sciences, computer science, engineering, and engineering technology [5] ABET accreditation is globally recognized for ensuring that college or university programs meet

the quality standards of the profession they prepare graduates for. One distinction between ABET and the NCAAA lies in their accreditation focus: while the NCAAA accredits entire universities, ABET accredits specific programs [29]. Another difference is that ABET prioritizes the attainment of student learning objectives over the instruction of methodologies and course standards. However, both agencies evaluate educational procedures within academic programs and investigate similar topics related to quality assurance, including program objectives, course learning outcomes, and individual student learning outcomes. Program Learning Outcomes (PLOs) are measurable statements that describe the knowledge or skills students acquire upon completing an academic program, while Course Learning Outcomes (CLOs) are specific statements that define the knowledge, skills, and attitudes learners will demonstrate upon completing a particular course. Assessing students' learning outcomes involves formulating questions using action verbs to measure their proficiency and understanding of the subject matter. Subsequently, students' comprehension is evaluated based on the grades they achieve for questions related to specific areas of knowledge or ability. Saudi institutions have the potential to obtain both national and international accreditation for academic programs, allowing them to pursue multiple accreditation approaches. Several research studies have suggested a potential connection between ABET and the NCAAA in the Kingdom of Saudi Arabia, with published evidence supporting these potential outcomes [30] [31].

IV. Bloom's Taxonomy

The importance of Bloom's Taxonomy in assessing knowledge [32] [33] [34] lies in its capacity to correlate specific verbs with the educational outcomes required for both ABET and NCAAA accreditation. This facilitates the use of standardized verbs in questions that meet the criteria of both accrediting bodies. Thus, when employing commonly used question verbs to evaluate the same educational outcomes, alignment with both ABET and NCAAA standards is ensured. The subsequent task involves categorizing the verbs used according to the areas required for NCAAA accreditation measurement. Bloom's Taxonomy functions to classify verbs assessing skills and comprehension into six distinct categories, as depicted in Fig. 1. These six domains of Bloom's Taxonomy have been further delineated to encompass areas specific to NCAAA accreditation. The top five domains, which are the most intricate, include the NCAAA domain of skills, covering processes such as application, assessment, evaluation, and production. The lowest and most intricate domain, understanding, remains a specialized area within the NCAAA domain. Verbs were categorized according to Bloom's Taxonomy, as illustrated in the figure depicting the mapping. Once the correlation between taxonomy verbs and NCAAA measure verbs that evaluate skills is established, they can be identified based on either the prescribed verbs for ABET certification or the NCAAA accreditation criteria.

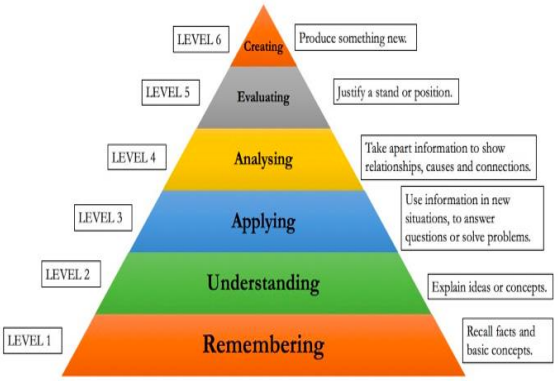


Fig. 1. The Bloom taxonomy verbs [40].

V. ABET Six Students' Outcomes of Information System Bachelor Program

The mapping of Bloom's Taxonomy to ABET is as follows:

SO 1 corresponds to the third-level verb in Bloom's Taxonomy:

Design, implement, and evaluate a computing-based solution to meet a given set of computing requirements in the context of the program's discipline.

SO 2 corresponds to levels 6, 5 and 4 based on the SO subpoint and the action verb in Bloom's Taxonomy: Communicate effectively in a variety of professional contexts.

SO 3 corresponds to special verb in Bloom's Taxonomy based on the target of the asked question and to the value domain of NCAAA:

Recognize professional responsibilities and make informed judgments in computing practice based on legal and ethical principles.

SO 4 corresponds to a special verb in Bloom's Taxonomy based on the target of the asked question and to the value domain verb in NCAAA:

Function effectively as a member or leader of a team engaged in activities appropriate to the program's discipline.

SO 5 corresponds to a special verb in Bloom's Taxonomy based on the target of the asked question and to the value domain verb in NCAAA (Affective Learning):

Support the delivery, use, and management of information systems within an information systems environment.

SO 6 corresponds to the lower level (Understanding) of Bloom's Taxonomy based on the target of the asked question and to the SKILL domain verb in NCAAA. For each SO, there are special verbs for the one director who must ask the question. To find a method to connect the verbs of each ABET's SO with the NCAAA's domains, the six ABET outcomes were divided into the six Bloom's Taxomy domains, which were then divided into the three NCAAA domains taxonomy.

VI. MAPPING BLOOM'S TAXONOMY WITH ABET AND NCAAA

Fig. 2 shows the mapping between Bloom's Taxonomy & NCAAA. Table I summarizes the mapping between the accreditations ABET & NCAAA and Bloom's Taxonomy.

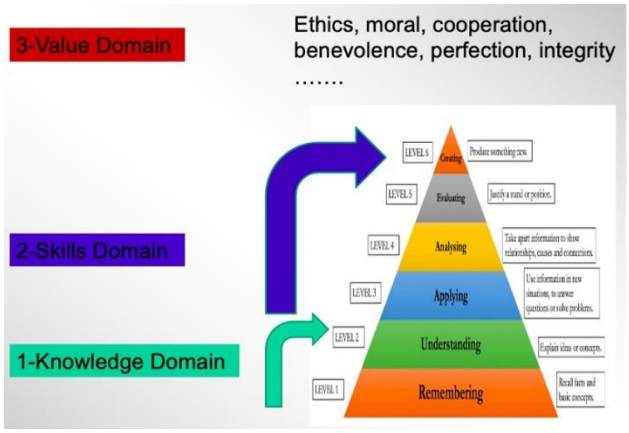


Fig. 2. Mapping Bloom's taxonomy to NCAAA three domains.

TABLE I. THE MAPPING BETWEEN THE ACCREDITATIONS AND THE BLOOM'S TAXONOMY. [22]

ABET SO	NCAAA Domains	Bloom's Taxonomy Level
SO1	Skills domain	L3, L5, L6
SO2	Skills domain	L5, L6
SO3	Value domain	L2, L3
SO4	Value domain	L2, L3
SO5	Value domain	L2, L3
SO6	Skills domain	L2, L3

Correlating the Verb Used in a Question to the ABET and Bloom's Taxonomy. ABET specializes in specific verbs to request information about each subject object. This tool facilitates the creation of questions that align with accreditations. The NCAAA mandates the use of specific verbs. The utilization of these verbs is essential for the automated production and verification of questions utilizing ChatGPT technology. ChatGPT technology is user-friendly when the computer is supplied and educated via generative AI. These identical verbs might be utilized to generate inquiries pertaining to NCAAA. They collectively fulfilling the criteria of both ABET and NCAAA. Therefore, if the academic program receives two distinct accreditations, the questions developed will be suitable for both accreditations. Here are some instances of these verbs. Table II displays the question verb that ABET has designated to assess each Student Outcome (SO). The table also displays the verb mapped to the Bloom's Taxonomy. The question verbs are expected to be the same for both ABET and NCAAA [22].

Whereas the subpoints 1.1, 1.2, 2.1, 2.2, 2.3, 3.1-3.3, 4.1-4.3, 5.1, 5.2-5.3, 6.1, 6.2 and 6.3 are subpoints of the major 6 SO and they measure the following SO.

1.1: An ability to analyze a complex computing problem. (Analyzing).

1.2: An ability to apply principles of computing and other relevant disciplines to identify solutions. (Applying).

2.1 An ability to design a computer-based system, process, component, or program to meet desired needs.

2.2: An ability to implement a computer-based system, process, component, or program to meet desired needs.

2.3: An ability to evaluate a computer-based system, process, component, or program to meet desired needs.

3.1: An ability to conduct an oral presentation using effective communication skills. (Applying).

3.2: An ability to write in a clear, concise, grammatically correct and organized manner. (Applying).

3.3: An ability to develop appropriate illustrations including hand sketches, computer generated drawings/graphs and pictures. (Applying).

TABLE II. QUESTION VERBS MAPPED TO NCAAA AND ABET SO

The question verb	The Bloom's Taxonomy VERB LEVEL	The ABET SO number
Appraise, assess, evaluate, compare, contrast, criticize, differentiate, discriminate, distinguish, examine, experiment, question, test	[Analyzing]	1.1
Choose, demonstrate, employ, illustrate, interpret, operate, schedule, sketch, draw, solve, use, write.	[Applying]	1.2
An ability to design a computer-based system, process, component, or program to meet desired needs.	[Creating]	2.1
An ability to implement a computer-based system, process, component, or program to meet desired needs.	[Applying]	2.2
An ability to evaluate a computer-based system, process, component, or program to meet desired needs.	[Evaluating]	2.3
Choose, demonstrate, employ, illustrate, interpret, operate, schedule, sketch, draw, solve, use, write.	[Applying]	3.1-3.3
Classify, describe, discuss, explain, identify, locate, recognize, report, select, translate, paraphrase	[Understanding]	4.1-4.3
Choose, demonstrate, employ, illustrate, interpret, operate, schedule, sketch, draw, solve, use, write.	[Applying]	5.1
[Affective Learning] Appreciate, accept, attempt, challenge, defend, dispute, join, judge, justify, question, share, support	Any verb level which should be determined by the SO of the topic.	5.2-5.3
Appraise, assess, evaluate, compare, contrast, criticize, differentiate, discriminate, distinguish, examine, experiment, question, test	[Analyzing]	6.1
Choose, demonstrate, employ, illustrate, interpret, operate, schedule, sketch, draw, solve, use, write.	[Applying]	6.2
Classify, describe, discuss, explain, identify, locate, recognize, report, select, translate, paraphrase.	[Understanding]	6.3

4.1: Understanding of professional responsibilities, ethical theories, legal and social issues. (Understanding).

4.2: Understanding of cyber security threats and corresponding procedures to mitigate these threats. (Understanding).

4.3: Understanding of risk management, security policies and audit procedures. (Understanding).

5.1: An ability to prepare a work schedule for the assigned task and complete it within the appropriate deadlines. (Applying).

5.2: An ability to participate in team meetings with full preparedness for providing useful input. (Affective Learning).

5.3: An ability to share ideas among the team and promote good communication among the team members. (Affective Learning).

6.1 Support the delivery of information systems within an information Systems environments.

6.2 Support the use of information system within an information Systems environments.

6.3 Support the management of Information Systems within an information Systems environments.

Studying the extent to which faculty members accept the use of ChatGPT technology. In order to produce student assessment questions and tests. These subpoints are extracted from ABET official documents from IS department of FCIT of King Abdulaziz University [35].

VII. Design a Custom ChatGPT

Proficiency in programming is essential for developing a customized ChatGPT application. However, beneficiaries may find platforms offering pre-made tools for design applications. Typically, the beneficiary must ascertain three fundamental aspects:

- The application's objective.
- Enumerate the characteristics of it.
- Selecting the platform, whether it is a mobile or web-based chat application.

To integrate GPT, the user must determine the appropriate version of GPT to utilize, such as GPT-3.5, developed by OpenAI. The designer requires an API key, which can be obtained by accessing the OpenAI platform. They can use either the library or the requests module in their chosen programming language, such as Python. In this research paper, the OpenAI platform is employed to correct questions based on accreditation requirements. Two accreditations were examined: ABET and NCAAA. The implementation test focused on the course COIS 492, which is part of the fifth level of the Information Systems Department in the bachelor's degree program at the Faculty of Computing Science & Information Technology, King Abdulaziz University, Rabigh Branch. The course adheres to ABET accreditation requirements for measuring Student Learning Outcomes (SLOs) and aims to fulfill NCAAA requirements, as national NCAAA accreditation is mandated by the Ministry of Education of Saudi Arabia [36]. The questions in the course assessments must meet ABET standards. In this research, the implementation of the ChatGPT application ensured that both NCAAA and ABET requirements were met before approving a question generated by the application. Several researchers in Saudi Arabia have studied the alignment of ABET [37] and NCAAA [38] [39] accreditation requirements as a unified and mutually satisfactory solution.

In the proposed ChatGPT application, the most crucial condition involves using the correct question verb for a specific Student Outcome within the appropriate domain. The course COIS 492 focuses on SO 2, 4, and 6. Table II illustrates the question verbs mapped to NCAAA and ABET SOs. The content of this table was integrated into the ChatGPT application to ensure the accuracy of the questions. The application's role is to verify the correctness of questions posed by the instructor for a specific SO. If the question is correct, it is approved. ChatGPT was specifically developed to test questions submitted by faculty members, ensuring their adherence to accreditation requirements such as ABET and NCAAA. The application is equipped with two files: an SO file and a relationship table between NCAAA and ABET. The application has been named "QUESTION CHECKER," as depicted in Fig. 3, which illustrates the app interface.



Fig. 3. Question checker application interface.

A. Role and Goal

This GPT serves as an advanced question checker, specifically designed for educators setting short answer and essay questions. It verifies questions for correctness based on the relation between the question's verbs and the intended student outcomes. It also offers suggestions for improving question clarity and alignment with educational goals.

B. Constraints

The checker provides feedback on question grammar, recommends more suitable verbs based on a table of verb relationships and student outcomes, and provides constructive feedback to refine questions. It proactively suggests improvements and awaits explicit queries before offering feedback, aiming to enhance the clarity and educational value of questions.

VIII. The Application Test

The application's ability to correct questions provided by instructors to align with ABET and NCAAA accreditation requirements has been tested. This includes verifying the accuracy of the question's action verb to test the appropriate Student Outcome (SO) and the relevant domain intended for assessment by the question. The application was used to test the accuracy of a diverse set of questions, identifying both incorrect and correct questions. The application was tested for two purposes: assisting in creating accurate questions and correcting erroneous questions while providing comments and suggestions.

A. To Assist in Creating Accurate Questions

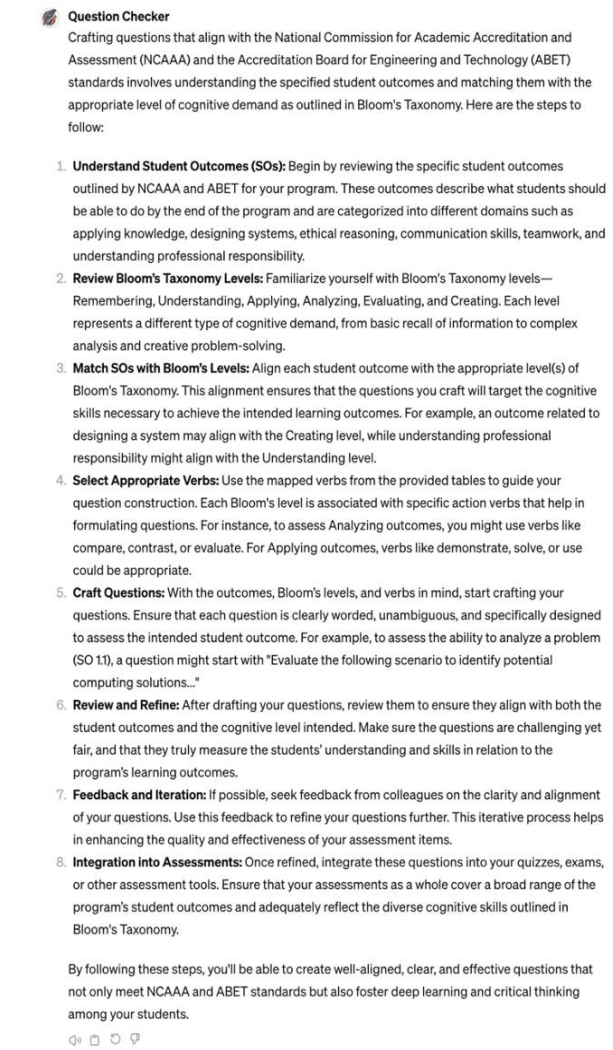
After providing the application with the accreditation criteria in terms of verbs and scope, the application explained how to formulate questions as a guide for faculty members. The application provides users with steps to write correct questions as showing in the output of the application which is illustrated in Fig. 4. Fig. 4 shows Question Checker Q. C. explaining how to create appropriate questions that meet NCAAA & ABET criteria. This explanation can be edited by developers.

B. To Validate and Correct the Questions

Faculty members can seek evaluation of their questions using any method they find suitable. A set of questions from various topics and courses was input into the Question Checker App to assess its efficiency. For this paper, a subset of questions was selected from the Web Design & Development course, with each question targeting a specific Student Outcome (SO). The course has three basic ABET Student Outcomes: 1, 2, and 6. The application's responses were as follows:

Questions 1 through 8 were entered into the app to validate test questions targeting Student Outcomes SO: 1.1, SO: 1.2, SO: 2.1, SO: 2.2, SO: 2.3, SO: 6.2, SO: 6.1, and SO: 6.3, respectively. Fig. 5, 6, 7, 8, 9, 10, 11, and 12 show the instructor's questions and the responses of the Question Checker app. These questions were introduced to test the efficiency of the QC application in evaluating the eight student outcomes. The first question was: Q1. "Differentiate among the validators: Range validator and Regular Expression validator in terms of purpose, syntax, and an example."

The QC application validates questions, provides reasons for validation, and offers enhancement suggestions. The question and the app's response are illustrated in Fig. 5. As shown in Fig. 5, the ABET and NCAAA domains were used to evaluate the question. The QC app can either validate or invalidate a question and suggest updates for both cases. For example, consider question Q2: "Use ASP.NET to create a master page including title, dynamic navigation, and copyright." The QC application provided suggestions to improve the question without giving explicit validation responses. Fig. 6(a) shows the instructor's question and the QC application's response with suggestions. The app validated the question but suggested enhancements to clarify it. Fig. 6(b) displays the validation of the revised question.



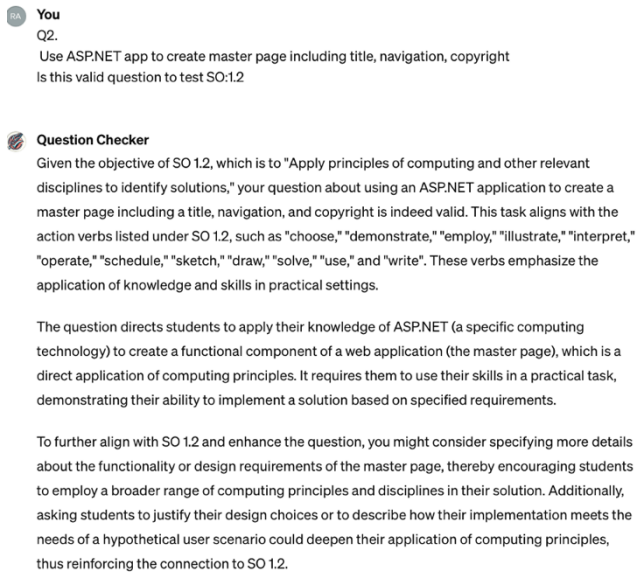
Question Checker

Crafting questions that align with the National Commission for Academic Accreditation and Assessment (NCAAA) and the Accreditation Board for Engineering and Technology (ABET) standards involves understanding the specified student outcomes and matching them with the appropriate level of cognitive demand as outlined in Bloom's Taxonomy. Here are the steps to follow:

1. **Understand Student Outcomes (SOs):** Begin by reviewing the specific student outcomes outlined by NCAAA and ABET for your program. These outcomes describe what students should be able to do by the end of the program and are categorized into different domains such as applying knowledge, designing systems, ethical reasoning, communication skills, teamwork, and understanding professional responsibility.
2. **Review Bloom's Taxonomy Levels:** Familiarize yourself with Bloom's Taxonomy levels—Remembering, Understanding, Applying, Analyzing, Evaluating, and Creating. Each level represents a different type of cognitive demand, from basic recall of information to complex analysis and creative problem-solving.
3. **Match SOs with Bloom's Levels:** Align each student outcome with the appropriate level(s) of Bloom's Taxonomy. This alignment ensures that the questions you craft will target the cognitive skills necessary to achieve the intended learning outcomes. For example, an outcome related to designing a system may align with the Creating level, while understanding professional responsibility might align with the Understanding level.
4. **Select Appropriate Verbs:** Use the mapped verbs from the provided tables to guide your question construction. Each Bloom's level is associated with specific action verbs that help in formulating questions. For instance, to assess Analyzing outcomes, you might use verbs like compare, contrast, or evaluate. For Applying outcomes, verbs like demonstrate, solve, or use could be appropriate.
5. **Craft Questions:** With the outcomes, Bloom's levels, and verbs in mind, start crafting your questions. Ensure that each question is clearly worded, unambiguous, and specifically designed to assess the intended student outcome. For example, to assess the ability to analyze a problem (SO 1.1), a question might start with "Evaluate the following scenario to identify potential computing solutions..."
6. **Review and Refine:** After drafting your questions, review them to ensure they align with both the student outcomes and the cognitive level intended. Make sure the questions are challenging yet fair, and that they truly measure the students' understanding and skills in relation to the program's learning outcomes.
7. **Feedback and Iteration:** If possible, seek feedback from colleagues on the clarity and alignment of your questions. Use this feedback to refine your questions further. This iterative process helps in enhancing the quality and effectiveness of your assessment items.
8. **Integration into Assessments:** Once refined, integrate these questions into your quizzes, exams, or other assessment tools. Ensure that your assessments as a whole cover a broad range of the program's student outcomes and adequately reflect the diverse cognitive skills outlined in Bloom's Taxonomy.

By following these steps, you'll be able to create well-aligned, clear, and effective questions that not only meet NCAAA and ABET standards but also foster deep learning and critical thinking among your students.

Fig. 4. Question checker guideline for creating a correct question.



You

Q2.
Use ASP.NET app to create master page including title, navigation, copyright
Is this valid question to test SO:1.2

Question Checker

Given the objective of SO 1.2, which is to "Apply principles of computing and other relevant disciplines to identify solutions," your question about using an ASP.NET application to create a master page including a title, navigation, and copyright is indeed valid. This task aligns with the action verbs listed under SO 1.2, such as "choose," "demonstrate," "employ," "illustrate," "interpret," "operate," "schedule," "sketch," "draw," "solve," "use," and "write". These verbs emphasize the application of knowledge and skills in practical settings.

The question directs students to apply their knowledge of ASP.NET (a specific computing technology) to create a functional component of a web application (the master page), which is a direct application of computing principles. It requires them to use their skills in a practical task, demonstrating their ability to implement a solution based on specified requirements.

To further align with SO 1.2 and enhance the question, you might consider specifying more details about the functionality or design requirements of the master page, thereby encouraging students to employ a broader range of computing principles and disciplines in their solution. Additionally, asking students to justify their design choices or to describe how their implementation meets the needs of a hypothetical user scenario could deepen their application of computing principles, thus reinforcing the connection to SO 1.2.

Fig. 5. The Validation of question 1 testing SO 1.1.

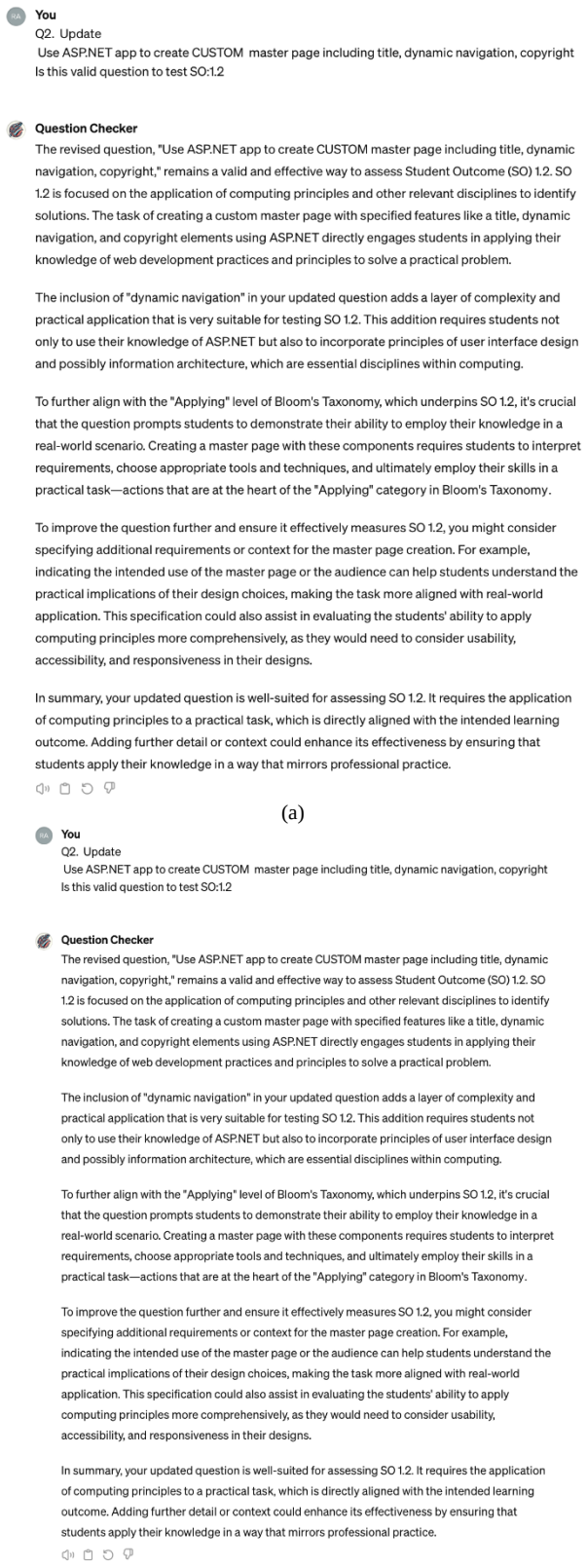


Fig. 6. (a) Suggestion of enhancement question 2 testing SO 1.2. (b) The Validation of revised question 2 to test SO: 1.2.

Another example is question Q4: "Write the correct syntax" for testing SO 2.2. The QC application provided suggestions to improve the question without explicit validation responses. Fig. 8(a) shows the instructor's question and the QC application's response with suggestions. The app validated the question but suggested enhancements for clarity. The instructor implemented the QC suggestions and updated the question. Fig. 8(b) shows the updated question and the app's response.

Question 3 explicitly asked about SO: 2.1. The QC application validated the question and offered suggestions for enhancement. Fig. 7 shows the question and the app's response. Q4 tests the efficiency of the application in validating SO: 2.2. The application suggested adding more information to better target SO: 2.2. Fig. 8(a) shows the question and the QC application's response, while Fig. 8(b) displays the validation of the revised question as per the application's suggestions.

Fig. 9 shows the validation of question Q5, which tests SO: 2.3. As illustrated in Fig. 11, the question targets SO: 6.1, and the QC application directly validated it. Question Q8 tests the target SO: 6.3, and the validation with the application is shown in Fig. 12. Question Q6 was identified as invalid and will be discussed later. A summary of the introduced questions and the application responses is provided in Table III.

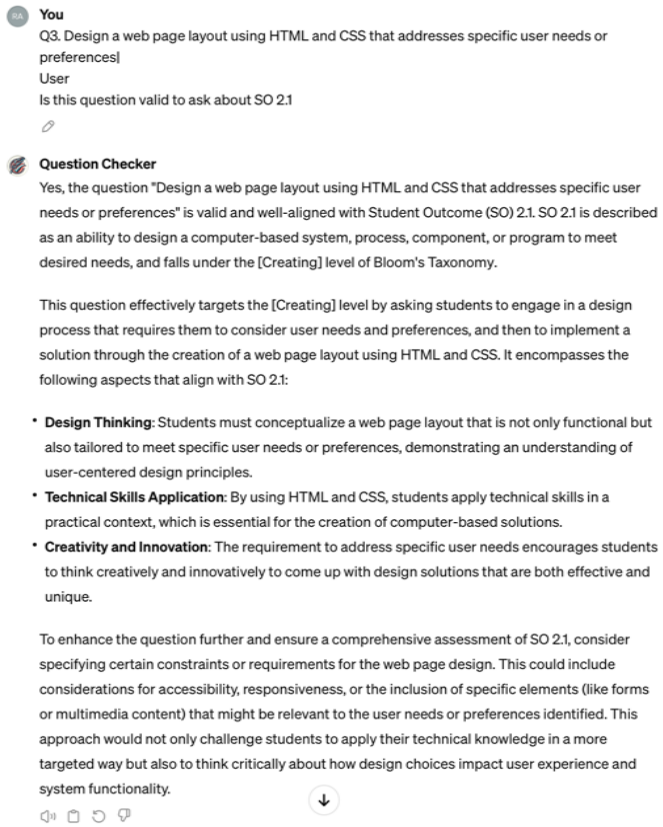
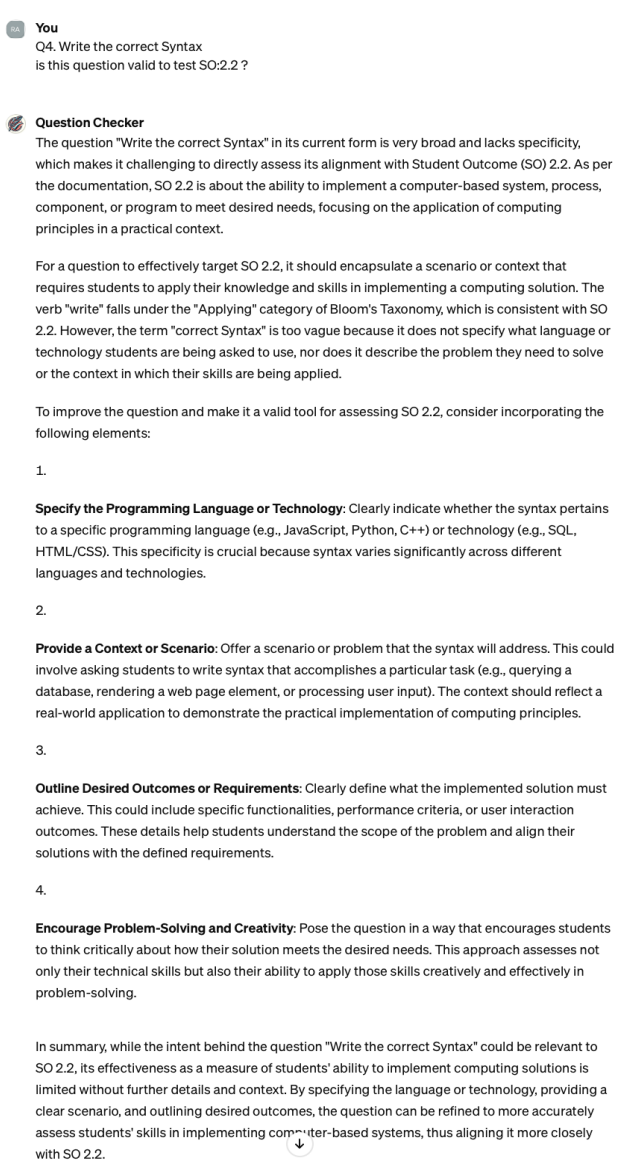


Fig. 7. Validation of question 3 to test SO 2.1.



You
Q4. Write the correct Syntax
is this question valid to test SO:2.2 ?

Question Checker
The question "Write the correct Syntax" in its current form is very broad and lacks specificity, which makes it challenging to directly assess its alignment with Student Outcome (SO) 2.2. As per the documentation, SO 2.2 is about the ability to implement a computer-based system, process, component, or program to meet desired needs, focusing on the application of computing principles in a practical context.

For a question to effectively target SO 2.2, it should encapsulate a scenario or context that requires students to apply their knowledge and skills in implementing a computing solution. The verb "write" falls under the "Applying" category of Bloom's Taxonomy, which is consistent with SO 2.2. However, the term "correct Syntax" is too vague because it does not specify what language or technology students are being asked to use, nor does it describe the problem they need to solve or the context in which their skills are being applied.

To improve the question and make it a valid tool for assessing SO 2.2, consider incorporating the following elements:

1.

Specify the Programming Language or Technology: Clearly indicate whether the syntax pertains to a specific programming language (e.g., JavaScript, Python, C++) or technology (e.g., SQL, HTML/CSS). This specificity is crucial because syntax varies significantly across different languages and technologies.

2.

Provide a Context or Scenario: Offer a scenario or problem that the syntax will address. This could involve asking students to write syntax that accomplishes a particular task (e.g., querying a database, rendering a web page element, or processing user input). The context should reflect a real-world application to demonstrate the practical implementation of computing principles.

3.

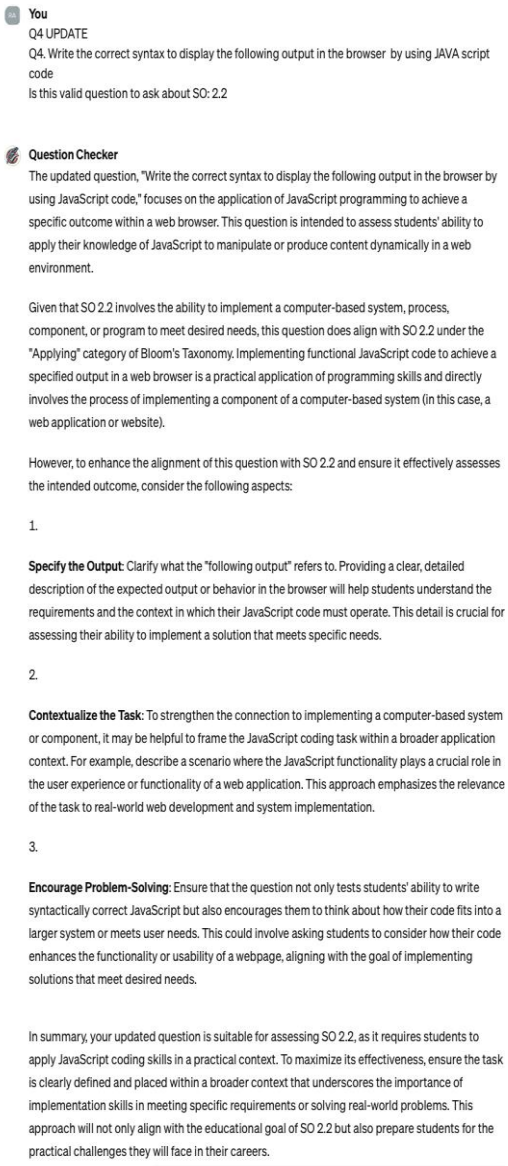
Outline Desired Outcomes or Requirements: Clearly define what the implemented solution must achieve. This could include specific functionalities, performance criteria, or user interaction outcomes. These details help students understand the scope of the problem and align their solutions with the defined requirements.

4.

Encourage Problem-Solving and Creativity: Pose the question in a way that encourages students to think critically about how their solution meets the desired needs. This approach assesses not only their technical skills but also their ability to apply those skills creatively and effectively in problem-solving.

In summary, while the intent behind the question "Write the correct Syntax" could be relevant to SO 2.2, its effectiveness as a measure of students' ability to implement computing solutions is limited without further details and context. By specifying the language or technology, providing a clear scenario, and outlining desired outcomes, the question can be refined to more accurately assess students' skills in implementing computer-based systems, thus aligning it more closely with SO 2.2.

(a)



You
Q4 UPDATE
Q4. Write the correct syntax to display the following output in the browser by using JAVA script code
Is this valid question to ask about SO: 2.2

Question Checker
The updated question, "Write the correct syntax to display the following output in the browser by using JavaScript code," focuses on the application of JavaScript programming to achieve a specific outcome within a web browser. This question is intended to assess students' ability to apply their knowledge of JavaScript to manipulate or produce content dynamically in a web environment.

Given that SO 2.2 involves the ability to implement a computer-based system, process, component, or program to meet desired needs, this question does align with SO 2.2 under the "Applying" category of Bloom's Taxonomy. Implementing functional JavaScript code to achieve a specified output in a web browser is a practical application of programming skills and directly involves the process of implementing a component of a computer-based system (in this case, a web application or website).

However, to enhance the alignment of this question with SO 2.2 and ensure it effectively assesses the intended outcome, consider the following aspects:

1.

Specify the Output: Clarify what the "following output" refers to. Providing a clear, detailed description of the expected output or behavior in the browser will help students understand the requirements and the context in which their JavaScript code must operate. This detail is crucial for assessing their ability to implement a solution that meets specific needs.

2.

Contextualize the Task: To strengthen the connection to implementing a computer-based system or component, it may be helpful to frame the JavaScript coding task within a broader application context. For example, describe a scenario where the JavaScript functionality plays a crucial role in the user experience or functionality of a web application. This approach emphasizes the relevance of the task to real-world web development and system implementation.

3.

Encourage Problem-Solving: Ensure that the question not only tests students' ability to write syntactically correct JavaScript but also encourages them to think about how their code fits into a larger system or meets user needs. This could involve asking students to consider how their code enhances the functionality or usability of a webpage, aligning with the goal of implementing solutions that meet desired needs.

In summary, your updated question is suitable for assessing SO 2.2, as it requires students to apply JavaScript coding skills in a practical context. To maximize its effectiveness, ensure the task is clearly defined and placed within a broader context that underscores the importance of implementation skills in meeting specific requirements or solving real-world problems. This approach will not only align with the educational goal of SO 2.2 but also prepare students for the practical challenges they will face in their careers.

(b)

Fig. 8. (a) Suggestion of enhancement for question 4 to test SO 2.2. (b) The Validation of Revised question 4 to test SO: 2.

TABLE III. SUMMARY OF THE QC APPLICATION RESPONSES TO VALIDATE THE INTRODUCED QUESTIONS

THE QUESTION #	THE TARGET SO	TYPE OF QUESTION TO TEST SO	QC VALIDATION	IMPROVEMENT SUGGESTION	THE UPDATED Q	THE VAIDATION OF REVISED QUESTION	FIGURE #
question 1	1.1	Explicit	YES	NA	NA	NA	5
question 2	1.2	Explicit	YES	YES	YES	Yes	6
question 3	2.1	Explicit	YES	NA	NA	NA	7-1
question 4	2.2	Explicit	YES	Yes	YES	Yes	8-1, 8-2 update
question 5	2.3	Explicit	Yes	Yes	NA	NA	9
question 6	6.2	Explicit	NO	YES	YES	YES	10-1, 10-2
question 7	6.1	Explicit	YES	YES	NA	YES	11
question 8	6.3	Explicit	YES	NA	NA	NA	12
question 9	2.2	Implicit	NA	NA	NA	NA	13

C. Examples of Question Checker Evaluation to Invalid Questions and Enhancement

Question Q6 was selected to test the application's efficiency in identifying invalid questions. Fig. 10(a) and (b) depict the evaluation of the Question Checker app for questions related to SO 6.2. The original question, Q6, asked for the four methods of a command class, with the instructor seeking validation that it targeted SO 6.2. However, the app's QC deemed the question invalid for testing SO 6.2, providing reasons and suggestions for improvement. Consequently, the question was revised based on the QC's suggestions, resulting in an updated question: Q6: "Explain how the Command class can be used to support information systems within specific environments."

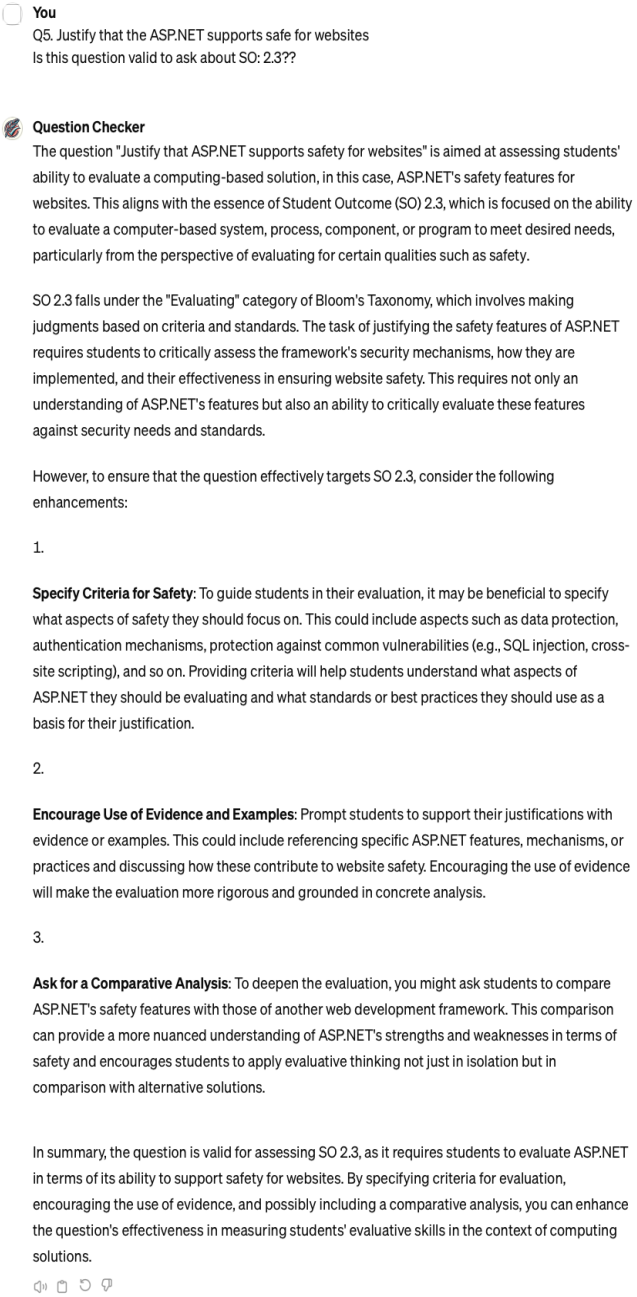
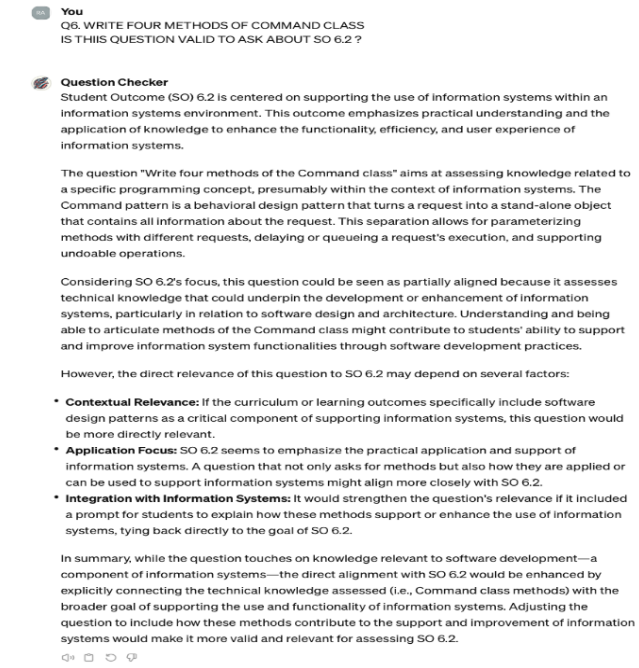
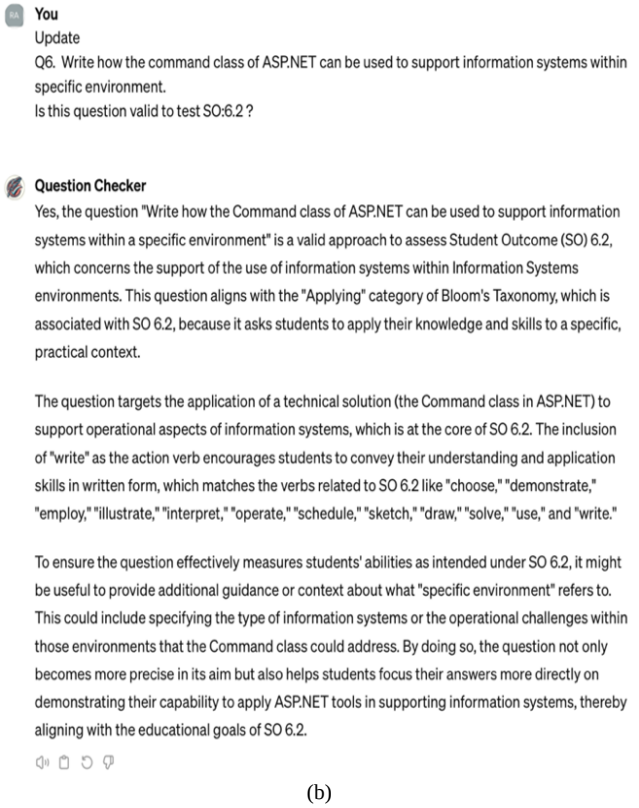


Fig. 9. The Validation of question 5 to test SO: 2.3

The QC suggested adjusting the question to include how these methods contribute to the support and improvement of information systems, which would make it more valid and relevant for assessing SO 6.2. Fig. 10(b) shows the validation of the revised question.



(a)



(b)

Fig. 10. (a) The Suggestion enhancement of Q6 to test SO 6.2. (b) 2-The Validation of Q6 to test SO 6.2 after considering the Question Checker application enhancement suggestion.

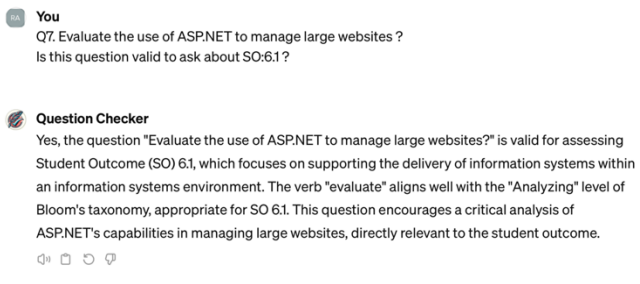


Fig. 11. The Validation of question 7 to test SO: 6.1

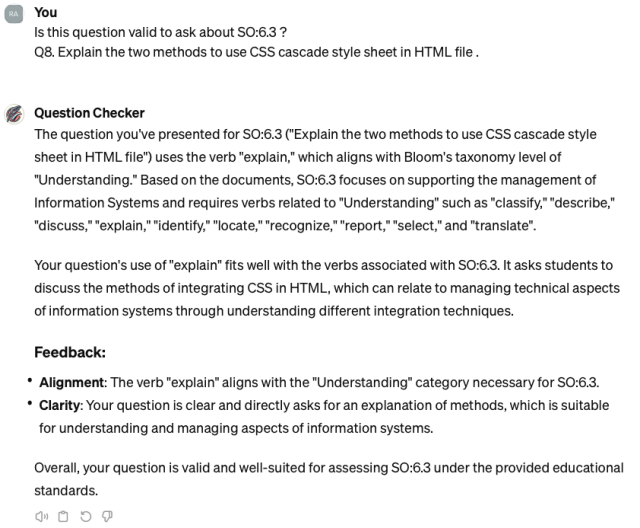


Fig. 12. The Validation of the explicit question Q8 to test SO 6.3.

D. To Suggest the Right SO

The instructor is free to ask the Question Checker application to evaluate a question using any method. For example, the instructor could request that the app suggest suitable student outcomes rather than inquire about a specific student outcome.

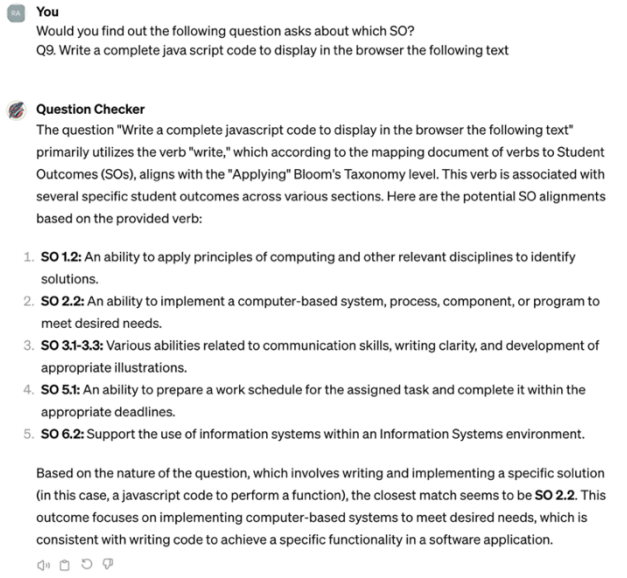


Fig. 13. The Validation of implicitly question 9 to ask about SO 2.2.

Question 9, "Write a complete JavaScript code to display the following text in the browser," is shown in Fig. 13 along with the app's response. As illustrated in Fig. 13, the app validated the question as targeting SO: 2.2 and provided the reason.

The efficiency of the QC application was demonstrated by asking a question that implied a student outcome (SO). The application was used to validate a question aimed at testing an SO implicitly—not by directly asking the app about a specific SO, but by having it identify the correct SO. Fig. 13 shows the question and the app's response with the correct SO. The summary of the tested SO targets is illustrated in Table III.

IX. QUESTIONNAIRE

A survey was conducted to assess academic teachers' acceptance of using artificial intelligence (AI) applications to guide instructors in creating high-quality questions that meet academic accreditation standards and to correct questions submitted by faculty members. The survey was administered in two stages: the first stage occurred before the application was created, and the second stage took place after the application was developed and its efficiency was tested. WhatsApp was utilized to distribute the survey to faculty members across various Saudi universities. In the first stage, the survey received 45 responses, all of which supported using AI tools to guide question creation aligned with accreditation standards. However, 11% of respondents were opposed to allowing the application to correct their questions. Fig. 14 and 15 illustrate the acceptance percentages from the first stage. In the second stage, the survey received 50 responses.

Usability and Effectiveness: Of the 43 responses, 7% did not accept the application's usability. It is possible that those who rejected it are not accustomed to using electronic applications (see Fig. 16 for the usability acceptance rate).

Assistance Rates: The acceptance rate for the application's assistance in correcting questions to meet academic accreditation requirements was 49% (see Fig. 17 for the acceptance rate). However, the acceptance rate for the application's assistance in enhancing assessment effectiveness based on academic accreditation was 33% (see Fig. 18 for the acceptance rate). It is possible that the 17% of respondents who rejected this feature prefer complete independence in question creation and do not want electronic intervention, except by accreditation committees.

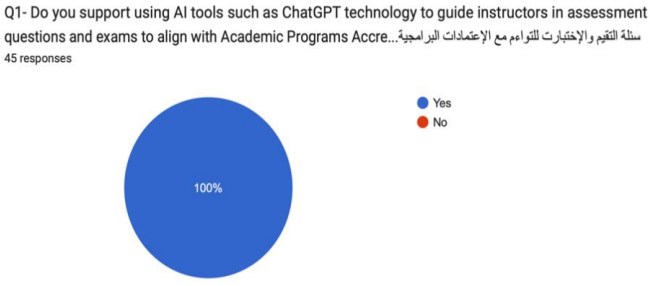


Fig. 14. The Acceptance of using artificial intelligence tools guide in creating question.

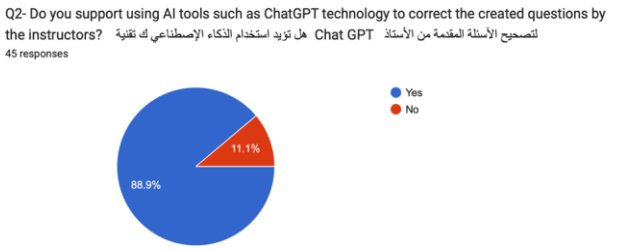


Fig. 15. The Acceptance of using artificial intelligence to correct the questions.

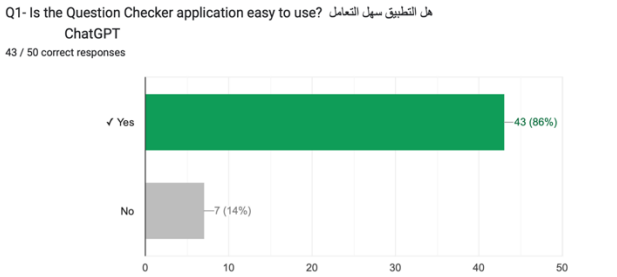


Fig. 16. The Usability agreement percentage of the Application Question Checker.

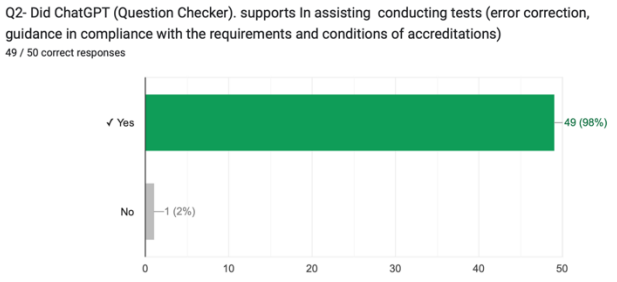


Fig. 17. The Agreement percentage of assisting in correction and guidance of the Application Question Checker QC.

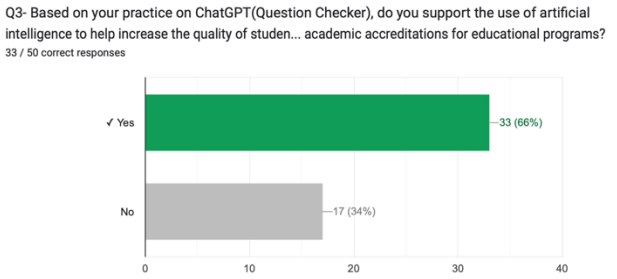


Fig. 18. The Percentage agreement of the application Question Checker QC. in enhancing quality of the as-assessment.

X. Discussion

The study most closely related to this research is referenced in [22]. It investigated the alignment between the two accreditations, NCAAA and ABET, and Bloom's Taxonomy, with the findings presented in Table I. The process of associating question verbs with the levels of Bloom's Taxonomy and ABET Student Outcomes (SO) was accomplished and is shown in Table II. The previous work [22] primarily focused on aligning educational content with the standards of both the NCAAA (National Commission for Academic Accreditation and Assessment) and ABET (Accreditation Board for Engineering and Technology).

In contrast, the proposed research advances this by introducing generative AI tools, specifically ChatGPT, to enhance the quality of test questions, ensuring they meet the rigorous requirements of both accreditations simultaneously. This study involves testing the efficacy of a custom application, named Question Checker, designed to validate and improve questions in alignment with these accreditation standards.

A key innovation of this research is the mapping of question verbs to Student Learning Outcomes (SLOs), which is critical for ensuring accurate measurement through appropriate questions. The Question Checker (QC) application was developed as a custom ChatGPT tool to verify the compatibility of questions with academic accreditation standards. The application's efficiency in validating questions based on ABET and NCAAA's SOs was rigorously tested. Furthermore, the acceptance of using this technology was assessed, with 100% of participants willing to use the technology for guidance, and 88.9% agreeing to allow the application to correct their questions.

The application successfully provided suggestions for any question aligned with a specific educational outcome. It demonstrated its effectiveness in confirming or rejecting questions submitted to it across all three basic educational outcomes of the applied subject, adhering to alignment conditions with both local and international accreditation standards. The application offered suggestions for all submitted questions, explaining the reasons for acceptance or rejection based on alignment with the quality standards of both local and international accreditation.

Additionally, the program was tested in two scenarios to verify its effectiveness:

- The teacher specifies the Student Outcome (SO), and the application either confirms or rejects it.
- The teacher presents the question without specifying the SO, and the application infers the appropriate SO.

In experiments, the program succeeded in both either accepting or rejecting questions based on a predefined SO and in identifying the appropriate SO for questions presented without a specified outcome.

XI. Conclusion and Future Work

This study introduces a mechanism for using ChatGPT to assist teachers in generating high-quality questions, thereby saving time and providing an effective means of assessing student learning outcomes. The ChatGPT application was successfully developed and thoroughly tested. This research presents a framework for utilizing generative AI applications to enhance educational assessment tools and promote assessment

equity. Specifically, the ChatGPT application for evaluating questions in IS courses was created and tested. The application's efficiency was demonstrated by its ability to assist in creating appropriate questions, provide steps for crafting questions, validate and correct questions, and suggest the correct Student Outcome (SO) for implicit questions. In future work, this study will be extended by testing the application for generating questions across different courses within the same field. Additionally, a comparison of results across various courses or open programs will be conducted to evaluate the application's effectiveness and adaptability in diverse educational contexts.

The application link is: https://chatgpt.com/g/g-7iUiGMgOD-question-checker.

ACKNOWLEDGMENT

The author would like to acknowledge King Abdulaziz University, the Faculty of Computing & Information Technology in Rabigh, the Deanship of Quality and Academic Accreditation, and all the anonymous faculty and staff who participated in this work.

REFERENCES

[1] "wikipedia.org,"[Online].Available: https://en.wikipedia.org/wiki/OpenAI.

[2] O. C, "ChatGPT," OPEN AI, January 2015–2024. [Online]. Available: https://openai.com/chatgpt. [Accessed 1 March 2023].

[3] V. Božić and I. Poola, Chat GPT and education, Preprint., 2023.

[4] "National Center for accademic accrediation and evaluation," Education and Training Evaluation Commission (ETEC), 2021. [Online]. Available: https://etec.gov.sa/en/About/Centers/Pages/Accreditation.aspx. [Accessed 1 November 2021].

[5] "ABET the Accreditation Board for Engineering and Technology," the Engineers' Council for Professional Development (ECPD), 2021. [Online]. Available: https://www.abet.org/about-abet/history/. [Accessed 18 AUGUST 2022].

[6] R. A. Aboalela, "An Assessment of Knowledge by Pedagogical Computation on Cognitive Level mapped Concept Graphs," Ohio Library and Information Network (OhioLINK) , 2017.

[7] R. A. Aboalela, "inferring of Cognitive Skill Zones in Concept Space of Knowledge Assessment," International Journal of Advanced Computer Science and Applications, vol. 9, no. 1, p. DOI: 10.14569/IJACSA.2018.090102, January 2018.

[8] R. Aboalela and J. Khan, "Are We Asking the Right Questions to Grade Our Students In a Knowledge-State Space Analysis?," IEEE Eighth International Conference on Technology for Education (T4E), vol. DOI: 10.1109/T4E.2016.037, 2016.

[9] S. Tingiris and B. Kinsella, Exploring GPT-3., Packt Publishing., 2021.

[10] B. Williamson, F. Macgilchrist and J. Potter, "Re-examining AI, automation and datafication in education.," Learning, media and technology, vol. 48, no. 1, pp. 1-5, 2023.

[11] I. l. u. b. g. pre-training., "Radford, A.; Narasimhan, K., Salimans, T.; Sutskever, I.," pp. Accessible online, IRL: https://cdn.openai.com/research-covers/language-unsupervised/language_understanding_paper.pdf, 2018.

[12] S. Biswas, "Role of Chat GPT in Education.," ENT surgery research, vol. 1, no. 1, pp. 1-3, 2023.

[13] M. D. &. S. J. Xames, "ChatGPT for research and publication: Opportunities and challenges.," Journal of Applied Learning and Teaching, vol. 6, no. 1, pp. 390- 395, 2023.

[14] M. M. Patil, R. PM, A. Solanki, A. Nayya and B. Qureshi, "Performing Big data analytics using swarm-based Long short-term memory neural network for temperature forecasting Computers," Materials & Continua , vol. 71, no. 2, pp. 2347-2361, https://doi.org/10.32604/cmc.2022.021447, 2022.

[15] C. Li and W. Xing, "Natural language generation using deep learning to support MOOC learners.," International Journal of Artificial Intelligence in Education. , vol. 31, no. 1, pp. 186-214. https://doi.org/10.1007/s40593-020-00235-x, 2021.

[16] D. R. Cotton, P. A. Cotton and J. R. Shipway, " Chatting and cheating: Ensuring academic integrity in the era of ChatGPT.," Innovations in Education and Teaching International, vol. https://doi.org/10.1080/14703297.2023.2190148, pp. 1-12, 2023.

[17] M. Sallam, "ChatGPT utility in healthcare education, research, and practice: systematic review on the promising perspectives and valid concerns," In Healthcare MDPI. https://doi.org/10.3390/healthcare11060887, vol. 6, p. 887, 2023.

[18] E. A. Van Dis, J. Bollen, W. Zuidema, R. van Rooij and C. L. Bockting, "ChatGPT: five priorities for research.," Nature, (7947), https://doi.org/10.1038/d41586-023-00288-7, vol. 614, pp. 224-226., 2023.

[19] J. Crawford, M. Cowling and K. Allen, " Leadership is needed for ethical ChatGPT: Character, assessment, and learning using artificial intelligence (AI).," Journal of University Teaching & Learning Practice https://doi.org/10.53761/1.20.3.02, vol. 20, no. 3, 2023.

[20] B. Jin and M. Kruppa, "Cheating with ChatGPT: Can an AI chatbot pass AP Lit?," The Wall Street Journal, pp. https://www.wsj.com/articles/chatgpt-creatoropenai-is-in-talks-for-tender-offer-that-would-value-it-at29-billion-11672949279, 2023, January 5.

[21] M. D. G. B. Marietto, R. V. de Aguiar, G. D. O. Barbosa, W. T. Botelho, E. Pimentel, R. D. S. França and d. S. V. L., "Artificial intelligence markup language: a brief tutorial," arXiv preprint, https://arxiv.org/abs/1307.3091, pp. 1-19, 2013.

[22] R. Aboalela, "ChatGPT for generating questions and assessments based on accreditations," in 13th International Conference on Advances in Computing and Information Technology (ACITY 2023), London, November 25 ~ 26, 2023. This presentation was published in arXiv

[23] T. L. Weng, Y. M. Wang, S. Chang, T. J. Chen and S. J. Hwang, "ChatGPT failed Taiwan's family medicine board exam," Journal of the Chinese Medical Association, vol. 86, no. 8, pp. 762- 766 , 2023.

[24] I. Cribben and Y. Zeinali, "The benefits and limitations of ChatGPT in business education and research: A focus on management science, operations management and data analytics.," Operations Management and Data Analytics, 2023.

[25] S. Neendoor, "hurixdigital," ChatGPT: Pros and Cons of Using ChatGPT in Higher Education, 20 September 2023. [Online]. Available: https://www.hurix.com/chat-gpt-pros-and-cons-of-using-chatgpt-in-higher-education/. [Accessed 12 March 2024].

[26] E. B. School, "Eu Business School," The Challenges of Chat GPT in Higher Education, 21 July 2023. [Online]. Available: https://www.euruni.edu/blog/the-challenges-of-chat-gpt-in-higher-education/. [Accessed 12 March 2024].

[27] R. Aboalela, "The allignment between untenational and national academic accrediations -An application in information systems bachelor program at Kingdom of Saudy Arabia. International Journal of Computer Science & Information Technology (IJCSIT), vol. 15, no. 6, pp. 27-51, 2023.

[28] K. A. University, "Quality Assurance and Accreditation," 22 August 8/22/2021 11:24:08 AM. [Online]. Available: https://drive.google.com/file/d/1ve4FxAqEmOBRKdshUa4QsyD7slXBIdpG/view. [Accessed July 2022].

[29] A. Rabaa'i, A. Rababaah and S. Al-Maati, "Comprehensive guidelines for ABET accreditation of a computer science program: The case of the American University of Kuwait. Int. J. Teach. Case Stud. 2017, 8, 151–191.".

[30] h. Taleb, A. Namoun and M. Benaida, "A Holistic Ouality Assurance Framework to Acquire National and International EducationalAccreditation: The Case of Saudi Arabia," Journal ofEngineering and Applied Sciences, vol. 14, no. 18, pp. 6685-6698, ISSN: 1816-949X © Medwell Journals, 2019, 2019.

[31] A. M. A. 2. ,. Y. A. B. Saqib Saeed 1, D. A. Alabaad, H. Gull, M. Saqib, S. Z. Iqba and A. A. Salam, "Sustainable Program Assessment Practices: A Review of the ABET and NCAAA Computer Information Systems

Accreditation Process," International Journal of Environmental Research and Public Health (MDPI), vol. 18, no. 12691, p. https:// doi.org/10.3390/ijerph182312691, 2021.

[32] W. A. e. a. Lorin, A taxonomy for learning, teaching, and assessing.: A revision of Bloom's taxonomy of educational objectives, New York, 2001.

[33] R. Aboalela and J. khan, "Inferring of cognitive skill zones in concept space of knowledge assessment," International Journal of Advanced Computer Science and Applications, vol. 9, no. 1, pp. 11-17 https://doi.org/10.14569/IJACSA.2018.090102, 2018.

[34] R. Aboalela and J. Khan, "Model of Learning Assessment to Measure Student Learning: Inferring of Concept State of Cognitive Skill Level in Concept Space," 3rd International Conference on Soft Computing & Machine Intelligence (ISCMI), vol. IEEE Xplore: 05 October , pp. 189-195, doi: 10.1109/ISCMI.2016.26, 2017.

[35] K. A. University, " Tha faculty of Computing and Information Technology in Rabigh Department of Information System," All Rights reserved King Abdulaziz University 2022, August 2023. [Online]. Available: https://fcitr-is.kau.edu.sa/Default-830002-ar. [Accessed 10 August 2023].

[36] M. o. E. –. K. o. S. Arabia, "Ministry Of Education of Saudi Arabia," ©Copyrights, Ministry of Education – Kingdom of Saudi Arabia , 30 January 2015. [Online]. Available: https://moe.gov.sa/ar/pages/default.aspx. [Accessed 16 07 2023].

[37] ABET, "ABET org.," ABET org., 2021. [Online]. Available: https://www.abet.org/accreditation/. [Accessed 1 November 2021].

[38] T. N. C. f. A. Accreditation, Education and Training Evaluation Commission (ETEC) , 2021. [Online]. Available: https://etec.gov.sa/en/About/Centers/Pages/Accreditation.aspx. [Accessed 1 November 2021].

[39] S. Saeed, A. M. Almuhaideb, Y. A. Bamarouf, D. A. Alabaad, H. Gull and M. Saqib, "Sustainable Program Assessment Practices: A Review of the ABET and NCAAA Computer Information Systems Accreditation Process".

[40] W. Fastiggi, "Technology of learner," Technology of learner LtD, 2022. [Online]. Available: https://technologyforlearners.com/applying-blooms-taxonomy-to-the-classroom/. [Accessed 1 February 2022].

Enhancing Digital Financial Security with LSTM and Blockchain Technology

Thanyah Aldaham[1], Hedi HAMDI[2]
Department of Computer Science, Jouf University. Sakkaka, Saudi Arabia[1, 2]
Manouba University, Manouba, Tunisia[2]

***Abstract*—The growing dependence on digital financial and banking transactions has brought about a significant focus on implementing strong security protocols. Blockchain technology has proved itself throughout the years to be a reliable solution upon which transactions can safely take place. This study explores the use of blockchain technology, specifically Ethereum Classic (ETC), to enhance the security of digital financial and banking transactions. The aim is to develop a system using an LSTM model to predict and detect anomalies in transaction data. The proposed LSTM model was trained before being tested and the results prove that the proposed model can effectively enhance the security, especially when compared to other studies in the same domain. The proposed model achieved a prediction accuracy of 99.5%, demonstrating its effectiveness in enhancing security by preventing overfitting and identifying potential threats in network activities. The results suggest significant improvements in digital transaction security, enhancing both the traceability and transparency of blockchain transactions while reducing fraud rates. Future work will extend this model's applicability to larger-scale decentralized finance systems.**

***Keywords*—Digital financing; block chain; ETC; security; anomaly detection; machine learning; LSTM**

I. Introduction

Financial transactions are essential to both the national and global economy. Each day, trillions of dollars are traded in the global financial networks that serve a vast number of individuals. Nakamoto [1] claims that although the financial system still uses an underlying trust-based methodology, it works well enough for the majority of transactions. He said that because financial organizations must arbitrate disputes, irrevocable transactions are not feasible. The banking sector is heavily regulated and conservative, and the revenue model hasn't changed in many years. Financial institutions already deal with a number of problems that impair their effectiveness and performance, including high transaction costs, high fraud rates, centralized control that might be challenged by pirates, and a lack of traceability and transparency [2]. Blockchain technology may boost a company's level of trust and control. Performance is impacted when banking institutions attempt to adapt to new client registration and money transfer procedures. Recently, crypto currencies have attracted the attention of both industry and academia. According to Coinmarketcap [3], the capital market for Bitcoin which is the original cryptocurrency, is expected to reach $880 billion. Thus, the influence of blockchain technology acceptance on financial transactions and implementation concerns in the banking sector are determined by this research.

ETC which is known today as legitimate and famous decentralized platform for cryptocurrency other than ETH. In particular, as balanced books are maintained and computers do everything very quickly, the system create such a secure digital stamp, just from the blockchain [4].

Digital technology in the form that has already existed in the pre-digital era and has been introduced and popularized can change and acquire new functions such that it can involve completely new tools and services [5]. Cybersecurity is already integrated in many aspects of digital forensics, which poses as a necessary cornerstone to achieving desirable level of security in Internet of Things [6]. Nevertheless, banking industry represents the bunker of all different kinds of money and private conversations and the secrete storage for people's monetary resources. Competitive factors such as efficiency, performance enhancement, and deposit security have significantly propelled the financial sector forward. However, there is a risk that grows in proportion to the increasing number of users for whom the system is designed or as the system becomes more sophisticated. This situation is not a fair play because people started exploring the system's flaws [7].

The identification of errors in blockchain networks is an enormous task because you may find similar issues if you are looking for attacks or fraudulent activities. Anomaly detection plays out as a key element in blockchain security, allowing for deviations to encrypted content or other unexpected events through the monitoring of blockchain data. A quick recognition and response to the anomaly help minimize the possible damage by attackers and safeguard the whole web [8].

Although Blockchain technology has significantly enhanced security in financial transactions, several gaps remain, particularly in detecting sophisticated anomalies such as 51% attacks. Existing machine learning methods have not fully addressed these gaps, often struggling with overfitting and real-time detection issues. This research seeks to bridge this gap by leveraging an Encoder-Decoder LSTM model within the Ethereum Classic blockchain ecosystem to improve anomaly detection and enhance transaction security.

In this paper, the following contributions can be considered:

- Use of Encoder-Decoder LSTM approach for Ethereum Classic Blockchain (ETC) attack detection enhances blockchain security.

- LSTM model's ability to identify sequential dependencies improves accuracy in anomaly detection.
- Encoder-Decoder LSTM model excels in learning from serialized data.
- Application of recurrent neural networks, particularly LSTM, enhances current blockchain security.
- Improved anomaly detection aids in early detection of threats.
- Enhancements contribute to the reliability of blockchain systems.

II. Literature Review

Blockchain technology has been utilized by the majority of today's companies in order to improve the safety of their data. It is one of the newest technologies that is gaining the greatest traction in the field of protecting the digital world. This section explores a variety of approaches, surveys, strategies, and procedures that have been used in blockchain to address concerns around data sharing and security.

Javaid et al. [9], explored the potential applications of blockchain technology for financial service providers seeking to improve risk management, authenticity, and security. In order to create smart contracts, improve efficiency and transparency, and open up new revenue streams, a lot of organizations are aggressively integrating blockchain into trade and finance systems. The adoption of blockchain-enabled IDs is growing widespread in the banking industry, as the unique recordkeeping capabilities of blockchain render traditional clearing and settlement procedures obsolete. In addition to stressing the transfer of asset ownership and the need of keeping accurate financial ledgers, the study highlights the significance of enterprises anticipating upcoming trends in financial blockchain applications. The measurement, communication, and analysis of financial data are the main areas of concentration for accounting experts. The paper focuses on the importance of blockchain technology for financial services by methodically locating and analyzing pertinent papers. It also explores a range of tools, tactics, and featured services. At the end, major applications of blockchain technology in financial services are identified and evaluated, demonstrating the technology's superior security in credit reporting and its potential to open up new markets, cut costs for issuers, and reduce counterparty risk by customizing digital financial instruments. Blockchain provides a single trustworthy source of truth for network users, making it simpler for members of the business network to collaborate, handle data, and reach consensus by utilizing mutualized standards, protocols, and shared procedures.

Trivedi et al. [10], focused in their study on how blockchain technology is used in the financial and e-finance industries. Research questions about the technology's development, acceptance obstacles, and useful applications are examined. After conducting a thorough analysis of 76 scholarly articles, the study narrowed its attention to 59 articles and created a three-dimensional classification framework that encompasses blockchain development, obstacles, and financial sector applications. The results point to untapped blockchain potential in the finance industry and point to areas in need of technological advancement. The report highlights that the technology is now unregulated, suggesting that it is still in its early stages and that there is ample opportunity for further growth and research in this area.

Hartmann and Hasan [11] drew attention to the abundance of Decentralized Finance (DeFi) Peer-to-Peer (P2P) lending platforms that either demand collateral from users or use conventional credit scoring techniques based on variables like credit history. Some users may find these requirements burdensome, nevertheless. The authors suggest using social media, which has a wealth of publicly accessible personal data and is used by over 55% of the world's population, as an alternative risk mitigator for lending. A user's professional behavior and dependability can be inferred by examining their social media accounts, which results in the creation of a "social score". The study's major contribution is the creation of a fully decentralized lending network that is enabled by the Ethereum blockchain and depends on this social score. With the help of this cutting-edge platform, consumers can obtain a loan even in the event that they don't have enough credit or collateral. The study also explores privacy issues, offering an improved platform that is intended to safeguard the borrower's privacy.

Liao et al. [12] focused on the open banking (OB) adoption trend that financial institutions are currently experiencing for service innovation and integration. Third-party service providers (TSPs) can now access user financial data in an effort to improve user experiences and find the best offers. However, the OB ecosystem's success depends on public confidence in third parties, which raises questions regarding data sharing, privacy protection, and the integration of digital identities. Although there are already decentralized applications (DApps) that address these issues, their integration into a workable three-phase OB method is still lacking, especially in areas like Taiwan. The study presents a blockchain-based identity management and access control (BIMAC) framework and lists the main needs of OB participants. The BIMAC framework creates a trustworthy platform for personal information transaction security control (PITSC) by utilizing smart contracts and a stateless authentication method. This platform provides features like online bank account opening, decentralized third-party login (TPL), integrated payouts, data authorization/revocation, and TSP access monitoring. The evaluation's findings show that the suggested framework's frequently executed functions have less computational overhead than the typical Ethereum transaction cost.

Boughaci et al. [13] introduced, blockchain technology and its fundamental ideas opens the discussion. The paper explores machine learning as a sophisticated instrument for examining large datasets and spotting potentially harmful transactions in untrusted networks. For the purpose of making wise decisions in the fields of banking and finance, the synergy of these clever strategies is highlighted. The suggested method is applied to the Bitcoin system, using the Elliptic dataset available on Kaggle as a standard. Because the dataset is not fully labeled, unlabeled data is divided into two primary clusters using the kmeans technique, and labeled data is

allocated to the appropriate cluster. Four machine learning approaches are then used for a thorough classification of the data. The results show promise, especially when k-means and the random forest classifiers are combined, indicating the potential effectiveness of this integrated approach in boosting security precautions.

Song and Chen [14], conducted research on the security of digital financial transactions using blockchain technology. To begin, the security of sdte is examined, as well as the DoS attacks that each role may launch, the assaults that a single role may send, and the attacks that numerous roles may launch in cooperation. It demonstrates that sdte can withstand various assaults and has robust security. Then, the system test's environment is detailed. Then, performance testing and analysis are performed on the key security transmission, smart contract execution in the trusted environment SGX, and overall running time. The testing findings demonstrate that employing the k-nearest neighbor (KNN) method to process data takes less than 0.45 seconds. At the same time, the system's additional cost is acceptable.

With the advent of post-quantum cryptography, it has become increasingly important to stay informed about the latest developments in cryptographic techniques and systems. Post-quantum cryptography is a branch of cryptography that aims to develop cryptographic systems that are secure against quantum computers. Quantum computers have the potential to break many of the classical cryptographic systems currently in use, such as RSA and ECC, by solving the underlying mathematical problems (like integer factorization and discrete logarithms) much more efficiently. As a result, researchers and organizations are actively working on developing cryptographic algorithms that can withstand attacks from quantum computers. Several relevant papers discussing advances in post-quantum cryptography and related topics, such as low-cost S-box implementations for AES [15], a survey on quantum-resistant algorithms and their applications, and strategies for optimizing cryptographic systems to resist quantum attacks [16]. Additionally, studies on lightweight cryptographic techniques for resource-constrained environments and their relevance in the post-quantum era provide further insights into the field [17]. Jalali [18] offer insights into the development of efficient and secure post-quantum cryptographic algorithms. This work, for example, presents a constant-time software library for the CSIDH (Commutative Supersingular Isogeny Diffie-Hellman) protocol, optimized for 64-bit ARM processors, and discusses its potential in the quantum era, particularly regarding its resistance to timing attacks. Another paper by Koziel et al. [19] focuses on the optimization of cryptographic algorithms based on Binary Edwards Curves (BEC), which are designed to be both efficient and secure, particularly for resource-constrained environments such as embedded systems and IoT devices.

Side-channel attacks (SCA) pose a significant threat to the security of cryptographic implementations, particularly in lightweight cryptography designed for resource-constrained environments such as IoT devices. Lightweight cryptographic algorithms such as PRINCE and GIFT-128 offer efficiency in power and memory usage, making them suitable for applications with strict resource limitations. However, their compact designs often expose vulnerabilities to side-channel attacks. For instance, in the work by Xue et al. [20], an SCA was demonstrated against the PRINCE cipher, which utilizes an unrolled architecture optimized for low latency but requires careful handling to prevent leakages during encryption rounds. Similarly, Benjamin et al. [21] explored deep learning-based side-channel attacks on GIFT-128, revealing the effectiveness of neural networks like CNNs in recovering cryptographic keys, even in scenarios involving desynchronized traces. Furthermore, a comprehensive survey by Chao Su and Qingkai Zeng [22] provides an analysis of CPU cache-based side-channel attacks, discussing security models and mitigation strategies, emphasizing the need for resilient designs in modern cryptography. These studies highlight the pressing need for enhanced countermeasures to safeguard lightweight cryptographic algorithms from the growing threat of side-channel attacks.

Table I shows the comparison of the related work mentioned earlier.

TABLE I. COMPARISON OF RELATED WORK

Work	Method	Technologies	Advantages	Limitations
[9]	Integration, smart contracts, revenue	Blockchain technology	Improved risk management, authenticity, security	Privacy concerns, data verification challenges
[10]	Study, examination of development, acceptance, applications	Blockchain technology	Untapped potential, areas for advancement	Lack of regulations, need for further research
[11]	Risk mitigation, social score creation	Ethereum blockchain	Decentralized lending, accessibility for users without credit or collateral	Privacy issues, reliance on social media
[12]	Blockchain-based identity management, access control	Blockchain technology	Improved user experiences, data sharing, privacy protection	Lack of integration, regulatory challenges
[13]	Analysis of large datasets, identification of harmful transactions	Blockchain technology, machine learning	Enhanced decision-making in banking and finance	Limited labeled data, computational overhead
[14]	Analyzing the security of digital financial transactions using blockchain technology, using KNN algorithm to process data for digital financial transaction security.	Blockchain technology, machine learning.	Offering strong security against various attacks and acceptable performance costs.	Lack of extensive studies and established frameworks to build upon.

IV. BACKGROUND

A. Blockchain in Financial and Banking Transactions

Blockchain is a distributed ledger technology that makes it possible for all parties to check and agree on a transaction before it is added to the value chain [23]. The banking industry has tried out new ways to use technology to improve customer flexibility, the speed of transactions and efficiency. Blockchain technology, as part of Industry 4.0, has the potential to transform business operations across a wide range of industries. Blockchain has been widely adopted and used in the banking and finance industries. Many financial institutions are often run by trusted third parties who are in charge of their operations. In the last step of a digital payment, a bank, credit or debit card, or other service provider acts as a trusted central expert and charges a fee to complete the transaction. For this operation to work, it needs an infrastructure that is both expensive and inefficient. The largest financial institutions in the world are now using this technology (see Fig. 1).

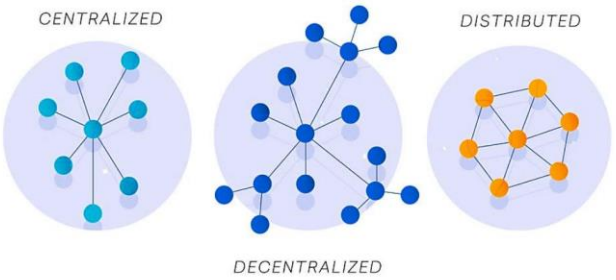


Fig. 1. Decentralization [24].

A blockchain can accommodate any new digital asset across multiple nodes. If a node fails, the data remains accessible and can be delivered by the other nodes. Since the blockchain is a public ledger, any sensitive personal information stored on it must be encrypted and can only be viewed by two parties. Data on the blockchain is encrypted using a public key and decrypted using a private key. Due to its consensus mechanism, the blockchain is immutable and cannot be duplicated. A block is added to the chain if there is consensus that the transactions within it are valid [25]. Despite this, blockchain is not yet widely adopted in the investment sector. However, industries are expected to quickly move towards implementing blockchain-integrated infrastructure in business organizations [26].

The primary advantages of blockchain in the banking sector include improved efficiency, enhanced security, immutable records, faster transaction times, and the elimination of third-party involvement, which reduces costs. One of the key benefits of blockchain is its history of unchangeable transactions—once a transaction is made, it cannot be undone, thereby reducing threats to financial institutions. Blockchain utilizes smart contracts, which are sets of rules agreed upon by the contracting parties. These contracts allow digital information to be stored, accessed, or altered only under specific conditions. Blockchain accelerates transaction processing and, due to its decentralized nature, reduces the need for financial intermediaries. This makes currency conversion cheaper and easier compared to traditional banking methods, while also protecting against scams, money laundering, and trust issues. Financial institutions are expected to adopt blockchain technology very soon, and the banking industry is planning for rapid growth in its use.

B. Ethereum Classic (ETC) Blockchain

Ethereum Classic functions as both a smart contract platform and a cryptocurrency. It's important to note that Ethereum Classic (ETC) should not be mistaken for Ethereum (ETH), despite their shared origins prior to a contentious disagreement that resulted in a split. Below, we delve into the factors that precipitated this divergence.

Ethereum Classic closely resembles Ethereum due to their shared origin. Both are blockchains that facilitate the development of other applications on top of them. These decentralized applications, often referred to as dapps, utilize smart contracts, enabling individuals to exchange money, property, or any other valuable assets without the need for intermediaries. ETC serves as the network's native currency. Additionally, the Ethereum Classic network allows dApps on its platform to create their own tokens, including NFTs [27].

In summary, Ethereum Classic is a decentralized public ledger based on proof-of-work, featuring an embedded Turing-complete programming language that enables the creation of smart contracts and decentralized applications [28][29].

The underlying principles of the Ethereum Classic blockchain closely resemble those of Bitcoin, which stands as the most renowned and prosperous cryptocurrency presently [30]. The consistency of a public ledger in a proof-of-work system is maintained through decentralized mining. Miners continually attempt to solve a complex computational puzzle to find a hash value lower than a specified target. Upon success, miners can generate a block and receive a reward from it.

Fig. 2 represents an example of the general scheme of a blockchain system.

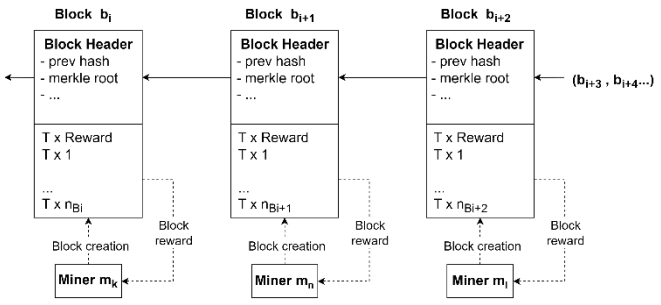


Fig. 2. General scheme of a blockchain system [31].

We won't delve deeply into the technical intricacies of the Ethereum platform here, but notable differences from Bitcoin include the use of accounts instead of UTXO, enhanced internal structures, and Turing-complete scripting languages. Those keen on exploring further can find comprehensive details in the original sources [28]. Instead, we'll focus on a few aspects relevant to the proposed treasury system.

Firstly, it's important to note that the average block time is approximately 14 seconds. This translates to approximately $B_{month} = 1851428$ blocks generated every 30 days.

Top of Form

Another significant distinction lies in the block reward system. Each block incorporates a special reward payment for the miner who mined it. Presently, in Ethereum Classic, this reward amounts to 5 newly created coins per block (uncle blocks excluded). This translates to approximately 9,257,140 coins generated per month $R_{month} = B_{month} \cdot 5 = 9257140$. Miners receive the entirety of these rewards, constituting the sole source of new coins within the system.

In summary, while Ethereum introduces advanced features such as a Turing-complete programming language, enhanced Merkle trees, and a modified GHOST protocol, its foundational principles remain akin to those employed in Bitcoin and other proof-of-work altcoins.

C. Ethereum Classic Security Challenges

In the early days, Ethereum stood alone. A collective known as The DAO (decentralized autonomous organization) utilized Ethereum to establish what essentially functioned as a venture capital fund. Ordinary individuals could invest using ETH, participate in decisions regarding asset allocation, and ideally, reap profits. The venture amassed over $100 million through token sales. However, a vulnerability in the fund's code was exploited, resulting in millions of dollars' worth of ETH being siphoned out and causing panic among investors. Developers had a 28-day window to devise a solution before the hackers could convert the tokens, representing a substantial portion of Ethereum's market capitalization at that time. The prevailing solution involved implementing a hard fork to nullify the hack and reimburse affected individuals. Although endorsed by Buterin and other prominent figures, this moves triggered backlash from purists advocating for the blockchain principle of non-interference with the ledger—arguing that the blockchain should persist with the theft intact. Those advocating for maintaining the status quo remained on the original platform, renaming it Ethereum Classic. Meanwhile, the majority of miners, developers, and users migrated to the forked network, which retained the Ethereum name [27].

Similar to Ethereum, the Ethereum Classic blockchain operates on a "proof of work" mining mechanism, where individuals worldwide utilize hardware and software to validate transactions and maintain network security, earning ETC as a reward. Users can send ETC to each other, akin to Bitcoin or Ethereum transactions with BTC or ETH, respectively. Furthermore, ETC can be used to engage with applications on the Ethereum Classic network, including decentralized exchanges for token swapping. However, it's worth noting that the Ethereum Classic ecosystem isn't as vibrant as Ethereum or other smart contract networks like Solana. As of February 2022, Ethereum Classic exhibited minimal activity in decentralized finance applications, as reported by DeFi Llama. This lower usage rate has raised concerns. Blockchain security hinges on having a diverse group of users actively operating the network; insufficient participation can leave the blockchain susceptible to vulnerabilities. Between 2019 and 2020, the Ethereum Classic network faced several "51% attacks," allowing a hacker to seize control of the majority of the network's computational power. This enabled them to manipulate the ledger and acquire more ETC. Despite these challenges, Ethereum Classic enthusiasts persist in network maintenance and code updates. In December 2020, core developers enhanced the network to render 51% attacks economically unfeasible. The latest upgrade, the Mystique hard fork, occurred in 2022 [27].

D. Machine Learning in Anomaly Detection of Blockchain Transactions

The merging of both technologies: Machine Learning and Blockchain Technology, has the potential to provide outcomes that are strong and of practical value. This chapter provides an overview of distributed ledger technology and investigates the ways in which machine learning skills may be incorporated into a system that is based on distributed ledgers. In addition to that, it highlighted a number of well-known uses and instances of how this connected method might be used [32].

The capacities for learning that machine learning algorithms possess are very remarkable. These features can be implemented in the blockchain, which will result in the chain becoming wiser than it was in the past. This collaboration could be useful in helping to enhance the safety of the blockchain's shared ledger in some way. Also, the processing power of ML can be used to take advantage of the shorter time it takes to find the best nonce, and ML can also be used to improve how data is exchanged. Additionally, it is able to construct a great many improved models of machine learning by utilizing the decentralized design characteristics that distributed ledger technology offers [33].

The selfish mining assault, often referred to as a transaction holdback attack, is a deliberate effort to compromise the integrity of the decentralized network. Once one member of a mining pool tries to prevent a correctly verified block from being announced to the others in the mining group cluster, this is known only as "selfish miner assault." This selfish operator shows greater proof-of-work than all the other prospectors in the network as a consequence of hiding their correctly extracted block from the community before moving onto the next frame. By doing this, the community as a whole may accept their transaction methods while the self-centred node keeps the block benefits or cash benefits [32].

V. System Model and Problem Formulation

A. Problem Formulation

In today's world of digital finance and banking, the security and integrity of financial transactions are at risk due to increased reliance on technology and growing cyber risks. For example, consider a business owner who transfers funds between accounts frequently using online banking services. Suddenly he notices unauthorized transactions on his account, which indicates a violation in security. In addition to causing financial loss, this incident also damages people's trust in the banking sector.

The aim of the project is to use blockchain technology to improve the security of online banking services and financial transactions. The paper covers a number of important topics such as cyber risk prevention, protecting data privacy, maximizing business efficiency and assessing the pros and cons of integrating blockchain technology into the financial sector. Given the weaknesses of current digital financial transactions (lack of trust, inefficient data sharing, privacy concerns and immutable data silos) a comprehensive review is needed to build a robust and secure blockchain framework.

B. System Model

A number of processes are involved in the suggested methodology for anomaly detection in the Ethereum Classic Blockchain (ETC), beginning with database analysis, employing the Encoder-Decoder LSTM architecture, and evaluating the outcomes as shown in Fig. 3.

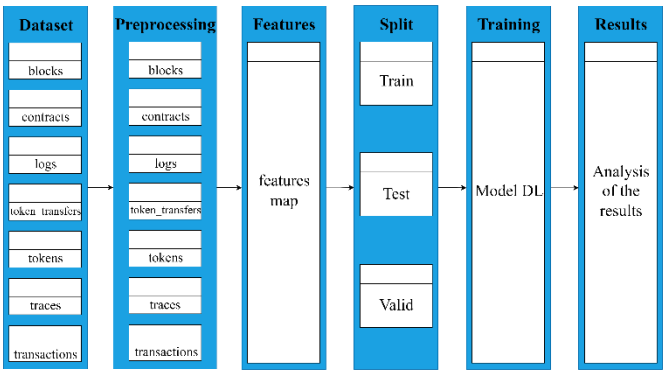


Fig. 3. Methodology used in the research.

VI. Proposed Solution

A. Introduction

Ethereum Classic Blockchain has risen to prominence as one of the leading decentralized platforms leveraging the immutable nature of a ledger to undertake safe and transparent transactions. As blockchain technology progresses and is used in more and more industries, the importance of securing and protecting the immutability of blockchain networks grows more important. Given that these systems contain significant monetary transactions and sensitive data, they are prime targets for cybercriminals looking for areas to exploit. Aside from cryptocurrency exchange platforms, another significant challenge that threatens the credibility of blockchain networks is maintaining the safety and security of blockchain networks.

Detecting anomalies is one of the most important challenges to keep blockchain networks trustworthy since these might suggest possible attacks or malicious actions. It is an essential part of proactive measures that help identify any movements or actions that are not consistent with normal blockchain behavior. Early detection and appropriate responses to such anomalies can minimize the consequences of an attack and protect the network.

Deep learning methods have shown impressive results recently in a number of fields, from natural language processing to computer vision. The Encoder-Decoder Long Short-Term Memory (LSTM) architecture in particular has become well-known due to its capacity to learn and represent intricate sequential patterns. This project focuses on using the Encoder-Decoder LSTM architecture to address the anomaly identification problem on the Ethereum Classic Blockchain by utilizing deep learning.

The Encoder-Decoder LSTM model is a good fit for jobs involving anomaly detection because it can efficiently identify temporal patterns and long-term dependencies in sequential data. The model can be trained on past ETC blockchain data to find patterns in the expected behavior of the network and then spot variations that might point to possible attacks or anomalies. The model effectively captures the subtle patterns that rule-based or statistical approaches may miss because of its capacity to encapsulate the input data and produce insightful representations.

B. Methodology

1) Database Analysis: Ethereum Classic is a public, open-source distributed computing platform built on the blockchain. It is notable for having smart contract capabilities, which enable scripts to run on the Ethereum Virtual Machine (EVM), a decentralized Turing-complete virtual machine. An international network of public nodes enables this functionality.

Ethereum Classic is notable for having a native value token called "ether." Ether is a cryptocurrency that may be held in wallets, transferred between users on the network, and used to pay nodes for the processing power they provide to the Ethereum platform.

Over the course of four years, from July 2015 to July 2019, we conducted tests on a section of the ETC blockchain as part of our research. The seven tables in the dataset we used are blocks, transactions, contracts, logs, token transfers, tokens, and traces. It can be accessed on Kaggle. These tables include important details regarding the blocks themselves, the operation of the network, and network use.

Multiple preprocessing stages were carried out in order to get the data ready for additional analysis. First, we carried out feature engineering, which included aggregation, correlation analysis, filtering, and the selection of the most relevant features. In order to rescale values and lessen the possibility of instability impacts during neural modeling, we secondly normalized the data. In addition, the process of normalization attempted to regularize the data by removing trending, cyclic, and seasonal irregularities. Using a shifting quantile ratio, we were able to achieve normalization.

The two parameters that the function first requires are {x}, which stands for the input data, and `window}, which indicates the rolling window's size (the default value is 20). Next, we make an object that rolls windows.

Next, we compute the first quartile. The value that divides the lowest 25% of the data from the remaining 75% is determined by passing the argument {0.25}. Furthermore, we set {interpolation='midpoint'} to ascertain the quartile value estimation technique.

In addition, we compute the third quartile. This time, the parameter {0.75} is passed in order to determine the value that divides the bottom 75% of the data from the top 25%. Lastly, we use the formula:

$$S= \{(x - q2) / (1.5 * (q3 - q1))\}$$

The original data is denoted by {df}, the median by {q2}, the first quartile by {q1}, and the third quartile by {q3}. Taking into consideration the interquartile range, the algorithm scales each value in the DataFrame according to how far it is from the median.

In a similar manner, we get the median by using the rolling window object's `median () ` function. The center figure that divides the data's upper and lower halves is known as the median.

2) Model architecture: The model's architecture is made up of multiple layers and hyperparameters that are specifically engineered to handle and evaluate data sequences as represented in Fig. 4. First, the training set is used to extract the length of the sequence and the number of features. One kind of recurrent neural network is the LSTM layer, which has 64 cells or neurons in its configuration. Additionally, the model has attention mechanisms with four attention heads, each with 64 dimensions. In addition, a convolutional neural network (CNN) layer with 64 filters and a kernel size of three is included in the architecture. Regularization methods like L1 and L2 regularization are used with lambda values of 0.2 to avoid overfitting.

Sequence support is defined for the input layer. Masking is applied to the input layer to manage sequences of varying lengths. To lessen overfitting, the model then incorporates numerous CNN layers with L2 regularization, dropout layers, and ReLU activation functions. To extract significant features and downsample the output, max pooling layers are used. The ReLU activation function and the designated number of cells are integrated into the LSTM layer. Furthermore, a multi-head attention layer is added to capture sequence relationships. Next come further CNN layers, dropout layers, and max pooling layers, then another LSTM layer.

To create the output sequence, a time-distributed dense layer is added last. The mean squared error loss function and Adam optimizer are used to construct the model. The model's summary is printed, together with information on its layers and parameter count. A predetermined path is used to save the trained model, and checkpoints are made to save the optimal model in accordance with validation loss. Additionally, a CSV logger is used to monitor the training progress.

The model has a batch size of 100, a validation split of 0.3, and is trained for 50 epochs on the training data. The model checkpoint and CSV logger are two of the callbacks that are used in the training process.

The architecture we've adopted to predict data, whether it's an attack or normal data. In our model, the green blocks represent the inputs and outputs. Then, there's the masking block, which allows us to handle data of different sizes. When we preprocess the data, we take a fixed-size sliding window. However, at the end of each time sequence, we have a set of data whose size is smaller than the window's size. So, instead of discarding this data, we add it to the model. Thus, the data won't have a fixed size altogether, and there won't be data of different sizes. Therefore, we have to deal with this data, and that's where we use the masking block.

Moving on, the blue block represents the convolutional layer, and the yellow block represents the dropout layer. Here, we'll use two sets, each consisting of two convolutional layers followed by a dropout layer. In the first set, the dropout will be 0.5, and in the second set, it will be 0.8. After finishing the second set, we'll transition to the LSTM layer. Then, after the LSTM layer, we'll apply a multi-head attention layer. At the end of the architecture, in the last group, we have two LSTM layers, and in between, there's a multi-head attention layer. After the LSTM layer, we'll enter two more sets, each comprising two convolutional layers and a dropout layer. The dropout rate will be reversed here, with the first set at 0.8 and the second at 0.5. The transition between them will be done through max pooling, and finally, we'll have the output layer.

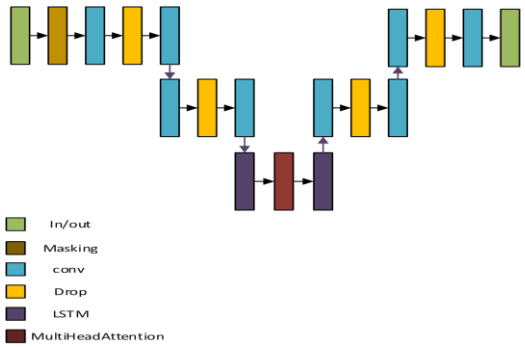


Fig. 4. Architecture of the proposed model.

a) LSTM layer: One kind of recurrent neural network (RNN) layer that is frequently utilized for sequence modeling is the Long Short-Term Memory (LSTM) layer [34]. The LSTM layer in this architecture is set up with 64 cells, or neurons. By preserving a memory state, LSTM cells are made to detect long-term dependencies in sequential input. Rectified Linear Unit (ReLU), the activation function utilized in this layer, gives the output non-linearity.

The LSTM (Long Short-Term Memory) layer employs a gating mechanism to regulate the memorization process. Through gates that open and close, you can store, write, or read information within LSTMs. An LSTM layer comprises the following components as can be seen in Fig. 5:

Forget gate: Responsible for deciding what information to retain and what to discard.

Input gate: Updates the cell state by incorporating information from the current input state (x) and the previous hidden state (h).

Cell state: Stores information based on the previous cell state (c) and new layer state. The current cell state is denoted as g.

Output gate: Determines the value of the next hidden state (h).

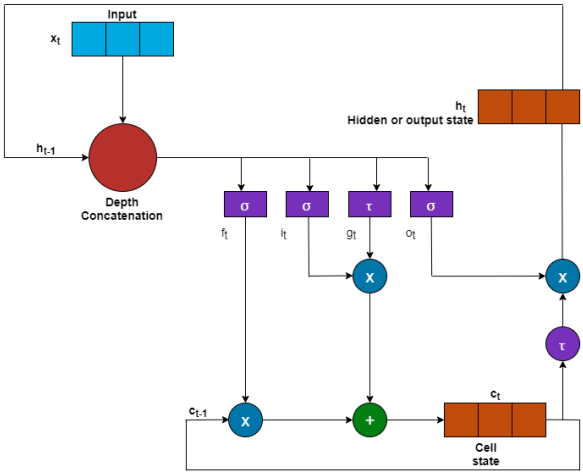


Fig. 5. The components of an LSTM layer [34].

b) Multi-Head attention layer: For the purpose of identifying linkages within the sequence, the Multi-Head Attention layer [35] is essential. It makes use of an attention mechanism with several attention heads, each of which focuses on a distinct segment of the input sequence. This enables the model to extract useful features by concentrating on pertinent data. With four attention heads and a key dimension of 64, the Multi-Head Attention layer in this design is applied to the LSTM layer's output.

At their core, they consist of keys (k) and values (v). We can create queries (q) to interact with these (k,v) pairs in a way that remains valid regardless of the size of the database.

The same query can yield varied responses depending on the database's contents. Let $D = \{(k_1, v_1), \dots, (k_m, v_m)\}$ represent a database of key-value pairs, and denote a query. We can define attention over D as:

$$Attention(q, D) = \sum_{i=1}^{m} \propto (q, k_i)\ v_i \quad (2)$$

Where $\propto (q, k_i) \in \mathbb{R}\ (i = 1, \dots, m)$ represents scalar attention weights, with the operation commonly known as attention pooling. The term "attention" stems from the focus placed on terms with significant weights $\propto$, implying larger values. Consequently, attention over D produces a linear combination of database values. Notably, this encompasses the earlier example as a special case where all weights except one are zero.

Given a query $q \in \mathbb{R}^{p_q}$, a key $k \in \mathbb{R}^{p_k}$, and a value $v \in \mathbb{R}^{p_v}$, each attention head $h_i (i = 1, \dots h)$ is computed as:

$$h_i = f\left(W_i^q q, W_i^k k, W_i^v v\right) \in \mathbb{R}^{p_v}$$

Where $W_i^q \in \mathbb{R}^{p_q \times d_q}$, $W_i^k \in \mathbb{R}^{p_k \times d_k}$, $W_i^v \in \mathbb{R}^{p_v \times d_v}$ are learnable parameters and is attention pooling, such as additive attention and scaled dot product attention .The multi-head attention output is another linear transformation via learnable parameters $W_o \in \mathbb{R}^{p_o \times h p_v}$ of the concatenation of heads:

$$W_o \begin{bmatrix} h_1 \\ \vdots \\ h_h \end{bmatrix} \mathbb{R}^{p_o}$$

According to this structure, each head has the ability to focus on distinct segments of the input, which allows for the expression of more complex functions beyond simple weighted averages.

The components of a Multi-Head Attention layer are shown in Fig. 6.

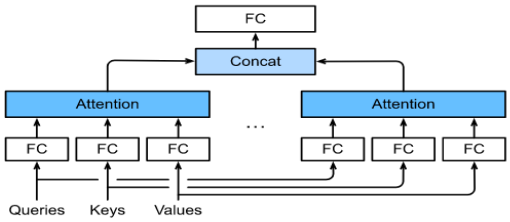


Fig. 6. The components of a Multi-Head Attention layer [36].

c) Convolutional Neural Network (CNN) layers: When attempting to extract geographical and temporal information from data, CNN layers are frequently employed. CNN layers are used in this model to examine the sequence data. Multiple CNN layers with the same configuration are part of the architecture. A predetermined number of filters make up each CNN layer, and these filters are in charge of identifying various patterns and features in the input. When a kernel size of three is employed, the CNN layer scans the input sequence using a three-size window. ReLU is the activation function used in these layers, which gives the output non-linearity. Furthermore, padding is set to'same' to guarantee that the length of the output and the input sequence match.

d) Dropout layers: A regularization method called dropout layers is employed to stop overfitting. During training, they arbitrarily deactivate a set of the neurons, which compels the model to acquire more resilient and comprehensive representations. Dropout layers with a dropout rate of 0.5 or 0.8 are placed after specific CNN layers in this architecture. Dropout layers are strategically placed to improve the model's generalization ability and lessen its sensitivity to noise.

e) Max pooling layers: The most notable aspects of the data are captured by downsampling the output using max pooling layers. They remove less significant data by dividing the input into non-overlapping parts and keeping just the largest value within each zone. By keeping the most important attributes, this downsampling aids in lowering the dimensionality of the data. In order to help with relevant information extraction and efficient feature representation, this design uses many max pooling layers after certain CNN layers.

f) Time-Distributed dense layer: The output sequence is produced at the end of the architecture using the Time-Distributed Dense layer. It separately applies a dense (completely linked) layer to every time step. As a result, the temporal relationships in the data are captured by the model,

which can now generate predictions for every element in the sequence. Since the aim in this scenario is to predict a single value for each element in the sequence, a dense layer with a single unit is used.

VII. RESULTS

The results that were obtained in this study are divided into three categories, the first of which is the output of the Preprocessing stage, where the data scanning process was applied. The second category, which is model training, is summarized using the loss function. Finally, the third stage is predicting attack data.

There are seven parameters in the database, the result of the Preprocessing for each variable will be presented separately. Regarding the attack data, it will be identified the same for all the data, so that it is possible to see at what timestamp it was determined to be attack data. To show the experimental results, initially the results of data processing are displayed after being pre-processed such that the horizontal axis represents the timestamps, and the vertical axis represents the different variables. Fig. 7 plots the normalized average gas provided along the time window to show the effect of the preprocessing processes. The red lines within the graphs indicate the attack-prone periods.

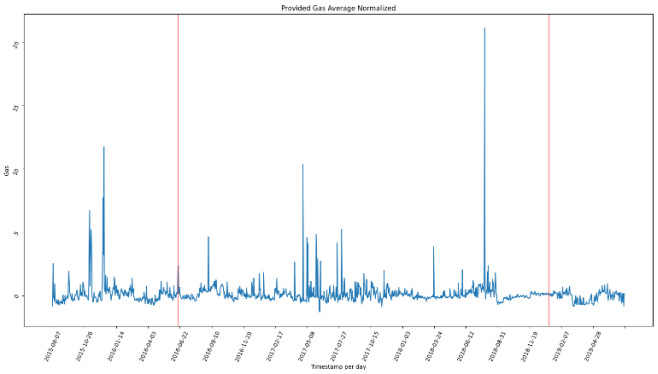


Fig. 7. Relationship of provided gas average with timestamp after normalization.

Similarly, Fig. 8 shows the timestamps represented on the horizontal axis, while the normalized transaction number is represented on the vertical axis. The red lines also represent attack data at these timestamps.

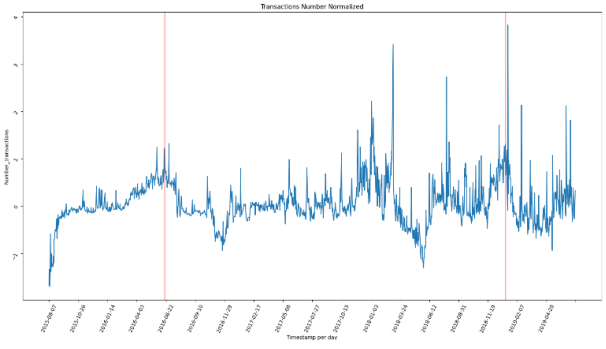


Fig. 8. Relationship of Transactions Number with Timestamp after normalization.

Fig. 9 shows the relationship between Block Difficult Average and Timestamp after normalization, where the horizontal axis represents the timestamps, while the vertical axis represents the normalized Block difficult average.

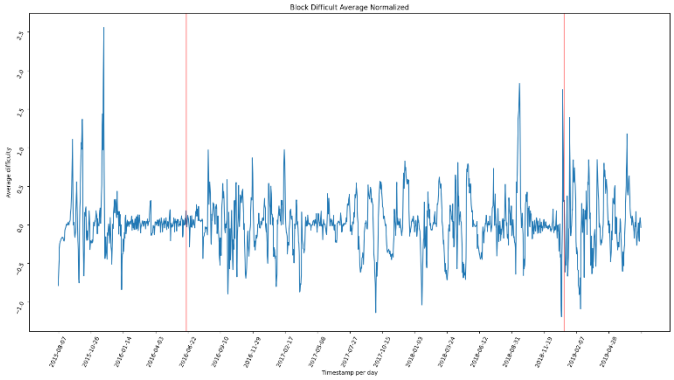


Fig. 9. Relationship of Block Difficult Average with Timestamp after normalization.

Another figure, Fig. 10 illustrates the relationship between block size average and timestamp after normalization, where timestamps are represented on the horizontal axis, whereas the block size average is represented on the vertical axis.

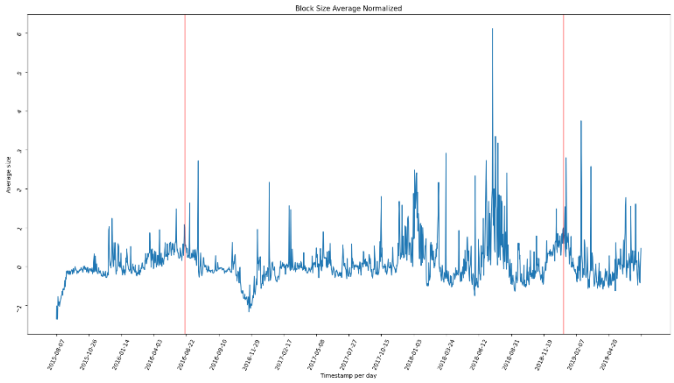


Fig. 10. Relationship of Block Size Average with Timestamp after normalization.

Fig. 11 shows the relationship between the gas used sum and the timestamps, where the latter is plotted on the horizontal axis, whereas the sum of used gas is plotted on the vertical axis.

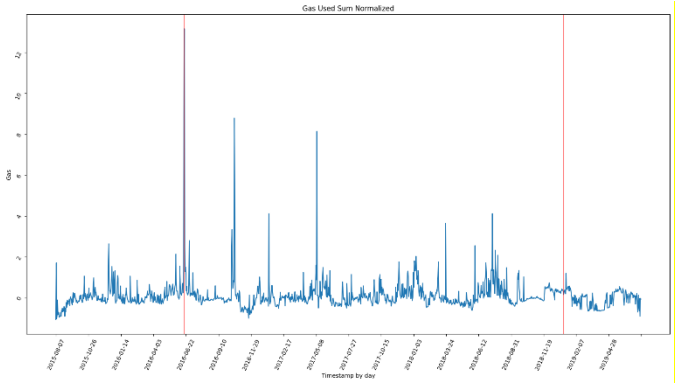


Fig. 11. Relationship of Gas Used Sum with Timestamp after normalization.

Fig. 12 and Fig. 13 show the relationship between the timestamps and the transaction average per plot, and the gas average per transaction respectively.

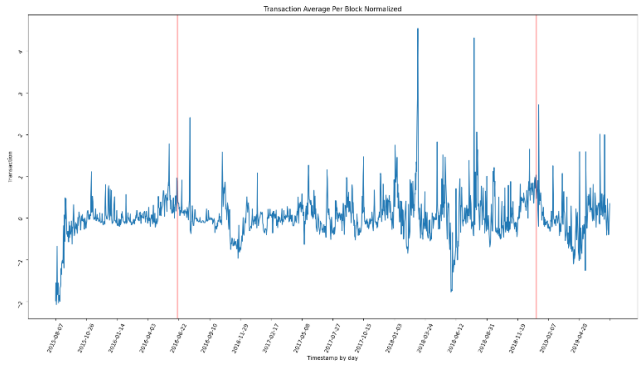


Fig. 12. Relationship of Transaction Average per Block with Timestamp after normalization.

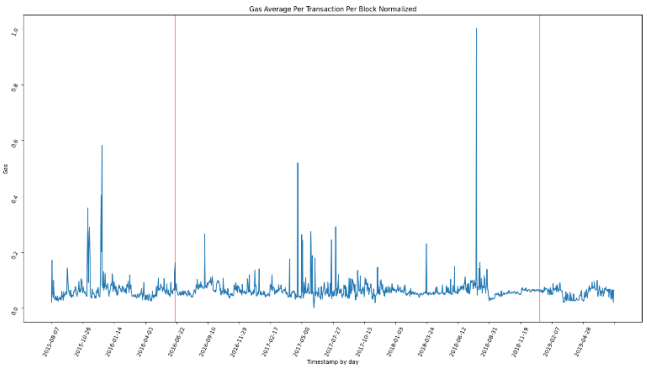


Fig. 13. Relationship of Gas Average Per Transaction Per Block with Timestamp after normalization.

The second part of the results involves the model training process. Fig. 14 shows the training loss for both training (orange) and testing (blue), where it is noticed that the presented model completely handles overfitting as a result of the L2 regularization with a value of 0.02 and the two dropout layers. A dropout rate of 0.8 is considered high enough to address overfitting, so the overfitting ratio was nearly zero from the beginning to the end of the training. However, this will affect the model's accuracy, making it lower than usual or expected, which is also compensated for by data scanning and pre-processing. Therefore, the model with almost have no overfitting and achieve high accuracy.

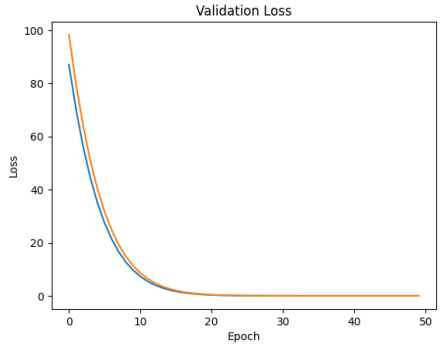


Fig. 14. Training and Validation Loss.

Fig. 15 shows the results of testing the proposed model on the test data, which includes attack data consistent with the reference study 1 [37]. This data includes a set of natural data over a period of time. Within these time periods, there will be attack data. This model output shows that the data marked by the red lines are attack data, while the rest of the data is natural. The test results show that the pre-processing and the proposed model give good performance for detecting logger anomalies on the network. Within the test data, it is worth noting that the test accuracy of the proposed model is 0.995.

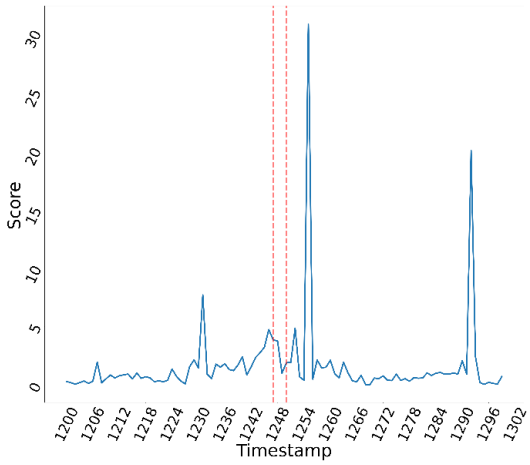


Fig. 15. Results obtained by the proposed model.

VIII. DISCUSSION

Fig. 16 and Fig. 17 represent a comparison between the results obtained in this study and the results found in study 1 [37] and study 2 [38]. The red line represents attack data, while the other data points represent normal data. In both our model and the reference study, the horizontal axis represents the timestamp, while the vertical axis represents a parameter from the database, such as average gas. As we can see, there is a difference because we applied preprocessing to our data, while the reference study used a different preprocessing method. Therefore, there is a slight difference in the data representation. However, both studies practically cover the same time period. We also observed that both models identify attack data during the same time periods. The difference lies in the fact that our model achieved higher accuracy in testing, i.e. in classifying this data as either attack or natural.

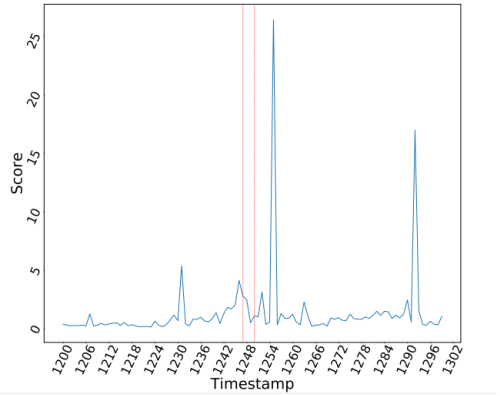


Fig. 16. Comparison of Results with reference Study 1 [37].

The anomaly detection results of the proposed model and results of Study 2 are shown in Fig. 17.

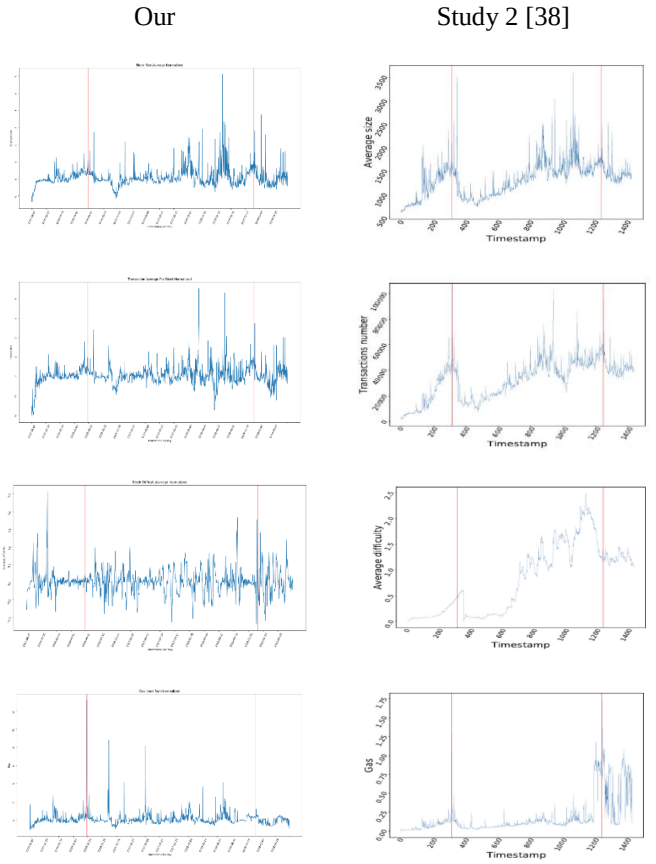


Fig. 17. Comparison of results with reference study 2 [38].

As a result, the proposed model shows the ability to predict the detection of anomalies for activities recorded on the network. It also efficiently addresses the challenge of overfitting during model training. It also achieves a prediction accuracy rate of 0.995 for the model on test data. Compared with reference studies, we find that the proposed model has the ability to capture dependencies. It is time efficient and has the ability to detect attacks on the network. For further development, we recommend using more general training data, as well as testing transformers on this type of challenges.

IX. Open Issues and Research Challenges

Open issues and research challenges in the field of enhancing the security of digital financial and banking transactions through blockchain-enabled approaches, particularly the implementation of a Long Short-Term Memory (LSTM) model, remain to be addressed. One of the key challenges is the need to develop more sophisticated anomaly detection techniques to effectively identify and mitigate potential threats in network-recorded activities. Additionally, there is a requirement for further exploration of the scalability and performance implications of using blockchain technology in large-scale financial systems, as well as the development of robust security systems to withstand evolving cyberattacks. Furthermore, the integration of blockchain technology with existing regulatory frameworks and compliance standards poses legal and regulatory challenges that need to be addressed for widespread adoption. These open issues call for continued research and collaboration among academia, industry, and regulatory bodies to ensure the successful implementation and utilization of blockchain-enabled security solutions in the realm of digital financial and banking transactions.

X. Conclusion

Blockchain has shown to be a revolutionary technology, but its widespread adoption has been hampered by a number of restrictions. Our project focused on enhancing the security of digital financial and banking transactions using blockchain technology. It addresses the challenges faced by the banking industry, such as inefficiency, high fraud rates, and lack of transparency, and proposes a solution through the implementation of blockchain. The research aims to develop an analytical model capable of detecting attacks and anomalies on the Ethereum Classic (ETC) blockchain by employing an Encoder-Decoder LSTM architecture. The study emphasizes the importance of cybersecurity in the banking sector and the potential of blockchain technology to revolutionize the industry by providing a secure, efficient, and transparent platform for financial transactions. The thesis outlines the methodology used, the results obtained, and the contributions made to the field of blockchain security. It concludes with suggestions for future research directions, highlighting the ongoing need for innovation in the realm of digital financial security.

The project's findings revealed that adding machine learning and blockchain technology may significantly improve and refine numerous security sectors. However, this study is simply the beginning of a broader and more extensive inquiry into this type of integration, underlining the need for more research that looks into numerous authentication elements across diverse datasets to balance security and usability.

Future work will focus on addressing the scalability of the proposed LSTM model within larger decentralized financial ecosystems, particularly those operating on multiple blockchain platforms. Additionally, further research will explore the integration of reinforcement learning techniques to enhance real-time anomaly detection. Another promising avenue is the application of this model to emerging blockchain networks to determine its effectiveness in varied contexts. Extending the study to multi-blockchain scenarios could also provide insights into cross-network security enhancements.

Acknowledgment

The authors would like to thank the Deanship of Graduate Studies and Scientific Research at Jouf University for funding and supporting this research through the initiative of DGSR, Graduate Students Research Support (GSR) at Jouf University, Saudi Arabia.

References

[1] S. Nakamoto, "Bitcoin: A peer-to-peer electronic cash system," 2008. [Online]. Available: https://bitcoin.org/bitcoin.pdf.

[2] M. Casey and P. Vigna, "In blockchain we trust," MIT Technol. Rev., 2018.

[3] "Bitcoin BTC," Coinmarketcap, 2022. [Online]. Available: https://coinmarketcap.com/.

[4] T. I. Team, "Ethereum Classic (ETC) definition, history, future," Investopedia, 31 May 2023. [Online]. Available: https://www.investopedia.com/terms/e/ethereum-classic.asp. [Accessed: Mar. 21, 2024].

[5] M. Paige, "The evolution of digital transformation: From pre-internet to post-pandemic," Hatchworks, 3 Feb 2023. [Online]. Available: https://hatchworks.com/blog/product-design/history-digital-transformation/. [Accessed: Mar. 21, 2024].

[6] M. Kirmani and M. T. Banday, "Digital forensics in the context of the Internet of Things," in Cryptographic Security Solutions for the Internet of Things, 2019, pp. 296-324.

[7] A. Demirgüç-Kunt, "Is bank competition a threat to financial stability?," World Bank, 10 Apr 2012. [Online]. Available: https://blogs.worldbank.org/en/allaboutfinance/is-bank-competition-a-threat-to-financial-stability. [Accessed: Mar. 21, 2024].

[8] M. U. Hassan, M. H. Rehmani, and J. Chen, "Anomaly detection in blockchain networks: A comprehensive survey," arXiv, 2022.

[9] M. Javaid, A. Haleem, R. P. Singh, R. Suman, and S. Khan, "A review of blockchain technology applications for financial services," BenchCouncil Trans. Benchmarks, Standards, and Evaluations, vol. 2, 2022.

[10] S. Trivedi, K. Mehta, and R. Sharma, "Systematic literature review on application of blockchain technology in E-finance and financial services," J. Technol. Manage. Innov., vol. 16, no. 3, pp. 89-102, 2021.

[11] J. Hartmann and O. Hasan, "Privacy considerations for a decentralized finance (DeFi) loans platform," Cluster Comput., vol. 26, no. 4, pp. 2147-2161, 2023.

[12] C. H. Liao, X. Q. Guan, J. H. Cheng, and S. M. Yuan, "Blockchain-based identity management and access control framework for open banking ecosystem," Future Gener. Comput. Syst., vol. 135, pp. 450-466, 2022.

[13] D. Boughaci and A. A. Alkhawaldeh, "Enhancing the security of financial transactions in blockchain by using machine learning techniques: Towards a sophisticated security tool for banking and finance," in Proc. 2020 1st Int. Conf. Smart Syst. Emerg. Technol. (SMARTTECH), 2020, pp. 110-115.

[14] H. Song and Y. Chen, "Digital financial transaction security based on blockchain," J. Phys.: Conf. Ser., vol. 1744, 2021.

[15] M. Mozaffari-Kermani and A. Reyhani-Masoleh, "A low-cost S-box for the advanced encryption standard using normal basis," in Proc. IEEE Int. Conf. Electro/Inf. Technol., 2009.

[16] K.-K. R. Choo, M. M. Kermani, R. Azarderakhsh, and M. Govindarasu, "Emerging embedded and cyber physical system security challenges and innovations," IEEE Trans. Depend. Secure Comput., vol. 14, no. 3, pp. 235-246, 2017.

[17] M. Mozaffari-Kermani, R. Azarderakhsh, C.-Y. Lee, and S. Bayat-Sarmadi, "Reliable concurrent error detection architectures for extended Euclidean-based division over GF(2m)," IEEE Trans. Very Large Scale Integr. (VLSI) Syst., vol. 22, no. 5, pp. 995-1003, 2014.

[18] A. Jalali, R. Azarderakhsh, M. M. Kermani, and D. Jao, "Towards optimized and constant-time CSIDH on embedded devices," in Proc. Springer, vol. 11421, pp. 215–231, 2019.

[19] B. Koziel, R. Azarderakhsh, and M. Mozaffari-Kermani, "Low-resource and fast binary Edwards curves cryptography," in Proc. Springer, vol. 9462, pp. 347–369, 2015.

[20] J. Xue, X. Jiang, P. Li, W. Xi, C. Xu, and K. Huang, "Side-channel attack of lightweight cryptography based on MixColumn: Case study of PRINCE," Electronics, vol. 12, no. 544, 2023.

[21] A. Benjamin, J. Herzoff, L. Babinkostova, and E. Serra, "Deep learning based side channel attacks on lightweight cryptography (student abstract)," in Proc. AAAI Conf. Artif. Intell., 2022.

[22] C. Su and Q. Zeng, "Survey of CPU cache-based side-channel attacks: Systematic analysis, security models, and countermeasures," Secur. Commun. Netw., vol. 2021, no. 1, pp. 1-15, 2021.

[23] R. Walters, Blockchain technology and the future of banking, Robert Walters, 2021.

[24] PinkExc, "PinkExc," Twitter, 11 Oct 2019. [Online]. Available: https://twitter.com/pinkexc/status/1182550209944944640. [Accessed: May 2024].

[25] W. Kersten, T. Blecker, and C. Ringle, "Digitalization in supply chain management and logistics: Smart and digital solutions for an Industry 4.0 environment," in Hamburg Int. Conf. Logist. (HICL), Berlin, 2017.

[26] CIP, "CIU's must use blockchain technology for working together and share data, citizen by investment," CIP J., 2018.

[27] K. C. Tran and J. Benson, "What is Ethereum Classic?," Decrypt, 2 Feb 2022. [Online]. Available: https://decrypt.co/resources/what-is-ethereum-classic-explained-guide-cryptocurrency. [Accessed: Mar. 27, 2024].

[28] E. Erdmann, Strengths and drawbacks of voting methods for political elections, University of Minnesota, 2011.

[29] S. Park and R. Rivest, "Towards secure quadratic voting," Eprint, 2016.

[30] M. Hasan, M. S. Rahman, H. Janicke, and I. H. Sarker, "Detecting anomalies in blockchain transactions using machine learning classifiers and explainability analysis," arXiv, 2024.

[31] D. Kaidalov, L. Kovalchuk, A. Nastenko, M. Rodinko, O. Shevtsov, and R. Oliynykov, "Ethereum Classic treasury system proposal," Input | Output, 2017.

[32] J. Li, C. Gu, F. Wei, and X. Chen, "A survey on blockchain anomaly detection using data mining techniques," in Blockchain and Trustworthy Systems, 2020, pp. 491-504.

[33] G. BigQuery, M. Risdal, A. Day, and Y. Khoury, "Ethereum Classic blockchain," Kaggle, 2019. [Online]. Available: https://www.kaggle.com/datasets/bigquery/crypto-ethereum-classic?select=transactions. [Accessed: Apr. 22, 2024].

[34] MathWorks, "How Deep Learning HDL Toolbox compiles the LSTM layer," [Online]. Available: https://www.mathworks.com/help/deep-learning-hdl/ug/how-deep-learning-hdl-toolbox-compiles-the-lstm-layer.html. [Accessed: Apr. 22, 2024].

[35] A. Zhang, Z. C. Lipton, M. Li, and A. J. Smola, "Multi-head attention," in Dive into Deep Learning, Cambridge University Press, 2023.

[36] Google Colab, "Multihead-attention.ipynb," 2021. [Online]. Available: https://colab.research.google.com/github/deepjavalibrary/d2l-java/blob/colab/chapter_attention-mechanisms/multihead-attention.ipynb. [Accessed: Apr. 22, 2024].

[37] F. Scicchitano, A. Liguori, M. Guarascio, E. Ritacco, and G. Manco, "A deep learning approach for detecting security attacks on blockchain," in Proc. 4th Italian Conf. Cyber Secur. (ITASEC), Ancona, Italy, 2020.

[38] S. Dhandapani and G. Maragatham, "Design of blockchain-enabled intrusion detection model for detecting security attacks using deep learning," Pattern Recognit. Lett., vol. 153, 2021.

Sketch and Size Orient Malicious Activity Monitoring for Efficient Video Surveillance Using CNN

K. Lokesh[1], M. Baskar[2]*
Research Scholar, Department of Computer Science and Engineering-School of Computing, College of Engineering and Technology, SRM Institute of Science and Technology, Kattankulathur, Chengalpattu, Tamilnadu 603 203, India[1]
Associate Professor, Department of Computing Technologies-School of Computing, College of Engineering and Technology, SRM Institute of Science and Technology, Kattankulathur, Chengalpattu, Tamilnadu 603 203, India[2]

***Abstract*—Towards malicious activity monitoring in organizations, there exist several techniques and suffer with poor accuracy. To handle this issue, an efficient Sketch and Size orient malicious activity monitoring (SSMAM) is presented in this article. The model captures the video frames and performs segmentation to extract the features of frames as shapes and size. The video frames are enhanced for its quality by applying High Level Intensity Analysis algorithm. The quality improved image has been segmented with Color Quantization Segmentation. Using the segmented image, the feature are extracted and applied with scaling and rotation for different number of size and angle. Such features extracted have been trained with convolution neural network. The CNN model is designed to perform convolution on two levels and pooling as well. At the test phase, the method extract the same set of features and performs convolution to obtain same set of feature lengths and the neurons are designed computes the value of Sketch Support Measure (SSM) towards various class of activity. According the value of SSM, the method classifies the user activity towards efficient video surveillance. The proposed approach improves the performance in activity monitoring and video surveillance.**

Keywords*—Video surveillance; deep learning; activity monitoring; malicious activity; SSMAM; SSM

I. Introduction

Growth of information and communication technology has been adapted to various scientific, medical and security problems. Industrial security is the most concern in recent days and they tends to enforce the human monitoring in the units of organization. Video surveillance is the process of monitoring the human activity in various locations and units of any organization. In this way, they can monitor the activity of human, worker towards enforcing security and maintaining the human resource [5, 6]. By monitoring the human activity, the malicious event happening within or outside the organization can be tracked and used towards enforcing security.

The video frames captured through set of video devices and cameras are the key input for the video surveillance system of any organization. By using the images of video, the human object can be tracked in successive frames and their activity can be classified. Image processing plays vital role in identifying the human texture and can be used to extract the features from the frame to support the classification of activity. In order to perform event or activity classification, the system has to maintain number of frames and objects of humans about different activities. By maintaining such set of frames and objects, the input frame can be classified for its activity [7].

To perform video surveillance there are numbers of machine learning algorithms described in literature. For example, support vector machine (SVM), Decision Tree, Bayesian Classifier, Neural Network are used in earlier articles [11]. The methods capture the video frame and perform background subtraction which is learned from the different previous frames. With the subtracted image, the method would identify the human object to extract the object and train the model. The issue with the machine learning algorithms is the dimensionality and missing features. In order to achieve higher performance in video surveillance, it is necessary to maintain the set of human objects under different activities. With the trace, the method can identify the class of any video frame for its activity. The machine learning algorithms are little uncomfortable to maintain and handling such huge volume of frames. This is where the deep learning algorithms come to play.

The deep learning algorithms are capable of converting the huge volume of features into small set without the loss of feature and data. For example, Convolution neural network is more popular on handling such higher volume data. It would convert the features of video frames into tiny feature set and used to perform classification. With all these consideration, this article presents a novel Sketch and Size Orient Malicious Activity Monitoring (SSMAM) model. The model concentrates on extracting the human features and sketch features. By extracting the sketch and size features, the model has been trained with number of features at each class with convolution neural network. The SSMAM-CNN model is capable of handling various class of activities and support the detection of malicious activity in the premises. By adapting SSMAM-CNN model the human activity can be greatly monitored.

A. Problem Statement

The methods discussed in literature uses variety of features from object, graph, edge, etc. towards human activity monitoring. But still they suffer to achieve higher accuracy towards malicious activity monitoring. The accuracy of video surveillance and activity monitoring is greatly depending on the kind of feature considered. The human sketch is the most effective one which can be used in activity monitoring to detect any malicious activity happening in the venue.

B. Contributions

- The problem of activity monitoring is approached with sketch and size features along with other features.
- CNN model is designed for training and classification.
- Sketch Support Measure (SSM) is estimated to perform classification.

The article is structured to provide general introduction about the problem in Section I. Section II briefs the literature review around the problem and Section III presents the detailed working of the model. Section IV is dedicated to present the experimental result and Section V discuss the conclusion.

II. Related Work

Different methods are analyzed around video surveillance and activity monitoring. Set of methods around the problem is discussed here.

A Gaussian Mixture Model with Universal attribute model is presented in [1], which monitor the violent activities by computing super action vector. A Growing self-organizing map (GSOM) is presented in [2], which uses traditional deep learning and multistream learning to use unlabeled data to handle the over fitting problem to identify malicious activity.

An Efficient Marine Organism Detector (EMOD) is discussed in [3], which uses attention relation to detect malicious activity. A deep learning based video surveillance model is presented in [4], which classify the objects under different categories and identify the region in the frame to perform supervised learning. A human action recognition (HAR) model is presented to monitor the activity in aircraft surveillance. The method uses temporal features and use LSTM to perform classification.

A visual surveillance framework is presented in [6], to monitor the human fall which uses deep CNN and uses aggregated heuristic visual features in detecting the occurrence of falls. An audio video based activity recognition model is presented in [7], towards video surveillance.

A human activity recognition model is sketched in [8], which uses data enhancement techniques to collect discriminative features from various activities. The features captured are transformed with existing model to perform recognition. A hybrid visual geometry group based bidirectional short term memory model is presented in [9], towards monitoring the moments of animals and produces alarm according to the activity of the animals. A transformer network based LSTM model is presented in [10], towards recognizing the activity of the human.

A multipedestrian multicamera tracking (IMPMCT) model is presented in [11], which uses spatial and temporal features towards monitoring the pedestrian monitoring. An Interactive video surveillance as an edge service (I-ViSE) model is presented in [12], which works according to the feature queries. The method adapts features of human body, color and cloths in activity monitoring. An LSTM based baby activity recognition model is presented in [13], which uses pose features of baby towards recognition.

A Class-privacy Preserved Collaborative Representation (CPPCR) based multi-modal human action recognition model is presented in [14], which uses temporal structure in recognizing the activity. An LSTM with 3DCNN model is presented in [15], towards detecting illegal activity in the environment.

A realtime surveillance model is presented in [16], towards detecting malicious behavior using deep transfer learning. A posture recognition system is presented in [17], which uses mobilenetV2 towards estimating the human posture and LSTM is used to extract the features. Machine learning algorithm is used for classification. A secure surveillance scheme is presented in [18], to support healthcare system with enabled internet of Things. A deep learning multi layered CNN with LSTM is presented in [19], to capture the physical activities of persons and performs activity classification.

A machine learning based harmful weapon detection model is presented in [20], which uses different pistol classes and objects. The method uses sliding window classification model to perform classification. Lightweight Deep Neural Network (LDNN) with Convolution Long Short Term Memory (ConvLSTM) model is presented in [21], towards detecting abnormal activity in the environment. A Graph Convolution Network with 3DCNN is presented in [22], towards detecting abnormal behavior. 3DNN is used to extract the features and GCN has been used to perform classification. An audio and visual based activity detection model is sketched in [23], which uses both audio and video data in detecting abnormal activity in video frames. The method uses PSO and social force model towards feature extraction and classification. A Faster Region-Based Convolutional Neural Networks (Faster RCNN) is presented in [24], to detect the firearms in organization. The model uses ensemble learning to detect the human face and guns using Weighted Box Fusion techniques.

A spatio temporal attention fusion slowfast network based model is presented in [25], which uses spatial and temporal features to perform classification. A 2D and 3D feature based HAR is presented in [26], which extract the geometric features using deep and machine learning algorithms like CNN to classify the activity of human using LSTM and SoftMax classifier. A multi perspective abnormal posture recognition model is presented in [27], which works according to multi view cross information and posture features towards detecting illegal activity.

All the above discussed approaches suffer to achieve higher performance in activity monitoring and video surveillance.

Research Gap: According to the literature, the existing works does not consider the sketch and size features with the combination of other object features. This affects the performance of activity monitoring.

III. Sketch and Size Orient Malicious Activity Monitoring with CNN (SSMAM-CNN)

The model captures the video frames and performs segmentation to extract the features of frames as shapes and size. The video frames are enhanced for its quality by applying High Level Intensity Analysis algorithm. The quality improved image has been segmented with Color Quantization

Segmentation. Using the segmented image, the feature are extracted and applied with scaling and rotation for different number of size and angle. Such features extracted have been trained with convolution neural network. The CNN model is designed to perform convolution on two levels and pooling as well. At the test phase, the method extract the same set of features and performs convolution to obtain same set of feature lengths and the neurons are designed computes the value of Sketch Support Measure (SSM) towards various class of activity. According the value of SSM, the method classifies the user activity towards efficient video surveillance.

The functional architecture of SSMAM-CNN model has been presented in Fig. 1 and the functions of the model are briefed in this section.

A. High Level Intensity Analysis Preprocessing

The high level intensity analysis algorithm enhances the quality of video frame given. To perform this, the method fetch the image and extract the RGB layers. Among the three layers, the method use the red layer features to enhance the image quality. The method initializes a window size and crops the features from the red layer. Among the window feature, the method computes the max intensity value (MIV) and least intensity value (LIV). Using these two, the method traverse through each pixel and computes the Max Intensity Distance and Least Intensity Distance. Based on the value of distances, the method identifies the closest intensity sector and computes the Intensity Normalization Value. Based on that, the pixel has been adjusted for its intensity. The quality enhanced image has been used to perform segmentation and activity monitoring.

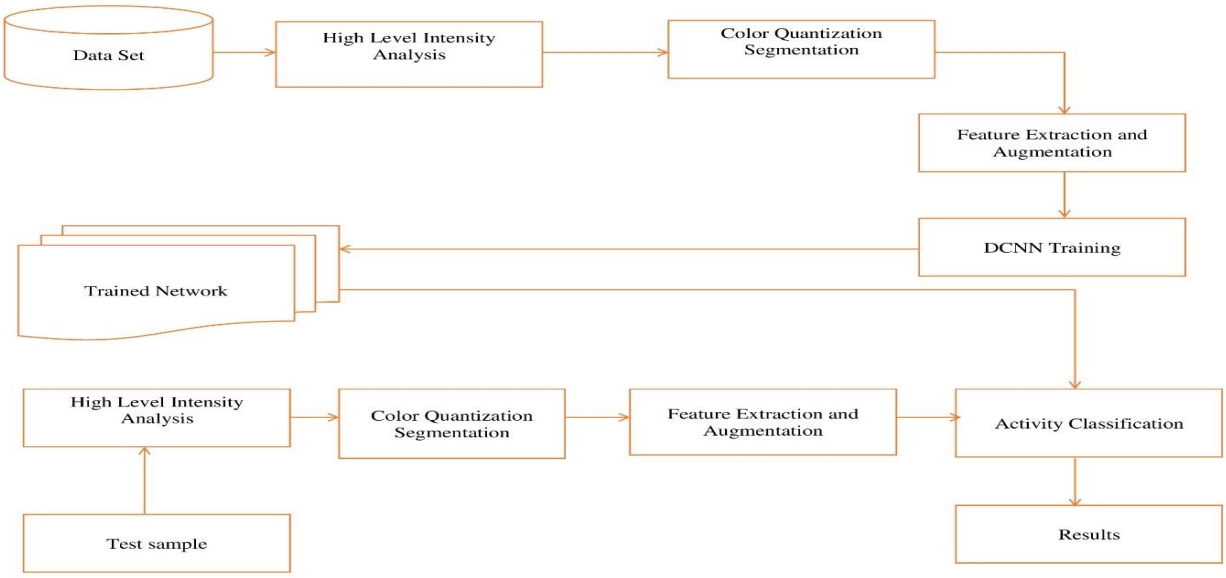


Fig. 1. Architecture SSMAM-CNN model.

Algorithm:

Given: Video Frame Vf
Obtain: Enhanced Video Frame Evf.
Start
Read Vf.
[RGB]= Extract RGB Layers (Vf)
Initialize window size Ws = 5
For each window w
Compute Max Intensity Value Miv = $Max(W(R(i))_{i=1}^{Size(w)}$
Compute least intensity value LIV = $Min(W(R(i))_{i=1}^{Size(w)}$

For each pixel p
Compute Max intensity Distance Mid = Dist(p(R), Miv)
Compute Least intensity Distance Lid = Dist (P (R), LIV)
if Mid<Lid then
Compute intensity normalization value Inv = p(R)+Mid/3
Evf(W(p(R))) = Inv
End
End
End
Stop

The high level intensity normalization algorithm computes the INV value for different pixels of red layers of the frame to adjust the quality of the frame. The enhanced image has been used towards segmentation and activity monitoring.

B. Colour Quantization Segmentation

The colour quantization algorithm segments the image according to the colour features. The quantization is performed based on the RGB values obtained from the frame given. To perform this, the method reads the red green black layer features and for each layer feature, the method computes the Quantization Measure (QM) which is measured based on the overall feature value and identifies the non-dominant value. To obtain this, the method generates the histogram for each layer

and identifies the feature value with same numbers. Further, the method segments the identified pixels with same quantization measure to perform segment the image. The pixels selected with the same quantization value has been grouped and produces segmented image.

The colour quantization segmentation algorithm groups the pixels according to the QM value identified from the histogram of all layer features to produce segmented image.

Algorithm:

Given: Enhanced Image Eimg
Obtain: Segmented Image Simg
Start
Read Eimg
[R,G,B]= Extract RGB features.
For each layer R, G, B
Histogram Histo = Generate Histogram (R, G, B)
For each histogram value h
For each other layer histogram Hi
Compute QM = $\sum_{i=1}^{size(Hi)} Hi(i) == h$
End
End
End
For each pixel pi
If Pi(R,G,B)== QM(hi) then
Add to group
Simg(pi)= 0
End
End
Stop

IV. Feature Extraction and Augmentation

The proposed model extract the sketch features from the segmented image and computes the size of the sketch. According to the sketch feature, the method generates various size of sketch to produce data augmentation. Augmented feature set has been used to perform training and testing.

A. DCNN Training

The activity data set given has been fetched and the frames of the data set has been used to perform high level intensity analysis which enhance the image. Further, the method applies color quantization segmentation to segment the image. From the segmented image, the method extract the sketch and size features and produces augmented data. With the augmented data, the method generates the convolution neural network, which comes with two level of convolution. At the first level, the sketch features are convolved to one dimension feature and at the second convolution, the method generates unique size of feature set. The max pooling layers adjust the values of the feature. The method generates number of neurons according to the number samples and augmented data given. Each has been initialized with the unique sample given. At the first convolution layer, the sketch features are convolved to unique standard size and at the second convolution layer, the neurons convert the features into one dimensional array. Finally, the size features are append to the feature set. Generated network has been used to perform activity classification.

B. Activity Classification

At the test phase, the method applies high level intensity analysis and color quantization segmentation on the given video frame. Further, the method extract the same set of features and performs convolution to obtain same set of feature lengths and the neurons are designed computes the value of Sketch Support Measure (SSM) towards various class of activity. According the value of SSM, the method classifies the user activity towards efficient video surveillance.

Algorithm:

Given: Video Frame Vf
Obtain: Activity_Class Ac
Start
Read Vf and CNN.
Eimg = Perform high level intensity analysis on Vf.
Simg = Apply color quantization segmentation (Eimg)
[sketch, size] = Feature Extraction and Augmentation on Simg.
Pass through CNN
At convolution layer one
Convert sketch feature to standard size.
Perform Max pooling
At convolution layer two
Convert sketch feature to one dimension value.
Perform max pooling
At output layer
For each class of activity
For each neuron n
Compute Sketch support measure SSM.

$$SSM = \frac{\sum_{i=1}^{size(n(f))} count(n(f).value == Vf(f(i)))}{size(Vf(f))}$$

End
Compute cumulative SSM = $\frac{\sum SSM}{size(neurons)}$
End
Activity_Class = Choose class with maximum SSM.
Stop

The activity classification algorithm computes the SSM value for various class of activity according to the feature considered. Based on the value of SSM, the method performs classification.

V. Results and Discussion

The sketch and size based malicious activity monitoring model is hardcoded with MATLAB. The performance of the method is evaluated under various factors. The method uses various data sets with different poses and shapes. The results obtained have been compared with the results of other approaches.

The evaluation detail used towards evaluating the performance of various methods are measured and presented in this section. The methods discussed in literature are evaluated for their performance with different data sets according to Table I. Their performances are evaluated under various parameters and the results produced by the approaches are compared with others in this section.

TABLE I. EVALUATION DETAILS

Parameter	Value
Number of Activities	20
Total Images	10000
Tool	MATLAB
No of users	700

TABLE II. ANALYSIS OF MALICIOUS ACTIVITY DETECTION ACCURACY

Malicious Activity Detection Accuracy % vs No of Activities			
	5 Activities	10 Activities	20 Activities
DMPMAM	82	88	95
DFI_SVQA	72	75	79
MulVIS	76	81	85
SSOcT	71	77	82
SSMAM_CNN	85	91	97

The accuracy of methods in detecting malicious activity has been counted and analyzed in Table II, where the SSMAM_CNN scheme stimulates higher accuracy than other techniques.

The accuracy in detecting malicious activity is measured and compared in Fig. 2. The SSMAM_CNN model improves the performance at the increasing number of activity classes and videos.

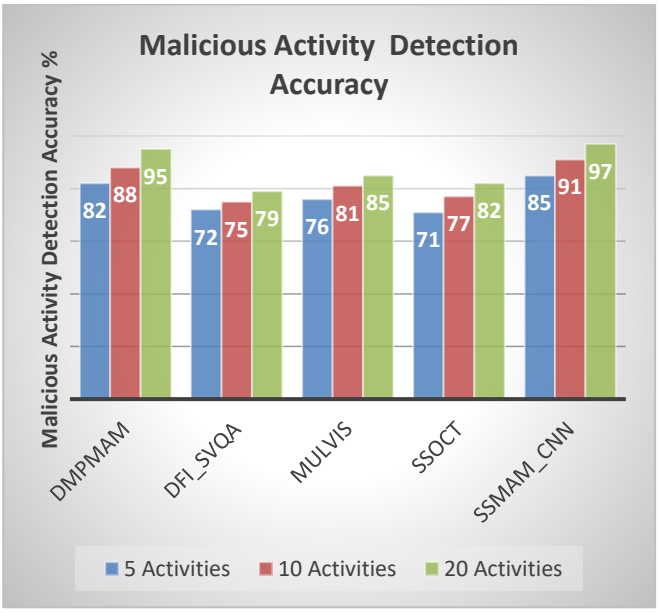


Fig. 2. Malicious activity detection accuracy.

The ratio of false detection is measured and compared in Table III, where the SSMAM_CNN approach has produced fewer ratios than other techniques.

TABLE III. FALSE RATIO IN MALICIOUS ACTIVITY DETECTION

False Ratio % vs No of Activities			
	5 Activities	10 Activities	20 Activities
DMPMAM	18	12	5
DFI_SVQA	18	25	21
MulVIS	24	19	15
SSOcT	29	23	18
SSMAM_CNN	15	9	3

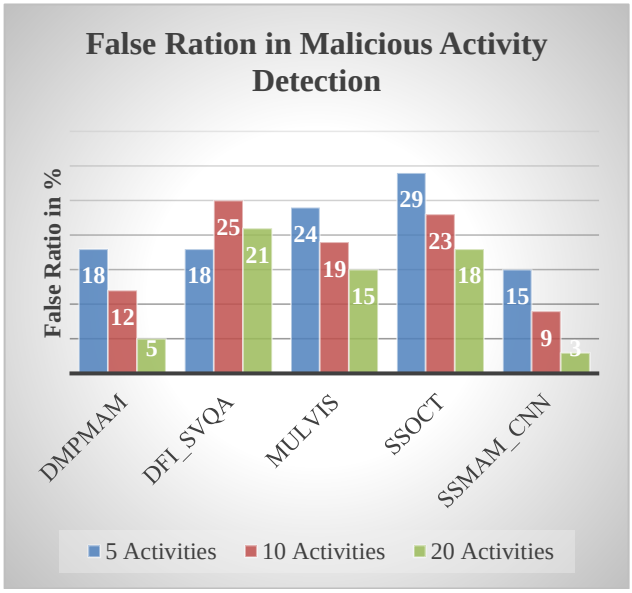


Fig. 3. False ratio in malicious activity detection.

The false classification ratio introduced by various approaches in malicious activity detection is measured & compared in Fig. 3. The SSMAM_CNN models introduces negligible false detection ratio compare to others in all the cases.

TABLE IV. TIME COMPLEXITY IN MALICIOUS ACTIVITY DETECTION

Time Complexity in Malicious Activity Detection % vs No of Activities			
	5 Activities	10 Activities	20 Activities
DFI_SVQA	67	77	88
DMPMAM	21	32	45
MulVIS	58	73	83
SSOcT	48	62	77
SSMAM_CNN	18	26	37

The time complexity in Millie seconds produced by various methods in detecting the malicious activity is measured and compared in Table IV, where the SSMAM_CNN approach produced less time complexity compare to other approaches.

The time complexity produced by various approaches in malicious activity detection is measured and compared in Fig. 4. The SSMAM_CNN model has produces less time complexity in all the test cases.

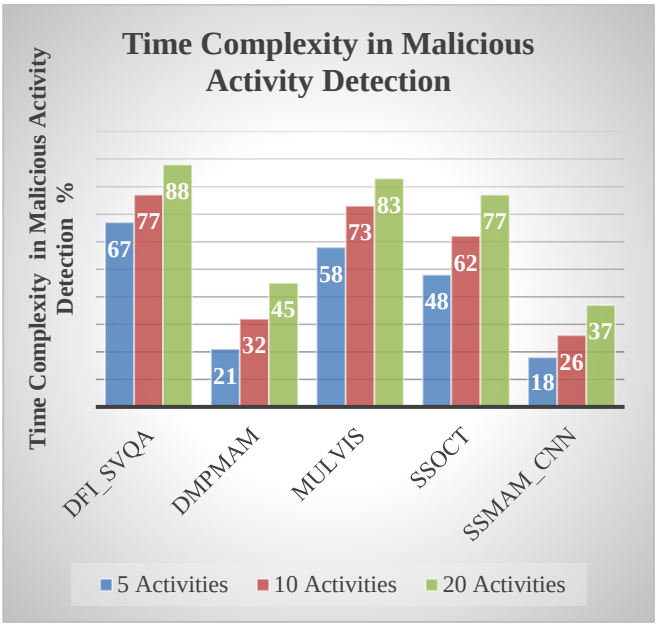


Fig. 4. Analysis time complexity in malicious activity detection.

VI. Conclusion

To maximize the accuracy of malicious activity detection, this article presented a novel sketch and size based malicious activity detection model with CNN (SSMAM_CNN). Apart from using object and shape features towards the problem, this approach uses the sketch and size features of the person to classify the activity. By adapting sketch and size features with the problem, the accuracy has been greatly improved. The model applies high level intensity analysis to enhance the quality of the image. Further, the method applies color quantization segmentation to segment the image. Next, the features are extracted and augmented data has been produced to train the convolution neural network. At the test phase, the method computes the sketch support measure (SSM) towards various class of activity features. According to the value of SSM, the class with maximum SSM value has been selected to perform the classification. The proposed model improves the performance of malicious activity detection accuracy up to 97%.

Further, the research can be extended by adapting invariant position features to stimulate the performance.

VII. Funding Statement

The authors received no specific funding for this study.

Conflicts of Interest

The authors declare they have no conflicts of interest to report regarding the present study.

References

[1] M. Mudgal, D. Punj and A. Pillai, "Suspicious Action Detection in Intelligent Surveillance System Using Action Attribute Modelling," in Journal of Web Engineering, vol. 20, no. 1, pp. 129-146, January 2021, doi: 10.13052/jwe1540-9589.2017.

[2] R. Nawaratne, D. Alahakoon, D. De Silva, H. Kumara and X. Yu, "Hierarchical Two-Stream Growing Self-Organizing Maps With Transience for Human Activity Recognition," in IEEE Transactions on Industrial Informatics, vol. 16, no. 12, pp. 7756-7764, Dec. 2020, doi: 10.1109/TII.2019.2957454.

[3] Z. Shi et al., "Detecting Marine Organisms Via Joint Attention-Relation Learning for Marine Video Surveillance," in IEEE Journal of Oceanic Engineering, vol. 47, no. 4, pp. 959-974, Oct. 2022, doi: 10.1109/JOE.2022.3162864.

[4] S. A. Ahmed et al., "Query-Based Video Synopsis for Intelligent Traffic Monitoring Applications," in IEEE Transactions on Intelligent Transportation Systems, vol. 21, no. 8, pp. 3457-3468, Aug. 2020, doi: 10.1109/TITS.2019.2929618.

[5] M. Ding, Y. Ding, L. Wei, Y. Xu and Y. Cao, "Individual Surveillance Around Parked Aircraft at Nighttime: Thermal Infrared Vision-Based Human Action Recognition," in IEEE Transactions on Systems, Man, and Cybernetics: Systems, vol. 53, no. 2, pp. 1084-1094, Feb. 2023, doi: 10.1109/TSMC.2022.3192017.

[6] Y. Zhang, X. Zheng, W. Liang, S. Zhang and X. Yuan, "Visual Surveillance for Human Fall Detection in Healthcare IoT," in IEEE MultiMedia, vol. 29, no. 1, pp. 36-46, 1 Jan.-March 2022, doi: 10.1109/MMUL.2022.3155768.

[7] S. Cristina, V. Despotovic, R. Pérez-Rodríguez and S. Aleksic, "Audio- and Video-Based Human Activity Recognition Systems in Healthcare," in IEEE Access, vol. 12, pp. 8230-8245, 2024, doi: 10.1109/ACCESS.2024.3353138.

[8] Amarendra Reddy Panyala, M. Baskar, "Real-time GB pattern convolution neural network-based brain image classification", AIP Conf. Proc. 3075, 020056 (2024), https://doi.org/10.1063/5.0218626.

[9] B. Natarajan, R. Elakkiya, R. Bhuvaneswari, K. Saleem, D. Chaudhary and S. H. Samsudeen, "Creating Alert Messages Based on Wild Animal Activity Detection Using Hybrid Deep Neural Networks," in IEEE Access, vol. 11, pp. 67308-67321, 2023, doi: 10.1109/ACCESS.2023.3289586.

[10] S. Juraev, A. Ghimire, J. Alikhanov, V. Kakani and H. Kim, "Exploring Human Pose Estimation and the Usage of Synthetic Data for Elderly Fall Detection in Real-World Surveillance," in IEEE Access, vol. 10, pp. 94249-94261, 2022, doi: 10.1109/ACCESS.2022.3203174.

[11] W. Liu, G. Wei, Y. Wang and R. Wu, "Indoor Multipedestrian Multicamera Tracking Based on Fine Spatiotemporal Constraints," in IEEE Internet of Things Journal, vol. 10, no. 11, pp. 10012-10023, 1 June1, 2023, doi: 10.1109/JIOT.2023.3235148.

[12] S. Y. Nikouei, Y. Chen, A. J. Aved and E. Blasch, "I-ViSE: Interactive Video Surveillance as an Edge Service Using Unsupervised Feature Queries," in IEEE Internet of Things Journal, vol. 8, no. 21, pp. 16181-16190, 1 Nov.1, 2021, doi: 10.1109/JIOT.2020.3016825.

[13] M. M. E. Yurtsever and S. Eken, "BabyPose: Real-Time Decoding of Baby's Non-Verbal Communication Using 2D Video-Based Pose Estimation," in IEEE Sensors Journal, vol. 22, no. 14, pp. 13776-13784, 15 July15, 2022, doi: 10.1109/JSEN.2022.3183502.

[14] C. Liang, D. Liu, L. Qi and L. Guan, "Multi-Modal Human Action Recognition With Sub-Action Exploiting and Class-Privacy Preserved Collaborative Representation Learning," in IEEE Access, vol. 8, pp. 39920-39933, 2020, doi: 10.1109/ACCESS.2020.2976496.

[15] C. Gupta et al., "A Real-Time 3-Dimensional Object Detection Based Human Action Recognition Model," in IEEE Open Journal of the Computer Society, vol. 5, pp. 14-26, 2024, doi: 10.1109/OJCS.2023.3334528.

[16] K. Rezaee, M. R. Khosravi and M. S. Anari, "Deep-Transfer-Learning-Based Abnormal Behavior Recognition Using Internet of Drones for Crowded Scenes," in IEEE Internet of Things Magazine, vol. 5, no. 2, pp. 41-44, June 2022, doi: 10.1109/IOTM.001.2100138.

[17] P. Nguyen Huu, N. Nguyen Thi and T. P. Ngoc, "Proposing Posture Recognition System Combining MobilenetV2 and LSTM for Medical Surveillance," in IEEE Access, vol. 10, pp. 1839-1849, 2022, doi: 10.1109/ACCESS.2021.3138778.

[18] J. Khan et al., "SMSH: Secure Surveillance Mechanism on Smart Healthcare IoT System With Probabilistic Image Encryption," in IEEE Access, vol. 8, pp. 15747-15767, 2020, doi: 10.1109/ACCESS.2020.2966656.

[19] A. Manocha and M. Bhatia, "IoT-Inspired Monitoring Framework for Real-Time Stereotypic Movement Analysis," in IEEE Systems Journal,

vol. 16, no. 2, pp. 2788-2796, June 2022, doi: 10.1109/JSYST.2021.3084370.

[20] M. T. Bhatti, M. G. Khan, M. Aslam and M. J. Fiaz, "Weapon Detection in Real-Time CCTV Videos Using Deep Learning," in IEEE Access, vol. 9, pp. 34366-34382, 2021, doi: 10.1109/ACCESS.2021.3059170.

[21] Reddy Panyala, A., & Manickam, B. (2024). Generative adversarial network for Multimodal Contrastive Domain Sharing based on efficient invariant feature-centric growth analysis improved brain tumor classification. Electromagnetic Biology and Medicine, 1–15. https://doi.org/10.1080/15368378.2024.2375266.

[22] Y. Hao et al., "An End-to-End Human Abnormal Behavior Recognition Framework for Crowds With Mentally Disordered Individuals," in IEEE Journal of Biomedical and Health Informatics, vol. 26, no. 8, pp. 3618-3625, Aug. 2022, doi: 10.1109/JBHI.2021.3122463.

[23] A. -U. Rehman, H. S. Ullah, H. Farooq, M. S. Khan, T. Mahmood and H. O. A. Khan, "Multi-Modal Anomaly Detection by Using Audio and Visual Cues," in IEEE Access, vol. 9, pp. 30587-30603, 2021, doi: 10.1109/ACCESS.2021.3059519.

[24] R. Chatterjee, A. Chatterjee, M. R. Pradhan, B. Acharya and T. Choudhury, "A Deep Learning-Based Efficient Firearms Monitoring Technique for Building Secure Smart Cities," in IEEE Access, vol. 11, pp. 37515-37524, 2023, doi: 10.1109/ACCESS.2023.3266514.

[25] H. Wang, B. Dong, Q. Zhu, Z. Chen and Y. Chen, "Spatio-Temporal Attention Fusion SlowFast for Interrogation Violation Recognition," in IEEE Access, vol. 11, pp. 103801-103813, 2023, doi: 10.1109/ACCESS.2023.3316724.

[26] M. Waheed et al., "An LSTM-Based Approach for Understanding Human Interactions Using Hybrid Feature Descriptors Over Depth Sensors," in IEEE Access, vol. 9, pp. 167434-167446, 2021, doi: 10.1109/ACCESS.2021.3130613.

[27] M. Xu, L. Guo and H. -C. Wu, "Robust Abnormal Human-Posture Recognition Using OpenPose and Multiview Cross-Information," in IEEE Sensors Journal, vol. 23, no. 11, pp. 12370-12379, 1 June1, 2023, doi: 10.1109/JSEN.2023.3267300.

Enhancing Arabic Phishing Email Detection: A Hybrid Machine Learning Based on Genetic Algorithm Feature Selection

Amjad A. Alsuwaylimi
Department of Information Technology-Faculty of Computing and Information Technology,
Northern Border University, Rafha 91911, Saudi Arabia

***Abstract*—Recently, owing to widespread Internet use and technological breakthroughs, cyber-attacks have increased. One of the most common types of attacks is phishing, which is executed through email and is a leading cause of the recent surge in cyber-attacks. These attacks maliciously demand sensitive or private information from individuals and companies. Various methods have been employed to address this issue by classifying emails, such as feature-based classification and manual verification. However, these methods face significant challenges regarding computational efficiency and classification precision. This work presents a novel hybrid approach that combines machine learning and deep learning techniques to improve the identification of phishing emails containing Arabic content. A genetic algorithm is employed to optimize feature selection, thereby enhancing the performance of the model by effectively identifying the most relevant features. The novel dataset comprises 1,173 records categorized into two classes: phishing and legitimate. A number of empirical investigations were carried out to assess and contrast the performance outcomes of the proposed model. The findings reveal that the proposed hybrid model outperforms other machine learning classifiers and standalone deep learning models.**

Keywords*—Machine learning; phishing email; BiLSTM; Arabic content-based

I. INTRODUCTION

Phishing emails are powerful instruments for scammers who want to make money by taking advantage of their feelings and trust. Attackers can achieve this by pretending to be trustworthy organizations such as banks or government authorities, and creating a sense of urgency or panic in their victims [1-3]. Phishing attacks offer scammers the possibility of high returns on investment because they are inexpensive and easy to perform. These assaults exploit human susceptibilities, such as the need for self-preservation and curiosity, while avoiding security systems through emotional manipulation. Therefore, one must learn how phishing emails operate if they are to protect themselves from falling victims to these dishonest campaigns. Phishing emails come in all shapes and sizes but share the same end, outsmarting the target to perform their bidding. One type of phishing email is scam email, which disguises itself as a legitimate organization, often a bank or another well-known business, requesting the target to click on links or share personal information [4,5]. The other type is spear-phishing, which uses intimate and personal information about the recipient to appear more authentic and trustworthy. Furthermore, another type exists, which is called the clone phishing email, which copies a real email the victim has already received, and then sends it back directly with a malicious attachment or URL. Furthermore, cybercriminals imitate high-ranking officers through CEO fraud emails to deceive employees and send urgent payments or sensitive data to them. Meanwhile, phishing emails direct their victims to fake websites, where they attempt to acquire their financial details or log-in credentials. These kinds of deceitful electronic mail play on either trust, urgency, or curiosity to make unsuspecting individuals compromise their own safety.

Researchers have been investigating various techniques, such as Natural Language Processing (NLP) [6-8], Machine Learning (ML) [9-16] and Deep Learning (DL) [17-21] to deal with the significant challenge posed by phishing emails in the field of cybersecurity. NLP methods are used for analyzing the text of emails and detecting linguistic indications or patterns indicative of phishing attempts. ML algorithms can be trained with large datasets containing examples of phishing emails to identify the commonalities between them and subsequently automate the recognition and classification of new ones into predefined categories based on these observed regularities. DL, specifically Convolutional Neural Networks (CNNs) alongside Recurrent Neural Networks (RNNs), makes it possible to detect phishing emails more precisely by learning complex features from both email content and its metadata.

Despite these efforts, achieving time efficiency and accuracy with these techniques remains a significant challenge in the field. The processing of extensive features necessitates substantial memory and computational time, further complicating the development of effective email classification techniques. Furthermore, the proliferation of big data presents also significant challenges for these techniques, mainly due to increased training durations from the heightened computational demands of processing large datasets. Both the larger sample size and greater dimensionality of the data contribute to this issue. High dimensionality particularly affects inference times due to the added computational burden of feature extraction. In real-time phishing detection models, these factors can negatively impact user experience and compromise the effectiveness of deployed techniques. Consequently, optimizing these techniques to balance accuracy and computational efficiency is an ongoing area of research. To address these challenges, researchers and practitioners often use feature reduction or feature selection techniques. By

strategically selecting a subset of features, these methods reduce computational costs [22].

The existing literature describes several feature selection algorithms that are commonly used, such as tree-based methods [23], selectKBest [24], Recursive Feature Elimination [25], LASSO [26], Principal Component Analysis [27], and Evolutionary Algorithms [28]. Evolutionary algorithms utilize a population of candidate solutions that progressively adapt over time via number of operators: selection, mutation, and recombination mechanisms to identify optimal solutions. Their robustness, flexibility, ability address complex, non-differentiable, and non-linear problems with versatility and resilience, parallelization capabilities, and multi-objective optimization potential make them particularly advantageous for various optimization tasks, including feature selection.

Genetic Algorithms (GAs), a subset of EAs, emulate natural selection to solve optimization problems. GAs have demonstrated significant success in addressing diverse optimization challenges, including feature selection. GAs have been employed in feature selection tasks, whereby the features are represented as chromosomes, and various genetic operators, including selection, mutation, and crossover are applied to evolve candidate solutions. The efficacy of each proposed solution is evaluated using a predefined objective function, and the optimization procedure continues iteratively until a satisfactory feature subset is obtained.

In the context of our research, the primary aim is to enhance detection accuracy and recall while simultaneously reducing processing time by minimizing the feature set. This work introduces a new hybrid approach for email classification. Specifically, we propose a method that integrates machine learning (ML) and deep learning (DL) methodologies to detect and categorize content-based phishing emails in the Arabic language, utilizing a GA to identify and select the best features. Our research significantly advances the current state of knowledge in this domain through several key contributions. First, we develop a hybrid model combining Random Forest (RF) and Bidirectional Long Short-Term Memory (BiLSTM) techniques, enhancing the detection of Arabic-based phishing emails. Additionally, we create and introduce a novel dataset comprising 1,173 Arabic content-based emails, providing a valuable resource for future research in Arabic phishing email detection. Furthermore, we innovatively apply a genetic algorithm to optimize feature selection and reduce feature dimensionality, thereby improving the efficiency and accuracy of the model. Finally, we conduct a comprehensive evaluation of the impact of genetic algorithms on model performance, demonstrating their effectiveness in enhancing accuracy relative to ML classifiers and DL models.

The remainder of this paper is organized as follows: Section II reviews related studies on phishing detection techniques, emphasizing recent advancements and the integration of machine learning (ML) and deep learning (DL) methodologies. Section III provides a detailed formulation of the problem addressed in this research. Section IV outlines the methodology and materials used, describing the dataset, pre-processing techniques, and the proposed hybrid model combining Random Forest and BiLSTM with Genetic Algorithm Feature Selection. Section V presents the experimental results and discussion, comparing the performance of the proposed model with baseline machine learning classifiers and deep learning models. Finally, Section VI concludes the paper by summarizing the key findings and suggesting directions for future research in the field of phishing email detection.

II. Related Studies

A. Phishing Email Detection Approaches

The ongoing evolution and increasing prevalence of phishing attacks necessitate continued research in detection methodologies. As these threats become more diverse, studies examining detection methods are concurrently updated and enhanced to address emerging challenges. The advent of ML and DL techniques, having proven their effectiveness in various problem domains, has led to their adoption in phishing detection research [16]. Researchers in this field have increasingly employed these advanced computational methods due to their competence in detecting complex patterns and addressing emerging attack strategies. Current studies primarily concentrate on identifying phishing emails and websites, utilizing ML and DL methods to enhance detection accuracy and minimize false positives. This transition to more advanced analytical approaches reflects the field's adaptation to the growing sophistication of phishing attacks and highlights the need for ongoing innovation in cybersecurity defences.

The authors of [1] explored the effectiveness of a transformer model named Bidirectional Encoder Representations from Transformers (BERT) and word embeddings for spam email classification. The findings were compared with those of Deep Neural Network (B-DDN) model, which includes Naive Bayes classifiers, k-nearest Neighbors, the BiLSTM layer. The model was tested and trained using two public datasets. The BERT transformer model with English Wikipedia and BookCorpus was used as the training data, with F1-score of 98.66% and an accuracy of 98.67%. This work investigated spam email detection using contextual word embeddings, attention layers, and deep learning techniques.

The study by [11] explored the use of ML-based spam detection models. Various ML methods were employed to categorize SMS messages as legitimate or spam, including Naive Bayes (NB), Decision Trees (DT), Support Vector Machines (SVM), Convolutional Neural Networks (CNN), Random Forest (RF), and Artificial Neural Networks (ANN). The research utilized several datasets, such as the Spam SMS Dataset and UCI SMS database, along with a custom-crawled dataset. For real-world data, both English and Indonesian languages were considered. Application of these models to the datasets yielded promising accuracy rates: SVM achieved 97.4% precision, CNN demonstrated 99.19% accuracy, and Weka SVM exhibited 99.3% accuracy for spam classification. Preprocessing techniques, including tokenization, removal of stop words, and feature extraction, were identified as methods to enhance accuracy. Similarly, focusing on SMS spam detection, the authors in [17] introduced a Hybrid Convolutional Neural Network-Long Short-Term Memory (CNN-LSTM) approach for the detection of spam within SMS

messages in Arabic and English languages. This approach was compared to conventional machine learning algorithms, including K-Nearest Neighbors (KNN), Support Vector Machine (SVM) and DL models such as LSTM and CNN individually. The study utilized two datasets: the SMS Spam data collection obtained from the UCI Machine Learning Repository containing a collection of Arabic messages and English messages sourced from local mobile phones. This hybrid model was designed to handle mixed linguistic content in Arabic and English communications. The proposed approach achieved a notable accuracy of 98.37% in categorizing SMS messages as spam or legitimate, outperforming all other machine learning algorithms tested in the study.

The study of [12] introduced two novel algorithms, FMMPED and FMPED, specifically designed to enhance phishing email detection capabilities. These algorithms were developed utilizing undersampling techniques and ensemble learning methodologies. To simulate real-world email environments, the researchers employed a dataset with a 10:1 ratio of malicious to legitimate emails. The study, published in a highly technical journal focused on machine learning and cybersecurity, employs sophisticated terminology commensurate with its computer science orientation. The proposed algorithms demonstrated superior performance compared to traditional machine learning and deep learning approaches, accomplishing accuracy and an impressive F1-score of 0.9945. The authors of [16] employed 13 machine learning classifiers to differentiate between spam and non-spam emails, including Bayesian methods, Decision Trees, Random Forest, Support Vector Machines, Decision Tables, and Bagging. The evaluation methodology incorporated two datasets: Spam Corpus and Spambase. The Random Forest classifier demonstrated superior performance on the Spam Corpus dataset, surpassing all other classifiers implemented using the Python programming language, with an accuracy rate of 99.91%.

The study of [15] investigated the implementation of diverse machine learning approaches for spam email classification. The study utilized traditional spam detection methods, which include Support Vector Machine, Naïve Bayes, Random Forest, and K-Nearest Neighbour models. The research incorporated a comprehensive collection of email datasets and real-life scenarios of varying sizes and formats from multiple sources, such as Kaggle and Sklearn. The study emphasized document pre-processing, encompassing cleaning, integration, transformation, and reduction. Additionally, tokenization and stop word removal were considered. The problem statement effectively elucidated the significance of this research.

In the study by [29], they utilized email samples as the dataset and employed federated learning approaches with THEMIS and BERT models for phishing email detection. Due to architectural constraints, the BERT model focused specifically on the email body. Both training and evaluation of the models were conducted in English. THEMIS achieved a test accuracy of 97.9% for federated learning with five clients at epoch 45, while BERT attained a test accuracy of 96.1% for federated learning with five clients at epoch 15. It is noteworthy that accuracy rates varied depending on the client. The authors in [18] introduced a new Intelligent Cybersecurity Phishing Email Detection and Classification (ICSOA-DLPEC) model that leverages n-gram feature extraction, a Gated Recurrent Unit (GRU) model, and a Compressive Sensing (CS) algorithm for optimal parameter tuning. They evaluated its performance using a standard dataset, achieving impressive accuracy rates between 98.46% and 99.72% across different training and testing data volumes. This study also discusses online safety terms and deep learning concepts. When compared to other models such as LSTM, CNN, and THEMIS, the ICSOA-DLPEC model demonstrated superior performance in being correct, precise, able to recall, and in its F-score.

Tong et al. 2021 [30], proposed a capsule network model with long-short attention for Chinese spam detection. The authors used the Trec06c dataset and received an accuracy of 98.72% on an unbalanced dataset and a 99.30% accuracy on a balanced dataset. Similarly, Li et al. [31], introduced an LSTM based phishing email detection method and tested it on a dataset of 29,942,735 emails from an enterprise mail server with both Chinese and English content. The model got nearly 100% accuracy in classifying phishing emails. Wu et al. [32], evaluated ChatGPT's spam detection against baseline models like SVM, LR, NB and BERT on the English Email Spam Detection (ESD) dataset and the low-resourced Chinese Spam Dataset (CSD) and the results were achieved in accuracy was 83% and 86% accuracy on two different experiments.

B. Feature Selection Methods

The process of feature selection is a critical component in the fields of ML and data analytics. It involves identifying and extracting the most salient subset of attributes from the complete set of features present within a given dataset. The primary goals of feature selection are to reduce dimensionality, improve model performance, enhance generalization, and provide better interpretability of results [33]. In the context of phishing email detection, the feature selection process involves the identification and selection of the most informative characteristics of emails that effectively differentiate between legitimate and phishing messages. This process is essential for developing accurate and efficient phishing detection systems that can adapt to evolving threats. The selected features must be relevant to phishing indicators, adaptable to new techniques, and effective across different languages and cultural contexts, especially in multilingual environments [34-36].

Common types of features in phishing email detection include linguistic features, such as writing style and urgency indicators; structural features, like email header information and HTML content; contextual features, including sender reputation and domain age; and behavioral features, such as user interaction patterns and email sending times [37-40]. The selection process must strike a balance between content-based and metadata-based features to create a comprehensive and robust feature set.

Various methods are employed for feature selection in phishing detection. These include filter methods, which use statistical approaches to select features according to their relationship with the target variable; wrapper methods, which evaluate feature subsets using specific machine learning

models; embedded methods, which integrate feature selection into the model training process; and evolutionary algorithms, such as Genetic Algorithms, which optimize feature subsets based on multiple objectives [41,42].

Recent advancements in deep learning introduced methods that have the ability to automatically derive pertinent features directly from raw email data, sometimes mitigating the requirement for manual feature selection [43]. However, in many applications, especially those dealing with multilingual content or requiring model interpretability, careful feature selection remains a critical step in developing effective phishing detection systems. By focusing on the most discriminative features, researchers and cybersecurity professionals can develop more accurate, efficient, and adaptable phishing detection models. This approach not only enhances overall email security but also helps protect users from increasingly sophisticated phishing attacks. The ongoing challenge in this field is to continuously refine feature selection methods to keep pace with evolving phishing techniques and maintain robust detection capabilities across diverse linguistic and cultural contexts.

C. Genetic Algorithm-Based Feature Selection Approaches for Phishing Email Detection

Genetic Algorithms (GAs) are evolutionary optimization techniques that draw inspiration from the mechanisms of natural selection and genetics. GAs have become powerful tools for solving complex optimization problems across various domains [44]. In the context of ML and DL, GAs have shown remarkable efficacy in feature selection tasks, offering a robust approach to identifying optimal subsets of features from large and complex datasets [45]. The core premise of genetic algorithms (GAs) is their capacity to iteratively refine a population of candidate solutions over successive generations. Each solution, or ``chromosome" denotes a possible subset of features. The algorithm evaluates these chromosomes based on a fitness function, which typically measures the performance of a machine learning model using the selected features. Through processes mimicking genetic inheritance – selection, crossover, and mutation – GAs iteratively refine the population, converging towards an optimal or near-optimal feature subset [46].

In the domain of phishing email detection, the application of GAs for feature selection presents a promising approach to enhancing detection accuracy while optimizing computational efficiency. Phishing emails often contain subtle and evolving characteristics, making the selection of relevant features a critical and challenging task. Given these challenges, GAs can effectively navigate this complex feature space, considering various combinations of linguistic, structural, and behavioral email attributes to identify the most discriminative feature set [47]. The use of GAs in phishing email detection typically involves encoding email features as binary strings, where each element indicates the presence or absence of a specific feature. Moreover, the fitness function may incorporate multiple objectives, such as maximizing detection accuracy, minimizing false positives, and reducing the overall number of features used. Consequently, this multi-objective optimization approach allows for a balanced solution that meets the often-conflicting goals of high accuracy and computational efficiency [48].

Several studies have demonstrated the effectiveness of GA-based feature selection in phishing detection systems. For instance, the authors in [49] employed a GA to optimize feature selection for their email classification system, resulting in improved accuracy and reduced computational complexity. Similarly, the study in [50] utilized a GA-based approach to identify the most pertinent features for their phishing website detection model, achieving high accuracy rates with a compact feature set.

The adaptability of GAs makes them particularly suitable for the dynamic nature of phishing attacks. As cybercriminals continually evolve their techniques, GA-based feature selection can be periodically re-run on updated datasets, ensuring that the selected features remain relevant and effective against new phishing strategies [51]. This adaptability is crucial in maintaining the long-term effectiveness of phishing detection systems. Moreover, the interpretability of GA-selected feature subsets can provide valuable insights into the most significant indicators of phishing attempts. This transparency can aid security professionals in understanding evolving phishing tactics and developing more targeted prevention strategies [52]. Therefore, the application of Genetic Algorithms to feature selection in phishing email detection offers a powerful means of enhancing detection accuracy, improving computational efficiency, and adapting to evolving threats. As phishing attacks continue to grow in sophistication, the role of advanced feature selection techniques like GAs becomes increasingly critical in developing robust and effective defence mechanisms.

III. Problem Formulation

This study involves a problem-formation process for binary text classification. The problem can be framed as classifying emails into two distinct classes: phishing emails and legitimate emails. Let $\{D\}$ be a collection of emails $\{E\}$, known as a dataset. Let $\{D\}$ consist of E_{phishing} and $E_{\text{legitimate}}$. Feature matrices to be used as inputs for the models are derived from *D*, where rows represent emails (content-based) and columns represent features. Let X be the input feature and Y be the target variable, which can be represented as $D = \{X_1, Y_1, X_2, Y_2, X_3, Y_3 \dots \dots \dots \dots X_n, Y_n\}$, where n denotes the total number of words in V, X_1 is the feature vector (V), and Y_1 is the label.

The D model is divided into two parts, Dtraining and Dtesting. The Dtraining is used for training the model, whereas the Dtesting is used to assess and test it. In both cases, the learn a function F(X), which makes the decision in to detect emails whether the input E is legitimate or phishing, as shown in the mathematical formula (1).

$$F(X) = f(X, Pa) \tag{1}$$

where X is the input features and the Pa is the parameter of the model.

IV. METHODOLOGY AND MATERIALS

This section details the five-phase methodological framework employed in this study. Each phase contributes to the development and evaluation of the machine learning model for email classification. A visual representation of these stages is provided in Fig. 1.

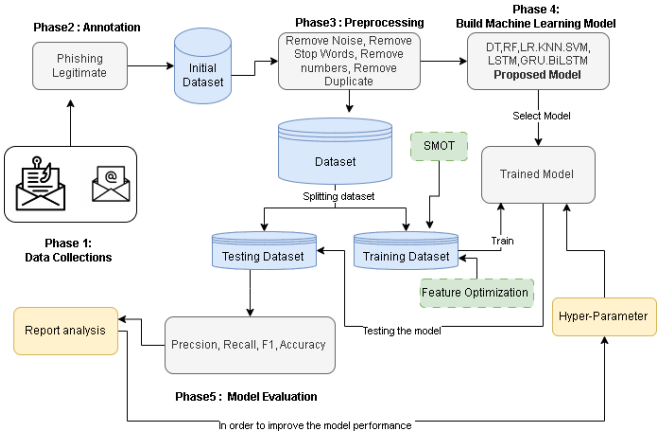


Fig. 1. Methods and Phases of this study.

A. Phase 1: Data Collection

Due to the scarcity of Arabic datasets for phishing emails, this study employed an initial method to construct the study's dataset. Data was gathered from a variety of personal email collections, ensuring a diverse range of sources. To further enrich the dataset, phishing emails were generated using ChatGPT, complemented by additional emails crafted manually. This compilation process ensured a diverse and representative dataset. Upon finalizing the collection, all emails were organized and stored in a CSV file, which was subsequently sent to annotators for detailed processing.

B. Phase 2: Annotation Process

In the annotation process, three native Arabic speakers assisted in the annotation process by classifying the emails into two categories: phishing and legitimate. In addition, Cohen's Kappa was utilized in order to measure the degree of agreement between the annotators [53,54] based on the mathematical formula (2).

$$K = \frac{P_o - P_e}{1 - P_e} \quad (2)$$

P_o is the proportion agreement between judges, while P_e is the expected agreement proportion by chance. Thus, the K is 81% which indicates prefect agreement between the judges (raters). In the end, Table I provides information on the size of the final dataset.

TABLE I. DATASET DESCRIPTION

Factors	No. of Emails	Maximum Words	Minimum Words
Phishing	861	166	23
Legitimate	312	178	46
Total	**1,173**		

C. Phase 3: Pre-Processing Steps

Pre-processing phase were performed on the Arabic dataset before it was used in the models. These steps included the removal of stop words from the Arabic language, noise such as HTML tags, and duplicate messages or emails.

All Arabic-specific stop words were systematically eliminated to enhance the efficacy of subsequent text analysis procedures. Stop words, defined as high-frequency lexical items that typically carry minimal semantic content, are routinely excluded to optimize the quality and relevance of the dataset. Table II presents a representative sample of common Arabic-specific stop words. Additionally, for noise removal, all irrelevant text, including HTML tags, was removed from the emails. This step ensured that only meaningful textual information was retained, making the data more suitable for analysis. Furthermore, duplicates were identified and removed to ensure that each email in the dataset was unique, thereby preventing redundancy and improving the accuracy of the analysis.

TABLE II. COMMON EXAMPLES OF ARABIC-SPECIFIC STOP WORDS

Arabic-specific stop words	Meaning in English
و	and
في	in
من	from
على	on
إلى	to
أو	Or

D. Phase 4: Build Machine Learning Model

Prior to detailing the process of this phase, we provide a brief overview of the ML classifiers and DL models and GA-based feature selection employed in this investigation to assess the efficacy of our proposed approach.

1) Machine learning classifiers: Support Vector Machine (SVM): SVM is a supervised learning algorithm that analyze data to perform regression and classification tasks. The algorithm operates by identifying the optimal hyperplane that best separates the given dataset into distinct classes. SVMs are particularly adept in high-dimensional feature spaces and have been widely employed in text categorization tasks including sentiment analysis and spam detection [55,56].

a) Decision Tree (DT): DT is a non-parametric supervised learning algorithm used for regression and classification. DT works by splitting the data into a number of subsets according to the most significant differentiators in the input features. Decision Trees are highly interpretable and easily understood, making them indispensable tools in decision-making processes [57].

b) Logistic Regression (LR): LR is a statistical technique that employs a logistic function to analyze a binary dependent variable in a model. LR is a widely adopted technique for binary classification problems, including applications such as email spam detection and medical diagnosis [58-60].

c) Random Forest (RF): RF is a widely utilized machine learning algorithm that generates numerous decision trees during the training process and predicts the class that is the statistical mode of the classes output by the individual trees. RF helps overcome the tendency of individual decision trees to overfit to the training data, resulting in a robust and accurate model [61,62].

2) Deep learning models:

a) Long Short-Term Memory (LSTM): LSTM is a specialized form of Recurrent Neural Network (RNN) that can effectively model and capture long-term temporal dependencies within sequential data. It addresses the vanishing gradient issue that conventional RNNs face by leveraging memory cells to retain information over extended durations. LSTMs have demonstrated effectiveness in tasks involving sequential data, including time-series forecasting and natural language processing [63].

b) Gated Recurrent Units (GRU): GRU is another variety of RNN which is similar to LSTM but with a simplified architecture. GRU integrates the forget and input gates into a single update gate, resulting in improved computational efficiency without compromising its overall performance. GRUs are used in various applications requiring sequence modeling [64].

c) Bidirectional Long Short-Term Memory (BiLSTM): BiLSTM is a bidirectional variant of the LSTM architecture, which enhances performance by analyzing input sequences in both the backward and forward temporal directions. BiLSTM enables the model to have access to future and past context information, making it particularly useful in tasks like machine translation and text generation [65,66].

3) Genetic algorithm for feature selection: Feature selection tasks for content-based email message analysis have proven to be highly effective when using GAs [67, 68], which are optimization techniques. As part of detecting phishing emails, feature selection involves recognizing the most relevant attributes present within the email content, including textual patterns, keywords, and structural characteristics. A genetic algorithm then generates a population of potential solutions and iteratively evolves them through processes such as selection, crossover, and mutation.

The GA efficiently narrows down the feature set to those features that contribute most significantly to distinguishing phishing emails from legitimate ones, by evaluating the fitness of each solution. This process enhances the efficiency of the model by reducing dimensionality and eliminating irrelevant or redundant features, leading to improved accuracy and faster computation. In the proposed model, the GA for feature selection played an important role in optimizing the input features, thus enhancing the overall performance and robustness of the Random Forest and BiLSTM hybrid approach.

To start the GA for detecting normal and phishing emails, a dataset containing both types of emails is loaded, and relevant features are extracted. An initial population of candidate individuals is generated, with each individual denoting a set of features. The classification accuracy is evaluated based on the fitness of these chromosomes. The algorithm selects parents, performs mutations and crossovers to create offspring, and evaluates their fitness, replacing the old population with new offspring until a satisfactory fitness level is achieved or a specified number of generations is reached. Ultimately, the GA outputs the highest-performing chromosome, representing the optimal feature set for email classification, as depicted in Fig. 2.

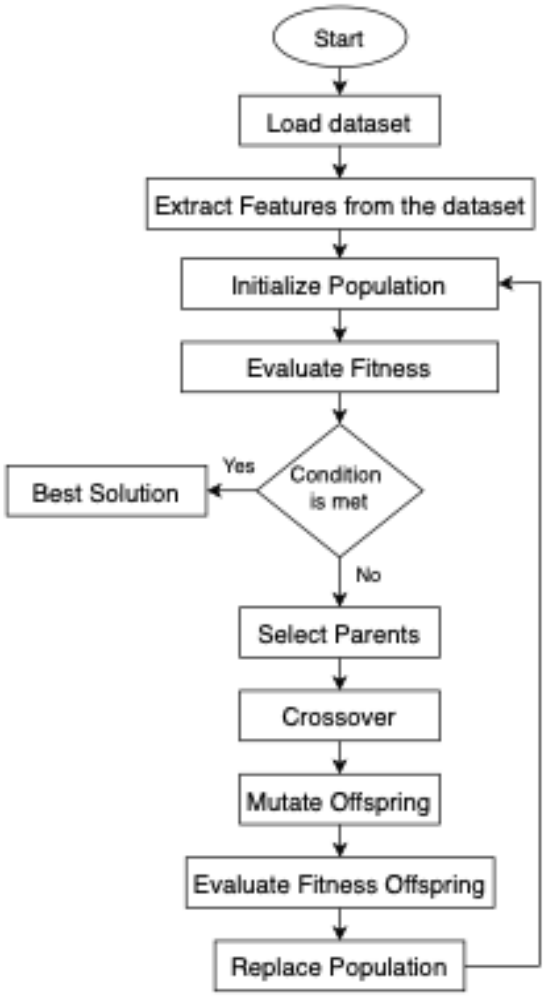


Fig. 2. Genetic algorithm flowchart.

4) The process of phase 4: In phase 4, the proposed model for Arabic detecting phishing emails is designed with a multi-stage process, leveraging both machine learning and deep learning techniques optimized through feature selection using a genetic algorithm, as shown in Fig. 3. The proposed hybrid approach provides a number of significant benefits. It integrates RF and BiLSTM with a GA for feature optimization. The model takes advantage of the advantages of both methods by combining the outputs of RF and BiLSTM, improving prediction performance by capturing a variety of data features. Sequential data is best captured by BiLSTM, unlike RF's ensemble approach reduces variance and minimizes overfitting to produce robust predictions. By

choosing the most pertinent features, the addition of GA for feature optimization further improves the model's efficiency and reduces its dimensionality, resulting in quicker training durations. By combining the advantages of both sequential pattern expertise from BiLSTM and structured data handling from RF, this hybrid technique also enhances generalization. The model performs well against noise, and is suitable for various data types. Its scalability and ability to handle large datasets make it a valuable tool in machine learning.

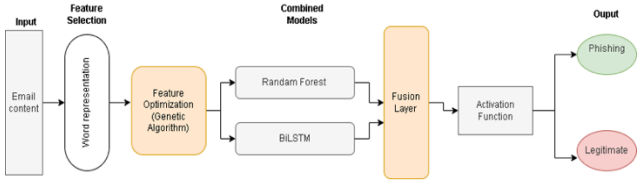


Fig. 3. Flowchart of the proposed model.

Initially, the process begins with the input of email content, which serves as the raw data for analysis. This content undergoes transformation into a word representation format, capturing the essential textual features needed for further processing. In the feature selection stage, a GA is employed to optimize the features derived from the word representation. This optimization step is crucial as it enhances the model's capacity to recognize relevant features that are indicative of phishing attempts.

The process begins with the input of email content, which is transformed into numerical vectors through word representation. This transformation allows the textual data to be processed effectively by subsequent machine learning algorithms. In the next stage, feature selection is carried out using a genetic algorithm. This step is crucial as it identifies and selects the most relevant features from the word representations, enhancing the model's capability to distinguish between phishing and legitimate emails.

Following feature optimization, the refined features are fed into two different machine learning models: BiLSTMb and RF. The RF model brings robustness and the capability to handle a large number of features effectively, while the BiLSTM model leverages its capacity to capture sequential dependencies within the email content, making it well-suited for analyzing the context and order of words [69].

The outputs from both the RF and BiLSTM models are then integrated in a fusion layer. This layer combines the strengths of both models, creating a more comprehensive understanding of the email content. The combined data is subsequently processed through an activation function, which refines the decision-making mechanisms employed by the model.

Finally, the model produces its output by classifying the email as either phishing or legitimate. This classification is the culmination of the entire process, providing a clear and actionable result based on the sophisticated analysis of the email content. Through this structured approach, the model effectively identifies phishing emails, ensuring robust detection by integrating feature optimization, machine learning, and deep learning methodologies.

E. Model Evaluation

This section introduces the most widely-used metrics to evaluate the performance of the proposed model. The metrics employed include Precision, Recall, F1-score, Accuracy, and Area Under the Receiver Operating Characteristic (AUC-ROC) curve.

Precision evaluates the accuracy of positive classifications, specifically the proportion of emails categorized as phishing that were correctly identified. This metric is crucial for evaluating the model's capacity to minimize the occurrence of false positive results. Mathematically, Precision is defined in the mathematical formula (3):

$$Precision = \frac{\text{Correct emails retrieved}}{All\ retrieved\ emails} \quad (3)$$

In this formula, “Correct emails retrieved” represents the number of emails accurately identified as phishing, and “All retrieved emails” represents the total number of emails classified as phishing by the model. The model's high level of precision is evidenced by a diminished rate of false positive outcomes, thus ensuring that most of the emails flagged as phishing are indeed malicious.

Recall metric assesses the model's capability to detect all pertinent instances of phishing emails. It represents the proportion of correctly identified phishing emails relative to the total number of actual phishing emails. A high recall value indicates that the model effectively identifies most phishing emails, thereby minimizing the occurrence of false negatives. The mathematical formula (4) is for calculating Recall:

$$Recall = \frac{\text{Correct emails retrieved}}{\text{All relvent emails}} \quad (4)$$

The F1-score is a composite metric that synthesizes the trade-off between precision and recall. It represents the harmonic mean of these two measures, offering a holistic evaluation that is particularly valuable in the context of unbalanced datasets. A high F1-score indicates that the model effectively balances the identification of phishing emails while minimizing both false negative and false positive classifications. The mathematical formula (5) calculates the F1-score:

$$F1 - Score = 2 \text{ x } \frac{Recall \text{ x } Precision}{Recall + Precision} \quad (5)$$

This metric is highly useful for comprehensively evaluating the model's performance in scenarios where both incorrectly identifying positive cases and failing to detect true positive cases are crucial considerations.

Accuracy metric evaluates the overall correctness of the model by quantifying the ratio of correctly classified emails (both phishing and legitimate) to the total number of emails. Accuracy is calculated by the mathematical formula (6):

$$Accuracy = \frac{Number\ of\ correct\ emails}{Total\ number\ of\ predication} \quad (6)$$

In this context, “Number of correct emails” includes both correctly identified phishing and legitimate emails, while “Total number of predictions” is the total number of emails that have been classified by the model. High accuracy denotes

that the model performs well in classifying both phishing and legitimate emails correctly.

In addition to the above metrics, the Area Under the Receiver Operating Characteristic (AUC-ROC) is employed to measure the model performance. The AUC-ROC provides a single value that represents the capability of the model to differentiate between classes across various threshold settings. The ROC curve graphically depicts the balance between the true positive rate and the false positive rate across varying decision criteria. A higher AUC value signifies improved model performance, with an AUC of 1.0 denoting a model with perfect classification performance.

V. Results and Discussion

This section details the experimental settings, findings, and analysis. The proposed model is evaluated in comparison to the baseline ML classifiers and DL models to distinguish between phishing and authentic emails for future identification.

A. Experimental Settings

The experiments were conducted using Google Colaboratory, which has been utilized for all the experiments on ML classifiers, DL models, and the proposed model. We used three Python libraries to conduct the experiments: Matplotlib, Scikit-learn (sklearn), and DEAP. Using Matplotlib, we created plots and charts to visualize the data and results. Sklearn was used for ML classifiers, pre-processing steps, and splitting the dataset. Finally, DEAP was used for the genetic algorithm. The hyperparameters for the machine learning (ML) classification algorithm are enumerated in Table III, while Table IV delineates the architectural and training parameters employed in the deep learning (DL) model.

TABLE III. Hyper-Parameter of ML Models

Model	Hyper-parameter	Default Value
Random Forest	n_estimators	100
	criterion	“gini”
	max_depth	None
	min_samples_split	2
	min_samples_leaf	1
Decision Tree	criterion	"gini"
	splitter	"best"
	max_depth	None
	min_samples_split	2
	min_samples_leaf	1
SVM	C	1.0
	kernel	"rbf"
	gamma	"scale"
Naive Bayes	priors	None
Logistic Regression	solver	"lbfgs"
	max_iter	100

TABLE IV. Hyper-Parameter of DL Models

Hyper-parameter	Value
Embedding Dimension	32
GR/LSTM Units	32
Batch Size	32
Sequence Length	100
Optimizer	"Adam"
Loss Function	"binary_crossentropy"
Metrics	["accuracy"]
Number of Epochs	30

B. Experimental Results

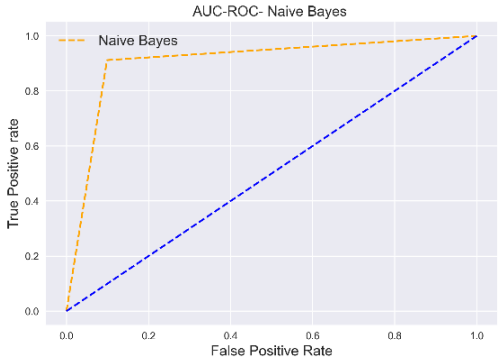
Numerous experiments have been conducted using ML, DL, and the proposed model. In these experiments, genetic algorithms are used to select the best features. Tables V and VI present a comparison of precision, recall, F1-score, and accuracy, respectively. Table V presents a comparative analysis of the experimental results of the ML classifiers without using genetic algorithms. The performance of various ML classifiers is used to detect phishing and legitimate emails with a focus on their accuracy. Among the classifiers evaluated, the Naive Bayes (NB) model demonstrated the most robust performance, attaining the highest accuracy rate of 90.91%, which underscores its substantial reliability and effectiveness. RF also performed exceptionally well, with an accuracy of 90.06%, making it a strong contender for detecting phishing emails. DT, LR, and SVM showed moderate accuracy, with values of 85.51%, 83.81%, and 84.38%, respectively, indicating that they are fairly accurate, but not as high-performing as NB and RF. In contrast, KNN showed the lowest accuracy among the classifiers, with an accuracy of 67.90%, indicating that it is less effective for this detection task. In summary, NB and RF are the top-performing models, whereas KNN has the lowest accuracy.

TABLE V. Experimental Results of ML Classifiers

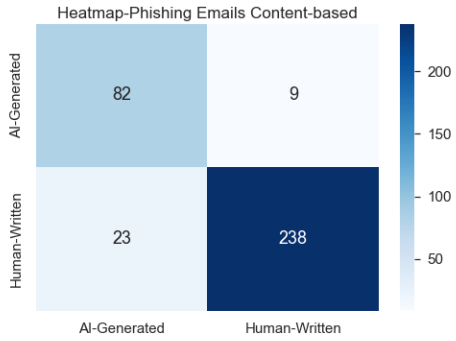
Classifier	Precision	Recall	F1-score	Accuracy
DT	91.83%	71.98%	76.08%	85.51%
KNN	65.35%	69.76%	64.76%	67.90%
LR	88.61%	69.40%	72.95%	83.81%
SVM	88.04%	70.85%	74.51%	84.38%
RF	92.68%	81.48%	85.25%	90.06%
NB	87.23%	90.65%	88.69%	**90.91%**

Fig. 4 shows the AUC-ROC and confusion matrix for the NB classifier, which achieved the highest accuracy. The figure illustrates the performance of a Naive Bayes classifier in detecting phishing emails based on content, consisting of a confusion matrix and an AUC-ROC curve. The AUC-ROC curve (a) shows the classifier's true positive rate (y-axis) against the false positive rate (x-axis), with the Naive Bayes model (orange dashed line) performing significantly better than random guessing (blue dashed line), indicating high sensitivity and specificity. The confusion matrix heatmap (b) highlights

classification results, with 82 AI-generated and 238 human-written emails correctly identified, while 9 AI-generated and 23 human-written emails were misclassified. The color intensity reflects the number of instances, demonstrating the classifier's overall accuracy and effectiveness.



(a)



(b)

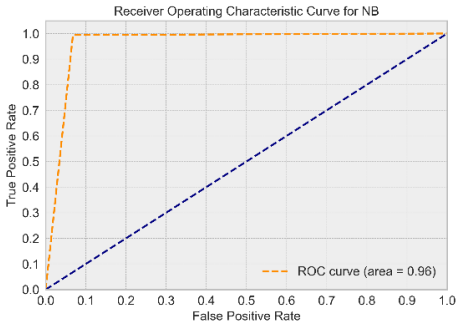
Fig. 4. AUC-ROC and confusion matrix of naive bayes. (a) AUC-ROC curve. (b) Confusion matrix showing correct (82 AI, 238 human) and incorrect (9 AI, 23 human) classifications.

Table VI summarizes the performance metrics for various classifiers used to detect phishing and legitimate emails, this time utilizing genetic algorithms for optimization. Notably, NB emerged as the top performer, with an accuracy of 97.87%, significantly improving from its previous accuracy of 90.91%. LR also saw a substantial increase in performance, achieving an accuracy of 94.47% compared to its earlier 83.81%. RF recorded an improved accuracy of 92.77%, up from 90.06%, thus maintaining its position as a strong classifier. DT demonstrated a marked improvement, with its accuracy increasing from 85.51% to 90.64%, indicating that genetic algorithms significantly enhanced its performance. In contrast, SVM showed a marginal improvement, with its accuracy slightly increasing from 84.38% to 85.53%. KNN improved its accuracy from 67.90% to 74.89%, achieving perfect recall; however, it remained less effective overall compared to other classifiers. In summary, the application of genetic algorithms for optimization led to performance improvements across most classifiers, particularly boosting the accuracy of NB and LR. Despite these enhancements, NB remains the top-performing model, while KNN still lags behind the others.

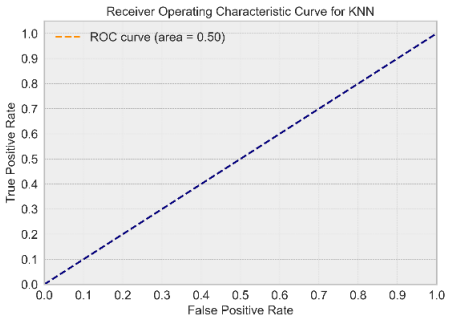
TABLE VI. EXPERIMENTAL RESULTS OF ML CLASSIFIERS USING GENETIC ALGORITHM

Classifier	Precision	Recall	F1-score	Accuracy
DT	87.55%	87.55%	87.55%	90.64%
KNN	74.89%	100.00%	85.64%	74.89%
LR	96.56%	88.98%	92.03%	94.47%
SVM	91.90%	71.19%	75.36%	85.53%
RF	94.68%	86.16%	89.43%	92.77%
NB	97.99%	96.33%	97.12%	**97.87%**

Fig. 5 shows the AUC-ROC for the best ML classifier and worst ML classifiers NB and KNN. The figure compares the ROC curves for two classifiers: NB and KNN. In subplot (a), the ROC curve for NB shows a high area under the curve (AUC) of 0.96, demonstrating outstanding performance with a high rate of correctly identified positives and a low rate of incorrectly identified positives. In contrast, subplot (b) displays the ROC curve for KNN, which has an AUC of 0.50, signifying performance equivalent to random guessing, as indicated by the diagonal line. This comparison highlights the superior effectiveness of the Naive Bayes classifier over the KNN classifier in this context.



(a)

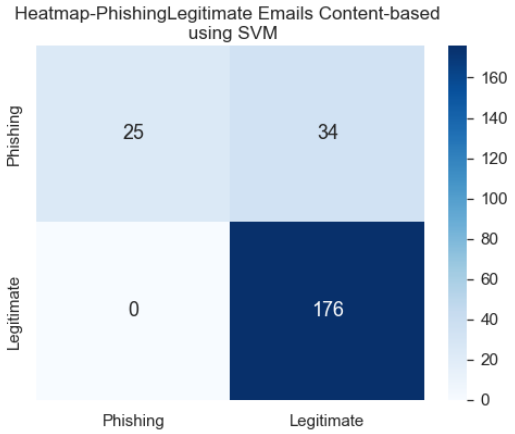


(b)

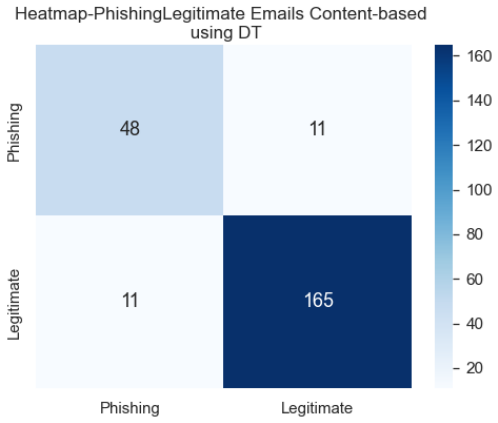
Fig. 5. AUC-ROC Curves for naive bayes and k-nearest neighbors; (a) Naive bayes (b) K-nearest neighbors.

Fig. 6 illustrates the confusion matrices for two machine learning classifiers, SVM and DT, applied to the classification of phishing and legitimate emails. Panel (a) shows the performance of the SVM classifier, which correctly identified 25 phishing emails and 176 legitimate emails. However, it misclassified 34 legitimate emails as phishing, with no false negatives (phishing emails classified as legitimate). Panel (b)

displays the results for the DT classifier, which correctly identified 48 phishing emails and 165 legitimate emails, with 11 false positives (legitimate emails classified as phishing) and 11 false negatives. The color intensity in both matrices represents the number of instances, providing a visual comparison of the classification accuracy and error distribution between the SVM and DT models. This analysis highlights the strengths and weaknesses of each classifier in distinguishing between phishing and legitimate emails.



(a)



(b)

Fig. 6. Confusion matrices (a) SVM, (b) DT.

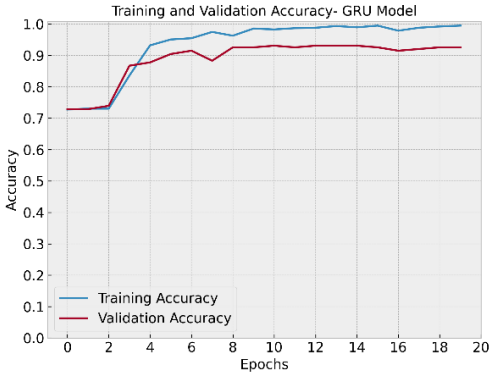
In the DL experiments, various models were evaluated based on precision, recall, F1-score, and accuracy in detecting phishing and legitimate emails. The models tested included LSTM, BiLSTM, GRU, and the proposed model utilizing Genetic Algorithm Feature Selection, as detailed in Table VII. The LSTM model achieved a precision of 94.97%, recall of 96.59%, F1-score of 95.77%, and accuracy of 93.62%. The BiLSTM model recorded a precision of 93.76%, recall of 88.98%, F1-score of 91.04%, and accuracy of 93.62%. The GRU model demonstrated an accuracy of 95.12%, with a precision of 95.28%, recall of 95.32%, and F1-score of 95.28%.

TABLE VII. EXPERIMENTAL RESULTS FOR THE DL MODELS AND PROPOSED MODEL

Classifier	Precision	Recall	F1-score	Accuracy
LSTM	94.97%	96.59%	95.77%	93.62%
BiSLTM	93.76%	88.98%	91.04%	93.62%
GRU	95.28%	95.32%	95.28%	95.12%
Proposed Model	96.77%	100.00%	98.36%	**97.90%**

In comparison, the proposed model outperformed the ML classifiers and DL models in terms of the common matrix and achieved a precision of 96.77%, recall of 100.00%, F1-score of 98.36%, and accuracy of 97.90% using Genetic Algorithm Feature Selection. Thus, the proposed model can select the most relevant features, thereby increasing its efficiency and accuracy. By combining RF with BiLSTM, we leveraged the strengths of both algorithms: RF's ability to handle high-dimensional data and reduce overfitting through ensemble learning, and BiLSTM's ability to capture sequential dependencies. This combination results in a robust model that can accurately detect phishing emails. In addition to the superior performance of the proposed model, it has also been demonstrated that when RF and BiLSTM are combined with Genetic Algorithm Feature Selection in order to enhance email security, this combination results in a superior performance that is reflected in all metrics evaluated.

Fig. 7 and 8 present the training and validation accuracy and loss functions for the GRU, BiLSTM, and proposed BiLSTM-RF models with Genetic Algorithm Feature Selection over 20 epochs. Fig. 7(a) and 7(c) show that both the GRU and BiLSTM models achieve high training and validation accuracy, with the GRU showing steady improvement and the BiLSTM demonstrating rapid early gains. Fig. 7(b) and 7(d) illustrate the loss functions for these models, indicating good generalization with minimal divergence between training and validation losses. In comparison, Fig. 8 highlights the proposed BiLSTM-RF model's performance, showing superior results: Fig. 8(a) demonstrates that it quickly reaches near-perfect accuracy for both training and validation, while Fig. 8(b) shows a sharp decline and stabilization in loss values, indicating efficient learning and minimal overfitting. Overall, the proposed model outperforms the GRU and BiLSTM models in accuracy and robustness, making it the most effective in detecting phishing and legitimate emails.



(a)

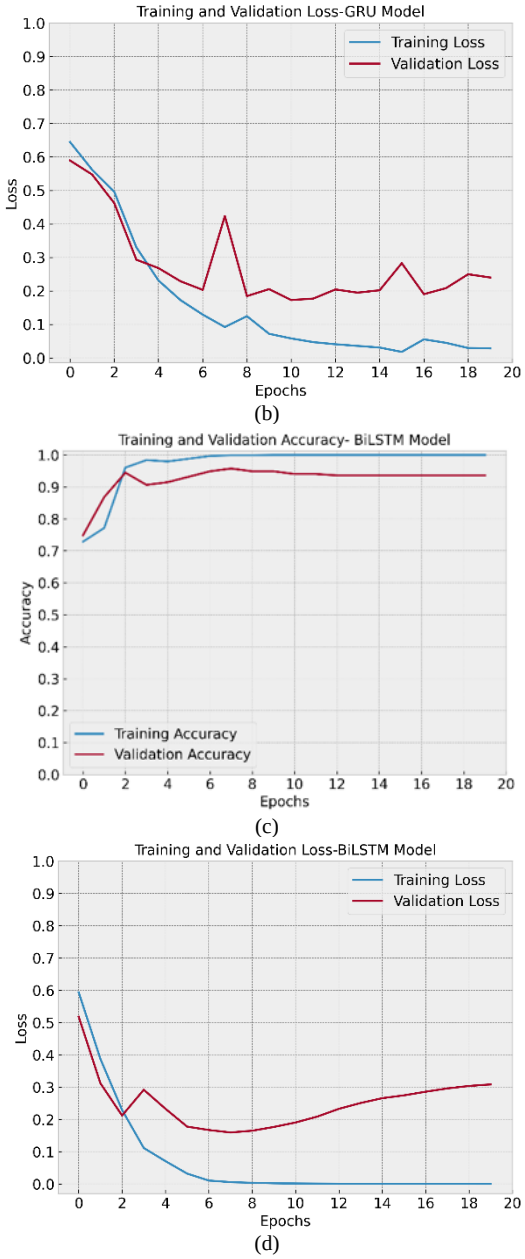


Fig. 7. Training accuracy, validation, and loss function of GRU and BiLSM.

In terms of precision, recall, and F1-score, the LSTM model achieved 94.97% precision, 96.59% recall, and 93.77% accuracy. A precision of 93.76%, recall of 88.98%, F1-score of 91.04%, and accuracy of 93.62% were recorded by the BiLSTM model. The GRU model demonstrated an accuracy of 95.12%, precision of 95.28 percent, recall of 95.32%, and F1-score of 95.28 percent.

The proposed model outperformed the ML classifiers and DL models in terms of the common matrix and achieved a precision of 96.77%, recall of 100.00%, F1-score of 98.36%, and accuracy of 97.90% using Genetic Algorithm Feature Selection. Thus, the proposed model can select the most relevant features, thereby increasing its efficiency and accuracy. By combining RF with BiLSTM, we leveraged the strengths of both algorithms: RF's ability to handle high-dimensional data and reduce overfitting through ensemble learning, and BiLSTM's ability to capture sequential dependencies. This combination results in a robust model that can accurately detect phishing emails. In addition to the superior performance of the proposed model, it has also been demonstrated that when RF and BiLSTM are combined with Genetic Algorithm Feature Selection in order to enhance email security, this combination results in a superior performance that is reflected in all metrics evaluated. Fig. 7 shows the training accuracy, validation, and loss function of the proposed model.

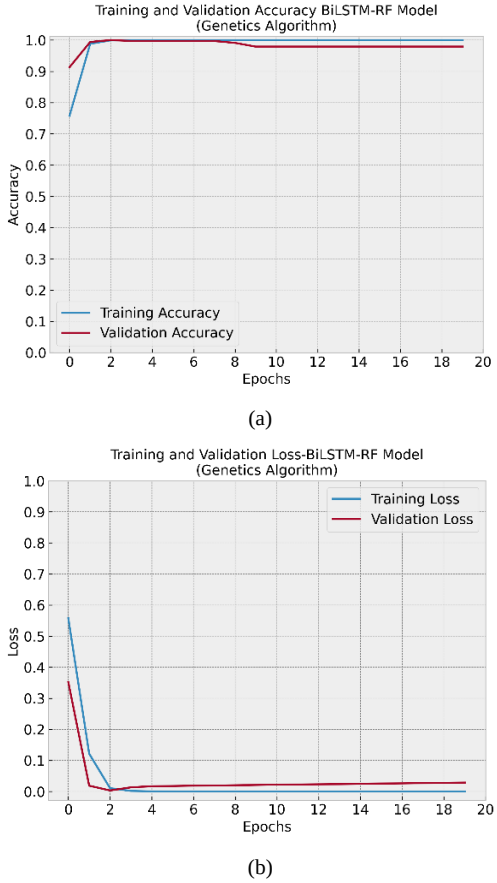


Fig. 8. Training accuracy, validation, and loss function of proposed model.

C. Discussions

The experiment findings demonstrate the effectiveness of feature optimization with a GA in identifying Arabic phishing emails. The classifiers performed differently in the first experiments conducted without GA. Certain models outperformed others in terms of precision, recall, and total accuracy, while others had difficulty reaching high effectiveness.

However, in the second set of experiments, all classifiers showed notable gains when the Genetic Algorithm was used for feature optimization. By picking the most pertinent features, the GA improved the models and improved their prediction power. This resulted in increased scores on all metrics overall. The significance of the GA in machine learning workflows, especially for intricate jobs like phishing

email detection, is highlighted by its efficacy in improving features. In addition to improving each classifier's performance, the GA also helped create detection systems that are more dependable and resilient by cutting down on noise and concentrating on the most important data points.

The proposed model has several advantages over conventional deep learning models like LSTM, BiLSTM, and GRU. It combines RF and BiLSTM with GA for feature optimization. This hybrid technique achieves improved precision, recall, F1-score, and accuracy in phishing email detection, demonstrating superior performance. A balanced and thorough feature representation is produced by combining the power of BiLSTM to capture sequential patterns with the capabilities of RF to evaluate feature relevance. By streamlining the feature set, lowering noise, and boosting overall effectiveness, the addition of GA improves the model even more. This results in a detection system that is more precise, dependable, and flexible and can adjust to the changing strategies used by phishing attempts. The proposed model's outstanding performance metrics underscore its effectiveness as a robust tool for enhancing cybersecurity.

VI. Conclusion

In this study, we proposed a hybrid model for phishing email detection, combining Random Forest (RF) and Bidirectional Long Short-Term Memory (BiLSTM) networks, augmented with Genetic Algorithm Feature Selection. The experimental results demonstrated that the proposed model significantly outperformed conventional approaches, including traditional machine learning classifiers, LSTM, BiLSTM, and Gated Recurrent Unit (GRU) models, across multiple performance metrics: Precision, Recall, F1 Score, and Accuracy. Specifically, the proposed model achieved an accuracy of 97.90%, recall of 100.00%, F1 score of 98.36%, and precision of 96.77%, illustrating its exceptional capability in correctly classifying phishing emails. The integration of RF and BiLSTM leveraged RF's proficiency in handling high-dimensional data and BiLSTM's capacity to capture sequential relationships, while the Genetic Algorithm Feature Selection ensured optimal feature subset identification.

Future research directions include expanding the dataset to encompass a broader range of phishing and legitimate emails, incorporating diverse linguistic and cultural variations. Additionally, we plan to explore advanced feature selection techniques such as Particle Swarm Optimization or Ant Colony Optimization. Furthermore, to capture more complex patterns in the data, we intend to investigate the integration of additional deep learning architectures, such as Transformers.

Acknowledgment

The authors extend their appreciation to the Deanship of Scientific Research at Northern Border University, Arar, KSA for funding this research work through the project number NBU-FFR-2024-1197-05.

References

[1] Yaseen, Q. (2021). Spam email detection using deep learning techniques. Procedia Computer Science, 184, 853-858. https://doi.org/10.1016/j.procs.2021.03.107.

[2] Salloum, S., Gaber, T., Vadera, S., & Shaalan, K. (2022). A systematic literature review on phishing email detection using natural language processing techniques. IEEE Access, 10, 65703-65727. https://doi.org/10.1109/ACCESS.2022.3172553.

[3] Wang, J., Li, Y., & Rao, H. R. (2016). Overconfidence in phishing email detection. Journal of the Association for Information Systems, 17(11), 1. https://doi.org/10.17705/1jais.00443.

[4] Valecha, R., Mandaokar, P., & Rao, H. R. (2021). Phishing email detection using persuasion cues. IEEE Transactions on Dependable and Secure Computing, 19(2), 747-756. https://doi.org/10.1109/TDSC.2021.3055228.

[5] Form, L. M., Chiew, K. L., & Tiong, W. K. (2015, August). Phishing email detection technique by using hybrid features. In 2015 9th International Conference on IT in Asia (CITA) (pp. 1-5). IEEE. https://doi.org/10.1109/CITA.2015.7349825.

[6] Egozi, G., & Verma, R. (2018, November). Phishing email detection using robust NLP techniques. In 2018 IEEE International Conference on Data Mining Workshops (ICDMW) (pp. 7-12). IEEE. https://doi.org/10.1109/ICDMW.2018.00010.

[7] Alhogail, A., & Alsabih, A. (2021). Applying machine learning and natural language processing to detect phishing email. Computers & Security, 110, 102414. https://doi.org/10.1016/j.cose.2021.102414.

[8] Verma, P., Goyal, A., & Gigras, Y. (2020). Email phishing: Text classification using natural language processing. Computer Science and Information Technologies, 1(1), 1-12.

[9] Harikrishnan, N. B., Vinayakumar, R., & Soman, K. P. (2018, March). A machine learning approach towards phishing email detection. In Proceedings of the anti-phishing pilot at ACM international workshop on security and privacy analytics (IWSPA AP) (Vol. 2013, pp. 455-468).

[10] Alhogail, A., & Alsabih, A. (2021). Applying machine learning and natural language processing to detect phishing email. Computers & Security, 110, 102414.

[11] Teja Nallamothu, P., & Shais Khan, M. (2023). Machine Learning for SPAM Detection. Asian Journal of Advances in Research, 6(1), 167-179. https://doi.org/10.9734/ajarr/2023/v6i113039.

[12] Qi, Q., Wang, Z., Xu, Y., Fang, Y., & Wang, C. (2023). Enhancing Phishing Email Detection through Ensemble Learning and Undersampling. Applied Sciences, 13(15), 8756. https://doi.org/10.3390/app13158756.

[13] Atlam, H. F., & Oluwatimilehin, O. (2022). Business email compromise phishing detection based on machine learning: A systematic literature review. Electronics, 12(1), 42. https://doi.org/10.3390/electronics12010042.

[14] Unnithan, N. A., Harikrishnan, N. B., Akarsh, S., Vinayakumar, R., & Soman, K. P. (2018). Machine learning based phishing e-mail detection. Security-CEN@ Amrita, 65-69.

[15] Kumar, N., & Sonowal, S. (2020, July). Email spam detection using machine learning algorithms. In 2020 Second International Conference on Inventive Research in Computing Applications (ICIRCA) (pp. 108-113). IEEE. https://doi.org/10.1109/ICIRCA48905.2020.9183145.

[16] Ghosh, A., & Senthilrajan, A. (2023). Comparison of machine learning techniques for spam detection. Multimedia Tools and Applications, 82(19), 29227-29254. https://doi.org/10.1007/s11042-023-14693-3.

[17] Ghourabi, A., Mahmood, M. A., & Alzubi, Q. M. (2020). A hybrid CNN-LSTM model for SMS spam detection in Arabic and English messages. Future Internet, 12(9), 156. https://doi.org/10.3390/fi12090156.

[18] Brindha, R., Nandagopal, S., Azath, H., Sathana, V., Joshi, G. P., & Kim, S. W. (2023). Intelligent Deep Learning Based Cybersecurity Phishing Email Detection and Classification. Computers, Materials & Continua, 74(3). https://doi.org/10.32604/cmc.2023.032386.

[19] Mughaid, A., AlZu'bi, S., Hnaif, A., Taamneh, S., Alnajjar, A., & Elsoud, E. A. (2022). An intelligent cyber security phishing detection system using deep learning techniques. Cluster Computing, 25(6), 3819-3828. https://doi.org/10.1007/s10586-022-03672-0.

[20] Fang, Y., Zhang, C., Huang, C., Liu, L., & Yang, Y. (2019). Phishing email detection using improved RCNN model with multilevel vectors and attention mechanism. IEEE Access, 7, 56329-56340. https://doi.org/10.1109/ACCESS.2019.2909314.

[21] Alsaidi, R. A., Yafooz, W. M., Alolofi, H., Taufiq-Hail, G. A. M., Emara, A. H. M., & Abdel-Wahab, A. (2022). Ransomware detection using machine and deep learning approaches. International Journal of Advanced Computer Science and Applications, 13(11). https://doi.org/10.14569/IJACSA.2022.0131132.

[22] Hasan, B. M. S., & Abdulazeez, A. M. (2021). A review of principal component analysis algorithm for dimensionality reduction. Journal of Soft Computing and Data Mining, 2(1), 20-30.

[23] Shobana, G.; Bushra, S.N. Classification of Myopia in Children using Machine Learning Models with Tree Based Feature Selection. In Proceedings of the 2020 4th International Conference on Electronics, Communication and Aerospace Technology (ICECA), Coimbatore, India, 5-7 November 2020; pp. 1599-1605. https://doi.org/10.1109/ICECA49313.2020.9297528.

[24] Gupta, K. Optimizing Performance: SelectKBest for Efficient Feature Selection in Machine Learning. 2020. Available online: https://medium.com/@Kavya2099/optimizing-performance-selectkbest-for-efficient-feature-selection-in-machine-learning-3b635905ed48 (accessed on 11 July 2024).

[25] Jeon, H.; Oh, S. Hybrid-Recursive Feature Elimination for Efficient Feature Selection. Appl. Sci. 2020, 10, 3211. https://doi.org/10.3390/app10093211.

[26] Li, F.; Lai, L.; Cui, S. On the Adversarial Robustness of LASSO Based Feature Selection. IEEE Trans. Signal Process. 2021, 69, 5555-5567. https://doi.org/10.1109/TSP.2021.3102136.

[27] Greenacre, M., Groenen, P. J., Hastie, T., d'Enza, A. I., Markos, A., & Tuzhilina, E. (2022). Principal component analysis. Nature Reviews Methods Primers, 2(1), 100.

[28] Rey, C.C.T.; García, V.S.; Villuendas-Rey, Y. Evolutionary feature selection for imbalanced data. In Proceedings of the 2023 Mexican International Conference on Computer Science (ENC), Guanajuato, Mexico, 11-13 September 2023; pp. 1-7. https://doi.org/10.1109/ENC57540.2023.10177468.

[29] Thapa, C., Tang, J. W., Abuadbba, A., Gao, Y., Camtepe, S., Nepal, S., ... & Zheng, Y. (2023). Evaluation of federated learning in phishing email detection. Sensors, 23(9), 4346. https://doi.org/10.3390/s23094346.

[30] Tong, X., Wang, J., Zhang, C., Wang, R., Ge, Z., Liu, W., & Zhao, Z. (2021). A content-based chinese spam detection method using a capsule network with long-short attention. IEEE Sensors Journal, 21(22), 25409-25420.

[31] Li, Q., Cheng, M., Wang, J., & Sun, B. (2020). LSTM based phishing detection for big email data. IEEE transactions on big data, 8(1), 278-288.

[32] Wu, Y., Si, S., Zhang, Y., Gu, J., & Wosik, J. (2024). Evaluating the performance of chatgpt for spam email detection. arXiv preprint arXiv:2402.15537.

[33] Sonowal, G. (2020). Phishing email detection based on binary search feature selection. SN Computer Science, 1(4), 191.

[34] Ablel-Rheem, D. M., Ibrahim, A. O., Kasim, S., Almazroi, A. A., & Ismail, M. A. (2020). Hybrid feature selection and ensemble learning method for spam email classification. International Journal, 9(1.4), 217-223.

[35] Saber, W. M., Ding, W., Sonne, C., & Abdelsalam, H. M. (2022). Email phishing detection: A systematic literature review. ACM Computing Surveys, 55(2), 1-37. https://doi.org/10.1145/3491207.

[36] [35] Ghourabi, A., Mahmood, M. A., & Alzubi, Q. M. (2020). A hybrid CNN-LSTM model for SMS spam detection in Arabic and English messages. Future Internet, 12(9), 156. https://doi.org/10.3390/fi12090156.

[37] [36] Valecha, R., Mandaokar, P., & Rao, H. R. (2021). Phishing email detection using persuasion cues. IEEE transactions on Dependable and secure computing, 19(2), 747-756. https://doi.org/10.1109/TDSC.2021.3050803.

[38] [37] Thapa, C., Tang, J. W., Abuadbba, A., Gao, Y., Camtepe, S., Nepal, S., ... & Zheng, Y. (2023). Evaluation of federated learning in phishing email detection. Sensors, 23(9), 4346. https://doi.org/10.3390/s23094346.

[39] [38] Harikrishnan, N. B., Vinayakumar, R., & Soman, K. P. (2018, March). A machine learning approach towards phishing email detection. In Proceedings of the anti-phishing pilot at ACM international workshop on security and privacy analytics (IWSPA AP) (Vol. 2013, pp. 455-468). https://doi.org/10.1145/3180445.3180635.

[40] [39] Unnithan, N. A., Harikrishnan, N. B., Akarsh, S., Vinayakumar, R., & Soman, K. P. (2018). Machine learning based phishing e-mail detection. Security-CEN@ Amrita, 65-69.

[41] Mughaid, A., AlZu'bi, S., Hnaif, A., Taamneh, S., Alnajjar, A., & Elsoud, E. A. (2022). An intelligent cyber security phishing detection system using deep learning techniques. Cluster Computing, 25(6), 3819-3828. https://doi.org/10.1007/s10586-022-03613-6.

[42] Yafooz, W., & Alsaeedi, A. (2024). Leveraging User-Generated Comments and Fused BiLSTM Models to Detect and Predict Issues with Mobile Apps. Computers, Materials & Continua, 79(1). https://doi.org/10.32604/cmc.2024.027108.

[43] Brindha, R., Nandagopal, S., Azath, H., Sathana, V., Joshi, G. P., & Kim, S. W. (2023). Intelligent Deep Learning Based Cybersecurity Phishing Email Detection and Classification. Computers, Materials & Continua, 74(3). https://doi.org/10.32604/cmc.2023.025628.

[44] Holland, J. H. (1992). Genetic algorithms. Scientific American, 267(1), 66-73. https://doi.org/10.1038/scientificamerican0792-66.

[45] Xue, B., Zhang, M., Browne, W. N., & Yao, X. (2016). A survey on evolutionary computation approaches to feature selection. IEEE Transactions on Evolutionary Computation, 20(4), 606-626. https://doi.org/10.1109/TEVC.2015.2504420.

[46] Goldberg, D. E. (1989). Genetic algorithms in search, optimization, and machine learning. Addison-Wesley Longman Publishing Co., Inc.

[47] Hamid, I. R. A., & Abawajy, J. (2011). Hybrid feature selection for phishing email detection. In International Conference on Algorithms and Architectures for Parallel Processing (pp. 266-275). Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-24669-2_25.

[48] Deb, K. (2001). Multi-objective optimization using evolutionary algorithms. John Wiley & Sons. https://doi.org/10.1002/9780470496947.

[49] Zareapoor, M., & Seeja, K. R. (2015). Feature extraction or feature selection for text classification: A case study on phishing email detection. International Journal of Information Engineering and Electronic Business, 7(2), 60-65. https://doi.org/10.5815/ijieeb.2015.02.08.

[50] Akinyelu, A. A., & Adewumi, A. O. (2014). Classification of phishing email using random forest machine learning technique. Journal of Applied Mathematics, 2014. https://doi.org/10.1155/2014/425731.

[51] Chowdhury, M., Colbert, J., Kabir, M., Sait, S. M., & Aslam, N. (2020). A multi-optimization based feature selection method for phishing detection using neural networks. IEEE Access, 8, 219616-219626. https://doi.org/10.1109/ACCESS.2020.3042717.

[52] Zou, Y., & Schaub, F. (2019). Beyond mandatory: Making data breach notifications useful for consumers. IEEE Security & Privacy, 17(2), 67-72. https://doi.org/10.1109/MSEC.2019.2905629.

[53] Wang, J., Yang, Y., & Xia, B. (2019). A simplified Cohen's Kappa for use in binary classification data annotation tasks. IEEE Access, 7, 164386-164397.

[54] Yafooz, W. M., Alsaeedi, A., & Emara, A. H. M. (2023, February). AraDS: Arabic datasets for text mining approaches. In 2023 International Conference on Smart Computing and Application (ICSCA) (pp. 1-6). IEEE. https://doi.org/10.1109/ICSCA57840.2023.10087675.

[55] Pisner, D. A., & Schnyer, D. M. (2020). Support vector machine. In Machine learning (pp. 101-121). Academic Press.

[56] Yafooz, W. M., Alsaeedi, A., Alluhaibi, R., & Abdel-Hamid, M. E. (2022). Enhancing multi-class web video categorization model using machine and deep learning approaches. Int. J. Electr. Comput. Eng, 12, 3176. https://doi.org/10.11591/ijece.v12i3.pp3176-3191.

[57] Charbuty, B., & Abdulazeez, A. (2021). Classification based on decision tree algorithm for machine learning. Journal of Applied Science and Technology Trends, 2(01), 20-28.

[58] Song, X., Liu, X., Liu, F., & Wang, C. (2021). Comparison of machine learning and logistic regression models in predicting acute kidney injury: A systematic review and meta-analysis. International journal of medical informatics, 151, 104484.

[59] Alhejaili, R., Alhazmi, E. S., Alsaeedi, A., & Yafooz, W. M. (2021, September). Sentiment analysis of the COVID-19 vaccine for Arabic tweets using machine learning. In 2021 9th International conference on reliability, infocom technologies and optimization (Trends and Future Directions)(ICRITO) (pp. 1-5). IEEE. https://doi.org/10.1109/ICRITO51393.2021.9596517.

[60] Nusinovici, S., Tham, Y. C., Yan, M. Y. C., Ting, D. S. W., Li, J., Sabanayagam, C., ... & Cheng, C. Y. (2020). Logistic regression was as good as machine learning for predicting major chronic diseases. Journal of clinical epidemiology, 122, 56-69.

[61] Boateng, E. Y., Otoo, J., & Abaye, D. A. (2020). Basic tenets of classification algorithms K-nearest-neighbor, support vector machine, random forest and neural network: A review. Journal of Data Analysis and Information Processing, 8(4), 341-357.

[62] Antoniadis, A., Lambert-Lacroix, S., & Poggi, J. M. (2021). Random forests for global sensitivity analysis: A selective review. Reliability Engineering & System Safety, 206, 107312.

[63] Yahya, A. E., Gharbi, A., Yafooz, W. M., & Al-Dhaqm, A. (2023). A novel hybrid deep learning model for detecting and classifying non-functional requirements of mobile apps issues. Electronics, 12(5), 1258. https://doi.org/10.3390/electronics12051258.

[64] Chung, J., Gulcehre, C., Cho, K., & Bengio, Y. (2014). Empirical evaluation of gated recurrent neural networks on sequence modeling. arXiv preprint arXiv:1412.3555. https://doi.org/10.48550/arXiv.1412.3555.

[65] Hameed, Z., & Garcia-Zapirain, B. (2020). Sentiment classification using a single-layered BiLSTM model. Ieee Access, 8, 73992-74001.

[66] Deng, J., Cheng, L., & Wang, Z. (2021). Attention-based BiLSTM fused CNN with gating mechanism model for Chinese long text classification. Computer Speech & Language, 68, 101182.

[67] Saibene, A.; Gasparini, F. Genetic algorithm for feature selection of EEG heterogeneous data. Expert Syst. Appl. 2023, 217, 119488. https://doi.org/10.1016/j.eswa.2023.119488.

[68] Catak, F.O. Genetic algorithm based feature selection in high dimensional text dataset classification. WSEAS Trans. Inf. Sci. Appl. 2015, 12, 290-296.

[69] Xue, B., Zhu, C., Wang, X., & Zhu, W. (2022, March). The study on the text classification based on graph convolutional network and BiLSTM. In Proceedings of the 8th International Conference on Computing and Artificial Intelligence (pp. 323-331).

A Feature Interaction Based Neural Network Approach: Predicting Job Turnover in Early Career Graduates in South Korea

Haewon Byeon
Department of AI-Software, Inje Medical Big Data Research Center, Inje University, Gimhae 50834, South Korea

***Abstract*—Predicting job turnover among early career university graduates is crucial for both employees and employers. This study introduced a Feature Interaction based Neural Network model designed to predict job turnover among university graduates in their 20s and 30s in South Korea within the first five years of employment. The FINN model leveraged the Graduates Occupational Mobility Survey dataset, which included detailed information on approximately 26,544 graduates. This rich dataset encompassed a wide range of variables, including personal attributes, employment characteristics, job satisfaction, and job preparation activities. The model combined an embedding layer to convert sparse features into dense vectors with a neural network component to capture high-order feature interactions. We compared the FINN model's performance against eight baseline models: Logistic Regression, Factorization Machines, Field-aware Factorization Machines, Support Vector Machine, Random Forest, Product-based Neural Networks, Wide & Deep, and DeepFM. Evaluation metrics used were Area Under the ROC Curve (AUC) and Log Loss. The results demonstrated that the FINN model outperformed all baseline models, achieving an AUC of 0.830 and a Log Loss of 0.370. The FINN model represents a significant advancement in predictive modeling for job turnover, providing valuable insights that can inform both individual career planning and organizational human resource practices. This research underscores the potential of advanced neural network architectures in employment data analysis and predictive modeling.**

Keywords*—Job turnover prediction; feature interaction based neural network; employment data analysis; predictive modeling; university graduates

I. Introduction

The advent of big data has revolutionized various sectors, including the labor market, where the analysis of employment data plays a crucial role in understanding workforce dynamics and predicting future trends [1,2]. This study focuses on the early career trajectories of university graduates in their 20s and 30s in South Korea, particularly their likelihood of job turnover within the first five years of employment. The accurate prediction of job turnover is paramount for both employees and employers [3-5]. Employees benefit by understanding potential career pitfalls, while employers can devise better retention strategies, thereby reducing recruitment and training costs [3,4].

Employees stand to gain significantly from insights into job turnover predictors. By identifying the factors that contribute to early job departure, graduates can better prepare themselves to meet the challenges of their initial employment experiences [6,7]. This knowledge empowers them to seek roles and environments that align more closely with their career aspirations and stability [8,9]. For employers, the implications of job turnover are substantial. High turnover rates can lead to increased recruitment costs, loss of organizational knowledge, and diminished productivity. By accurately predicting which employees are at risk of leaving, employers can implement targeted interventions to improve job satisfaction and engagement [10-12]. This proactive approach not only enhances employee retention but also fosters a more stable and committed workforce. The dual benefits of predicting job turnover—enhancing employee career stability and optimizing employer retention strategies—underscore the importance of this research. By leveraging advanced predictive models, this study seeks to provide valuable insights that can inform both individual career planning and organizational human resource practices.

Historically, regression analysis has been a popular method for studying job turnover. Regression models, such as logistic regression, have been widely used to identify factors influencing employee turnover by modeling the relationship between dependent and independent variables [13-16]. For instance, logistic regression can help in estimating the probability of an event (like job turnover) occurring, given a set of predictor variables (such as age, education level, job satisfaction, etc.) [13-16]. However, while regression analysis has its merits, it also has significant methodological limitations when applied to complex, high-dimensional datasets typical in employment studies [7]. One of the primary limitations of traditional regression analysis is its inability to effectively capture complex interactions between features [7]. In employment data, factors influencing job turnover are often interdependent. For example, the interaction between job satisfaction and work-life balance might significantly affect turnover rates, but such interactions can be challenging to model accurately using simple regression techniques. Regression models assume a linear or specific non-linear relationship between the independent and dependent variables, which may not hold true in real-world scenarios where relationships can be highly non-linear and intricate.

Moreover, regression models often require extensive feature engineering to improve their predictive accuracy [16]. Feature engineering involves the manual creation of new features from raw data to better represent the underlying patterns [17]. This process can be labor-intensive and requires

domain expertise to identify meaningful interactions and transformations [18]. Despite these efforts, the performance of regression models may still be limited due to their inherent inability to model complex, higher-order interactions between features [19].

To address these limitations, there has been a growing interest in leveraging advanced machine learning techniques, particularly deep learning models, which can automatically learn feature interactions from raw data without extensive manual intervention. Among these, Feature Interaction based Neural Networks (FINNs) have shown great promise [20]. FINNs enhance the capabilities of traditional deep neural networks (DNNs) by explicitly modeling feature interactions, thereby improving predictive accuracy in complex datasets [21].

Deep neural networks have achieved remarkable success in various fields, such as image classification, natural language processing (NLP), and speech recognition, due to their ability to learn hierarchical feature representations [22]. In the context of employment data, DNNs can be particularly useful as they can capture complex, non-linear relationships between features. However, one of the challenges in applying DNNs to employment data is the sparsity and high-dimensionality of the data. Employment datasets often contain categorical variables, such as job title, industry, and education level, which are typically converted into high-dimensional sparse features using techniques like one-hot encoding. These sparse features need to be transformed into dense representations before being fed into the neural network.

FINNs address this challenge by employing a feature embedding layer that converts sparse categorical features into dense vectors [20]. These embeddings are then used to model pairwise interactions between features, capturing the complex dependencies that influence job turnover [20]. By incorporating a feature interaction layer, FINNs can learn both low-order and high-order interactions, providing a more comprehensive understanding of the factors driving employee turnover.

The need for FINNs is underscored by the limitations of traditional methods. For example, factorization machines (FMs) have been proposed to model feature interactions via the inner product of feature embeddings, but they primarily capture only second-order interactions [21]. While FMs have been successful in some applications, they may not fully exploit the higher-order interactions present in employment data. In contrast, FINNs can model both second-order and higher-order interactions, providing a more robust framework for predicting job turnover [21]. Furthermore, the integration of deep learning components in FINNs allows for the modeling of non-linear interactions and complex feature hierarchies, which are often present in employment data. This capability is particularly important for understanding the multifaceted nature of job turnover, where factors such as job satisfaction, career development opportunities, and organizational culture interplay in intricate ways.

FINNs work by first employing an embedding layer to transform high-dimensional sparse features into dense vectors [20,21]. This transformation is crucial for handling the sparsity issue inherent in employment data. Once the features are embedded, FINNs apply a feature interaction layer that captures pairwise interactions between the dense vectors. This layer can utilize operations such as inner product or element-wise product to model the interactions. By doing so, FINNs can effectively represent the complex relationships between features, which traditional regression models might miss. Moreover, FINNs extend the capability of simple interaction models by incorporating deep neural network components that can capture higher-order interactions [21]. This means that after modeling the basic pairwise interactions, the network can further process these interactions through multiple layers to extract more complex patterns. This deep architecture allows FINNs to model non-linear relationships and hierarchies among features, enhancing their predictive power. The novelty of FINNs lies in their ability to combine the strengths of traditional interaction models and deep learning. Unlike traditional models that require extensive manual feature engineering to capture interactions, FINNs can automatically learn these interactions from data. This reduces the need for domain expertise and manual intervention, making the modeling process more efficient and scalable.

While traditional regression methods have provided valuable insights into employee turnover, their limitations necessitate the adoption of more advanced techniques like Feature Interaction based Neural Networks. FINNs offer a powerful alternative by automatically learning complex feature interactions from high-dimensional data, thereby enhancing predictive accuracy and providing deeper insights into the factors influencing job turnover. This study aims to leverage FINNs to analyze the early career trajectories of university graduates in South Korea, with the goal of identifying key predictors of job turnover and informing strategies to improve employee retention. The remainder of this paper is organized as follows: Section II reviews related works. Section III describes the details of the proposed model. Section IV presents the experimental analysis. Finally, Section V concludes the paper.

II. Related Work

Logistic Regression (LR) is a fundamental technique widely used in classification tasks, including click-through rate prediction. It is a linear model that solves an unconstrained convex optimization problem, ensuring efficient convergence to a globally optimal solution via gradient descent [23]. The primary advantage of LR is its interpretability; by examining the weights assigned to each feature, one can understand the significance and impact of these features on the prediction outcome [23]. This transparency makes LR particularly valuable in fields where interpretability is crucial, such as field of employment. However, the linear nature of LR limits its ability to capture complex relationships between features [24]. To overcome this, extensive feature engineering is often required, including polynomial features and interaction terms, to improve the model's expressiveness.

Another shallow method worth mentioning is Polynomial Regression, which extends linear regression by considering polynomial terms of the features [25]. By including polynomial terms, the model can capture non-linear relationships between

the features and the target variable. Polynomial regression can be particularly useful when the relationship between the features and the outcome is known to be non-linear. However, as the degree of the polynomial increases, the model becomes more complex and prone to overfitting, especially with limited data [25].

Decision Trees are another fundamental shallow method used for classification and regression tasks. A decision tree splits the data into subsets based on feature values, creating a tree-like model of decisions [26]. Each node represents a feature, each branch represents a decision rule, and each leaf node represents an outcome. Decision trees are easy to interpret and visualize, making them useful for understanding the decision-making process. However, they can be prone to overfitting, especially with deep trees [27]. Techniques such as pruning, bagging, and boosting are often used to mitigate overfitting and improve performance.

Ensemble methods, such as Random Forests and Gradient Boosting Machines (GBM), build on the strengths of decision trees while addressing their limitations [28]. Random Forests create multiple decision trees using different subsets of the data and features, and then aggregate their predictions [29,30]. This approach reduces overfitting and improves generalization. GBM, on the other hand, builds trees sequentially, with each tree attempting to correct the errors of the previous ones [31]. This method is highly effective for both classification and regression tasks but can be computationally intensive.

Field-aware Factorization Machines (FFM) extend the capabilities of FM by introducing the concept of fields [32-34]. In FM, a feature interacts with other features using the same vector, whereas in FFM, a feature uses different vectors to interact with features from different fields [33]. This distinction allows FFM to model interactions more precisely, enhancing the model's expressiveness [34]. For instance, in a recommendation system, user-related features and item-related features can interact differently depending on their respective fields. However, the enhanced expressiveness of FFM comes at the cost of increased memory requirements, which can be a significant limitation when dealing with very large datasets [32].

In summary, shallow methods like Logistic Regression, Factorization Machines, and Decision Trees provide foundational techniques for modeling interactions and making predictions. While they offer interpretability and simplicity, their expressiveness is often limited, necessitating extensive feature engineering or the use of ensemble techniques to capture complex relationships. The ongoing development of these methods continues to enhance their applicability across various domains, from recommendation systems to employment analytics.

III. PROPOSED METHOD

Our main objective in this study is to model the feature interaction representation more effectively to predict job turnover among early career university graduates in South Korea. To achieve this, we propose a Feature Interaction based Neural Network (FINN) tailored for employment data analysis.

A. Sparse Input and Embedding Layer

Unlike image classification or speech recognition tasks, the input data in employment prediction tasks are usually non-contiguous and categorical. These raw input features are typically converted into high-dimensional sparse features via one-hot encoding. One-hot encoding is a process that transforms categorical variables into a binary vector representation, where only one element is "hot" (i.e., set to 1) and all other elements are "cold" (i.e., set to 0). For instance, consider the following categorical variables:

User ID = 001, 002, ...

Job Type = Engineer, Teacher, ...

Gender = Male, Female

Using one-hot encoding, an input instance can be transformed as follows:

User ID = 001 → [1, 0, 0, ...]

Job Type = Engineer → [1, 0, 0, ...]

Gender = Female → [0, 1]

The dimension of these features, particularly the user ID and job type, becomes large after encoding. For instance, if there are 550 job types, the dimension of the job type feature increases to 550, with only one of these values being effective. This results in extremely sparse feature vectors, where the majority of the elements are zero. The sparsity of the coded feature suggests that deep neural networks (DNNs) are not directly applicable because DNNs typically require dense input vectors to perform effectively.

Therefore, these sparse features are embedded into a continuous, dense, real-value vector space with lower dimensions. The embedding layer transforms these high-dimensional sparse vectors into dense vectors. This transformation is achieved by mapping each categorical value to a dense vector of fixed size. The embedding layer can be represented as:

$$\left[E = [e_1, e_2, \ldots, e_l, \ldots, e_m] \right]$$

where (m) denotes the number of fields, (e_i\in \mathbb{R}^k) denotes the embedding vector of the (i)-th field, and (k) is the dimension of the embedding vector. The embedding process can be visualized as follows:

Each unique value in a categorical feature is assigned a unique dense vector.

These dense vectors are learned during the training process, allowing the model to capture semantic similarities between different categorical values.

The resulting embedded vectors are then concatenated to form a dense representation of the original sparse input.

For example, consider a feature vector with three categorical variables: user ID, job type, and gender. After one-hot encoding and embedding, the resulting dense vector might look like this:

$$\left[\{\text{User ID embedding}\} = [0.1, 0.3, 0.5] \right]$$

$$\left[\{\text{Job Type embedding}\} = [0.2,0.4,0.6]\right]$$

$$\left[\{\text{Gender embedding}\} = [0.3,0.7]\right]$$

These embeddings are then concatenated to form a single dense vector:

$$[E = [0.1,0.3,0.5,0.2,0.4,0.6,0.3,0.7]]$$

The dimension of the dense vector is much smaller than the original sparse vector, making it more suitable for input into a deep neural network. The embedding layer not only reduces the dimensionality of the input features but also captures latent relationships between different categorical values, which can be crucial for accurate prediction.

B. Feature-Interaction Layer

To improve prediction accuracy, it is crucial to model feature interactions after the embedding layer. The feature-interaction layer aims to model second-order feature relationships. Intuitively, we can represent the interaction of the (i)-th feature and the (j)-th feature using a vector (p_{ij}). However, due to data sparsity, training (p_{ij}) directly is impractical.

Instead, we use embedding vectors to calculate interaction vectors through methods like inner product and element-wise product. These methods are defined as:

$$\left[f_{\{\text{inner}\}}(x) = \sum_{\{i,j \in X\}(\mathbf{v}i \cdot \mathbf{v}j)x_i x_j}\right]$$

$$\left[f\{\text{element} - \text{wise}\}(x) = \sum\{i,j \in X\}\left(\mathbf{v}_i \circ \mathbf{v}_j\right)x_i x_j\right]$$

where (X)is the set of features, (\mathbf{v}_i)and (\mathbf{v}_j)are the embedding vectors, and (\circ) denotes the element-wise product. These methods can be too simple to effectively calculate feature interactions, so we propose a more sophisticated method:

$$[\mathbf{p}\{ij\}= [p\{ij\}^1,p_\{ij\}^2,\ldots,p_\{ij\}^l]]$$

$$[p_\{ij\}^u= \mathbf{v}_(i_j^(T\mathbf{W}^u\mathbf{v})))]$$

where (\mathbf{W}\in \mathbb{R}^{k \times k \times l}) is a three-dimensional tensor. Each slice (\mathbf{W}^u) represents a relation matrix.

C. Combination Layer and Deep Network

The interaction vectors (\mathbf{p})are concatenated and fed into a deep neural network (DNN). The combination layer merges the outputs of the feature-interaction layer:

$$[\mathbf{c}= [c_1,c_2,\ldots,c_k]]$$

The deep network captures higher-order interactions between features. The fully connected layers are defined as:

$$\left[\mathbf{h}^{(l)} = \sigma\left(\mathbf{W}^{(l)}\mathbf{h}^{(l-1)} + \mathbf{b}^{(l)}\right)\right]$$

where (σ)is the activation function, (mathbf{W}^{(l)})and (\mathbf{b}^{(l)}) are the weight matrix and bias vector of the (l)-th layer. The DNN captures high-order feature interactions through non-linear activation functions like ReLU, sigmoid, or tanh.

Finally, the output vector of the last neural network layer is used to calculate the prediction score:

$$\left[y_d = \sigma\left(\mathbf{W}^{(L+1)}\mathbf{h}^{(L)} + \mathbf{b}^{(L+1)}\right)\right]$$

D. Output Layer and Learning

The overall formulation of the FINN model output is:

$$[y = \sigma (w_0+ \sum_{i=1}^n w_i x_i+ y_d)]$$

where (y) is the predicted probability of job turnover, (\sigma) is the sigmoid function, (n) is the number of features, and (w_i) are the weights of the sparse features. The loss function we aim to minimize is the binary cross-entropy loss:

$$[\text{loss}=$$

$$-\sum_{x \in X}[y_(i(x)\log(((-)))^{y}_i(x))))$$

$$+ (1 - y_i(x))\log(1 - () ^{y}_i(x))]]$$

where ($y_{i(x)}$)is the ground truth, ($\{y\}_{i(x)}$) is the predicted value,and (X)is the set of training instances.

To optimize the model, we use Mini-Batch Gradient Descent with the Adam optimizer. Adam combines RMSProp and momentum methods, adjusting the learning rate adaptively:

$$\left[\mathbf{m}t = \beta_{1\mathbf{m}\{t-1\}} + (1 - \beta_1)\mathbf{g}_t\right]$$

$$\left[\mathbf{v}t = \beta_{2\mathbf{v}\{t-1\}} + (1 - \beta_2) \mathbf{g}_t^2\right]\left[\widehat{\{\mathbf{m}\}}_t = \frac{\mathbf{m}_t}{1 - \beta_1^t}\right]$$

$$\left[\widehat{\{\mathbf{v}\}}_t = \frac{\mathbf{v}t}{1 - \beta_2^t}\right]$$

$$\left[\theta_t = \theta\{t - 1\} - \alpha\frac{\widehat{\{\mathbf{m}\}}_t}{\left\{\sqrt{\widehat{\{\mathbf{v}\}}_t} + \epsilon\right\}}\right]$$

where (β_1)and (β_2)are decay rates, (α)is the learning rate, (ϵ)is mathbf{g}_t)is the gradient.

We also apply dropout and batch normalization to prevent overfitting and stabilize training. Dropout randomly drops neurons during training with a probability (p), and batch normalization normalizes intermediate layer outputs.

IV. EXPERIMENTS

A. Dataset and Participants

This study utilizes data from the Graduates Occupational Mobility Survey (GOMS) conducted by the Korea Employment Information Service. The dataset includes approximately 5% of the 500,000 students who graduated from two-year and four-year colleges between August 2014 and February 2015, resulting in a sample size of 28,549 individuals. The survey was conducted in September and October 2016. The GOMS dataset is comprehensive, encompassing a wide range of variables that influence labor market entry and retention. These variables include academic background, current economic activity, job characteristics, job search activities, and individual demographics. The dataset's richness allows for a detailed analysis of the factors influencing job turnover among recent graduates. Participants were selected based on the following criteria: (1) they must have graduated after January 2014 and have secured their first job post-graduation. (2) Furthermore, only those employed in regular, full-time positions without fixed-term contracts were included in the study. This selection criterion ensures that the analysis focuses on stable employment scenarios, eliminating the variability introduced by temporary or part-time positions.

The GOMS survey provides detailed information on various aspects of the participants' careers and educational backgrounds. It includes data on the type of institution they graduated from, their graduation date, current employment status, details about their current job, and information about their first job. Additionally, the survey collects data on job search activities, vocational training experiences, language training, and certifications obtained. This comprehensive dataset allows for a nuanced analysis of the factors that influence job turnover among recent graduates. Table I is a summary of the dataset statistics and Table II is the variables measured in the study.

TABLE I. SUMMARY OF THE DATASET

Dataset	Instances	Categories	Fields	Positive Ratio
GOMS	28,549	50+	20+	0.27

TABLE II. VARIABLES MEASURED

Variable Category	Variable Name	Description
Personal Attributes	Gender	Male, Female
	Age	Age at the time of turnover
Employment Characteristics	Industry	Industry sector of the job
	Job Type	Specific job role
	Company Size	Number of employees
Working Conditions	Weekly Working Hours	Total hours worked per week
	Monthly Salary	Average monthly income
	Union Membership	Whether the employee is a union member
Job Satisfaction	Satisfaction Level	11 items on a 5-point Likert scale
	Job Fit	4 items measuring the alignment of job with skills and interests
Benefits	Social Insurance	Dummy variable indicating social insurance coverage
	Welfare Benefits	Dummy variable indicating availability of welfare benefits
Job Preparation	Work Experience	Employment experience during school
	Job Search Experience	Experience in job searching
	Vocational Training	Participation in vocational training
	Certification	Whether the individual holds any professional certifications
Academic Performance	GPA	Grade point average on a 5-point scale

B. Baseline Methods

We compare FINN with eight baseline models in our experiments (Table III), all implemented with TensorFlow and trained using the Adam optimization algorithm.

TABLE III. THE EIGHT BASELINE MODELS OF STUDY

Model	Description
Logistic Regression (LR)	A classical model in classification tasks that predicts the probability of positive samples. It is a linear model that uses the logistic function to model a binary dependent variable.
Factorization Machines (FM)	Models feature interactions by learning a feature vector for each feature and using the inner product of two feature vectors. FM is effective in capturing second-order interactions.
Field-aware Factorization Machines (FFM)	An extension of FM that considers the field information of each feature, allowing for more precise interaction modeling.
Support Vector Machine (SVM)	A supervised learning model used for classification tasks. SVM constructs a hyperplane or set of hyperplanes in a high-dimensional space to separate different classes. It is effective in high-dimensional spaces and for cases where the number of dimensions exceeds the number of samples.
Random Forest	An ensemble learning method that constructs multiple decision trees during training and outputs the mode of the classes (classification) or mean prediction (regression) of the individual trees. It reduces overfitting by averaging multiple trees.
Product-based Neural Networks (PNN)	Uses product operations to perform pairwise interactions between features to capture interaction information. This model enhances the representation power by explicitly modeling feature interactions.
Wide & Deep	A hybrid model consisting of a single layer wide part and a multilayer deep part. The wide part captures memorization of feature interactions, while the deep part captures generalization.
DeepFM	Improves the Wide & Deep model by replacing the wide part with a factorization machine. DeepFM combines the strengths of FM and deep neural networks to capture both low-order and high-order feature interactions.

C. Evaluation Metrics

We use two primary evaluation metrics to assess model performance: Area Under the ROC Curve (AUC) and Log Loss.

- AUC: A widely used metric in binary classification that measures the ability of the model to distinguish between

positive and negative samples. A higher AUC indicates better performance.

- Log Loss: Measures the distance between the predicted probabilities and the actual labels. Lower log loss values indicate better performance.

D. Data Processing and Experimental Setup

For data preprocessing, categorical features are converted into one-hot encoded vectors. Numerical features are discretized by equal-size buckets. We also apply negative down-sampling to address the issue of class imbalance, ensuring that the positive sample ratio is approximately 0.5.

We implement all models using TensorFlow and train them using the Adam optimization algorithm with a mini-batch size of 1000. The learning rate is set to 0.0001. For deep models, the depth of layers is set to 5, with ReLU activation functions. The number of neurons per layer is set to 700. We initialize the DNN hidden layers using Xavier initialization and the embedding vectors from uniform distributions. The experiments are conducted on two GTX 4060 Ti GPUs.

E. Performance Comparison and Analysis

We compare the performance of the FINN model with baseline models using the GOMS dataset. The results are summarized in Table IV.

TABLE IV. PERFORMANCE COMPARISON OF DIFFERENT MODELS

Method	AUC	Log Loss
LR	0.751	0.449
FM	0.783	0.417
FFM	0.792	0.408
SVM	0.770	0.430
Random Forest	0.765	0.435
PNN	0.801	0.399
Wide & Deep	0.813	0.387
DeepFM	0.820	0.380
FINN	0.830	0.370

F. Analysis of Results

The experimental results show that our proposed FINN model outperforms all baseline models in terms of both AUC and Log Loss. The superior performance of FINN can be attributed to its ability to effectively capture complex feature interactions through its feature interaction layer. Traditional models like Logistic Regression and Factorization Machines are limited in their ability to model higher-order interactions, which are crucial for accurate predictions in employment data.

Neural network-based models such as FFM, PNN, and DeepFM show better performance compared to traditional models, highlighting the importance of modeling feature interactions. Among these, DeepFM performs well due to its ability to capture both low-order and high-order interactions. However, FINN surpasses DeepFM by employing a more sophisticated feature interaction mechanism that extends beyond simple inner product or element-wise product operations.

To further analyze the effectiveness of FINN, we conduct additional experiments varying the size of the embedding vectors and the number of hidden layers in the DNN. The results, illustrated in Fig. 1 and 2 indicate that FINN consistently outperforms other models across different configurations, demonstrating its robustness and generalizability.

G. Parameter Study

In this subsection, we conduct hyper-parameter investigations for our model, focusing on the embedding part, the DNN part, and the feature interaction part. Specifically, we change the following hyper-parameters: (1) the dimension of embeddings, (2) the depth of DNN, and (3) the dimension of the feature interaction vector.

1) Embedding part: We change the embedding sizes from 10 to 50 and summarize the experimental results in Fig. 1 and Table V. As the dimension expands from 10 to 50, our model shows substantial improvement. We find that an embedding size of 30 yields the best performance on the GOMS dataset. Enlarging the embedding size increases the number of parameters in the embedding layer and the DNN part. The optimal embedding size balances model complexity and performance.

2) DNN part: We investigate the impact of different DNN depths by varying the number of hidden layers. Increasing the number of layers initially improves model performance; however, performance degrades if the number of layers continues to increase due to overfitting. Fig. 2 and Table VI shows that a depth of 5 layers provides the best balance between model complexity and performance.

3) *Feature interaction part:* We change the feature interaction vector sizes from 10 to 40. Fig. 3 and Table VII shows that a vector size of 10 provides the best performance on the GOMS dataset. The performance remains stable as we increase the vector size, indicating that the model is robust to this hyper-parameter.

TABLE V. EMBEDDING SIZE OF STUDY

Embedding Size	AUC (GOMS)	Log Loss (GOMS)
10	0.815	0.380
20	0.825	0.375
30	0.830	0.370
40	0.832	0.368
50	0.831	0.369

TABLE VI. DNN LAYERS OF STUDY

Number of Layers	AUC (GOMS)	Log Loss (GOMS)
3	0.828	0.373
5	0.830	0.370
7	0.831	0.369
9	0.830	0.371

TABLE VII. FEATURE INTERACTION VECTOR SIZE OF STUDY

Interaction Vector Size	AUC (GOMS)	Log Loss (GOMS)
10	0.830	0.370
20	0.828	0.372
30	0.829	0.371
40	0.829	0.371

H. Variable Importance Analysis

To identify the most important variables influencing job turnover, we use the feature importance scores from the FINN model. The top four variables are Monthly Salary, Job Satisfaction, Company Size, and Weekly Working Hours.

The importance scores indicate the relative impact of each variable on the prediction of job turnover. Table VIII and Fig. 4 illustrates the importance scores of these variables.

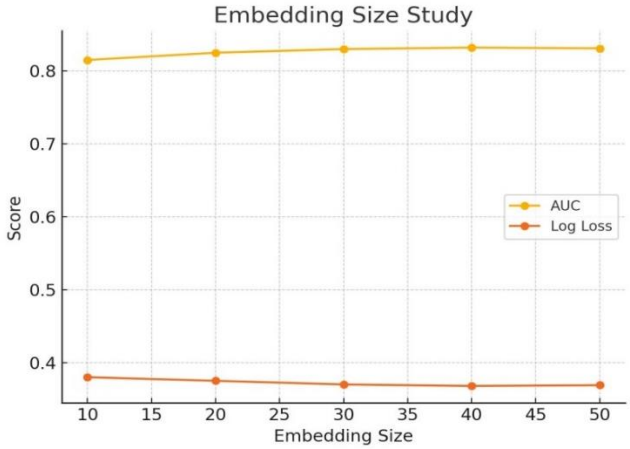


Fig. 1. Embedding size study.

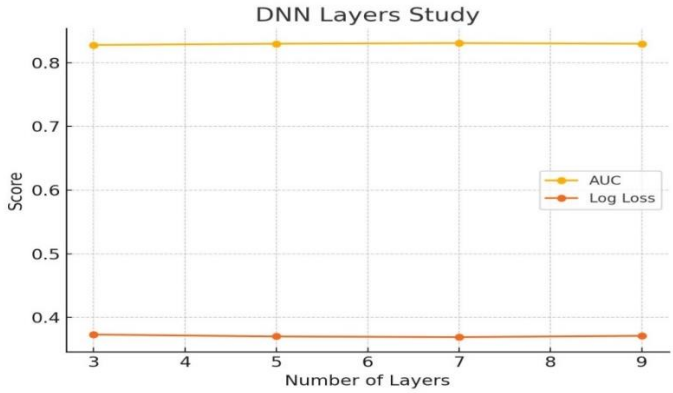


Fig. 2. DNN layers study.

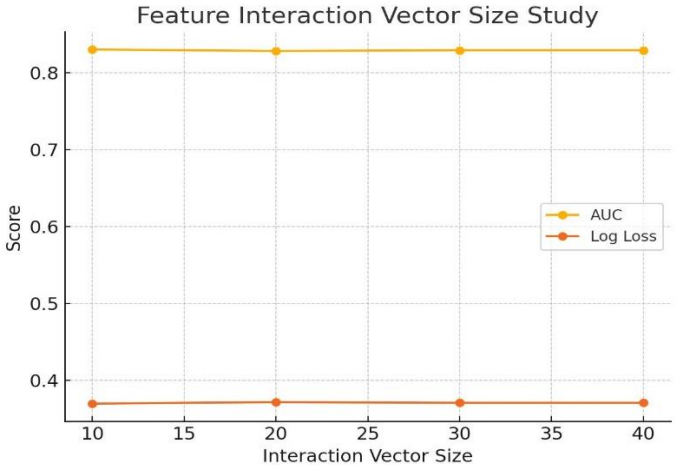


Fig. 3. Feature interaction vector size study.

TABLE VIII. VARIABLE IMPORTANCE OF STUDY

Variable	Importance Score
Monthly Salary	0.25
Job Satisfaction	0.22
Company Size	0.18
Weekly Working Hours	0.15

V. DISCUSSION

The prediction of job turnover among early career university graduates is a crucial task for both employees and employers. Accurate predictions can help employees navigate their career paths more effectively and assist employers in developing strategies to enhance employee retention, thereby reducing the costs associated with recruitment and training. This study proposes a Feature Interaction based Neural Network (FINN) model designed to address the complexities inherent in employment data and improve the accuracy of job turnover predictions.

In this paper, the results demonstrated that the FINN model outperforms all baseline models in terms of both AUC and Log Loss. Specifically, the FINN model achieved an AUC of 0.830 and a Log Loss of 0.370, indicating its superior ability to distinguish between employees who are likely to leave their jobs and those who are not. This performance can be attributed to the model's ability to effectively capture complex feature interactions through its feature interaction layer, which traditional models and even some advanced neural network models struggle to do [20, 21].

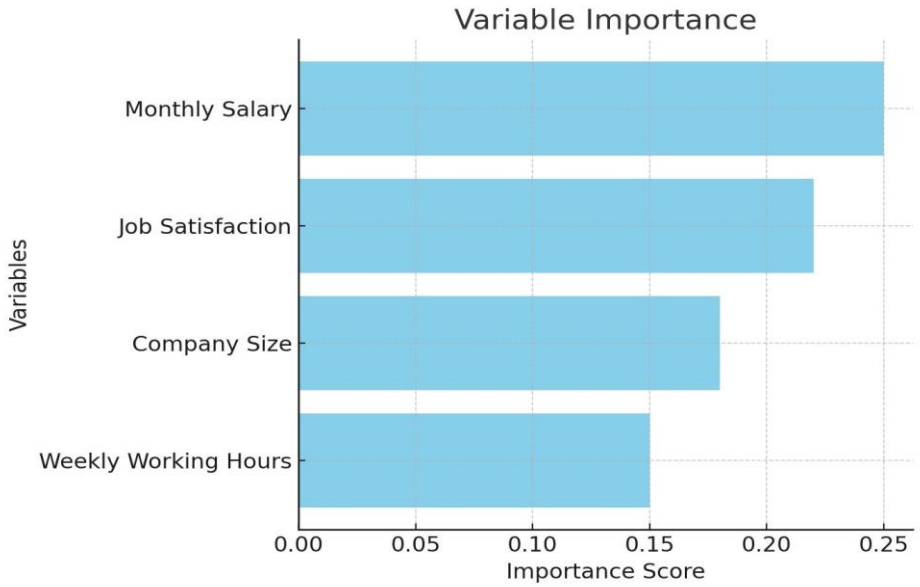


Fig. 4. Variable importance of study.

The study conducted an extensive parameter study to identify the optimal settings for the FINN model. This included varying the embedding sizes, the depth of the DNN, and the size of the feature interaction vector. The results provide valuable insights into the impact of these hyper-parameters on model performance [35,36]. Additionally, the analysis identified key variables that significantly influence job turnover, such as monthly salary, job satisfaction, company size, and weekly working hours. These findings can inform both policy and practice by highlighting areas where interventions might be most effective in reducing turnover rates. This study makes several significant contributions to the field of employment data analysis and predictive modeling. The primary contribution is the development and validation of the FINN model. This model enhances the predictive accuracy of job turnover by effectively modeling complex interactions between features. By leveraging the rich GOMS dataset, the study provides a detailed analysis of various factors influencing job turnover among early career graduates. This comprehensive dataset allows for a nuanced understanding of the predictors of job turnover. These findings have several practical implications. Employers can use the insights from the FINN model to develop targeted retention strategies. For instance, improving job satisfaction and offering competitive salaries could be effective measures to reduce turnover rates among early career employees. Career counselors and advisors can use the model's predictions to provide personalized guidance to graduates, helping them make informed decisions about their career paths and job choices. Policymakers can leverage the findings to design programs and policies aimed at improving job stability among young graduates. This could include initiatives to enhance job satisfaction and provide better working conditions.

While the FINN model has demonstrated significant improvements in predictive accuracy, there are several avenues for future research. First, future research could explore the integration of additional advanced techniques, such as attention mechanisms and sequence modeling, to further enhance the model's ability to capture complex feature interactions. Second, testing the FINN model on different datasets from various regions and industries could validate its generalizability and robustness across different contexts. Third, conducting longitudinal studies to track job turnover over a more extended

period could provide deeper insights into the long-term predictors of job stability and career success. Fourth, experimenting with different intervention strategies based on the model's predictions could help in identifying the most effective measures for reducing job turnover.

VI. Conclusion

In this paper, we have demonstrated that the FINN model represents a significant advancement in the field of predictive modeling for job turnover among early career graduates. By effectively capturing complex feature interactions and leveraging a rich dataset, the FINN model provides superior predictive performance compared to both traditional and contemporary models. The insights gained from this study have the potential to inform strategies and policies aimed at improving job retention and career outcomes for young professionals. As employment data analysis continues to advance, the FINN model is expected to provide a strong foundation for both future research and practical applications, enabling more precise and actionable predictions in the labor market.

Acknowledgment

This research Supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (NRF- RS-2023-00237287, NRF-2021S1A5A8062526) and local government-university cooperation-based regional innovation projects (2021RIS-003).

References

[1] J. M. Abowd and F. Kramarz, "The analysis of labor markets using matched employer-employee data," Handbook of Labor Economics, vol. 3, pp. 2629-2710, 1999.

[2] R. W. Crandall, W. Lehr, and R. E. Litan, "The effects of broadband deployment on output and employment: A cross-sectional analysis of US data," Washington, DC: Brookings Institution, vol. 6, 2007.

[3] M. Doede, "Race as a predictor of job satisfaction and turnover in US nurses," Journal of Nursing Management, vol. 25, no. 3, pp. 207-214, 2017.

[4] Y. Zhao, M. K. Hryniewicki, F. Cheng, B. Fu, and X. Zhu, "Employee turnover prediction with machine learning: A reliable approach," in Intelligent Systems and Applications: Proceedings of the 2018 Intelligent Systems Conference (IntelliSys) Volume 2, Springer International Publishing, pp. 737-758, 2019.

[5] M. Lazzari, J. M. Alvarez, and S. Ruggieri, "Predicting and explaining employee turnover intention," International Journal of Data Science and Analytics, vol. 14, no. 3, pp. 279-292, 2022.

[6] A. R. Skelton, D. Nattress, and R. J. Dwyer, "Predicting manufacturing employee turnover intentions," Journal of Economics, Finance and Administrative Science, vol. 25, no. 49, pp. 101-117, 2020.

[7] S. Sajjadiani, A. J. Sojourner, J. D. Kammeyer-Mueller, and E. Mykerezi, "Using machine learning to translate applicant work history into predictors of performance and turnover," Journal of Applied Psychology, vol. 104, no. 10, pp. 1207-1225, 2019.

[8] E. Rombaut and M. A. Guerry, "Predicting voluntary turnover through human resources database analysis," Management Research Review, vol. 41, no. 1, pp. 96-112, 2018.

[9] J. R. Carlson, D. S. Carlson, S. Zivnuska, R. B. Harris, and K. J. Harris, "Applying the job demands resources model to understand technology as a predictor of turnover intentions," Computers in Human Behavior, vol. 77, pp. 317-325, 2017.

[10] H. Min, B. Yang, D. G. Allen, A. A. Grandey, and M. Liu, "Wisdom from the crowd: Can recommender systems predict employee turnover and its destinations?," Personnel Psychology, vol. 77, no. 2, pp. 475-496, 2024.

[11] X. Gao, J. Wen, and C. Zhang, "An improved random forest algorithm for predicting employee turnover," Mathematical Problems in Engineering, vol. 2019, article ID 4140707, 2019.

[12] A. Coetzer, C. Inma, P. Poisat, J. Redmond, and C. Standing, "Does job embeddedness predict turnover intentions in SMEs?," International Journal of Productivity and Performance Management, vol. 68, no. 2, pp. 340-361, 2019.

[13] M. Akeyo and P. Wezel, "Factors Influencing Staff Turnover in Logistics Management," American Journal of Economics, vol. 1, no. 1, pp. 79-94, 2016.

[14] A. F. Schlechter, C. Syce, and M. Bussin, "Predicting voluntary turnover in employees using demographic characteristics: A South African case study," Acta Commercii, vol. 16, no. 1, pp. 1-10, 2016.

[15] H. J. Lee and Y. C. Cho, "Relationship between job satisfaction and turnover intention among nurses in general hospitals," Journal of the Korea Academia-Industrial Cooperation Society, vol. 15, no. 7, pp. 4404-4415, 2014.

[16] Y. Sun, Z. Luo, and P. Fang, "Factors influencing the turnover intention of Chinese community health service workers based on the investigation results of five provinces," Journal of Community Health, vol. 38, pp. 1058-1066, 2013.

[17] F. Horn, R. Pack, and M. Rieger, "The autofeat python library for automated feature engineering and selection," in Machine Learning and Knowledge Discovery in Databases: International Workshops of ECML PKDD 2019, Würzburg, Germany, September 16–20, 2019, Proceedings, Part I, Springer International Publishing, pp. 111-120, 2020.

[18] S. M. Fati, "A Loan Default Prediction Model Using Machine Learning and Feature Engineering," ICIC Express Letters, vol. 18, no. 1, pp. 27-37, 2024.

[19] A. Abdulhafedh, "Comparison between common statistical modeling techniques used in research, including: Discriminant analysis vs logistic regression, ridge regression vs LASSO, and decision tree vs random forest," Open Access Library Journal, vol. 9, no. 2, pp. 1-19, 2022.

[20] Y. Umuroglu, N. J. Fraser, G. Gambardella, M. Blott, P. Leong, M. Jahre, and K. Vissers, "Finn: A framework for fast, scalable binarized neural network inference," in Proceedings of the 2017 ACM/SIGDA international symposium on field-programmable gate arrays, pp. 65-74, 2017.

[21] F. Rezaei and A. Houmansadr, "FINN: Fingerprinting network flows using neural networks," in Proceedings of the 37th Annual Computer Security Applications Conference, pp. 1011-1024, 2021.

[22] M. M. Bejani and M. Ghatee, "A systematic review on overfitting control in shallow and deep neural networks," Artificial Intelligence Review, vol. 54, no. 8, pp. 6391-6438, 2021.

[23] S. Kost, O. Rheinbach, and H. Schaeben, "Using logistic regression model selection towards interpretable machine learning in mineral prospectivity modeling," Geochemistry, vol. 81, no. 4, p. 125826, 2021.

[24] R. Langone, A. Cuzzocrea, and N. Skantzos, "Interpretable Anomaly Prediction: Predicting anomalous behavior in industry 4.0 settings via regularized logistic regression tools," Data & Knowledge Engineering, vol. 130, p. 101850, 2020.

[25] J. Á. Martín-Baos, R. García-Ródenas, and L. Rodriguez-Benitez, "Revisiting kernel logistic regression under the random utility models perspective. An interpretable machine-learning approach," Transportation Letters, vol. 13, no. 3, pp. 151-162, 2021.

[26] Y. Izza, A. Ignatiev, and J. Marques-Silva, "On explaining decision trees," arXiv preprint arXiv:2010.11034, 2020.

[27] B. Charbuty and A. Abdulazeez, "Classification based on decision tree algorithm for machine learning," Journal of Applied Science and Technology Trends, vol. 2, no. 1, pp. 20-28, 2021.

[28] G. W. Cha, H. J. Moon, and Y. C. Kim, "Comparison of random forest and gradient boosting machine models for predicting demolition waste based on small datasets and categorical variables," International Journal of Environmental Research and Public Health, vol. 18, no. 16, p. 8530, 2021.

[29] H. Byeon, "Is the random forest algorithm suitable for predicting parkinson's disease with mild cognitive impairment out of parkinson's disease with normal cognition?," International Journal of Environmental Research and Public Health, vol. 17, no. 7, p. 2594, 2020.

[30] H. Byeon, "Can the random forests model improve the power to predict the intention of the elderly in a community to participate in a cognitive health promotion program?," Iranian Journal of Public Health, vol. 50, no. 2, pp. 315-324, 2021.

[31] H. Byeon, "Exploring the risk factors of impaired fasting glucose in middle-aged population living in South Korean communities by using categorical boosting machine," Frontiers in Endocrinology, vol. 13, p. 1013162, 2022.

[32] F. Lang, L. Liang, K. Huang, T. Chen, and S. Zhu, "Movie recommendation system for educational purposes based on field-aware factorization machine," Mobile Networks and Applications, vol. 26, no. 5, pp. 2199-2205, 2021.

[33] Y. Juan, Y. Zhuang, W. S. Chin, and C. J. Lin, "Field-aware factorization machines for CTR prediction," in Proceedings of the 10th ACM conference on recommender systems, pp. 43-50, 2016.

[34] Y. Juan, D. Lefortier, and O. Chapelle, "Field-aware factorization machines in a real-world online advertising system," in Proceedings of the 26th International Conference on World Wide Web Companion, pp. 680-688, 2017.

[35] A. Koutsoukas, K. J. Monaghan, X. Li, and J. Huan, "Deep-learning: investigating deep neural networks hyper-parameters and comparison of performance to shallow methods for modeling bioactivity data," Journal of Cheminformatics, vol. 9, pp. 1-13, 2017.

[36] N. Tran, J. G. Schneider, I. Weber, and A. K. Qin, "Hyper-parameter optimization in classification: To-do or not-to-do," Pattern Recognition, vol. 103, p. 107245, 2020.

A Systematic Review of Virtual Commerce Solutions for the Metaverse

Ghazala Bilquise[1], Khaled Shaalan[2], Manar Alkhatib[3]
Computer Information Science, Higher Colleges of Technology, Dubai, UAE[1]
Computer Science Department, The British University in Dubai, Dubai, UAE[2, 3]

***Abstract*—The metaverse, a rapidly evolving field, promises to transform online shopping through immersive technologies. This systematic review aims to explore and analyze the key design features of Virtual Commerce (v-commerce) solutions within this digital environment. By examining 24 studies that have developed immersive v-commerce applications, this review seeks to compile a taxonomy of essential design attributes necessary for creating effective and engaging v-commerce experiences. The review classifies these attributes into three primary dimensions: Product, Intelligent Services, and Functionality. The findings indicate that within the Augmented Reality (AR) category, product visualization and natural interaction were the most studied attributes. In the Virtual Reality (VR) category, intuitive affordances emerged as the most frequently investigated features. Meanwhile, Mixed Reality (MR) studies commonly focused on information quality, intuitive affordances, and shopping assistants. The insights from this review provide valuable guidance for researchers, developers, and practitioners aiming to enhance consumer engagement and satisfaction in the metaverse through well-designed v-commerce applications. By synthesizing the results of various studies, this review offers a comprehensive overview of the current state of v-commerce research, identifies existing gaps, and proposes potential directions for future development in the field.**

Keywords*—*Metaverse; v-commerce; immersive technologies; design attributes

I. INTRODUCTION

Over the past two decades, the online retail market has seen a significant increase, primarily driven by the rise of electronic commerce (e-commerce). The COVID-19 pandemic further accelerated this growth, leading to unprecedented levels of global e-commerce sales, which reached $5.7 trillion in 2022, representing a rise in market share from 18.8% to 19.7% [1]. Analysts predict that by 2026, the online segment will grow to $8.1 trillion, nearly a quarter of the total retail market share [2], [3], indicating a lasting shift in consumer behavior towards online shopping [4]. This shift is fueled by advancements in mobile and internet technologies, necessitating businesses to continuously adapt to stay competitive and enhance consumer shopping experiences.

The advent of the metaverse promises to transform online shopping into a more immersive, engaging, and seamless experience. Unlike traditional e-commerce, the metaverse allows users to interact through digital avatars in virtual storefronts, offering an enriched user experience powered by technologies like Augmented Reality (AR), Virtual Reality (VR), and Mixed Reality (MR) [5]. These immersive technologies, along with Artificial Intelligence (AI) powered chatbots and Internet of Things (IoT), enable personalized shopping experiences and provide retailers with valuable data analytics to tailor offerings to customer preferences [6]. For organizations to fully exploit these capabilities, it is crucial to select optimal v-commerce solutions that integrate the real and virtual worlds, enhance consumer engagement, and support business success. Furthermore, v-commerce attributes can also be used to assess, evaluate and compare multiple v-commerce solutions against each other to select the most optimal one.

To leverage the potential of the metaverse, organizations need to adopt v-commerce solutions that meet their specific needs and enhance consumer engagement and trust. However, there is no consensus among researchers on the effective design elements of these solutions [7]. Studies in Information Systems (IS) have explored factors influencing consumer behavior in immersive shopping environments [7], [8], [9], [10]. However, these studies often focus on behavioral intentions using self-reported data, which can be biased and require validation through more robust methodologies.

To effectively identify the essential design elements of metaverse v-commerce solutions, an investigation of Design Science (DS) research is recommended [11]. DS approach emphasizes the creation and evaluation of innovative IT artefacts and can provide a more rigorous foundation for understanding the impact of design attributes on consumer behavior [12], [13]. To the best of our knowledge, no study has comprehensively combined and presented design attributes for v-commerce solutions.

Several systematic reviews on v-commerce exist in literature [7], [9], [14]. However, the current systematic review diverges from the previous studies in several key aspects. First, the previous reviews focus on IS studies in v-commerce reporting behavioral constructs for the adoption of v-commerce solutions, and the theories applied to predict consumer purchase intentions. The aim is to inform practitioners about the essential factors needed for a v-commerce system to achieve its intended purpose. Second, previous reviews focus on a single immersive technology such as VR [7], or AR [9], [14]. This study seeks to investigate DS research to present the v-commerce attributes for all immersive environments focusing all the immersive technologies (AR, VR and MR). Furthermore, this review aims to identify the design elements for developing and evaluating v-commerce solutions.

This study employs a systematic approach to review and synthesize an analysis based on empirical studies that have

implemented v-commerce solutions. The study aims to categorize literature based on immersive technologies to reveal a taxonomy of the design attributes implemented in v-commerce solutions for consumer engagement. Additionally, it aims to outline future directions and research opportunities in this field.

The rest of the study is organized as follows. Section II provides background information on the metaverse and v-commerce. Section III presents the methodology of the systematic review, while Section IV presents the results and discussion of the review. Section V presents the conclusion of the study along with the research limitations and future work.

II. Background Information

A. Metaverse

The metaverse, initially coined in Neal Stephenson's 1992 science fiction novel "Snow Crash," is a collective virtual shared space created by the convergence of physical and virtual reality [15]. Over the past three decades, technological advancements have laid the foundation for a more tangible metaverse experience [16]. While still in its infancy, ongoing research and development are shaping its potential as the successor to the Internet.

The realization of the metaverse is made possible by various technologies, including immersive technologies, 3D computing, 5G and edge computing, AI, blockchain, and IoT [6], [17], [18]. These advancements offer high-speed, low-latency connectivity, personalized user experiences, secure digital identities, and seamless interactions within the metaverse. AI enhances user experiences by personalizing content and enabling intelligent interactions, while blockchain ensures secure transactions and digital ownership verification [6], [17], [18]. IoT bridges the physical and virtual worlds, mirroring real-world information within the metaverse [17], [19].

At the core of the metaverse experience are immersive technologies such as VR, AR and MR, collectively known as Extended Reality (XR). These technologies create a sense of presence in digital environments, allowing users to interact with and navigate the metaverse seamlessly. Low-immersion environments offer limited sensory stimulation, while high-immersion environments provide a rich, multi-sensory experience, enhancing the realism and interactivity of virtual objects and environments. The immersion spectrum of AR and VR technology is depicted in Fig. 1.

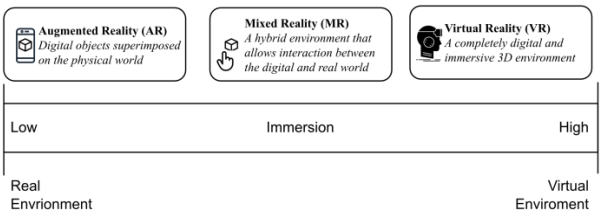


Fig. 1. Immersion spectrum of XR technologies.

Central to this immersive engagement are avatars, digital representations of users that facilitate social interactions within the metaverse [20]. Avatars range from simple 2D icons to complex 3D models and can perform various functions beyond visual representation, such as gestures and expressions. They enable users to communicate and collaborate in real-time, simulating real-world interactions [15]. The metaverse holds the potential to revolutionize socialization and interaction across various domains, including tourism [21], education, and retail [5], by providing immersive, interactive, and engaging experiences that integrate the physical and virtual worlds seamlessly. Designing v-commerce solutions in the metaverse requires addressing the challenge of creating unique, integrated experiences that enhance consumer engagement and business growth.

B. Virtual Commerce

The evolution of online shopping is rapidly transforming with the advent of new technologies that are quickly gaining traction among consumers. In this expansive digital landscape, v-commerce has emerged as a dynamic and transformative force, offering new opportunities and challenges for businesses and consumers alike. V-commerce is becoming a prominent feature of e-commerce, leveraging immersive XR technologies alongside AI to deliver captivating, three-dimensional shopping experiences [22]. These advanced digital technologies provide highly engaging and interactive experiences, immersing users in unique shopping environments [23].

V-commerce technologies allow consumers to explore virtual stores and examine products in a way that closely replicates the experience of shopping in person. This paradigm shift promises to revolutionize shopping practices by offering enhanced convenience and customer satisfaction [24]. Unlike traditional e-commerce, v-commerce provides immersive experiences that engage consumers through various sensory dimensions, including visual, auditory, kinetic, and tactile stimuli. This approach opens numerous opportunities for increased customer engagement, streamlined purchasing processes, stronger connections, and enhanced profitability.

The evolution of e-commerce towards virtual realms and encounters is evident as several stores facilitate immersive experiences for consumers. For instance, Ikea has developed a virtual reality app that allows customers to visualize furniture in their own rooms before purchasing [25]. Fashion retailer Zara introduced an AR shopping assistant app that displays clothing worn by models when customers point their cameras at the store. Additionally, brands like Sephora and L'Oreal use immersive technologies to let consumers digitally try on makeup products [26].

The convergence of v-commerce and the metaverse, which includes virtual environments where users can interact with various digital assets, offers countless possibilities for creating distinctive and immersive shopping experiences, such as virtual product explorations, virtual social interactions, virtual try-ons, and even virtual malls. In conclusion, v-commerce is poised to transform the online shopping landscape through rich, interactive, and immersive experiences.

III. Methodology

This paper uses a systematic approach to examine empirical research studies that have implemented v-commerce solutions.

For this purpose, the study adopts the systematic review framework established by [27]. This framework was selected over other frameworks, such as [28], since it provides guidelines specifically for conducting reviews in the technical field. Employing a rigorous theoretical framework is crucial for guiding the comprehensive data collection and analysis methods required for ensuring the reliability of the results. The systematic literature review guidelines by [27] offer a thorough approach to collecting, analyzing, and documenting findings from secondary data sources. By following this methodology, we aim to perform a thorough analysis of the included articles in the review and uncover the v-commerce design attributes to guide practitioners and researchers and further the study in this field.

The review process is segmented into five distinct phases: identifying the data sources, establishing the search strategy with the inclusion/exclusion criteria, selection of the articles using the PRISMA framework, performing the meta-analysis and finally reporting the results. The next subsections explain each phase and the outcomes of the phase. Results of the review are presented in Section IV.

A. Data Sources

Choosing the appropriate data sources is a crucial step in an Systematic Literature Review (SLR) process. To this end, three multidisciplinary databases were utilized as sources for collecting relevant articles – Web of Science (WoS), Scopus and ScienceDirect. The suitability and relevance of these databases has been acknowledged in numerous SLR articles published in well-regarded journals and in scientific research articles, demonstrating their resilience and scholarly evaluation [29]. All the selected databases encompass multiple publications across diverse scientific research domains and from various academic disciplines. Both Scopus and WoS are multidisciplinary databases encompassing peer-reviewed literature in science, medicine, social science, technology, arts and humanities, making them valuable for accessing a broad spectrum of research. ScienceDirect is a comprehensive resource for articles in the scientific, technical and medical fields.

B. Search Process

The search was conducted in two iterative rounds, the first in December 2023 and second in Jan 2024, to ensure that the latest articles are not left out. Additional studies were also sourced using a snowballing technique. The keywords representing immersive technologies such as ‘virtual reality’, ‘augmented reality’, ‘mixed reality’, ‘virtual world’ and ‘virtual environment’ were combined with the keywords ‘virtual commerce’, ‘virtual shopping’, ‘virtual store’ using Boolean operators to limit the search to the respective domain. Moreover, the search encompassed the title, abstract, and keywords to broaden the search space and ensure the inclusion of relevant studies. The formulated keywords search string is given below:

("virtual commerce" or "virtual shop*" or "virtual store") AND ((virtual AND (reality OR environment OR world)) OR ((mixed OR augmented) AND reality)).

A set of inclusion exclusion criteria, presented in Table I, was established to outline the crucial characteristics of the studies incorporated in the research.

TABLE I. INCLUSION / EXCLUSION CRITERIA

Inclusion Criteria	Exclusion Criteria
Must be an empirical study focused on the development of v-commerce applications.	Studies that are qualitative or unrelated to v-commerce application development
Must detail characteristics or design features of the developed application	Studies that do not describe or include any attribute of v-commerce application
Must be published in a peer-reviewed journal or conference proceedings	Publications such as books, book chapters, reviews, or articles not yet published
Must be written in English	Papers written in languages other than English
Must have been published between 2011 and 2024	Papers published before 2011

This study delves into the trends of research by analyzing keywords present in literature, particularly focusing on v-commerce solutions utilizing immersive technologies. After conducting the initial search, which yielded 525 results, the key terms from these findings were visualized using VOSviewer to identify prominent themes in the literature. VOSviewer is an advanced software tool designed for visualizing and analyzing scientific literature [30]. By examining the external characteristics of the data, it enables statistical and mathematical analysis to uncover trends and features within specific disciplines. Fig. 2 illustrates the visualization results. The diagram highlights the significance and interconnections among frequently occurring terms extracted from the abstracts, titles, and keywords of the search results. The size and label of each term indicate its importance, while the color represents clusters in the visualization. Each cluster comprises terms related to one another within the group, and the distance between clusters signifies their relatedness.

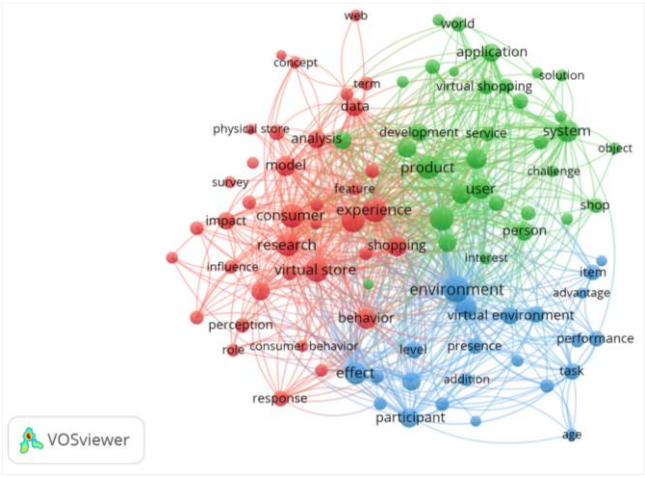


Fig. 2. Visualization of key terms in search results.

The visualization of key terms present in the titles and abstracts reveal three primary clusters, which clearly indicates the divide in the extant literature between IS and DS research. The red cluster encompasses keywords like consumer behavior, purchase intention, perception, influence,

relationship, and model, indicating a research focus on behavioral outcomes concerning v-commerce and purchase intentions arising from IS studies. The green cluster, on the other hand, features terms such as product, solution, service, application, technology, interaction, and development, showcasing a trend towards DS research for v-commerce solution development. Lastly, the blue cluster includes terms like performance, time, advantage, reflecting the evaluation of both IS and DS studies. This differentiation between IS and DS research underscores the need for mapping findings between the two domains.

C. Study Selection

The selection of articles for the systematic review was processed using the PRISMA framework [31] depicted in Fig. 3. This framework offers comprehensive guidelines and an organized method for reviewing documents. It consists of four main phases after performing the keyword search. The first phase is the identification phase in which the articles are retrieved from the identified databases and the duplicate articles are removed. Second is the initial screening phase, in which the inclusion/exclusion criteria are utilized to screen the articles based on the abstracts. Third is the eligibility phase in which a thorough full-text screening is performed to assess the relevance of the articles and to ensure that the article meets the requirements of the SLR. Finally, the last phase involves performing a quality assessment of the articles to ensure reliability of the findings. The application of each of the four phases using the PRISMA model is discussed next.

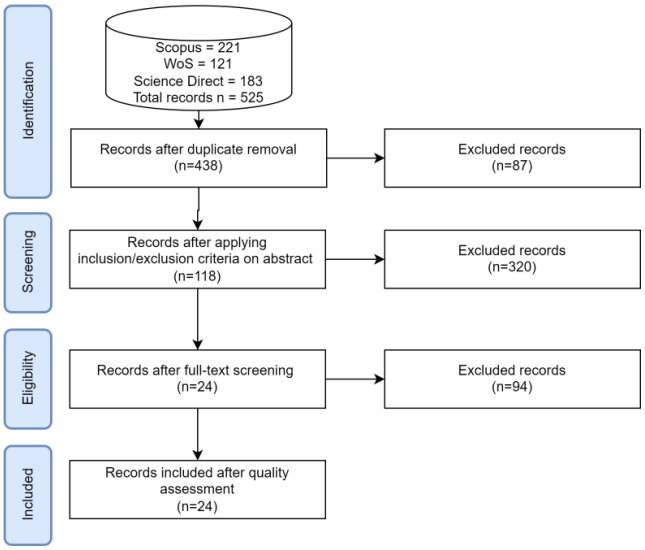


Fig. 3. PRISMA framework.

First, the search yielded 525 results across all databases. The removal of duplicate articles reduced the number of studies to n=438. Next, the abstract screening process was used to screen irrelevant articles which further reduced the number of studies to n=118. During the abstract screening process, numerous studies were excluded as they did not contribute to the context of this research such as the studies which were not empirical research in the field of v-commerce solutions development. In the eligibility phase a full-text screening of the articles led to the removal of all irrelevant studies that did not meet the inclusion criteria requirement thereby reducing the number of articles to n=24. A through screening of the articles revealed that several studies were irrelevant due to multiple reasons such as review articles, articles not in English, unavailability of full-text, articles not related to v-commerce application or solution development.

Finally, a quality assessment was conducted to ensure the reliability and validity of the included articles. It is important to note that the purpose of the quality assessment is to evaluate the relevance of the selected articles to this study's objectives, rather than to critique individual studies or their findings. Ensuring the validity of results and minimizing bias are crucial aspects of systematic reviews, highlighting the importance of quality assessment [32]. Additionally, quality assessment helps refine inclusion and exclusion criteria and contributes to the overall rigor of the review process [27]. The quality assessment checklist, was adapted from an extant SLR study [29]. It comprises of criteria that are crucial for the encoding process such as detailed description of the design elements and evaluation of the application design. The quality assessment also considers the credibility of the source and relevance of the publication to the research community. Each reviewed article was assessed using a quality checklist by assigning a score of 1 when the criteria is met, 0.5 if it is partially met or 0 if it is not met. The total score was converted to a percentage value. Only studies scoring 75% and above are considered for the review. Notably, all the studies passed the quality assessment check and were included in the meta-analysis.

D. Data Analysis and Coding

The objective of this phase is to thoroughly document the systematic review findings by gathering meta-data from the primary studies included in the review, with a focus on the research questions. A thorough data analysis of the relevant features identified during the planning phase is conducted to achieve this objective. The meta-data analysis encompasses various characteristics crucial for achieving the objective of the study such as identifying the design attribute implemented in the study and categorizing the attribute to its respective dimension.

E. Reporting the Review

In the concluding phase of the systematic review, the study results are unveiled. Section IV presents a comprehensive analysis of the meta-data extracted from the full-text review, with the aim of revealing the v-commerce design attribute implemented in the studies. This analysis presents and discusses the insights from the collected data, providing a thorough understanding of the findings and their implications within the context of the objectives of the study.

IV. Results and Discussion

The scope of the SLR included a qualitative analysis of 24 studies. The studies were categorized based on the immersive technologies utilized in the experimental design. Out of the 24 studies, nine studies developed AR solutions; nine studies developed VR solutions while six studies developed MR solutions. The meta-analysis aimed to uncover the design features in v-commerce applications implemented for respective immersive technologies. This section presents the

design attributes extracted from all the reviewed studies and presents a critical review of the implementation leading to a taxonomy of v-commerce design attributes.

A. V-Commerce Design Attributes

The results of the analysis revealed a total of 13 design attributes that were used to enhance shopping experience in v-commerce solutions. The design attributes were further categorized into three main dimensions: Product, Intelligent Services and Functionality. The definition of each attribute within each dimension is summarized below.

1) Product attributes: Product attributes encompass the richness and presentation of product information within an immersive environment. These design features are crucial for users to access detailed product information and interact effectively, enabling informed decisions. Through the literature review product attributes were categorized into four sub-attributes namely information quality (IQ), product visualization (PV), intuitive affordance (IA), and realistic modeling (RM). Each sub-attribute is described below.

a) Information Quality (IQ): Refers to the richness and presentation of the product information in the application. Information quality in a 3D environment may be defined as the quality of product information provided as direct information as well as indirect information revealed through interactions [33]. The richness and abundance of product information empowers consumers to make well-informed purchasing decisions. Utilizing 3D environments, simulations, or augmented reality can optimize the depth of information provided and elevate brand interaction and purchasing intent by improving perceived information accuracy and fostering positive brand perceptions [34]. Interaction with the product can also assist the consumer to gather more information while in an immersive environment.

b) Product Visualization (PV): PV, often implemented as Virtual try-on feature, in an immersive environment fosters customer involvement and provides an opportunity to test the items in order to make an informed purchase decision by reducing product ambiguity [35]. Several studies use algorithms to track the position of the surroundings or the human body to accurately display virtual accessories such as eyeglasses and clothing.

c) Intuitive Affordance (IA): IA refers to designing user interfaces that facilitate natural interactions with virtual objects, including actions like gaze, pinch, touch, grab, and hand gesture recognition, among others [36]. An immersive environment engages multiple human senses. To fully engage users in an immersive environment, the application must cater to their natural expectations for varied sensory experiences [37]. Therefore, virtual interactions that aligned with human perceptions of physical interactions create a more natural interaction.

d) Realistic Modelling (RM): Methods to render a product and/or the virtual environment in such a way that it is perceived to be realistic by users. Realistic modelling enhances perceived presence in the immersive environment [38], [39] and enhances the level of satisfaction [39].

2) Intelligent services: Intelligent Services within v-commerce applications leverage AI, machine learning, and data analytics to enhance the shopping experience. These services encompass the integration of agents that aid users in navigating the online environment, recommend products, and simulate the assistance provided by in-store human assistants. Additionally, AI technologies like computer vision and geolocation features can be integrated to offer a more intuitive and seamless interactive experience within the environment. Five sub-attributes were identified in this category, which include Natural Interaction (NI), Navigation Agent (NA), Recommendation Agent (RA), Shopping Assistant (SA) and Personalization (PS). Each sub-attribute is described below.

a) Natural Interaction (NI): NI involves utilizing technologies that enable users to engage with the application in intuitive and expected ways, for example using computer vision for face or body detection instead of using markers or silhouettes in AR [40], facial expression detection, natural language conversations, gesture recognition in a VR environment [41].

b) Navigation Agent (NA): A smart digital assistant that aids users in navigating a virtual space by providing guidance and directing them on how to interact with objects [42]. Navigation agents may take the shape of an embodied agent with an avatar or a virtual tour.

c) Shopping Assistant (SA): A virtual agent that provides information about the products or services in the virtual environment. The goal of the shopping assistant is to provide a new digital dimension to shopping mimicking the in-store assistants by answering basic FAQs, providing information on deals, coupons, and subscriptions [43].

d) Recommendation Agent (RA): RA utilize AI technologies to suggest products based on user preferences and behaviors. [44]. Product recommendations offer an efficient search option for end users and thereby enhance the brand attitude, user satisfaction and attitude towards technology [45].

e) Personalization (PS): Providing customized information to users based on characteristics, or preferences. Studies reveal that tailoring content to an individual's needs and interests is associated has a greater appeal and convincing impact on the consumer [46]. Moreover, increased personalization enhances brand attitude and user satisfaction [46].

3) Functionality design: Functionality attributes in v-commerce applications refer to the design features that aim to enhance interactivity, engagement, and control in an immersive environment. Interactive features are known to positively influence consumer satisfaction [47], [48] and provide an enjoyable shopping experience [38], [49], [50]. Four design attributes were identified in this category, namely Customization (CS), Navigability (NV), Seller Reputation

(SR) and Social Commerce (SC). Each sub-attribute is described below.

a) Customization (CS): CS is a feature of immersive technology that allows the user to manipulate and customize the product and/or the virtual environment. Incorporating VR technology for product customization is expected to improve task-related factors and enhance the perceived value of the user experience [51]. Moreover, these systems may enable users to personalize a 3D layout and adjust specific properties of virtual objects.

b) Navigability (NV): NV refers to the design of functions to enable user navigation or movement through the immersive environment [11]. It enhances the system's ease of use and is essential for creating a sense of physical presence within the immersive environment [52], [53], [54].

c) Seller Reputation (SR): SR is a technique for objectively measuring a seller's credibility based on customer feedback and personal ratings. A mechanism to rate seller's reputation promotes transparency and enhances repurchase intentions in an online environment [55].

d) Social Commerce (SC): Enabling social presence by adding features of collaborative shopping and social activities such as sharing reviews, voting, likes and more [56]. Social commerce uses technologies to blend shopping with social activities to develop a sense of connection with the e-retailer and thereby enhancing the degree of social presences in a digital environment.

B. Critical Review of V-Commerce Attribute Implementation

This section presents a summary of the design attribute implementation within the respective immersive technology (AR, VR, MR) and develops a taxonomy of the design attributes. Out of the 24 reviewed studies, nine studies belonged to the AR category, nine to the VR category and six to the MR category. Fig. 4 presents the taxonomy, which illustrates the mapping of the design attributes implemented in the reviewed studies.

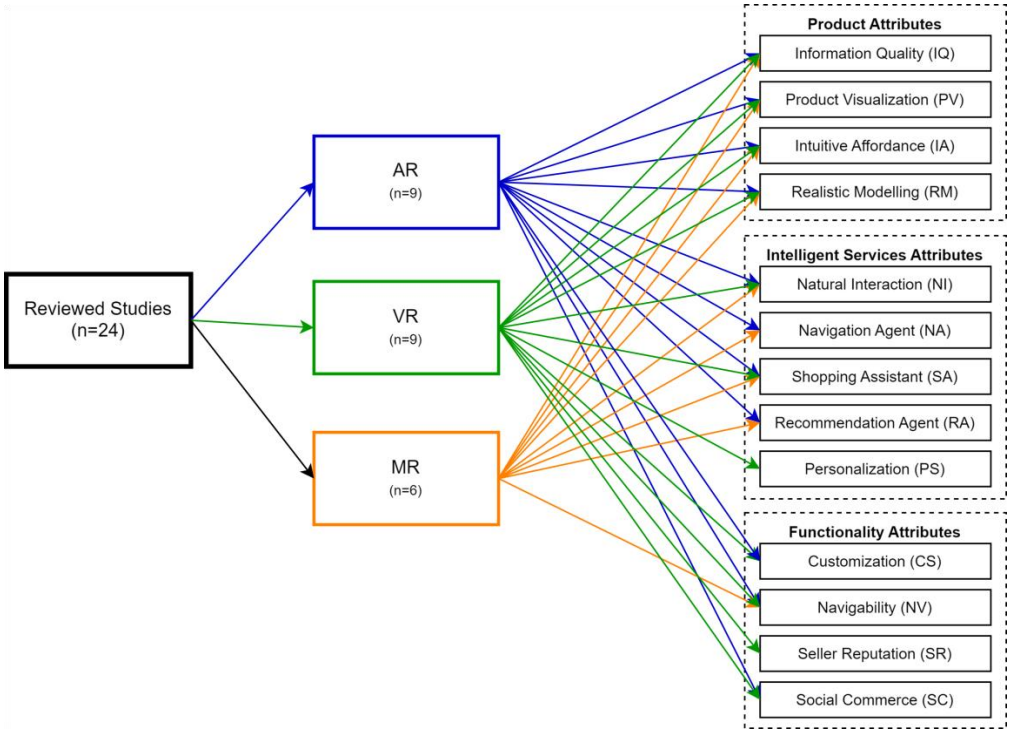


Fig. 4. Taxonomy of v-commerce design attributes.

1) AR studies: AR technologies superimpose digital information onto the real physical environment creating a new, combined experience that can be interactive and dynamic [57]. AR uses various devices such as smartphones, tablets, or specialized glasses to display computer-generated images, videos, or other virtual content in real time [14]. AR has the potential to offer new and innovative ways of engaging users in the shopping experience through product visualization, virtual try-ons and more [38], thereby providing consumers with a more immersive, engaging, and interactive shopping experience, leading to increased customer satisfaction and sales.

An analysis of the reviewed studies under the AR category revealed that numerous researchers focused on multiple design features of AR applications to enhance the consumer's shopping experience. Under the product dimension, product visualization and virtual try-ons (PV) were the most investigated design attributes by multiple researchers in the form of trying on clothes, glasses and visualizing furniture and toys [35], [40], [58], [59], [60]. Several studies enhanced the Information Quality (IQ) of the products by providing detailed product information [58], [61], information placement, and provision of additional relevant information [61], [62], while [63] enhanced Information quality by using explainable recommendations, and product comparisons. Intuitive Affordance (IA) was a focus of four out of nine studies in AR.

The research in [62] and [60] implemented intuitive affordance features by adding product rotating, zooming capabilities. The research in [58] implemented a direct selection feature that allows consumers to pick and try on glasses published by other users. While [63] added intuitive visual elements to ease navigation and discovery. Only two studies [62], [64] focused on enhancing the 3D quality ad Realistic Modelling (RM) of the product.

In the Intelligent Services dimension, Natural Interaction (NI) was among the highly researched design areas, and the focus of six out of nine studies. Computer vision used to naturally detect users' body [40], facial data [35], [58], or perform a product search [60]. On the other hand, [62] used deep learning to detect users current in-store location. The research in [60] also focused on other intelligent services such as a Navigation Agent (NA), shopping assistant (SA). Both [60] and [63] added product recommendations using Recommendation Agents (RA) to enhance the user's shopping experience.

In the functionality dimension three studies [40], [59], [62] added Customization (CS) features on the AR shopping app such by allowing users to customize the product size, color, type or more. Four out of nine studies focused on navigability (NV) by providing easy search features and filter features for products [58], [59], [60] or easy in-store navigation [62]. The research in [40], [58] and [59] enhanced the social presence in the AR shopping app by adding the feature allowing users to share their try-ons with others, thus promoting Social Commerce (SC).

Overall, in the AR category, NA and PV were among the most investigated design attributes with 67% (n=6) and 56% (n=5) studies experimenting with these features respectively. The IQ, IA attributes were studied by 44% (n=4) studies, while the PS and SR attribute were not investigated by any study.

2) VR studies: VR technologies simulate realistic and immersive experience in a three-dimensional environment. VR allows consumers to see and interact with a product in a virtual environment, providing a more immersive and realistic experience than traditional 2D images or videos. With VR, businesses can create virtual storefronts and showrooms, allowing customers to browse and providing a multisensory purchase experience [10].

Among the reviewed studies in the VR category, several researchers focused on multiple VR design attributes to applications to enrich the consumer's shopping experience. In the product dimension, Intuitive Affordance (IA) and information quality (IQ) were the most investigated criteria. Three studies, [65], [66] and [67] focused on product presentation and information to enhance the information quality of the user experience. In addition, [67] also provided detailed product information with price comparisons to enhance the quality of information. Intuitive affordance was investigated by four studies by incorporating features such as grabbing, picking, rotating and adding items to the cart [65], [66], [67], [68]. Two studies focused on Realistic Modelling (RM) of the products [66], [68], while only one study incorporated the product visualization (PV) feature as virtual try on for clothes [67].

In the intelligent services dimension, natural interaction (NI) in a shopping environment was investigated by five researchers in the form of using speech, eye and gaze interaction and virtual touch features [65], [66], [68], [69]. Moreover, [67], [70] also included the feature of an intelligent shopping assistant (SA) in the virtual environment. The research in [71] incorporated the feature of a personalization (PS) based on users shopping profile. Navigation agents (NA) and recommendation agents (RA) were not studied in a virtual shopping environment by any of the studies.

Navigability (NV) was the most investigated criteria in the functionality dimension. Navigability was implemented with the use of controllers to ease navigation [67], simple and easy environment with clear signs [68], 3D navigation for ease of movement and turns [65], [66]. The research in [72] investigated the feature of customizing (CS) the virtual store. Fang et al. (2014) implemented the seller reputation (SR) feature by using a five senses reputation mechanism to assess the reputation of the seller in the virtual environment. The research in [67] added a social commerce feature by providing users a sense of shopping with other users by showing their online presence.

In the VR category, the NI design attribute is one of the highest investigated design features investigated in 44% (n=4) studies. In addition, IA and navigability NV were also investigated by 44% (n=4) studies. The IQ feature was investigated by 33% (n=3) studies. Furthermore, in the VR studies category the NA and RA attributes were not investigated by any study.

3) MR studies: MR is an emerging technology that integrates both virtual and augmented reality elements to create a blended, hybrid environment which includes both real and virtual objects [74]. Unlike AR, MR enables virtual objects to interact with the physical world. MR offers all the benefits of AR with an addition of interaction. Consumers can not just visualize or try the product before purchase, but also interact with it in real-time. Thus, MR allows for a more natural and intuitive interaction between the consumers and the digital products, leading to a more immersive experience.

From the reviewed studies in the MR category, four studies investigated the online shopping experience using MR technology. The research in [75] implemented a total of six design attributes. The study included three product attributes in the form of virtual try-ons (PV), intuitive affordance (IA) using with a full hand manipulation feature and enhanced product information. Furthermore, [75] also added intelligent services with natural interaction (NI) using computer vision for a smart product detection feature and recommendation engine (RA). Navigability (NV) was enhanced by incorporating ease of searching.

The research in [76] applied eight design attributes to an immersive shopping environment. In the product category, all the attributes were implemented (information quality (IQ), virtual try on (PV), intuitive affordance (IA), realistic

modelling (RM)). The research in [76] also applied intelligent services to the shopping application in the form of a navigation agent (NA) that provides a tour of the application, a shopping assistant (SA) that answers users queries and a product recommendation (RA) system that suggests products based on user's traits. Finally, navigability (NV) of the system was enhanced by an avatar-based guidance system and providing a 360 view of shopping center with free walking and movement.

The research in [77] on the other hand implemented three design attributes (product visualization (PV), natural interaction (NI) and a voice-based shopping assistant (SA)) for a fashion store. The research in [78] implemented two design attributes, intuitive affordance (IA) using a feature to interact with a virtual hand, and realistic modelling (RM) by incorporating a sense of agency on the virtual hand to make it appear realistic. The research in [79] developed an MR shopping assistant using Microsoft HoloLens, which offers product information, reviews, and recommendations to the shopper. In a later study [80] enhanced the shopping assistant application to include more features. Their prototype incorporated features such as providing product information using textual and video-based formats to enhance the Information Quality (IQ), product reviews from other consumers (SC), recommendations (RA). The study also Intuitive Affordance (IA) using gesture-based interactions such as air tap, touch, drag and hand menu. Moreover, natural interaction (NI) was employed using computer vision technology to recognize images in the user's field of vision.

In the MR category, IQ, IA and SA were the most investigated design attributes by 67% (n=4) of the studies, while the attribute of PV was implemented by 50% (n=3) of the studies. The attributes of CS and SR were not investigated in any of the MR application design studies. Table II summarizes the design attributes implemented in the reviewed studies.

TABLE II. SUMMARY OF IMPLEMENTED V-COMMERCE ATTRIBUTES

		Product				Intelligent Services					Functionality			
	Reviewed Studies	*IQ*	*PV*	*IA*	*RM*	*NI*	*NA*	*SA*	*RA*	*PS*	*CS*	*NV*	*SR*	*SC*
AR	[40]	×	✓	×	×	✓	×	×	×	×	✓	×	×	✓
	[61]	✓	×	×	×	×	×	×	×	×	×	×	×	×
	[35]	×	✓	×	×	✓	×	×	×	×	×	×	×	×
	[58]	✓	✓	✓	×	✓	×	×	×	×	×	✓	×	✓
	[64]	×	×	×	✓	✓	×	×	×	×	×	×	×	×
	[62]	✓	×	✓	✓	✓	×	×	×	×	✓	✓	×	×
	[59]	×	✓	×	×	×	×	×	×	×	✓	✓	×	✓
	[63]	✓	×	✓	×	×	×	×	✓	×	×	×	×	×
	[60]	×	✓	✓	×	✓	✓	✓	✓	×	×	✓	×	×
VR	[65]	✓	×	✓	×	×	×	×	×	×	×	✓	×	×
	[73]	×	×	×	×	×	×	×	×	×	×	×	✓	×
	[68]	×	×	✓	✓	✓	×	×	×	×	×	✓	×	×
	[66]	✓	×	✓	✓	✓	×	×	×	×	×	✓	×	×
	[71]	×	×	×	×	×	×	×	×	✓	×	×	×	×
	[67]	✓	✓	✓	×	✓	×	✓	×	×	×	✓	×	✓
	[70]	×	×	×	×	×	×	✓	×	×	×	×	×	×
	[69]	×	×	×	×	✓	×	×	×	×	×	×	×	×
	[81]	×	×	×	×	×	×	×	×	×	✓	×	×	×
MR	[75]	✓	✓	✓	×	✓	×	×	✓	×	×	✓	×	×
	[77]	×	✓	×	×	✓	×	✓	×	×	×	×	×	×
	[76]	✓	✓	✓	✓	×	✓	✓	✓	×	×	✓	×	×
	[78]	×	×	✓	✓	×	×	×	×	×	×	×	×	×
	[79]	✓	×	×	×	×	×	✓	✓	×	×	×	×	✓
	[80]	✓	×	✓	×	✓	×	✓	✓	✓	×	×	×	✓

V. CONCLUSION

This systematic review has identified and synthesized the essential design attributes of v-commerce solutions within the metaverse. The study reviewed twenty-four empirical studies which have implemented v-commerce solutions and categorized them based on the three immersive technologies – AR, VR, and MR. The meta-analysis of the reviewed studies revealed thirteen v-commerce design attributes. The attributes were further classified into these attributes into three main

dimensions: Product, Intelligent Services, and Functionality. Each dimension encompasses specific design features that collectively enhance consumer engagement, satisfaction, and overall shopping experience in immersive environments. Results of the systematic review revealed that In AR category, the most investigated attributes were PV and NI. In the VR category, IA and NI were the most frequently investigated features. Lastly, in the MR studies, IQ, IA, and SA were the most common.

This comprehensive taxonomy of design attributes not only provides valuable insights for researchers and practitioners but also sets the stage for future innovations in v-commerce. By leveraging these attributes, businesses can create more engaging and effective v-commerce solutions that meet the evolving needs of consumers in the metaverse.

A. Limitations and Future Directions

Despite the comprehensive nature of this review, several limitations must be acknowledged. First, the scope of the review was limited to empirical studies published between 2011 and 2024, which may exclude relevant research outside this timeframe. Second, the reliance on peer-reviewed journal articles and conference proceedings may have omitted valuable insights from industry reports, white papers, and other non-academic sources. Third, the review primarily focused on studies written in English, potentially overlooking significant contributions in other languages. Additionally, the variability in methodologies and contexts of the included studies may affect the generalizability of the findings.

Future research should address these limitations by expanding the scope to include a broader range of sources and languages. Further research could focus on developing and testing comprehensive frameworks that combine design science and information systems research can provide more robust insights into the effective design and implementation of v-commerce solutions in the metaverse. Finally, incorporating the design attributes in decision-making processes can help businesses create more engaging and effective v-commerce solutions. Researchers can apply decision making methods to evaluate v-commerce solutions using the identified attributes as evaluation criteria. By addressing these future directions, researchers and practitioners can build on the foundation laid by this review, advancing v-commerce and shaping the future of online shopping through immersive and interactive experiences.

REFERENCES

[1] M. Keenan, “Global Ecommerce Explained: Stats and Trends to Watch,” Global Commerce, Industry Insights and Trends. Accessed: Mar. 15, 2023. [Online]. Available: https://www.shopify.com/enterprise/global-ecommerce-statistics.

[2] Obrelo, “eCommerce Share of Retail Sales (2021-2026),” Obrelo. Accessed: Jan. 21, 2023. [Online]. Available: https://www.oberlo.com/statistics/ecommerce-share-of-retail-sales.

[3] Statista, “e-Commerce as percentage of total retail sales worldwide from 2015 to 2021 with forecasts from 2022 to 2026,” Statista. Accessed: Jan. 21, 2023. [Online]. Available: https://www.statista.com/statistics/534123/e-commerce-share-of-retail-sales-worldwide/.

[4] M. Stanley, “Here’s Why E-commerce Growth Can Stay Stronger for Longer,” MorganStanley. Accessed: Mar. 15, 2023. [Online]. Available: https://www.morganstanley.com/ideas/global-ecommerce-growth-forecast-2022.

[5] Y. K. Dwivedi et al., “Metaverse beyond the hype: Multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy,” Int. J. Inf. Manage., vol. 66, no. July, p. 102542, 2022, doi: 10.1016/j.ijinfomgt.2022.102542.

[6] M. Trunfio, “Advances in Metaverse Investigation : Streams of Research and Future Agenda,” pp. 103–129, 2022.

[7] L. Xue, C. J. Parker, and H. McCormick, A virtual reality and retailing literature review: Current focus, underlying themes and future directions. Manchester: Springer International Publishing, 2018. doi: 10.3233/978-1-60750-873-1-327.

[8] B. Shen, W. Tan, J. Guo, L. Zhao, and P. Qin, “How to promote user purchase in metaverse? A systematic literature review on consumer behavior research and virtual commerce application design,” Appl. Sci., vol. 11, no. 23, pp. 1–29, 2021, doi: 10.3390/app112311087.

[9] R. Chen, P. Perry, R. Boardman, and H. McCormick, “Augmented Reality in Retail: A Systematic Review of Research Focus and Future Research Agenda,” Manchester Metrop. Univ., vol. 50, pp. 498–518, 2022.

[10] N. Xi and J. Hamari, “Shopping in virtual reality: A literature review and future agenda,” J. Bus. Res., vol. 134, no. May, pp. 37–58, 2021, doi: 10.1016/j.jbusres.2021.04.075.

[11] E. Dincelli and A. Yayla, “Immersive virtual reality in the age of the Metaverse: A hybrid-narrative review based on the technology affordance perspective,” J. Strateg. Inf. Syst., vol. 31, no. 2, p. 101717, 2022.

[12] A. R. Hevner, S. T. March, J. Park, and S. Ram, “Design Sceince in Information Systems,” MIS Q., vol. 28, no. 1, pp. 75–105, 2004.

[13] K. Peffers, T. Tuunanen, M. A. Rothenberger, and S. Chatterjee, “A design science research methodology for information systems research,” J. Manag. Inf. Syst., vol. 24, no. 3, pp. 45–77, 2007.

[14] M. Riar, N. Xi, J. J. Korbel, R. Zarnekow, and J. Hamari, “Using augmented reality for shopping: a framework for AR induced consumer behavior, literature review and future agenda,” Internet Res., no. 210301, 2022, doi: 10.1108/INTR-08-2021-0611.

[15] S. M. Park and Y. G. Kim, “A Metaverse: Taxonomy, Components, Applications, and Open Challenges,” IEEE Access, vol. 10, pp. 4209–4251, 2022, doi: 10.1109/ACCESS.2021.3140175.

[16] R. Cheng, N. Wu, S. Chen, and B. Han, “Will Metaverse Be NextG Internet? Vision, Hype, and Reality,” IEEE Netw., vol. 36, no. 5, pp. 197–204, 2022, doi: 10.1109/MNET.117.2200055.

[17] H. Ning et al., “A Survey on Metaverse: the State-of-the-art, Technologies, Applications, and Challenges,” 2021, [Online]. Available: http://arxiv.org/abs/2111.09673.

[18] W. Y. B. Lim et al., “Realizing the metaverse with edge intelligence: A match made in heaven,” IEEE Wirel. Commun., 2022.

[19] M. Xu et al., “A full dive into realizing the edge-enabled metaverse: Visions, enabling technologies, and challenges,” IEEE Commun. Surv. Tutorials, vol. 25, no. 1, pp. 656–700, 2022.

[20] S. Seidel, N. Berente, J. Nickerson, and G. Yepes, “Designing the metaverse,” in Proceedings of the 55th Hawaii International Conference on System Sciences, 2022, pp. 6699–6708.

[21] D. Gursoy, S. Malodia, and A. Dhir, “The metaverse in the hospitality and tourism industry: An overview of current trends and future research directions,” J. Hosp. Mark. Manag., pp. 1–8, 2022.

[22] R. El Khatib, J. Bassett, C. Bryce, and A. Ungaretti, “The metaverse: Navigating the evolving risk landscape for retailers,” Lockton’s Retail Pract., 2023.

[23] C. Peukert, J. Pfeiffer, M. Meißner, T. Pfeiffer, and C. Weinhardt, “Shopping in Virtual Reality Stores: The Influence of Immersion on System Adoption,” J. Manag. Inf. Syst., vol. 36, no. 3, pp. 755–788, 2019, doi: 10.1080/07421222.2019.1628889.

[24] S. Bhatnagar and R. Yadav, “Determinants of customer experience, satisfaction and willingness to purchase from virtual tour of a retail store,” Int. J. Manag. Pract., vol. 16, no. 1, pp. 38–58, 2023.

[25] L. Xue, “Designing effective augmented reality platforms to enhance the consumer shopping experiences.” Loughborough University, 2022.

[26] Y. F. Wu and E. Y. Kim, "Users' perceptions of technological features in Augmented Reality (AR) and Virtual Reality (VR) in fashion retailing: A qualitative content analysis," Mob. Inf. Syst., vol. 2022, 2022.

[27] B. Kitchenham and S. Charters, "Guidelines for performing systematic literature reviews in software engineering," 2007.

[28] D. Tranfield, D. Denyer, and P. Smart, "Towards a methodology for developing evidence - informed management knowledge by means of systematic review," Br. J. Manag., vol. 14, no. 3, pp. 207‑222, 2003.

[29] G. Bilquise, S. Ibrahim, and K. Shaalan, "Emotionally Intelligent Chatbots: A Systematic Literature Review," Hum. Behav. Emerg. Technol., vol. 2022, 2022.

[30] N. J. Van Eck and L. Waltman, "VOSviewer manual," Leiden: Univeristeit Leiden, vol. 1, no. 1, pp. 1–53, 2013.

[31] D. Moher, A. Liberati, J. Tetzlaff, and D. G. Altman, "Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement," Int J Surg, vol. 8, no. 5, pp. 336–341, 2010.

[32] L. Yang et al., "Quality assessment in systematic literature reviews: A software engineering perspective," Inf. Softw. Technol., vol. 130, p. 106397, 2021.

[33] M. Q. Tran, S. Minocha, D. Roberts, A. Laing, and D. Langdridge, "A Means-End Analysis of Consumers' Perceptions of Virtual World Affordances for E-commerce," Proc. HCI 2011 - 25th BCS Conf. Hum. Comput. Interact., pp. 520–525, 2011, doi: 10.14236/ewic/hci2011.87.

[34] I. P. de Amorim, J. Guerreiro, S. Eloy, and S. M. C. Loureiro, "How augmented reality media richness influences consumer behaviour," Int. J. Consum. Stud., vol. 46, no. 6, pp. 2351–2366, 2022.

[35] A. Welivita, N. Nimalsiri, R. Wickramasinghe, U. Pathirana, and C. Gamage, "Virtual product try-on solution for E-commerce using mobile augmented reality," Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), vol. 10324 LNCS, pp. 438–447, 2017, doi: 10.1007/978-3-319-60922-5_34.

[36] M. Meißner, J. Pfeiffer, T. Pfeiffer, and H. Oppewal, "Combining virtual reality and mobile eye tracking to provide a naturalistic experimental environment for shopper research," J. Bus. Res., vol. 100, no. September 2017, pp. 445–458, 2019, doi: 10.1016/j.jbusres.2017.09.028.

[37] D. C. Gross, "Affordances in the design of virtual environments." University of Central Florida, 2004.

[38] M. Y. C. Yim, S. C. Chu, and P. L. Sauer, "Is Augmented Reality Technology an Effective Tool for E-commerce? An Interactivity and Vividness Perspective," J. Interact. Mark., vol. 39, pp. 89–103, 2017, doi: 10.1016/j.intmar.2017.04.001.

[39] D. Kim and Y. J. Ko, "The impact of virtual reality (VR) technology on sport spectators' flow experience and satisfaction," Comput. Human Behav., vol. 93, pp. 346–356, 2019.

[40] F. Pereira, C. Silva, and M. Alves, "Augmented Reality Techniques for e-Commerce," Techniques, vol. 220, pp. 62–71, 2011, [Online]. Available: http://download.springer.com.ezproxy.vub.ac.be:2048/static/pdf/981/chp%253A10.1007%252F978-3-642-24355-4_7.pdf?originUrl=http%3A%2F%2Flink.springer.com%2Fchapter%2F10.1007%2F978-3-642-24355-4_7&token2=exp=1492783356~acl=%2Fstatic%2Fpdf%2F981%2Fchp%25253A1.

[41] W. Shen, "Natural interaction technology in virtual reality," Proc. - 2021 Int. Symp. Artif. Intell. its Appl. Media, ISAIAM 2021, pp. 1–4, 2021, doi: 10.1109/ISAIAM53259.2021.00008.

[42] T. Cao, C. Cao, Y. Guo, G. Wu, and X. Shen, "Interactive Embodied Agent for Navigation in Virtual Environments," Proc. - 2021 IEEE Int. Symp. Mix. Augment. Real. Adjunct, ISMAR-Adjunct 2021, pp. 224–227, 2021, doi: 10.1109/ISMAR-Adjunct54149.2021.00053.

[43] T. Fornelos et al., "A Conversational Shopping Assistant for Online Virtual Stores," pp. 6994–6996, 2022, doi: 10.1145/3503161.3547738.

[44] V. Shankar, "How Artificial Intelligence (AI) is Reshaping Retailing," J. Retail., vol. 94, no. 4, pp. vi–xi, 2018, doi: https://doi.org/10.1016/S0022-4359(18)30076-9.

[45] R. E. Hostler, V. Y. Yoon, and T. Guimaraes, "Recommendation agent impact on consumer online shopping: The Movie Magic case study," Expert Syst. Appl., vol. 39, no. 3, pp. 2989–2999, 2012.

[46] A. R. Smink, E. A. van Reijmersdal, G. van Noort, and P. C. Neijens, "Shopping in augmented reality: The effects of spatial presence, personalization and intrusiveness on app and brand responses," J. Bus. Res., vol. 118, no. August 2019, pp. 474–485, 2020, doi: 10.1016/j.jbusres.2020.07.018.

[47] E. Cheon, "Energizing business transactions in virtual worlds: An empirical study of consumers' purchasing behaviors," Inf. Technol. Manag., vol. 14, no. 4, pp. 315–330, 2013, doi: 10.1007/s10799-013-0169-6.

[48] S. Papagiannidis, E. Pantano, E. W. K. See-To, C. Dennis, and M. Bourlakis, "To immerse or not? Experimenting with two virtual retail environments," Inf. Technol. People, vol. 30, no. 1, pp. 163–188, 2017, doi: 10.1108/ITP-03-2015-0069.

[49] Y. Jung and S. D. Pawlowski, "Virtual goods, real goals: Exploring means-end goal structures of consumers in social virtual worlds," Inf. Manag., vol. 51, no. 5, pp. 520–531, 2014, doi: 10.1016/j.im.2014.03.002.

[50] J. V. Chen, Q. A. Ha, and M. T. Vu, "The Influences of Virtual Reality Shopping Characteristics on Consumers' Impulse Buying Behavior," Int. J. Hum. Comput. Interact., vol. 0, no. 0, pp. 1–19, 2022, doi: 10.1080/10447318.2022.2098566.

[51] S. Altarteer and V. Charissis, "Technology Acceptance Model for 3D Virtual Reality System in Luxury Brands Online Stores," IEEE Access, vol. 7, pp. 64053–64062, 2019, doi: 10.1109/ACCESS.2019.2916353.

[52] S. S. Sundar, A. Oeldorf-hirsch, and A. K. Garga, "A Cognitive-Heuristics Approach to Understanding Presence in Virtual Environments," PRESENCE 2008 Proc. 11th Annu. Int. Work. Presence, no. November, pp. 219–228, 2008, [Online]. Available: http://temple.edu/ispr/prev_conferences/proceedings/2008/sundar.pdf.

[53] E. L. M. Bourhim and A. Cherkaoui, "Efficacy of virtual reality for studying people's pre-evacuation behavior under fire," Int. J. Hum. Comput. Stud., vol. 142, p. 102484, 2020.

[54] J. Lee, A. Eden, D. R. Ewoldsen, D. Beyea, and S. Lee, "Seeing possibilities for action: Orienting and exploratory behaviors in VR," Comput. Human Behav., vol. 98, pp. 158–165, 2019.

[55] F. Malak, J. B. Ferreira, R. Pessoa de Queiroz Falcão, and C. J. Giovannini, "Seller Reputation Within the B2C e-Marketplace and Impacts on Purchase Intention," Lat. Am. Bus. Rev., vol. 22, no. 3, pp. 287–307, 2021, doi: 10.1080/10978526.2021.1893182.

[56] B. Lu, W. Fan, and M. Zhou, "Social presence, trust, and social commerce purchase intention: An empirical research," Comput. Human Behav., vol. 56, pp. 225–237, 2016, doi: 10.1016/j.chb.2015.11.057.

[57] A. Javornik, "Augmented reality: Research agenda for studying the impact of its media characteristics on consumer behaviour," J. Retail. Consum. Serv., vol. 30, pp. 252–261, 2016, doi: 10.1016/j.jretconser.2016.02.004.

[58] B. Zhang, "Augmented reality virtual glasses try-on technology based on iOS platform," Eurasip J. Image Video Process., vol. 2018, no. 1, 2018, doi: 10.1186/s13640-018-0373-8.

[59] B. AlHarbi et al., "The design and implementation of an interactive mobile Augmented Reality application for an improved furniture shopping experience," Rev. Română Informatică și Autom., vol. 31, no. 3, pp. 69–80, 2021, doi: 10.33436/v31i3y202106.

[60] W. M. C. D. Wijayalath, R. M. T. T. Ranasinghe, S. Kumari, M. T. H. Thennakoon, H. D. Vithanage, and S. Chandrasiri, "Kidland: An Augmented Reality-based approach for Smart Ordering for Toy Store," Proc. 6th Int. Conf. Inf. Technol. Res. Digit. Resil. Reinvention, ICITR 2021, 2021, doi: 10.1109/ICITR54349.2021.9657326.

[61] S. F. Liu and M. H. Lee, "Mobile commerce system integrated with augmented reality and interactive multimedia," Prz. Elektrotechniczny, vol. 88, no. 9 B, pp. 100–103, 2012.

[62] E. Cruz et al., "An augmented reality application for improving shopping experience in large retail stores," Virtual Real., vol. 23, no. 3, pp. 281–291, 2019, doi: 10.1007/s10055-018-0338-3.

[63] R. Zimmermann et al., "Enhancing brick-and-mortar store shopping experience with an augmented reality shopping assistant application using personalized recommendations and explainable artificial intelligence," J. Res. Interact. Mark., 2022, doi: 10.1108/JRIM-09-2021-0237.

[64] C. Gallardo et al., "Augmented reality as a new marketing strategy," in International conference on augmented reality, virtual reality and computer graphics, Springer, 2018, pp. 351–362.

[65] U. Elordi, A. Segura, J. Goenetxea, A. Moreno, and J. Arambarri, "Virtual Reality Interfaces Applied to Web-Based 3D E-Commerce," in Engineering Systems Design and Analysis, 2012, pp. 341–350.

[66] V. K. Ketoma, P. Schäfer, and G. Meixner, "Development and evaluation of a virtual reality grocery shopping application using a multi-kinect walking-in-place approach," Adv. Intell. Syst. Comput., vol. 722, no. January, pp. 368–374, 2018, doi: 10.1007/978-3-319-73888-8_57.

[67] Y. C. Huang and S. Y. Liu, "Virtual reality online shopping (vros) platform," Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), vol. 12204 LNCS, no. July, pp. 339–353, 2020, doi: 10.1007/978-3-030-50341-3_27.

[68] M. Speicher, S. Cucerca, and A. Krüger, "VRShop: A Mobile Interactive Virtual Reality Shopping Environment Combining the Benefits of On- and Offline Shopping," Proc. ACM Interactive, Mobile, Wearable Ubiquitous Technol., vol. 1, no. 3, pp. 1–31, 2017.

[69] K. Pfeuffer et al., "ARtention: A design space for gaze-adaptive user interfaces in augmented reality," Comput. Graph., vol. 95, pp. 1–12, 2021, doi: 10.1016/j.cag.2021.01.001.

[70] R. Matsumura and M. Shiomi, "An Animation Character Robot That Increases Sales," Appl. Sci., pp. 124–134, 2022, doi: 10.4324/9781315232140-14.

[71] A. Elboudali, A. Aoussat, F. Mantelet, J. Bethomier, and F. Leray, "A customised virtual reality shopping experience framework based on consumer behaviour: 3DR3CO," Int. J. Interact. Des. Manuf., vol. 14, no. 2, pp. 551–563, 2020, doi: 10.1007/s12008-020-00645-0.

[72] J. Wu, B. R. Joo, A. S. Sina, S. Song, and C. H. Whang, "Personalizing 3D virtual fashion stores: an action research approach to modularity development," Int. J. Retail Distrib. Manag., vol. 50, no. 3, pp. 342–360, 2022, doi: 10.1108/IJRDM-08-2020-0298.

[73] H. Fang, J. Zhang, M. Şensoy, and N. Magnenat-Thalmann, "Reputation mechanism for e-commerce in virtual reality environments," Electron. Commer. Res. Appl., vol. 13, no. 6, pp. 409–422, 2014, doi: 10.1016/j.elerap.2014.08.002.

[74] S. Rokhsaritalemi, A. Sadeghi-Niaraki, and S. M. Choi, "A review on mixed reality: Current trends, challenges and prospects," Appl. Sci., vol. 10, no. 2, 2020, doi: 10.3390/app10020636.

[75] Z. Li et al., Augmented reality shopping system through image search and virtual shop generation, vol. 12184 LNCS. Springer International Publishing, 2020. doi: 10.1007/978-3-030-50020-7_26.

[76] S. R. Billewar et al., "The rise of 3D E-Commerce: the online shopping gets real with virtual reality and augmented reality during COVID-19," World J. Eng., vol. 19, no. 2, pp. 244–253, 2022, doi: 10.1108/WJE-06-2021-0338.

[77] E. Morotti, L. Stacchio, L. Donatiello, M. Roccetti, J. Tarabelli, and G. Marfia, "Exploiting fashion x-commerce through the empowerment of voice in the fashion virtual reality arena: Integrating voice assistant and virtual reality technologies for fashion communication," Virtual Real., vol. 26, no. 3, pp. 871–884, 2022, doi: 10.1007/s10055-021-00602-6.

[78] N. A. A. Rahim, M. A. Norasikin, and Z. Maksom, "Improving E-Commerce Application through Sense of Agency of a Calibrated Interactive VR Application," J. Inf. Commun. Technol., vol. 21, no. 3, pp. 315–335, 2022, doi: 10.32890/jict2022.21.3.2.

[79] S. Jain, T. Schweiss, S. Bender, and D. Werth, "Omnichannel retail customer experience with mixed-reality shopping assistant systems," in Advances in Visual Computing: 16th International Symposium, ISVC 2021, Virtual Event, October 4-6, 2021, Proceedings, Part I, Springer, 2021, pp. 504–517.

[80] S. Jain, G. Obermeier, A. Auinger, D. Werth, and G. Kiss, "Design Principles of a Mixed-Reality Shopping Assistant System in Omnichannel Retail," Appl. Sci., vol. 13, no. 3, p. 1384, 2023.

[81] J. Wu, B. R. Joo, A. Saquib Sina, S. Song, and C. Haesung Whang, "Personalizing 3D virtual fashion stores: an action research approach to modularity development," Int. J. Retail Distrib. Manag., vol. 50, no. 3, 2021.

DIAUTIS III: A Fuzzy and Affective Platform for Obtaining Autism Mental Models and Learning Aids

Mohamed El Alami[1], Sara El khabbazi[2], Fernando de Arriaga[3]
National School of Applied Sciences of Tangier, Abdelmalek Essaadi University, Tangier, Morocco[1, 2]
School of Telecommunication Engineering, Universidad Politécnica de Madrid, Madrid, Spain[3]

Abstract*—Autism spectrum disorders (ASD) are conditions characterized by social interaction and communication difficulties, atypical patterns of activities, and unusual reactions to sensations. Characteristics of autism may be detected in early childhood, but diagnosis is often delayed. The diagnosis of autistic children typically aligns with medical and psychological recommendations, but it does not evaluate all the problems, intensity, or changes in symptoms over time. It also does not identify the affective states associated with these deficiencies, making aid less effective. The mental model of autistic children contains their deficits, tasks, and intensities, beyond diagnostics. That why we enhance DIAUTIS platform for achieve our objectives related to helping children with ASD. DIAUTIS I is a platform that aim to diagnosing autism and identifying its severity using cognitive, fuzzy, and affective computing. It presents tests, evaluates results, and presents a final model. Then, we implemented DIAUTIS II by adding KASP methodology, a new methodology of designing serious games, based on knowledge, affect, sensory, and pedagogy, this tool allows to DIAUTIS II agents to designing over 80 games, considering a child's background. In this paper, we will present a new tool of formalization of the autism mental model based on fuzzy and affective computing. DIAUTIS III is the extension of DIAUTIS II platform aim to represent a cognitive fuzzy mental model with using the metrics of category theory. So far, no mental model has been developed for autism. Our mental model, can be obtained anytime as a fuzzy cognitive map or fuzzy graph and the use of affective computing. In addition, the mathematical theory of categories represent this mental model of an autistic child from a fuzzy graph, and it allows for operations like CONS and TRA to evaluate the difference between two mental models. Fuzzy cognitive mental model can be used to develop new techniques for improving the learning and integration of autistic children into social life, which is the focus of our immediate future.

Keywords*—Fuzzy computing; affective computing; mental models; learning aids; autism; category theory

I. Introduction

Autism, also known as Autism Spectrum Disorder (ASD), is a condition characterized by difficulties in social skills, repetitive behaviors, speech, and nonverbal communication [1].

Autism spectrum disorder (ASD) is a neurological and developmental disorder that affects social interactions, communication, learning, and behavior. Symptoms typically appear in the first two years of life, making it a "developmental disorder." The Diagnostic and Statistical Manual of Mental Disorders (DSM-5), used by healthcare providers, identifies ASD symptoms as difficulty with communication, restricted interests, repetitive behaviors, and affecting functioning in school, work, and other areas of life [2].

Autism is a spectrum disorder with varying symptoms across genders, races, ethnicities, and economic backgrounds. It can be lifelong, but treatments can improve symptoms and daily functioning. The American Academy of Pediatrics recommends all children receive autism screening, and caregivers should consult their child's healthcare provider about screening or evaluation [2].

Mental models are an analogical representation of knowledge, involving a direct correspondence between the entities and relations in the representation structure and the entities and relations of the real network. They represent a functional form of prior conceptions in relation to a specific and momentary goal, consisting of elements and their relationships that represent the state of things. Each model is predisposed in a way consistent with its intended use.

A mental model is a cognitive simulation of how things work and are related, built over time through learning and understanding. It helps people make sense of the world, make decisions, solve problems, and predict outcomes, enhancing their decision-making abilities.

So far, according to our knowledge, no mental model of autism has been developed, although we have studied the development of other non -mental models.

Our mental model emphasizes the importance of emotional aspects in children, as affective states can trigger learning problems and improve them. It includes all developed affective relationships and uses affective computing to determine and improve these states. This elaborate model can be classified as fuzzy and affective, as it can be used to treat autistic children and improve their emotional well-being. Affective computing began with the works of Rosalind Pickard, followed by a large number of articles and collaborations of different authors [3-4].

Category theory is a mathematical branch that formalizes mathematical structure and has powerful applications in mathematics and programming languages. Concepts like categorical semantics, monads, functors, and proof assistants are deeply connected to category theory, making them essential in various fields throughout one's career.

Category theory is a generalization of algebra that focuses on understanding mathematical objects through structure-preserving transformations called morphisms. It does not embed a

single group into a category, but studies groups through the category of all groups Grp, where objects are groups and morphisms are group homomorphism. Instead, we learn about group's relationships through morphisms rather than the elements that make them up.

In this perspective, the article first presents an overview of Autism Spectrum Disorder (ASD), its symptoms, and diagnosis. Moreover, it describe the main features and functionalities of DIAUTIS I and DIAUTIS II. Next, it introduces DIAUTIS III, which will use the mathematics of category theory for formalization of fuzzy cognitive mental model of autism and compare the differences between two mental models for helping autistic children to enhance their learning and exceed their learning disabilities.

II. Autism Spectrum Disorder

American Psychiatric Association define the Autism Spectrum Disorder (ASD) as a set of permanent neuro-developmental described by troubles with social interactivity and restricting and repeating behaviors [5], usually beginning in early childhood, Children with Autism Spectrum Disorder may learn, move, or pay attention with different styles.

Autistic children often have problems with: Keeping eye contact, not responding to their name (9 months of age), Ignoring other children and avoiding play with them (36 months of age), etc. Moreover, they may behave in ways that may seem unfamiliar. as: Repeat voices, words or phrases, Have obsessive hobbies, Flap hands, Spinning in circles, Locked body, Have unusual reactions to some of sounds, odors, flavors, or feels. In addition, there are other characteristics that autistic children may have which include hyperactive, unusual eating and sleeping patterns, anxiety, stress, Not getting scared or getting scared more than expected [6].

A. Diagnosis

Over the past few decades, the symptoms of autism spectrum disorders (ASD) have multiplied, leading to increasing recognition of this developmental disorder [7]. For a diagnosis of ASD to be made, children must exhibit characteristics related to both social communication difficulties and restricted, repetitive, and/or sensory behaviors from an early age. The diagnostic criteria for ASDs have evolved considerably over the years. Currently, diagnosis is mainly made using observational tracking tools that measure the child's social, behavioral and cognitive abilities [8]. There are many diagnostic tools, but the two principle ones used in the diagnosis of ASD are DSM-V and M-CHAT.

American Psychiatric Association is publishing a book titled "The Diagnostic and Statistical Manual of Mental Disorder" for presenting a classification of mental patterns with associated criteria, this association classifies autism, Asperger's syndrome and developmental disabilities as Pervasive Developmental Disorders (PDD). It has been recommended that the term "PDD" should turn into "DSM-V" with "Autism Spectrum Disorder". Autism is described by a range of signs, including the existence of permanent deficiencies in social communication, deficiencies in social interaction in a different situations, and restricted and repetitive patterns of behavior in DSM-IV-TR. It was combined into two categories in DSM-V: 1) fusing troubles in social communication and interaction; 2) restricted and repetitive behaviors [9]. The DSM-V is a tool that can be used by wide clinical professionals to help in the diagnosis of mental illness and developmental disabilities. This manual (DSM-5) is used by psychiatrists, psychologists, social employees, doctors and nurses for people with mental illness.

B. Mental Models: Autism Mental Model

The Mental model theory proposes a theory of inference that aims to explain different types of thinking according to how the mental model represents. A mental model is a powerful cognitive entity that organizes people's interpretation of their world and their actions in response to it [10]. We view the concept of mental models, described as the ideas people form about the world and the activities they undertake [11].

A Mental Model (MM) is a simulation of the reality that we are producing, with conscience or unintentionally, in our minds [12]. It plays a crucial role in various fields of our life, which a main aim of a MM is to help the model's user to choose the right action of a target system (Johnson-Laird; Norman, 1983) for making sense of, solving problems, making decisions, making plans, or learning/education…etc. A mental model is created, updated and stored through a three-phase process: observation, learning and experience [12].

The purpose of education is to provide access to knowledge. In order to assess the evolution of a learner's knowledge, we believe it is useful to understand the process of learning which has been a principal subject in the field of education [13]. Hence, mental models are a very important part in this process, which play a crucial role in understanding the development of knowledge and the actions of learners. A learner's knowledge passes by three successive stages; these are the Novice, Learner and Expert stages. One of the important contributions of cognitive psychology to learning is the notion of mental models. This notion will allow us to better differentiate between these three stages of learner development [14, 15].

Presenting a mental model visually and formally requires at least three phases: extraction, analysis and formalizing. Finding ways of modeling a mental model presents a challenge to any discipline using this concept as a tool to better understand individuals' internal representations of their environment.

Our mental models of autistic children contain all the imperfections and activities they do not know how to perform, or which they perform clumsily. This is why their mental models are so important, because they show all their limitations, which can be quite different. These can then serve to identify techniques or strategies for enhancing their learning.

C. Learning Aids

Teaching and learning process needs three essential components, identifying the input elements (students, teachers, teaching and learning materials), the procedure, and the output elements (graduation, orientation, skills, diplomats ...). The quality of teaching and learning can be compromised by unfavorable general conditions. It is crucial to provide a sophisticated environment in which to teach and learn effective skills. Using of

learning aids help students to hear, see, or perform with efficiency during the learning process that become more enjoyable and less tedious thanks to learning aids [16].

Previous research has repeatedly demonstrated that the use of learning aids can greatly enhance learning from multimedia materials. For example, Renkl (2002) studied learning from examples worked in a computerized learning environment on probability calculations. He found that students learned better not only when they explained themselves, but also when they took advantage of pedagogical explanations offered through an online help system [17].

In this twenty-first century, the technological revolution plays a vital role in the learning process. The development of smart phones and other ICT devices has become increasingly popular in the field of communication, literacy and entertainment. These technologies can contribute to the diverse educational development areas of the autism community [18].

Children with autism cannot learn new things easily and cannot actively receive new information, but they can benefit from various types of learning aids specially designed for autistic children and other children with learning disabilities. There are many different types of beneficial learning aids available today, including sensory and auditory learning aids, as well as motor development aids. These tools enable autistic children to catch on quickly and increase their chances of leading a normal life. [19].

III. DIAUTIS I: A Multi-Agent Platform for the Diagnosis of Autism

M. El Alami, N. Tahiri and F. de Arriaga have proposed an approach based on agents, fuzzy logic and affective computing to build an automatic and autonomous autism diagnosis platform [20]. This platform will offer to clinical teams, doctors, parents, guardians and schools a set of tests that help them to get a model of the kid that allows them to evaluate the gravity of autism and the signs of autistic children. In this section, we will describe with details the features of this platform.

A. Features

This research is the beginning for what could be a long project. Over time, it may lead to developing a tool for helping professionals diagnose autism. Consequently, flexibility was highlighted as an important feature of this work. The aim of this flexibility is to anticipate, as far as possible, unknown elements that could discredit the work carried out to date. In line with these ideas, the method developed comprises the following phases [20]:

- Acquiring knowledge about autism in its various aspects.
- Definition of criteria and tools for diagnosing autism.
- Design of a diagnostic computer model based on a group of tests.
- Selection of a platform to build agents based on its flexibility and conception of the DIAUTIS architecture.
- Determination the DIAUTIS functional proofs assigned to the engineering staff.
- Conduct additional testing with observers, medical and clinical staff.

DIAUTIS proposes a set of tests grouped into seven kinds linked to the troubles identified in DSM-V manual. Each kind contains a set of elements to be evaluated and the evaluating agents that intervene in DIAUTIS (control agent, design agent, cognitive agents for sound, gaze and movement, affective agents, pedagogical agents, rule learning agents, evaluating agents, group evaluating agent, interface agent).

DIAUTIS also uses cameras and sensors to receive information on gaze, face, gestures and movements, as well as software developed in Python to capture emotions and analyze pronunciation and intonation to facilitate analysis of the information provided by the collection of tests proposed by DIAUTIS and approved by medical teams to determine the child's condition. These tests are based on ENT (the test collection generation criterion), fuzzy logic techniques to assess the child's condition and the OOC model introduced in NEOCAMPUS [21] to analyze emotions. Then this platform can facilitate the child's previous diagnostic results and generate a report of each test, indicating the times used, the child's age and the result, which can provide information on the test's difficulty, suitability for a particular age or relationship with the severity of autism.

DIAUTIS agents are independent, intelligent and can work autonomously or with other agents. They have learning capabilities thanks to neural networks and machine learning techniques, as well as specific knowledge to reach affective or cognitive goals. They can collaborate with others when objectives surpass their capabilities, or when another agent needs them. They can be cloned or deleted if necessary, have a natural understanding of language and can receive information from external sensors.

The features of NEOCAMPUS for cooperating and controlling agents [22], [23] have been enhanced to take into account the important set of agents that can try to act simultaneously and the specific roles they perform in DAUTIS.

B. Diagnosis Model

The child's diagnosis prepared by DIAUTIS I is represented by a fuzzy graph consisting of a main node (0) that contains the child's personal details and characteristics, connected to seven other nodes (1-7) representing different test kinds. Each test kind is further divided into subcategories, and the evaluation results of each test are linked to the corresponding category or subcategory node.

These fuzzy sets are then integrated into a single fuzzy set, which is then joined to the main node. The final child's model is obtained by using multi-criteria and incorporated into node 0. The system keeps a record of all diagnosis meetings, allowing analysis and comparison of previous sessions. The integration fuzzy set can be defuzzified into a single value when necessary, making it easy for clinical personnel, tutors, and parents to understand. The system primarily communicates with doctors or clinical teams, who set parameters for the tests and receive immediate results and the final child's model. Communication with parents and tutors is also possible through interviews stored in the system.

DIAUTIS I regularly updates the computer team on system performance and agent operation, including the number and type of agents, cooperation, learning, response time, and diagnosis stages.

DIAUTIS I, in summary, is a software platform that meets the objectives of autism diagnostic aid. However, the use of the results can still be clarified. The information from a diagnostic session provides medical equipment with a clear view of the results obtained and the severity of the disease. This is the essential objective of the platform. Nevertheless, DIAUTIS I can also provide other aids such as:

- Facilitating the comparison of the child's previous diagnostic results.
- Facilitate the history of each test, indicating the periods used, the age of the children who took them and the result they obtained, individually or statistically. In this way, it can provide information on the difficulty of the test, its suitability for a certain age or its relationship with the severity of autism.
- In the light of the last paragraph, DIAUTIS I may in the future begin to develop collections of standardized tests in order to produce test protocols, as indicated by doctors and medical teams, leading to eventual standards.

IV. DIAUTIS II: A Multi-Agent Platform for the Diagnosis of Autism and the Design of Serious Games

A. Features

DIAUTIS II extends the capabilities of DIAUTIS I by adding a methodology for designing serious games based on four dimensions: Knowledge, Affect, Senses, and Pedagogy. The KASP methodology enables DIAUTIS II agents to design games from a set of more than 80 templates, considering the child's background [24].

The results of the tests carried out through DIAUTIS are analyzed in order to deduce the child's basic abilities by developing a result containing five specifications: motor, behavioral, cognitive, socio-emotional and sensory. Using machine-learning techniques, an intelligent system will be set up to classify these specifications, which will indicate the type of disorder the diagnosed child suffers from. The authors propose using SVM, a supervised type algorithm, to classify a child's disorder using an intelligent classifier on selected diagnostic specifications.

DIAUTIS I has been enhanced with KASP (Knowledge, Affect, Sensory, Pedagogy) [25], a new approach to the conception of serious learning games. This is a new solution to aid children with learning disabilities to exceed these barriers by considering their cognitive, emotional and sensory aspects.

B. Serious Game as an Evaluation Tool for the Autistic Children

Headings, A game is a narrative that combines an intention, genre, main concept, and objectives. The balance between the objective and the concepts is crucial for a coherent and interesting story. To assess a child's assimilation of a concept with Autism Spectrum Disorder (ASD), it is necessary to consider the sub-concepts building the main concept. Mathematically, comparing the child's sub-concept weights to the expert's weights helps identify actions that cause significant degradation. The child's actions influence the sub-concepts' weights, generating a new vector with new values. The expert is then encouraged to offer a serious game to assess a concept in children with ASD, consisting of several scenes [26]:

$$\text{Game } S = \{S\alpha, S\beta, S\gamma, \cdots\} \text{ where Si a game scene}$$

The game's main objective is the same as the expert's assessment. Each scene consists of multiple objects representing sub-concepts, which interact with each other based on the child's actions:

$S\gamma = (C\gamma, A\gamma)$ where $\{C\gamma = \{ C\gamma1, C\gamma2, C\gamma3, \ldots\}\}$ is the set of sub-concepts of the scene Sγ and

$A\gamma = \{A\gamma1, A\gamma2, A\gamma3, \ldots \}$ is the set of the actions of the scene Sγ.

Each action is identified by a weight. The Likert scale (Table I) is a reliable tool for measuring perceptions, measures the influence of an action on a concept through a weight, consisting of five levels [27].

$$\forall A\gamma i \Longrightarrow W\gamma i \in [1, 5] \quad (1)$$

Likert's scale can also be normalysed by dividing its values by 5. So {0.2, 0.4, 0.6, 0.8, 1}.

TABLE I. Likert's Scale

1	2	3	4	5
Very bad action	Bad action	Acceptable action	Correct Action	Very good action

In addition, the expert attributes a weight to each concept (Table II) to describe its importance in relation to the main concept described on DIAUTIS [21]. The weighting varies from one scene to another:

$$\forall C\gamma i \text{ ->} V\gamma i \in [1, 2, 3, 4, 5] \quad (2)$$

TABLE II. Weights Attributed by the Expert

Concepts	Cγ1	Cγ2	Cγ3	Cγ4
Weight/Concept principal	Vγ1	Vγ2	Vγ3	Vγ4

The expert creates a semantic network of the scene (Fig. 1), representing all concepts and actions, and their weights. A node represents each concept, with an oriented arc connecting two nodes indicating the influence of action on the destination concept. The expert must specify at least five actions for each concept.

After a game session, a child's semantic subnet is created (Fig. 2), inheriting the expert's network and containing only the child's actions and associated concepts.

After that, the semantic network generates two matrices: the expert's matrix (Table III) and the child's matrix (Table IV). The expert's matrix contains all types of actions and concepts, while the child's matrix is a submatrix derived from the expert's matrix. The expert and child's action weights are typically different, as

the expert's matrix contains all the varieties of actions and concepts, while the child's matrix is a submatrix derived from the expert's matrix.

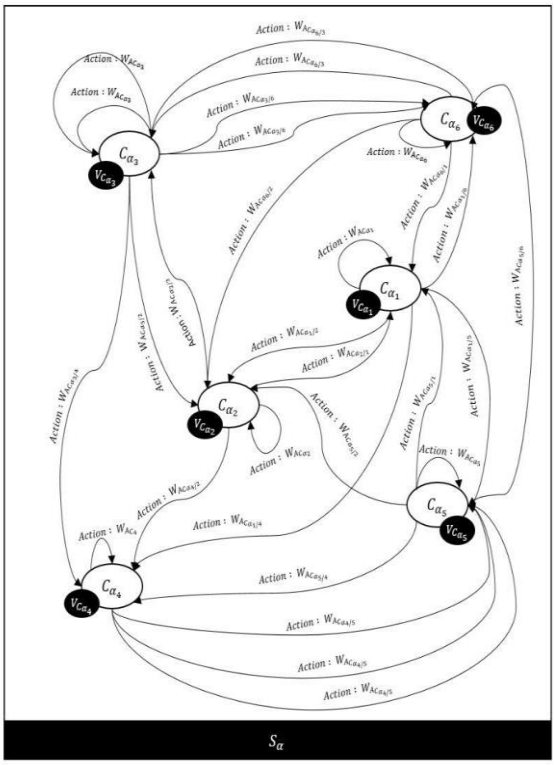


Fig. 1. Expert's semantic network: Scene Sα.

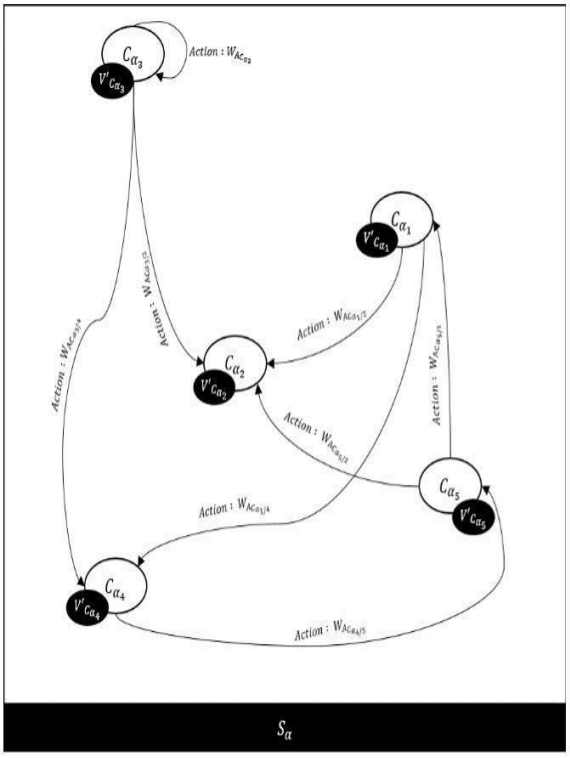


Fig. 2. Child's semantic network: Scene Sα.

TABLE III. EXPERT'S MATRIX: SCENE SA

	Cα1	Cα2	Cα3	Cα4	Cα5	Cα6
Action	WACα1	WACα2	WACα3	WACα4	WACα5	WACα6
Action	WACα1	WACα2	WACα3	WACα4	WACα5	WACα6
Action	WACα1	WACα2	WACα3	WACα4	WACα5	WACα6
Action	WACα1	WACα2	WACα3	WACα4	WACα5	WACα6
Action	WACα1	WACα2	WACα3	WACα4	WACα5	WACα6
…	…	…	…	…	…	…
Action	WACα1	WACα2	WACα3	WACα4	WACα5	WACα6
Action	WACα1	WACα2	WACα3	WACα4	WACα5	WACα6

TABLE IV. CHILD'S MATRIX: SCENE SA

	Cα1	Cα2	Cα3	Cα4	Cα5
Action	WACα1	0	0	0	0
Action	0	WACα2	0	0	0
Action	0	WACα2	0	0	0
Action	0	WACα2	0	0	0
Action	0	0	WACα3	0	0
Action	0	0	0	WACα4	0
Action	0	0	0	WACα4	0
Action	0	0	0	0	WACα5

Then, we have received:

- The expert assigns weights Vαi to scene sub-concepts, which can be represented as an n-dimensional vector, representing the number of scene sub-concepts (Table II).
- The child's sub-concept weights V'αi can now be expressed as: V'αi=Wαj × Wαk × Wαl ×…× Vαi= Π Wαj × Vαi

The Wαj and Vαi are the normalized child's and expert weights; they are vectors of n dimensions, with some components potentially being 0, based on normalized Likert's scale.

- The cosine of two vectors, a real number between -1 and 1, measures the child's ability to assimilate key concepts.

The study found that an acute angle between two vectors leads to higher understanding in a child, while an obtuse angle makes the main concept unclear. This allows experts to better analyze a child's perception of a concept using ADS measurements, enhancing their understanding of the concept.

- We identify actions causing significant sub-concept weight degradation, selecting actions with a weight WACαi ≤ 3 for each scene, ensuring accuracy:
  - If WACαi ≤ 3, a significant degradation is applied to the sub-concept.
  - If 4≤ WACαi <5, The sub-concept is subjected to low degradation.
  - If WACαi=5, The weighting of the sub-concept remains unchanged.

The expert will identify difficult concepts for each child, allowing tutors to adapt learning programs and highlight gaps in

the proposed technique to strengthen understanding for the child with ADS.

V. DIAUTIS III: A Fuzzy and Affective Platform for Obtaining Autism Mental Models and Learning Aids

A. Features

DIAUTIS III extends the capabilities of DIAUTIS II by adding a methodology for formalization of the elaborated mental model of autism (fuzzy graph) based on mathematical theory of categories. That way autism mental models can be automatically obtained. The theory of categories will be used to formalize the graph as well as its operations. We will explain in next sections the steps of this formalization.

The DIAUTIS III tests, based on KASP methodology, analyze the fuzzy mental model's elements. A collection of serious games is designed to cover all deficiencies in the seven families of the model. These games are divided into two kinds: basic discovery of deficiencies and deep exploration of problems, revealing details, intensities, features, and concept assimilation. This allows for a deeper understanding of learning problems and their evolution. The games are displayed on a computer, allowing children to see moving objects, colors, and sounds.

DIAUTIS III has been enhanced with the theory of categories which will be used to formalize the fuzzy and affective mental model of autism, a new approach to the presentation of the abstract concepts. The mathematical theory of categories is a suitable formalism for representing the mental model of an autistic child from a blurred graph, allowing operations like CONS and TRA to evaluate the difference between two mental models or the trajectory of a collection. This mental model allows for the creation of games that require low concentration levels and gradually increase them, thereby facilitating learning and allowing the child's mental model to evolve at any time, reflecting their overall learning experience.

B. Mathematical Theory of Categories

There are various perspectives on the nature and purpose of category theory. Carlos Polanco [28] defines category theory as a mathematical branch that focuses on abstract and unified mathematical structures and their relationships. This Theory is a fundamental aspect of modern mathematics, offering a unified framework for understanding and formalizing mathematical concepts and theories.

Hence, Hoare [29] describes category theory as a broad and abstract branch of pure mathematics, providing little assistance in solving specific problems within its sub-disciplines. As a generalist tool, it offers little benefit to practitioners, making it a valuable tool for generalists.

Category theory is a crucial tool in mathematics, organizing and unifying many fields such as algebraic topology, homological algebra, homotopy theory, representation theory, arithmetic geometry, and algebraic geometry [30]. It is essential for the development of new graphical notations and different levels of abstraction in contemporary mathematics. Category theory and its mathematical disciplines use commutative diagrams, leading to philosophical explorations. However, category theorists have also developed systematic and formal graphical languages to express various forms of argumentations. Mathematics has evolved from being done "up to isomorphism" to "up to equivalence" or "bi-equivalence" or even "n-equivalence." This shift in approach has led to the development of systematic and formal graphical languages to express various forms of argumentations. The level of abstraction in mathematics has evolved from "up to isomorphism" to "up to equivalence" or "bi-equivalence" or "n-equivalence".

Theory of categories focuses on the relationships between structures, offering a highly abstract yet powerful approach to mathematics, focusing on objects and morphisms. The composition operator is a central element, but reducing the theory to a refinement of it is oversimplification and fails for full-featured capture its richness and depth [28]:

1) Abstraction: Category theory is a mathematical field that emphasizes the relations between structures, rather than individual elements. It defines categories by their objects and morphisms, rather than their internal aspects. This abstraction allows for generalization across various mathematical contexts, enabling the transfer of concepts and results across seemingly disparate areas of Mathematics.

2) Unification: Category theory is a powerful tool that unifies various mathematical fields by drawing parallels between seemingly unrelated concepts. It uses factors and natural transformations to connect different categories, demonstrating that different structures and theories are manifestations of underlying ideas, leading to improved understanding and advancements in various mathematical fields.

3) General concepts: Category theory introduces concepts like limits, colimits, and adjunctions, which provide abstract perspectives applicable across various mathematical contexts. Examples include products in abstract algebra and products in topology. This generality enables mathematicians to apply intuitions and results from one area to another, leading to discoveries and advancements in various topics.

4) Applications in other areas: Category theory, a fundamental concept in mathematics, has broad applications in fields like computer science, physics, logic, and philosophy. It is utilized in type theory and programming language denotational semantics, while toposes provide a new framework for intuitionistic logic and set theory in logic.

A category is an abstraction based on objects and morphisms, often studied in groups, rings, and topological spaces. Category theory shifts focus from object elements to morphisms between objects. The axioms of a category do not require objects to be sets, making it unnecessary to speak of an object's elements [31].

A category is a collection of data, denoted by C = (Objects(C), Morphisms(C), ◦, Id) consisting of these items:

1) A set of objects Objects(C), which contains all the items in the category.

2) A set of Morphisms(C), containing the arrows (or Morphisms) between category objects.

3) An operation called composition ◦ that joins Morphisms to form another Morphism, and is associative: $(f \circ g) \circ h = f \circ (g \circ h)$, for any f, g, and h in Morphisms(C).

4) Identity Morphisms Id for each object in Objects(C).

C. Obtainment of the Child's Fuzzy Mental Model

Autism is a group of basic behaviors characterized by difficulties in social reciprocity, communication, and behavioral flexibility. Children with autism spectrum disorder (ASD) struggle to understand others' emotions, feelings, beliefs, and thoughts. Diagnostics for autistic children typically align with existing medical and psychological recommendations, but they do not evaluate all the problems, severities, or individual intensity of the child's symptoms. As they are not repeated, they cannot determine the appearance of new symptoms or deficiencies with age or changes in intensity. Additionally, the diagnosis does not determine the affective states associated with these deficiencies, making aid for autistic children ineffective due to ignorance of these factors. A mental model for autistic children contains all deficits and tasks they cannot perform, but no autism mental model has been developed to date.

In our last article, we propose a fuzzy affective mental model of autism, based on the DSM-V that considers a child's affective states. The model is included their specific deficiencies, intensities, frequencies, and associated affective states. It uses affective computation and fuzzy logic to account for affection and uncertainty. The model can be obtained using the KASP Methodology, and can be obtained in the form of a fuzzy cognitive map or graph. This model can be used to improve learning and social integration for autistic children.

The mental model of an autistic child contains their imperfections, awkward actions, and associated affective states. This model is crucial as it reveals their varied limitations and can help establish techniques or strategies to improve their learning. It is essential for understanding and addressing these difficulties in order to improve their overall development.

1) Autism: New Fuzzy Affective Mental Model: The proposed new conceptual fuzzy and affective model of autism will guide the creation of a specific model for each child, encompassing their specific deficiencies, intensities, frequencies, and associated affective states [32].

The mental model serves as a child's mental radiography, allowing for the identification and treatment of each present deficiency and its intensity. This article does not address this crucial issue, but the model goes beyond diagnosis, as per the DSM-V [33]:

- Persistent deficits in the ability to socialize and interact in a wide variety of contexts.
- Iterative patterns of restricted action, attention, or activity manifesting in at least two contexts.
- These symptoms have to be present in the first stage of the development of the disease.
- These symptoms are the result of a loss of distance in one or more areas of normal functioning.
- An intellectual or general developmental problem does not seem to be the best explanation for these disorders.

The mental model should consider a child's deficiencies and their characteristics to provide appropriate techniques. It also emphasizes the importance of the affective states of the autistic child in learning. Analyzing the link between a deficiency and a specific affective state is crucial for providing emotional and cognitive aids to improve learning deficiencies.

A universal fuzzy model cannot be developed for specific learning deficiencies, as they vary with autistic children's age and present unique characteristics. However, a list of general and concrete deficiencies can guide in expanding, detailing, or specifying them in each case.

The first list of deficiencies has been grouped into seven families:

1) Affective or emotional: This family contains deficiencies related to emotions of the autistic child and her/his affective state when her/his environment changed or meeting people.

2) Verbal and language comprehension: The second family includes all the deficiencies related to the communications and expression that the child with ASD suffers from in her/his daily life and/or in classroom.

3) Monitoring of visual objects: Researchers have identified eight visual processing disorders, each affecting different abilities and presenting unique challenges. These disorders include visual discrimination, optical sequencing, visual figure-ground differentiation, pictorial memory, visual-spatial relation, visual closing, letter and symbol reversal, and visual-motor processing.

4) Attention and response to basic sounds: This family consists of the difficulties related to the identification of sounds, the memory of oral instructions or the sensitivity to the noise…etc.

5) Other social behaviors: The fifth family presents common disabilities and atypical social behaviors of autistic children.

6) Combined tests and deficiencies.

7) Group behaviors.

To better represent the autistic child's specific conditions, each family should include specific deficiencies, rather than general ones, which may belong to multiple families based on the child's characteristics or decompose in multiple families.

Each deficiency has a fuzzy set associated (Table V), with elements as below:

- Severity or intensity: Described by a number that belongs to the whole [0, 1].
- Frequency with which this defect occurs: A number belonging to the range [0, 1] marks it.
- The most commonly associated emotional state. It is defined first by the number of the affective state that belongs to Family 1.
- The intensity of this association is a real number that belongs to [0.1].

2) Obtaining and Formalization the Fuzzy Mental Model: The autistic fuzzy mental model is typically diagnosed through a student progress study to identify learning issues.

To do this, we have implemented the KASP methodology, already tested and used, enables the development of tests based on the theory of serious games. Serious games, which aim beyond entertainment, have significant potential due to the impact of emotions on learning. However, there is a significant academic literature on the nature, composition, and effects of these games.

This methodology evaluates learning deficiencies in autistic children by measuring preschool children's concept assimilation, offering advantages such as identifying continuous learning strategies. It also aids in assessing their learning abilities.

A collection of serious games using KASP methodology has been designed to cover all deficiencies in the seven families of the model. Divided into two categories, the first allows basic discovery of deficiencies, while the second allows deep exploration of problems, revealing details, intensities, features, and concept assimilation. The games are displayed on a computer, allowing children to see moving objects, colors, and sounds.

The process of obtaining the mental model passes via these phases:

- Through the first kind of serious games, or diagnosing, various crucial deficiencies are detected.
- In order to identify new learning issues, serious games of the first kind are used.
- The games of the second kind are used to go deeper into the problems found in phase 1. Getting their intensities, severities and associated emotional states.
- The games of the second kind are used to go deeper into the problems found in phase 2. Getting their intensities, severities and associated emotional states.
- Creation of the fuzzy cognitive map (Fig. 3).

The mental model that is described can be formalized as a fuzzy cognitive map or a fuzzy graph (Fig. 3), which is characterized by its levels:

Level 1: The name of the child with autism, her/his birth date.

Level 2: The seven families of deficiencies presented above.

Level 3: List of deficiencies contained in each family [32].

Level 4: For each deficiency, the associated fuzzy set formed by: intensity or severity of the deficiency, frequency, associated emotional state, and intensity of that association.

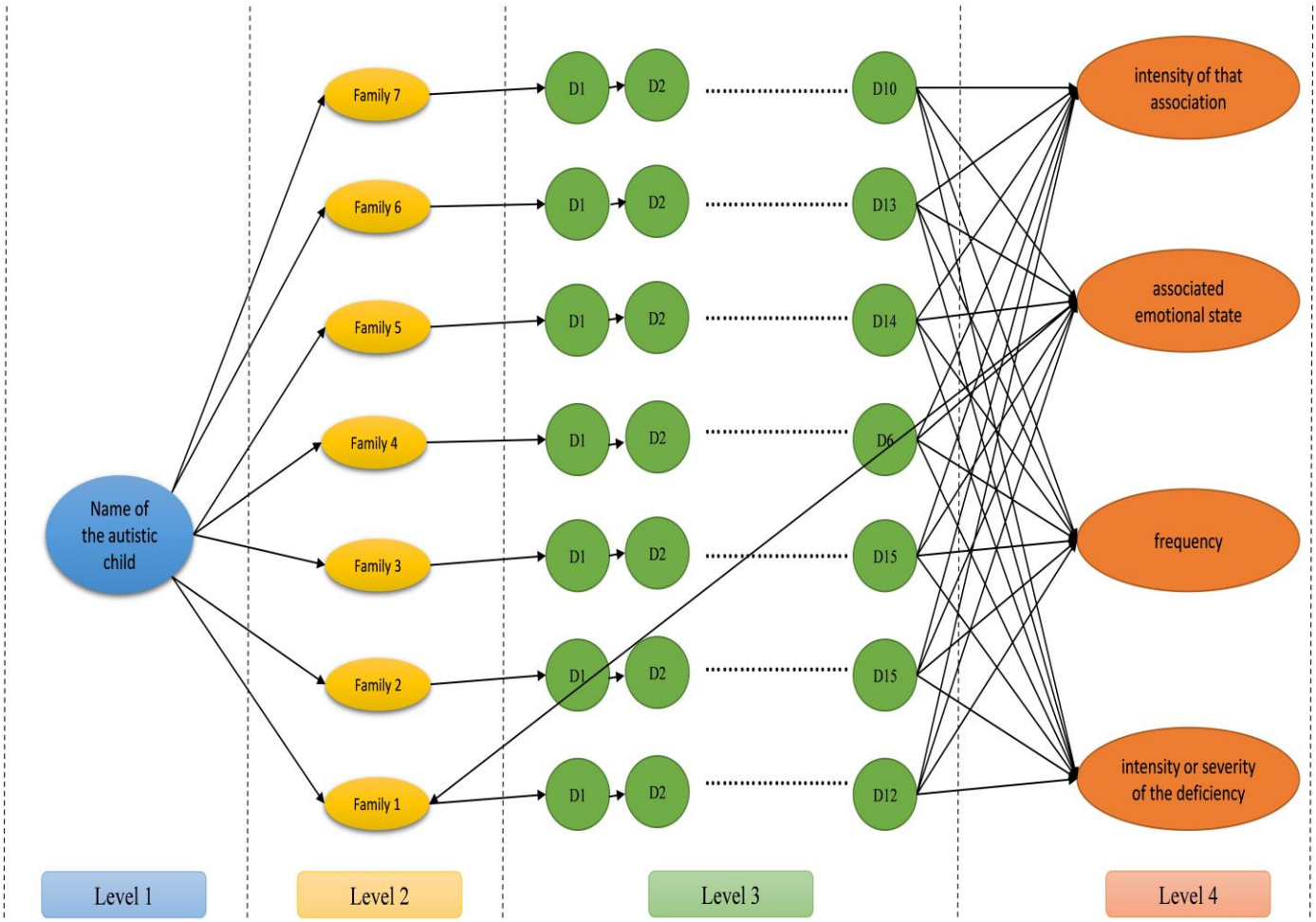


Fig. 3. Detail of the fuzzy cognitive map.

TABLE V. THE ASSOCIATED FUZZY SET OF EACH ELEMENT OF DEFICIENCY

Elements	The associated fuzzy set			
Intensity of the deficiency	[0,0.25] (weak)	[0.25,0.5] (medium)	[0.5, 0.75] (important)	[0.75, 1] (very important)
Frequency of the deficiency	[0, 0.25] (weak)	[0.25,0.5] (medium)	[0.5, 0.75] (important)	[0.75, 1] (very important)
The most frequent emotional state associated	It is determined by the number of the emotional state that appears in Family 1 at the list of deficiencies.			
The intensity of association	[0, 0.25] (light)	[0.25,0.5] (medium intense)	[0.5,0.75] (intense)	[0.75, 1] (very intense)

D. Metrics for the Evaluation of the Difference Between Fuzzy Mental Models of the Same Child

So far, we have used the fuzzy graph as a formalization of the elaborated mental model of autism, composed of a three-level graph. The first, of a single node, is constituted by the name of the child and the date on which the mental model is carried out; the second level has as many nodes as families of learning problems we have considered. The third level is formed by the nodes that represent the different learning problems, in total n, that we have found in the child, corresponding to each family (Fig. 3).

This fuzzy graph will be used as the basis of the mathematical concepts we are going to introduce. Next, we will consider first the theory of categories [34], to formalize the graph as well as its operations. We will try to explain in an intuitive manner the steps of this formalization. Subsequently, we will incorporate topological notions, vector spaces and basic algebraic structures.

1) Concept of mental model: The mental model j of the autistic child i, which we represent by Mentij, is composed of a set of three sub-sets, which are:

a) A subset Dij = (a,b,c,..s) of some (or all) of the nodes of level 3 of the fuzzy, which are the difficulties presented by the autistic child. In the particular case adopted in this paper, the total number of learning problems considered in the seven families is:

$$N = 10+13+14+6+15+15+12 = 85 \quad (3)$$

The cardinal of Dij, i.e. the number of elements containing this subset is s, which will vary in each case of autism. This number s will always be less than or equal to N.

b) A set ℭij =(Cija, Cijb,..., Cijs) of s fuzzy sets that hang from the nodes of the level 3 of the fuzzy graph, corresponding to the problems of the child. In our case, as we have already said, s<=85.

c) A set of Morphisms and morphisms, Mij = (Morfija, Morfijb,...Morfijs, morfijr,n,... morfijh,t); the Morphisms, called Morfijq apply the fuzzy set Cijq to the node q of level 3. These Morphisms, (with capital letter), are the first-class morphisms; besides, there are other morphisms, (with lower case): morfijr,n,.....,morfijh,t, which swap the fuzzy sets Cijr and Cijn hanging on the nodes r and n.

So, the mental model is composed of:

$$Ment^{ij} = (Dij, ℭij, Mij) ; \ Dij = (a,b,c,..r) ; \text{ its cardinal } s<=N ;$$

$$ℭij =(Cija, Cijb,..., Cijr) ; \text{ its cardinal is } s<=N; \quad (4)$$

Mij= (Morfija. Morfijb,...Morfijs, morfijr,n,...,morfijh,t) ;

If we consider that morfijr,n equivale a morfijn,r ,its cardinal is s+(s-1)+...+2+1= (s+1).s/2

d) Moreover, it has two operations, dom and codom, assigned to each Morphism or morphism. For Morphisms, Morfijs, the dom is the fuzzy set to apply, which is Cijs, and the codom is s, the node of the learning problem to which it is applied. For morphisms, morfijh,t the dom is the starting node, h, and the codom is its new position, t, of the node. In short:

$$dom (Morfijs) = Cijs , \quad codom (Morfijs) = s, \ s∈Dij$$

$$dom (morfijh,t) = h, \ h∈Dij , \ codom (morfih,t) = t, \ t∈Dij \quad (5)$$

The first Morphisms are those that assign a fuzzy set to a node; without them there is no mental model. The second morphisms exchange the nodes and the fuzzy sets hanging from them. Although dom(morfijh,t) and codom(morfijh,t) are different from dom(morfijt,h) and codom (morfijt,h) we consider both morphisms equivalent, because their effect on the fuzzy sets is the same.

e) It also has an id operation, which assigns to each node b of level 3, or to each fuzzy set h, the morphism idh which is the identity of h, so that its dom and codom coincide with h. In addition, it is verified:

- Identity law: idh.Morfijh= Morfijh, when h is a node; Morfijh.idh= Morfijh when h is a fuzzy set; in this last case the composition idh.Morfijh cannot be done
- Identity law: idh.morfijr,h= morfijr,h, when h is a node; morfijh,t.idh=morfijh,t when h is a node; when h is a fuzzy set the compositions idh.morfijr,h or morfijr,h.idr cannot be done.

f) It also has a composition operation of two morphisms (f.r) such that:

- The dom of (f.r) is the dom of r, and
- The codom of (f.r) is the codom of f.

It is worth remembering that, as in other mathematical formalisms, composite expressions or formulas are read from right to left, that is, in the case of (f.r), first acts r, and then f. In order for this composition to exist, the morphisms f and r must have a particular nature; r must be a Morphism (with capital letter) to be able to act on f, which must be one morphism with lower case. This way:

$$dom (morfijh,t.Morfijh) = dom (Morfijh) ; \ codom (morfijh,t.Morfijh) = codom (morf ijh,t) \quad (6)$$

g) Composition is also complied with:

Associative law (morfijg,h.morfijb,g).Morfijb = morfijg,h.(morfijb,g.Morfijb)

When, as previously verified

codom(Morfijb)= dom(morfijb,g) and codom(morfijb,g)=co-dom(morfijg,h) (7)

In this associative law must appear a Morphism (with capital letter), because without it there is no mental model; only one can appear, because if more are included the conditions of equality required of doms and codoms cannot be fulfilled. On the other hand, this Morphism must be the first on the right, for the same reason of equality of doms and codoms.

The associative law is also fulfilled with three morphisms (with lower case), when the appropriate conditions of doms and codoms are verified, since as there is no Morphism, there is not any fuzzy set in any node, therefore, the changes of the material hung in the nodes are inoperative. The mental model does not actually exist because the learning problem nodes do not have any fuzzy set to evaluate them.

We will elementally check that this associative law, which we will describe intuitively in a particular case, is complied with.

Suppose a set of five segments followed by numbers 1 to 6; these numbers represent the positions of the third-level nodes of the fuzzy graph, i.e., the learning problems observed in this autistic case:

1-----2-----3-----4-----5-----6

Let's imagine that, initially, under each node hangs an empty box or contains a fuzzy set in the box. Each box, though empty, contains the number of the node it initially hangs from

1-----2-----3-----4-----5-----6

! ! ! ! ! !

U1 U2 U3 U4 U5 U6

If we consider the associative law (morf4,6.morf2,4).Morf2 = morf4,6.(morf2,4.Morf2) we have:

When performing the right hand member of this equation we find:

- *According to the parenthesis,* Morf2 hangs the fuzzy set C2 in the box U2. And subsequently morf2,4 transfers C2 to the box U4.
- *Subsequently morf4,6* moves the fuzzy set C2 from the box U4 to the boxU6 . That's the final state.

By now carrying out the left hand member of the equation, we have:

- Morf2 hangs the fuzzy set C2 in the box U2.
- *According to the parenthesis,* morf2,4 changes C2 from the box U2 to U4, and morf4,6 also to U6.

Both end states corresponding to the left hand and right hand members of the last equation coincide.

The formal demonstration of this general associative law would follow the same steps commented, only that instead of using concrete numbers for nodes like 2, 4, and 6, we would use generic denominations like r, s, and t. and an arbitrary number of nodes. Accordingly, it can be said:

Theory 1. The Mentij mental model is a finite category.

Theory 2. The number of Morphisms (with capital letter) is s.

Theory 3. The number of morphisms (with lower case) also depends on s and is s+(s-1)+ (N-2)+...1= s(s+1)/2

It can also be demonstrated that:

Theory 4. The set of Morphisms constitutes an isomorphism between the subset of learning problems (nodes), and the subset of fuzzy sets that assess the intensity and other characteristics of these problems.

For if the doms of two Morphisms are different, so are their codoms and reciprocally.

Theory 5. The set of morphisms constitutes an epimorphism between the set of learning problems (nodes), and itself.

Indeed, if we remember that in monomorphisms, if the doms are different, the codoms must also be, we see that in this case it is not fulfilled, because several doms can lead to the same codom. In the case of isomorphism, all of them with different domains, have different codoms, and also different codoms also correspond to different doms. It is an epimorphism. Whose morphisms cover all nodes as doms and codoms.

Let us now present a concrete example to demonstrate this formalism.

2) Example: mental model of the autistic child Juanito: Since we will only consider one mental model for the time being, we can remove the first subindex, i, of Mentij, and place the name of the child in place of the second sub-index, j. So, we will have as the name of this mental model MentJuanito.

Suppose now that the autistic child Juanito has: problems learning grammar (family 2, number 7), he experiences disgust when writing (family 3, problem 6), and he does not allow or maintain social relationships with his peers (problem 13 of family 5). These problems are associated respectively with a tendency to be distracted (problem 2 of family 1), with irritability and physical over activity (problem 9 of family 1) and feelings of anxiety (problem 4 of family 1).

In this case, the fuzzy graph of Juanito's mental model (Fig. 4) has only three nodes at level 3, which are his three learning problems, which correspond with the numbers 19= (12+7), 36= (12+18+6), and 64= (12+ 18+15+6+13). We have replaced the numbering by the family and their respective number, by a total numbering that already includes the family number.

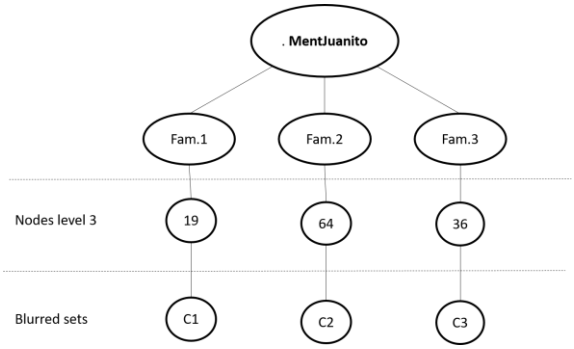


Fig. 4. Juanito's fuzzy graph.

In addition, three fuzzy sets have been obtained that reflect the intensity of the three problems, and are:

a) C1=[intensity 0.4, frequency 0.4, associated affective problem 2, association intensity 0.], for node 19.

b) C2=[intensity 0.5, frequency 0.5, associated affective problem 9, association intensity 0.4], for node 36.

c) C3=[intention 0.6, frequence 0.4, associated affective problem 4, association intensity 0,4], for node 64.

The Morphisms, (with capital letter), available in this case, which we could intuitively consider as arrows, are three: the first, MorfC1-19, takes the first fuzzy set and hangs it on node 19; the second, morfC2-36, takes the second fuzzy set and hangs it on node 36, and the third, morphC3-64, does the same with the third fuzzy set and hangs it on node 64. These Morphisms create a bi-univocal relationship between the set of the three nodes of the mental model and the three fuzzy sets; therefore, they constitute an isomorphism between these two sets.

The morphisms, (with lower case) that we have, are like circles that swap the position and hanging content of two nodes. Therefore, for example, the morf19-36, exchanges node 19 and what hangs from it with node 36. The number of existing morphisms in this example is 3: the morf19-36, the morph19-64, and the morp36-64; as we have already said, we consider morphism morf36-19 equivalent to morf19-36, because their effect on the fuzzy sets is the same (Fig. 5).

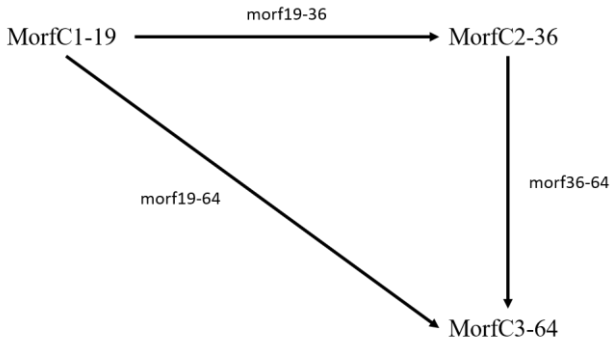


Fig. 5. Existing morphisms.

Furthermore, for each node, the 19 is considered as an example, the morphism id19 that does not change anything of that element, and that satisfies the:

$$\text{identity law: } id19.Morf1\text{-}19 = Morf1\text{-}19;\ Morf1\text{-}19.id19 = Morf1\text{-}19 \quad (8)$$

and the same would happen with the Morphisms associated with nodes 36 and 64.

Similarly, for existing morphisms it is verified:

$$\text{identity law: } id36.morf19\text{-}36 = morf19\text{-}36\ ;\ morf19\text{-}36 = morf19\text{-}36.id19 \quad (9)$$

Furthermore, the following properties are met for the morphism composition operation:

$$\text{Associative law } (morf19\text{-}36.morf19\text{-}36).MorfC1\text{-}19 = morf19\text{-}36\ (morf19\text{-}36.MorfC1\text{-}19) \quad (10)$$

Because here it is verified

$$codom(MorfC1\text{-}19) = dom(morf19\text{-}36), \text{ and } codom(morf19\text{-}36) = dom\ (morf19\text{-}36) \quad (11)$$

In this associative law must appear a Morphism of the first class, because without it there is no mental model; only one can appear, because if more are included the conditions of equality required of domains and codominions cannot be fulfilled. On the other hand, this first-class Morphism must be the last of the right, for the same reason of equality of domains and co-dominios.

The association law is also complied with:

$$(morf64\text{-}19.morf36\text{-}64).morf19\text{-}36 = morf64\text{-}19.(morf36\text{-}64.morf19\text{-}36) \quad (12)$$

However, it is a trivial law, because since no Morphism intervenes, no node has hanged a fuzzy set of evaluation; therefore, it does not produce the MentJuanito model.

Consequently, Juanito's mental model is a category.

3) New transactions with categories: Next, we will define several operations with mental models of the same autistic child. We will try to obtain metrics to evaluate the difference among them.

Value consequence of two mental models

Let us now go to the study of a new operation between two mental models of the same child, which somehow calculates the separation that exists between them. For this, it is necessary to introduce, in addition to the categories, notions of elementary topology, and algebraic structures; for the time being the arithmetic operations of numbers, and their algebraically structures, may suffice. Thus, it is possible to define as a value consequence of two mental models of a child, $CONS(Ment_{ij}, Ment_{il})$,as a new mental model $Ment_{ijl}$ defined as:

$$CONS(Ment_{ij}, Ment_{il}) = Ment_{ijl} \quad (13)$$

a) $Ment_{ijl}$ contains the same D_{ij} set of Level 3 nodes of $Ment_{ij}$ and $Ment_{il}$, which reflect their learning problems.

b) It has the same morphisms and morfisms as the two previous models.

c) As for their collection of fuzzy sets, each of them is obtained as an arithmetic difference of the numerical values that make up them.

Thus, if the fuzzy set of a certain learning problem in the first model is C_{ija} = [0.5, 0.4, 9, 0.3], and in the second model is C_{ila} =[0.3, 0.3, 9, 0.2], CONS ($Ment_{ij}$, $Ment_{il}$) has as fuzzy set for that node, Cijla = [0,5-0.3, 0.4-0.3, 9-9, 0.3-0.2]= [0.2, 0.1, 0, 0.1]

The positive values of this latter set indicate the benefits or achievements obtained from the first mental model to the second.

However, the following abnormalities may occur:

a) One of the obtained values is negative. That must trigger an alarm that starts a collateral study of the causes that have motivated it, to try to overcome them.

b) The second mental model obtained contains more nodes of level 3 that reflect the difficulties of the autistic child, and therefore more fuzzy sets relative to these new nodes. This happens when new problems arise in his/her evolution. In this case, we have to equalize the number of nodes of the problems, incorporating new rows in the first mental model, as well as increasing the Morphisms and morphisms to include these new nodes.

The values of the fuzzy sets created, relative to these new problems of the first mental model are all null values, according to the operation to be performed (in this first case, arithmetic rest); therefore, they would all be zeros. When performing the rest of the numerical values, the values of the new blurred CONS(Mentij, Mentil) set appear with a negative sign causing the corresponding alarm for indicating the appearance of new learning problems.

This new operation has the following properties:

a) There is no commutative property, or rather, there is the anti-commutative property reflected by:

a.b = -b.a, that is:

$$CONS(Ment_{ij}, Ment_{il}) = -CONS(Ment_{il}, Ment_{ij}) \quad (14)$$

b) There is a neutral element only on the right, when $Ment_{il}$ matches $Ment_{ij}$. which has all its elements equal to 0.

$$CONS(Ment_{ij}, Ment_{ij}) = Mentnulo$$

$$CONS(Ment_{ij}, Mentnulo) = Ment_{ij} \quad (15)$$

This null element only acts on the right, because:

$$CONS(Mentnulo, Ment_{ij}) = -Ment_{ij} \quad (16)$$

c) There is an associative property:

$$CONS(Ment_{ik}, CONS(Ment_{ij}, Ment_{il})) = CONS(CONS(Ment_{ik}, Ment_{ij}), Ment_{il}), \quad (17)$$

because,

$$CONS(Ment_{ik}, Ment_{ijl}) = Ment_{ikjl}; \; CONS(Ment_{ikj}, Ment_{il}) = Ment_{ikjl}$$

d) There is the reverse of every mental model, which is himself, as we have seen in (2). Accordingly, it can be said:

Theory 6. The CONS operation between categories constitutes a non-Abelic semigroup.

Estimation of the trajectory of several mental models of the same child

It is also possible to analyze a whole trajectory of mental models of a child,

$$TRA(Ment_{ij}, Ment_{ik}, Ment_{il}, , , , , Ment_{is}) = Ment_{ikl...s}$$

to evaluate the overall performance. To do this we will consider the summary of all your mental models as the following mental model, Mentikl...s such that:

a) Contains the same set of nodes of level 3 as the mental models $Ment_{ij}$, $Ment_{ik}$, $Ment_{il}$,..., $Ment_{is}$

b) It has the same Morphisms and morphisms as the previous models.

With regard to the contained fuzzy sets, there are several ways to get them. One would be to obtain the arithmetic average of the corresponding values of the mental models of the trajectory. However, after studying several practical cases, it seems more appropriate to use the geometric average of these values. In any case, both averages can be used. This new TRA operation has the following properties:

a) It is commutative, because it is the product or sum of the numbers that make up the mental models of the trajectory.

b) It is associative, for the same reason above, the associativity of the sum or product of numbers.

c) It has an identity, formed by the mental model whose numerical components are all equal to d) It does not have an inverse element, because there are no fuzzy sets with numbers greater than 1.

Accordingly:

Theory 7. The TRA operation constitutes an Abelian semigroup with the arithmetic operation + (sum) or x (product).

Example: Juanito's mental model

Suppose now that Juanito has two mental models; in the first one he shows two learning problems corresponding to nodes 19 and 36 of the third level of the fuzzy graph, and their associated fuzzy sets are:

C1= [0.5, 0.4, 9, 0.3]; C2= [0.4, 0.4, 13, 0.2]

In the second mental model the associated fuzzy sets are:

C1= [0.3, 0.3, 9, 0.2]; C2= [0.3, 0.2, 13, 0.1]

The CONS model obtained from these two previous models has as its first fuzzy set:

C1= [0.5-0.3, 0.4-0.3, 9-9, 0.3-0.2]= [0.2, 0.1, 0, 0.1]

and as the second set:

C2= [0.4-0.3, 0.4-0.2, 13-13, 0.2-0.1]= [0.1, 0.2, 0, 0.1]

The positive values of these fuzzy sets indicate the benefits or achievements obtained from the first mental model to the second.

Let us now apply the TRA operation, assuming that Juanito has three models of his autistic trajectory, with two learning problems associated with nodes 19 and 36.

If the fuzzy sets associated with these two problems are:

First model: C1 = [0.6, 0.5, 9, 0.4], C2= [0.5, 0.5, 13, 0.3]

Second model: C1= [0.6, 0.4, 9, 0.3], C2= [0,5, 0.4, 13, 0.2]

Third model: C1= [0.4, 0.4, 9, 0.3], C2= [0.4, 0,3, 13, 0.2]

The final values of the fuzzy sets of TRA will be, if we use the geometric average:

C1=[(0.6x0.6x0.4)1/3, (0.5x0.4x0.4)1/3, (9x9x9)1/3, (0.4x0.4x0.3)1/3]= [0.524, 0.43, 9, 0.363]

C2=[(0.5x0.5x0.4)1/3, (0.5, 0.4, 0.3)1/3, (13x13x13)1/3, (0.3x0.2 x0.2) 1/3]= [0.033, 0.02, 13, 0.04]

a) Vector Formalism: Another formal way of considering the mental model is to use vector spaces. In this case it can be assimilated to a column vector in which the fuzzy sets has been aligned within that column; thus the vector has as number of rows 4xs, being s the number of learning problems corresponding to that model.

Using this formalism it is easy to carry out the operations set out in the previous paragraph, which are CONS and TRA. So the result of CONS($Ment_{ij}$, $Ment_{il}$)= $Ment_{ijl}$ can be obtained as the difference between the numerical values of both vectors.

Given that all numerical values are <=1 except the third number of each fuzzy set that is >=1, it is convenient, though not necessary, to normalize that learning problem number by dividing it by the total number of mental problems of your family, and thus all of them will be <=1. The difference vector shows us in its positive terms the advantages achieved, and in the possible negative ones, the alarms for the emergence of new problems or deterioration of the existing ones.

New metric offered by vector space. An integral or metric evaluation that allows directly to obtain this formalism is the obtaining of the cosine of the angle that the component models of CONS present. For this we have to take into account that, given two vectors A, and B, it is verified that their scalar product A.B is equal to the product of their modules by the cosine of the angle they form; from that expression we can get the cosine of the angle they form.

4) Implementation of these new operations: If we consider each fuzzy set as a column vector, whose components are real numbers belonging to the set [0,1], except for the third which is an integer number, we can construct the matrix formed by these vector columns of learning problems of the autistic child. Thus, each column of the matrix contains all the column vectors of the different learning problems of a mental model.

Considering the previous example of a trajectory consisting of three mental models and their fuzzy sets, which in this case are two, corresponding to two mental problems, the matrix obtained is (Table VI):

TABLE VI. CHILD'S MATRIX

Fuzzy sets	Mental models		
	Mentij	***Mentil***	***Mentis***
Fuzzy set 1	0.6	0.6	0.4
	0.5	0.4	0.4
	8	8	8
	0.4	0.3	0.3
Fuzzy set 2	0.7	0.3	0.3
	0.5	0.4	0.2
	6	6	6
	0.5	0.2	0.2

As you can see, each column vector represents a mental model of the child and contains the two fuzzy sets, related to the two learning problems.

Now we can apply the TRA operation, which using the arithmetic average produces us the vector column: [0.533 0.4333 8 0.333 0.4 333 0.3666 6 0.3]

Similarly, another column vector would be obtained using the geometric average, which would be the cubic root of the product of the terms of each row.

E. DIAUTIS III: Agent Architecture

The various types of agent used in DIAUTIS III are as follows:

1) The control agent is responsible for cloning or eliminating agents, coordinating performance, and resolving conflicts. It grants control to affective or pedagogic agents when requested, centralizes communication with users, and handles initial interviews with parents or instructors. It can communicate with any agent.

2) The design agent is responsible for designing tests based on clinical team indications and criteria. They maintain a database of tests and prepare test records. They communicate with control agents, interface, affective, cognitive, and possibly pedagogic agents, and direct diagnosis achievement. They also design future normalized diagnosis protocols based on qualifications, experience, and medical equipment indications, including the possibility of incorporating affective and pedagogic agents.

3) Cognitive agents (type A) manage voice and sound analysis, collaborating with other pedagogic agents. They receive test information and have a sonorous model of the world. They communicate evaluation inputs from negative elements, design agent, and children's group.

4) Cognitive agents (type B) analyze and answer child's appearance tests, collaborating with other agents and control and design agents. They communicate with evaluation and group evaluation agents, ensuring test evaluation accuracy.

5) Cognitive agents (type C) monitor child movements, analyze movements, and use intelligent toys. They have a spatial model and perform tasks and communications similar to agents A and B.

6) Affective agents analyze a child's emotional state during tests using sensor data, communicate with cognitive agents, and sometimes, under clinical advice, improve their emotional state.

7) Pedagogic agents collaborate on test demonstrations, assuming test voice or recommendations, using simple voices, animals, or friendly objects to gain child confidence and interest.

8) Rule learning agents present various tests using videos, images, or intelligent simulations, similar to cognitive agents, with tasks and communication obligations.

9) The evaluation agent is responsible for obtaining tests, integrating evaluations, and defuzzing them into categories, while maintaining a child's cognitive and affective record and diverse diagnosis possibilities.

*10)*The group evaluation agent, when present, performs the same tasks as the evaluation agent but differentiates the child suffering from the rest of the group members.

*11)*The interface agent personalizes the interface based on test situations and the child's state, using elements like screen color, scene background, potential pedagogic agents, messages, and sounds.

F. Validation

Previous experience in assessing Intelligent E-learning Systems [35, 36, 37], has been initially applied and enhanced to take into account the specific features of this platform. According to our methodology already established [39,39], quality assessment of DIAUTIS III has been obtained by working at two different levels: the functional evaluation level and the overall evaluation level. So far, no child has participated in the quality assessment tests. Instead, some errors o anomalous behaviours, randomly chosen, similar to those presented by autistic children, have been introduced as the child's response to the tests, in order to simulate the diagnosis process.

The first level or functional level, with a more reduced scope, follows closely traditional methodologies. The following tests have been carried out:

1) Experimental cross-check of functions (agents) and of auxiliary hardware: design of follow-up exercises with several objects, language questions, social or affective behaviour, by four groups of five observers.

2) Experimental cross-check of the design of simple tests according to doctors' indication and initial test qualification such as: colour, object, words, questions, etc.

3) Experimental cross-check of the design of collection of tests from doctor's indications by using ENT and its built-in criteria.

4) Experimental cross-check of the fuzzy tests evaluations by five observers and members of the clinical equipment.

5) Experimental cross-check of the tests integration into the category fuzzy set and into the child's model by five clinical equipments.

The second level or overall evaluation requires a more creative approach. It includes the evaluation of four different aspects: overall functionality tested by six groups of experts and clinical members, reliability considering that the system learning capability will allow the platform to change parameters according to its experience, evidential validity [40, 41, 42]. In addition, consequential validity [43, 44].

G. Results

The information highlighted in this paper, shows that DIAUTIS III is a software framework that achieves the following objectives:

- Capable of getting for the first time the mental model of an autistic child as a fuzzy graph, by means of affective computing and fuzzy logic.
- Facilitating of the comparison of previous child diagnoses and facilitating their results.
- Providing a comprehensive history of each test, including its usage, age, and results, to understand test difficulty, adequacy at specific ages, and its correlation with autism severity.
- Providing a technique for assessing learning deficits in autistic children that allows the degree of assimilation of a concept by a preschool child to be measured, based on our KASP methodology, that already tested and used, allows the elaboration of the tests, based on the theory of serious games.
- Incorporating the application of the mathematical theory of categories to establish the formalization of the mental model and several other applications.
- Evaluating the difference existing between two mental models, or the trajectory of a collection of mental models of the Autistic child using the metrics of the categories theory.

VI. FUTURE WORKS

The authors suggest using a fuzzy logic approach to establish an efficient evaluation of the KASP based learning system and with the aid of metrics of category theory can extend the functionalities of DIAUTIS III.

For this purpose, we suggest as future works:

- Layout the intensive tests for children, involving doctors, medical equipment, schools, families, and associations. This system can provide crucial aid to all individuals involved with autism, as assessed in previous assessments.
- Design of normalized protocols or tests for autism diagnosis may eventually contribute to the adoption of standards, with the relevance and discriminant power of these tests determined by their history and results.
- Enhance DIAUTIS III with new sensors, software, functionalities, diagnostic tests, and algorithms of artificial intelligence, presenting a new world of possibilities due to rapid technological advancements.
- Use the category theory for a better, deeper, and balanced understanding, of the dimensions and characteristics of autism Among the many ways in which this better understanding could be achieved, it is necessary to identify patterns, relationships, and underlying structures in the behavior and cognition of autistic children.
- Another line of future work is aimed at the analysis of feelings. This is a recent research issue that opens with

great possibilities. So far it has been applied to the analysis of the feelings contained in the written language that is usually sent to social networks to extract opinions and even future actions. It could also be applied to extract deep motivations from the autistic child, analyze them, and even try to modify them where appropriate.

VII. CONCLUSIONS

Multi-agent systems with learning, fuzzy, and affective skills have great potential in autism diagnosis and patient assistance, particularly in learning, as they can help with various issues.

DIAUTIS III is a platform designed to assist individuals with autism, particularly in diagnosing this condition.

DIAUTIS III is a diagnostic tool that considers a child's affective and anomalous behavior to accurately diagnose and understand the severity of their illness.

DIAUTIS III is a framework that provides:

1) The mathematical theory of categories constitutes an appropriate formalism to represent the mental model of the autistic child from the blurred graph, and formulate operations such as CONS and TRA that allow evaluating the difference existing between two mental models, or the trajectory of a collection of mental models of the Autistic child.

2) Other formalisms, such as the vector spaces, also allow obtaining the difference between two mental models, as it may be from the cosine of the angle formed by the vectors representing them.

3) The vector formalism also allows a convenient implementation to obtain the CONS and TRA final values.

However, some limitations concerning the results of the platform have to be established:

- DIAUTIS III cannot be applied for the study of other syndromes besides autism.
- It works according to the DSM-V Manual, therefore the platform has to assume all future changes advised by APA.

REFERENCES

[1] "Autism spectrum disorder (ASD) | Autism Speaks," Autism Speaks. https://www.autismspeaks.org/what-autism

[2] "Autism Spectrum Disorder," National Institute of Mental Health (NIMH). https://www.nimh.nih.gov/health/topics/autism-spectrum-disorders-asd

[3] R. Picard, Affective Computing, Cambridge, MIT Press, 1997.

[4] De Arriaga, M. El Alami, "Affective Computing and Intelligent Systems," Proceedings IADAT International Conference on Education e-2006, Barcelone, 2006, pp. 115-120.

[5] American Psychiatric Association, "Diagnostic and statistical manual of mental disorders: DSM-5," vol. 5, no. 5, Washington, DC: American Psychiatric Association, 2013. https://doi.org/10.1176/appi.books.9780890425596

[6] "Signs and Symptoms of Autism Spectrum Disorder," Autism Spectrum Disorder (ASD), Jan. 25, 2024. https://www.cdc.gov/autism/signs-symptoms/?CDC_AAref_Val=https://www.cdc.gov/ncbddd/autism/signs.html

[7] E. Gordon-Lipkin, J. Foster, and G. Peacock, "Whittling Down the Wait Time," Pediatric Clinics of North America, vol. 63, no. 5, pp. 851–859, Oct. 2016,

[8] R. Lordan, C. Storni, and C. A. De Benedictis, "Autism Spectrum Disorders: Diagnosis and Treatment," Exon Publications eBooks, Aug. 23, 2021. https://www.ncbi.nlm.nih.gov/books/NBK573609/

[9] Blázquez Hinojosa, L. Lázaro Garcia, O. Puig Navarro, E. Varela Bondelle, and R. Calvo Escalona, "Sensitivity and specificity of DSM-5 diagnostic criteria for autism spectrum disorder in a child and adolescent sample," Revista de Psiquiatría y Salud Mental (English Edition), vol. 14, no. 4, pp. 202–211, Oct. 2021

[10] P. N. Johnson-Laird, Mental Models. Harvard University Press, 1983. [Online]. Available:http://books.google.ie/booksid=FS3zSKAfLGMC&printsec=frontcover&dq=Mental+Models+Towards+a+Cognitive+Science+of+Language+Inference+and+Consciousness&hl=&cd=1&source=gbs_api

[11] Gentner, D. and A.L. Stevens, "Mental models," 2014: Psychology Press. (Taylor and Francis) https://doi.org/10.4324/9781315802725

[12] Madiha Anjum, "Understanding, Formalizing, and Reconstructing Mental Models with an Online Tool for Serious Discussions," University of Technology Sydney, Faculty of Engineering and Information Technology, July 2022.

[13] R. Bhalwankar and J. Treur, "Modeling the development of internal mental models by an adaptive network model," Proceedings Computer Science, vol. 190, pp. 90–101, 2021,

[14] Carroll, John M., and Judith Reitman Olson. "Mental models in human-computer interaction." Handbook of human-computer interaction (1988): 45-65.

[15] W. B. Rouse and N. M. Morris, "On looking into the black box: Prospects and limits in the search for mental models.," Psychological Bulletin, vol. 100, no. 3, pp. 349–363, 1986,

[16] Uchechi Bel-Ann Ordu, "The Role of Teaching and Learning Aids/Methods in a Changing World," in New Challenges to Education: Lessons from Around the World, Vol 19, Sofia: Bulgarian Comparative Education Society, 2021, pp. 210-216.

[17] T. Ruf and R. Ploetzner, "One click away is too far! How the presentation of cognitive learning aids influences their use in multimedia learning environments," Computers in Human Behavior, vol. 38, pp. 229–239, Sep. 2014,

[18] T. Zaki et al., "Towards developing a learning tool for children with autism," 2017 6th International Conference on Informatics, Electronics and Vision & 2017 7th International Symposium in Computational Medical and Health Technology (ICIEV-ISCMHT), Sep. 2017,

[19] T. Shine, "Help Them Shine," Help Them Shine. https://www.helpthemshine.com/blogs/learning-aids-for-autistic-children

[20] M. Alami, N. Tahiri, and F. Arriaga, "DIAUTIS: A Fuzzy and Affective Multi-agent Platform for the Diagnosis of Autism," British Journal of Applied Science & Technology, vol. 21, no. 4, pp. 1–28, Jan. 2017,

[21] De Arriaga F, El Alami M., "Multi-agent platform for educational research on intelligent e-learning," Journal of Advanced Technology on Education, 2005, vol. 1, no. 4, pp. 101-106.

[22] De Arriaga F, El Alami M., "Agents control for intelligent e-learning systems. Proceedings IEEE International Conference IAWTIC'05 on Intelligent Agents, Web Technology and ECommerce," 2005, vol. 2, pp. 877-884.

[23] Laureano-Cruces AL, Ramírez J, De Arriaga F, Escarela R. "Agents control in intelligent learning systems: The case of reactive characteristics. Interactive Learning Environments," vol. 14, no. 2, 2006, pp. 95-118.

[24] El-Alami, M., El-Khabbazi, S. ., Tahiri , N. ., & Arriaga, F. de . (2022). DIAUTIS II: A Multi-agent Platform for the Diagnosis of Autism and the Design of Serious Games. Current Overview on Science and Technology Research, vol. 4, pp. 81–134. https://doi.org/10.9734/bpi/costr/v4/3088C

[25] Tahiri N., El Alami M. "KASP: A Cognitive-Affective Methodology for Designing Serious Learning Games," International Journal of Advanced Computer Science and Applications, vol 9, no 11, 2018, pp. 719-729.

[26] N. Tahiri and M. El Alami, "A New Evaluation Technique Through Serious Games for Children with ASD," International Journal of Emerging Technologies in Learning (iJET), vol. 15, no. 11, p. 202, Jun. 2020.

[27] R. Likert. (1932).A Technique for the measurement of attitudes. New York.

[28] Carlos Polanco. (2023). Foundations of Category Theory: An Introduction to Structures, Transformations, and Applications.

[29] Hoare, C.A.R. (1989). Notes on an Approach to Category Theory for Computer Scientists. In: Broy, M. (eds) Constructive Methods in Computing Science. NATO ASI Series, vol 55. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-74884-4_9.

[30] "Category Theory," Aug. 29, 2019. https://plato.stanford.edu/entries/category-theory/

[31] Randall R. Holmes. (2019). Category Theory.

[32] M. El Alami, S. El khabbazi, F. de Arriaga , "Autism Spectrum Disorder: A New Fuzzy and Affective Mental Model," Universal Journal of Public Health, Vol. 12, No. 2, pp. 383 - 395, 2024.

[33] M. Colleen, M. Harker, W. L. Stone, "Comparison of the Diagnostic Criteria for Autism Spectrum Disorder Across DSM-5, DSM-IV-TR, and the Individuals with Disabilities Education Act (IDEA) Definition of Autism," 2014.

[34] Asperti, G. Longo, Categories, Types and Structures, Cambridge, The MIT Press, 1991.

[35] Tahiri N, El Alami M. An intelligent E-learning system for autistic children: multi-agent architecture. InInternational Conference on Advanced Intelligent Systems for SustainableDevelopment 2019 Jul 8,. 83-90.. Springer.

[36] De Arriaga F, El Alami M. Guidelines for the evaluation of intelligent elearning systems. In Technological Advances applied to Theoretical and Practical Teaching, Iadat. 2005;142-147.

[37] De Arriaga F, El Alami M. Fuzzy intelligent e-learning systems: Assessment. Journal of Advanced Technology on Education. 2005;1(12):228-233.

[38] De Arriaga F, El Alami M. Evaluation of Fuzzy intelligent learning systems. In Recent Research Developments in Learning Technologies. ed: Méndez A., Mesa J., Formatex. 2005;1:109-114.

[39] El Alami M, De Arriaga F. Fuzzy assessment for affective and cognitive Computing in intelligent e-learning systems. International Journal of Computer Applications. 2014;100(10):40-46.

[40] Van Lehn K, Martin J. Evaluation of an assessment system based on Bayesian student modelling. International Journal of Artificial Intelligence in Education. 1997;8: 179-22.

[41] Mitrovic A, Ohlsson S. Evaluation of constraint-based tutor for a database language. International Journal of Artificial Intelligence in Education. 1999;10:230-256.

[42] Messick S. The interplay of evidence and consequences in the validation of performance assessment. Educational Researcher. 1994;23(2):13-23.

[43] Messick S. Validity, in educational measurement. Linn R. (ed.), Macmillan. 1989;13-103.

[44] Linn RL, et al. Complex, performance-based assessment: Expectations andvalidation criteria. Educational Researcher.1991;20(8):15-21.

Stock Price Forecasting with Optimized Long Short-Term Memory Network with Manta Ray Foraging Optimization

Zhongpo Gao[1], Junwen Jing[2]*
School of Economics and Management, Harbin University, Harbin 150086, Heilongjiang, China[1]
School of Economics, Harbin University of Commerce, Harbin 150028, Heilongjiang, China[1]
School of Mathematical Physics, Xi 'an Jiaotong-Liverpool University, Suzhou 215028, Jiangsu, China[2]

***Abstract*—The stock market is a financial marketplace where investors may participate through the acquisition and sale of stocks in publicly traded companies. Predicting stock prices in the securities sector may be challenging due to the intricate nature of the subject, which necessitates a comprehensive grasp of several interconnected factors. Numerous factors, including politics, society, as well as the economy, have an impact on the stock market. The primary objective of financial market investing is to exploit larger profits. Financial markets provide many opportunities for market analysts, investors, and researchers in several industries due to significant technology advancements. Conventional approaches encounter difficulties in capturing the complex, non-linear connections that exist in market data, which requires the implementation of sophisticated artificial intelligence models. This paper presents a new approach to tackling certain issues by suggesting a unique model. It combines the long short-term memory method and Empirical Mode Decomposition with the Manta Ray Foraging Optimization. When tested in the current study's dynamic stock market, the EMD-MRFO-LSTM model outperformed other models regarding performance and efficiency. The Nasdaq index data from January 2, 2015, to June 29, 2023, were used in this study. The findings demonstrate how the suggested model is capable of making precise stock price predictions. The suggested model offers a workable approach to studying and predicting stock price time series by obtaining values of 0.9973, 91.99, 71.54, and 0.57, for coefficient of determination (R^2), root means square error (RMSE), mean absolute error (MAE), and mean absolute percentage error (MAPE), respectively. Compared to other methods currently in use, the proposed model has a higher accuracy in forecasting and is more physically relevant to the dynamic stock market, according to the outcomes of the experiment.**

Keywords*—Stock price; hybrid forecasting method; Manta Ray Foraging Optimization; empirical mode decomposition; Nasdaq index

I. INTRODUCTION

Having a thorough understanding of the stock market is essential for anyone working in the finance sector. To reduce risk and maximize returns, investors have to be able to correctly forecast stock prices. Macroeconomic policies, stock options, capital movements of significant firms, and changes in ownership are only a few of the important variables that might affect stock prices. Making precise predictions is difficult due to the unpredictable nature of price fluctuations, which are characterized by non-linear, non-stationary, stochastic noise, and fluctuating [1]. Although econometric and statistical methods could be employed to forecast asset market-related series statistics analysis is constrained by excessive noise and non-linearity.

Artificial intelligence models are preferred over traditional methods for learning complex and non-linear relationships. In the domain of quantitative analysis in finance, their popularity is on the rise because of their ability to extract valuable data from input variables. They are used to detect patterns in historical data and forecast future trends [2], [3]. Deep learning has developed a number of sophisticated structures that address a variety of issues when it comes to varied data sets. A well-known model of neural networks that only processes data in one direction is called feedforward. However, working with sequential data in which earlier occurrences are crucial to projecting future results, can be challenging. As a result, such models have trouble correctly forecasting outcomes in such situations. To handle sequential data more successfully, there are sophisticated neural network models like recurrent neural network (RNN) as well as long short-term memory (LSTM). Data can be conveyed from one step to the next thanks to the RNNs' loop-based structural design, which allows them to retain critical data across time [4]. To train an RNN, a labeled training dataset is utilized to compute the error or cost between the actual and predicted values. The network's biases and weights are then continually adjusted to lower the error until it reaches the lowest practicable level. A gradient is used in the training process to calculate how much each parameter raises the cost [5]. Utilizing the gradient as a guide, backpropagation is used to iteratively change the error surface's parameters. Errors are propagated through this procedure between the intake and output layers. The main issue with this strategy is that to compute the gradient, partial derivatives must be generated for each parameter. Vanishing gradients are a typical problem during neural network training when gradients disappear or get smaller as they go back through the networks [6], [7]. The vanishing gradient problem was addressed through recurrent neural network development, including LSTM. The fundamental advantage of LSTM is that it can sustain long-term memory, which makes it a great option for tasks requiring long-term memory [8]. The vanishing gradient

problem makes it impossible for traditional recurrent neural networks to retain long-term dependencies, however, LSTM is made to get around these limitations.

The process of decomposing time series data using empirical mode decomposition (EMD) involves separating the data into interpretable intrinsic mode functions (IMFs) and a residue that represents the trend [9]. Obtaining immediate frequency data from natural signals, which often exhibit nonlinear and nonstationary characteristics, is a technique supported by actual evidence.

As opposed to previous methods, the optimization process has witnessed breakthroughs recently, making it more efficient in handling challenges associated with confined, rigid, or unidentified search areas. The genetic algorithm (GA), a powerful computer-based technique, mimics natural selection to find the best solutions. GA utilizes a set of potential solutions called individuals and genetic operations involving selection, crossover, and mutation to produce new individuals [10]. A set of optimization techniques known as meta-heuristic algorithms was developed to overcome the constraints of mathematical computation, convergence issues, and the need for informed guesses [11]. By continually iterating through a collection of initial random replies, these distinct kinds of optimization strategies seek the best general solutions for specific problems. battle royale optimization (BRO), manta ray foraging optimization (MRFO), and grey wolf optimization (GWO) are three of the most well-known techniques in the field [12], [13], [14]. In response to gray wolves' social foraging behavior, the Gray Wolf Optimizer algorithm, a meta-heuristic optimization technique, was created [13].

The motivation for this research is multifaceted and is rooted in the complexity and non-linearity that are inherent in stock markets. These complex relationships are not adequately captured by conventional linear models, which is why advanced models are necessary to make more precise predictions. The unprecedented opportunity to improve stock price forecasting is presented by the accelerated advancements in artificial intelligence and machine learning, which can be employed to implement sophisticated algorithms such as LSTM networks and optimization techniques like MRFO. For investors, analysts, and financial institutions, accurate predictions are essential for making informed investment decisions, optimizing portfolios, mitigating risks, and maximizing returns. The dynamic and volatile nature of stock markets presents a challenge for conventional methods, underscoring the necessity of innovative approaches that provide reliable performance in real-world scenarios. The powerful combination of EMD and LSTM networks is achieved through the synergy between the two. EMD decomposes complex time series data into simpler components, enabling LSTM to accurately represent them. The efficacy of the model is further improved by the integration of advanced optimization methods, such as MRFO, for hyperparameter tuning. The value of accurate stock price prediction models in trading and investment contexts can be underscored by the EMD-MRFO-LSTM model's superior performance on Nasdaq index data, which can demonstrate their practical application and real-world relevance. Following are the contributions of the investigation:

- A novel hybrid model that integrates EMD, LSTM network, and MRFO is introduced in this research. The accuracy and robustness of stock price predictions are improved by this combination, which capitalizes on the assets of each method.
- This research contributes to a more profound comprehension of market dynamics by effectively capturing the complex, non-linear relationships in stock market data. The model's capacity to analyze and process complex financial data is crucial in identifying the factors that influence stock price fluctuations.
- The research offers a thorough assessment of a variety of stock price forecasting models, such as EMD-LSTM, LSTM, EMD-GA-LSTM, EMD-BRO-LSTM, and EMD-GWO-LSTM. The study provides vital insights into the relative performance of these models and the advantages of the proposed EMD-MRFO-LSTM model by benchmarking them.

The following text comprises the remaining contents of the paper. The background of the study is covered in Section II. Related works are specified in Section III. The materials, data gathering, decomposition, evaluation metrics, and methodology are detailed in Section IV. The experimental results are reported in Section V. The discussions of the results are presented in Section VI. In the concluding section, the study's findings are briefly discussed in Section VII. The prospects and challenges are discussed in Section VIII.

II. BACKGROUND

A multitude of determinants impact the stock market, which is a dynamic and intricate system. These determinants comprise investor sentiment, geopolitical events, and [15], [16], [17]. Accurately forecasting stock prices is critical in order to facilitate well-informed investment decision-making and proficient risk management. Conventional financial models, which heavily depend on technical and fundamental analysis, frequently fail to encompass the intricacies of market dynamics. Conventional approaches frequently encounter challenges in capturing the complex patterns and non-linear associations that are intrinsic to market data. The utilization of artificial intelligence models in finance is becoming more prevalent due to their capacity to extract valuable insights from historical data as well as to discover complex relationships among input variables, as discussed in this study. In addition to price forecasting, trend analysis, and anomaly detection, ML algorithms have been implemented in a variety of stock market prediction domains. These methods assist in discerning parallels and distinctions between equities, identifying market anomalies, and revealing concealed correlations that could potentially impact price fluctuations. The article's primary contribution is the introduction of the GWO-LSTM hybrid model, which integrates the operational characteristics of LSTM and GWO to enhance the accuracy of stock price forecasts [18], [19]. To verify the efficacy of the hybrid model, the research utilizes a stringent methodology

that includes data analysis, model evaluation, and comparison with alternative techniques.

III. Related Works

The Shanghai Index's close price for the next day was predicted by Lu et al. [20] using the convolutional neural network (CNN) and LSTM approach. CNN's primary objective was to identify the most valuable features in the data, as well as the closing stock price was predicted using the LSTM approach. The problem stemmed from CNN's inability to identify the optimal feature from the input data. Rezaei et al. [21] introduced two hybrid algorithms called EMD-CNN-LSTM for stock price prediction. On the historical data of the S&P 500, Dow Jones Industrial Average, and Hang Seng Index dataset, Qiu et al. [22] constructed an LSTM-based model. Using the stock trading data from the S&P 500, to forecast the stock price for the ensuing 1, 5, and 10 minutes, Lanbouri et al. [23] employed the LSTM model for the high-frequency. To forecast the close price of the National Stock Exchange and NIFTY50 index, Yadav et al. [24] employed deep learning using the LSTM-based approach. According to the findings, a stateless LSTM model was discovered to be preferable due to its increased stability for time series forecasting problems. For the purpose of stock forecasting using time series data, Dash et al. [25] developed a novel machine learning (ML) technique that makes use of an optimized form of support vector regression. Zhang et al. [26] developed a two-stage prediction methodology that can accurately forecast stock prices. Three machine learning models, a nonlinear ensemble technique, and a decomposition algorithm are all incorporated into this mode. In the first stage, they decomposed stock price time series into sub-series using variational mode decomposition (VMD) as well as then used extreme learning machine (ELM), support vector regression (SVR), and deep neural network (DNN) to forecast each sub-series. Rao et al. [27] addressed the challenge of accurate stock market forecasting by proposing a hybrid machine learning model for stock market prediction. Accounting et al. [28] employed LSTM for predicting the Tehran stock market. The CatBoost algorithm has been utilized to predict financial distress, and the dataset was gathered from the Chinese stock market between 2016 and 2020 by Zhao et al. [29]. Kumar et al. [30] suggested a hybrid deep learning model that combines adaptive particle swarm optimization (PSO) as well as LSTM network.

IV. Materials and Method

A. Data Description

This study employs time series data, which are distinguished by their temporal dependencies. It is crucial to look at the volume of financial data and the open, high, low, and close prices (OHLC) over a specific period to conduct a thorough analysis. Open price is the initial price agreed upon by vendors and purchasers to conduct business following the market's regular trading hours. The open price holds considerable importance as it establishes the security's initial value for the course of the trading day. The term high price refers to the maximum price that fluctuates during a particular trading session for a given security. It represents the highest value point that the security price has reached during that particular period. The high price is indicative of the peak level of investors' demand and enthusiasm for the security throughout the trading session. In the context of a given trading process, a low price denotes the lowest price at which a specific security was transacted. The expression close price denotes the ultimate price at which certain securities were exchanged after a trade. It is the final price at which a transaction occurs just before the market's closing time. Volume denotes the aggregate quantity of shares or contracts that have been traded. As a result, information was gathered between January 2, 2015, and June 29, 2023, from the Yahoo Finance website's Nasdaq index.

The aforementioned information is contained within the dataset used in the investigation. Following the acquisition of the dataset, a comprehensive process of data cleansing was executed in order to preserve the precision and uniformity of the forecasting models. The multi-step procedure was designed to protect the dataset's integrity and avoid any inaccurate or incomplete information from being added that would cause problems. One of the critical stages required a careful analysis of the data to identify any outliers, anomalies, or discrepancies that can potentially undermine the validity of the outcomes. The information was preprocessed and cleaned using a variety of methods to ensure its suitability for use. To prevent gradient errors and inconsistent training results, the data was scaled and normalized. The data were normalized before training using the Min-Max-Scaler method, which helped to guarantee a stable model and avoid having too high weight values. Prices and volume for OHLC were used as the training data, which was fed into the model. High price, low price, open price, as well as volume data were given to the model for testing. The data were divided into 20% for testing and 80% for training.

By preprocessing the data to preserve its accuracy and consistency, it can be guaranteed that the models can accurately learn from historical stock price trends and make precise predictions. The inclusion of OHLC prices and volume data provided a comprehensive view of market activities, allowing the model to capture intricate patterns and dependencies in the stock market data.

B. Empirical Mode Decomposition

The empirical mode decomposition is a technique utilized to break down time series data into two components: a residual component that represents the trend, and a collection of interpretable intrinsic mode functions (IMFs) [31]. It is a technique that is backed by empirical evidence and used to obtain immediate frequency data from natural signals, which frequently display nonlinear and nonstationary properties. An IMF is a mathematical function that has an average value of zero and one extreme value between each time it crosses zero. Fig. 1 to 5 demonstrate the decomposition of a variety of stock market characteristics, including the Open, High, Low, Volume, and Close prices, using EMD. Each figure commences with the original time series data at the top, followed by 11 IMFs that capture various frequency components of the data. The highest frequency is captured in IMF 1, and subsequent IMFs progress to lower frequencies. After extracting the IMFs, the residual component is represented by the bottom plot in each figure, which illustrates

the long-term trend. In addition to enhancing prediction models, this decomposition process also facilitates a detailed analysis by isolating various patterns within the stock market data.

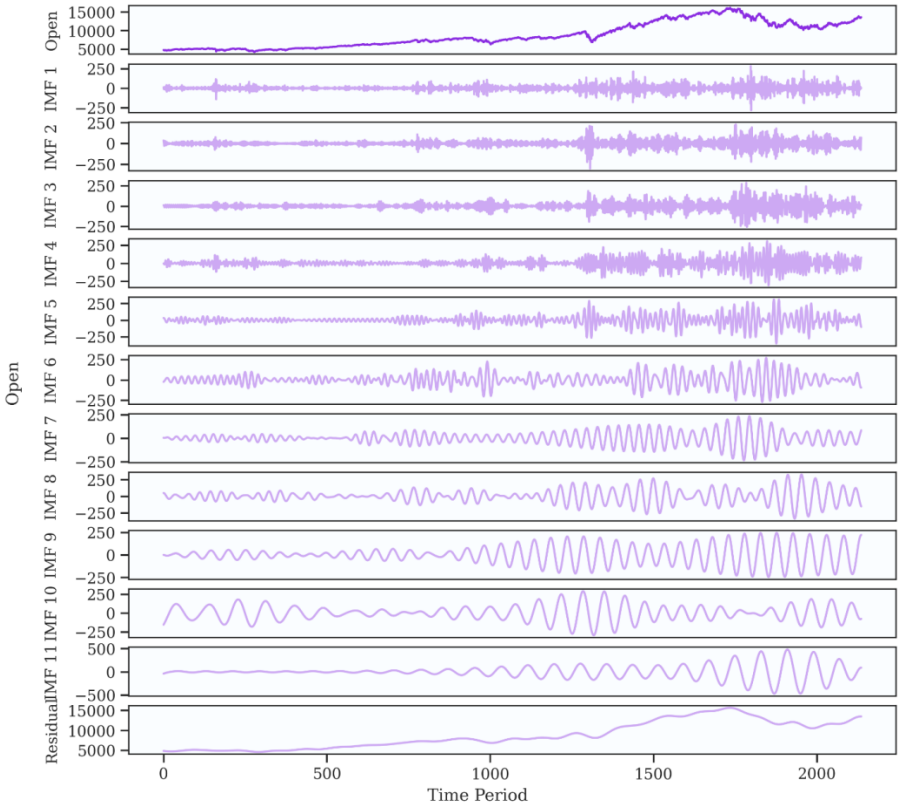


Fig. 1. The decomposition of Open price by EMD.

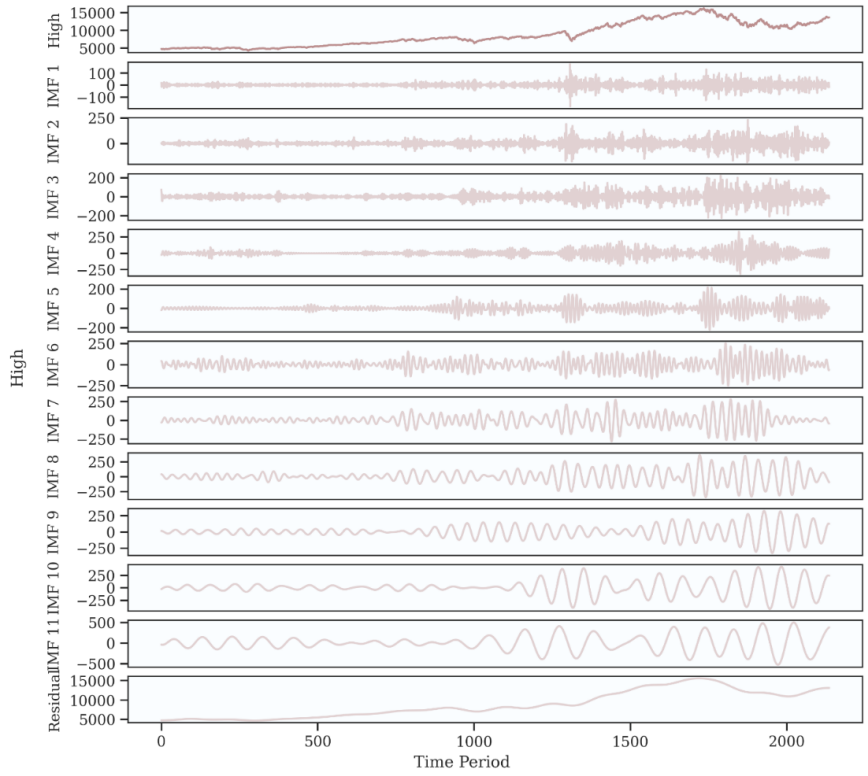


Fig. 2. The breakdown of High price using EMD.

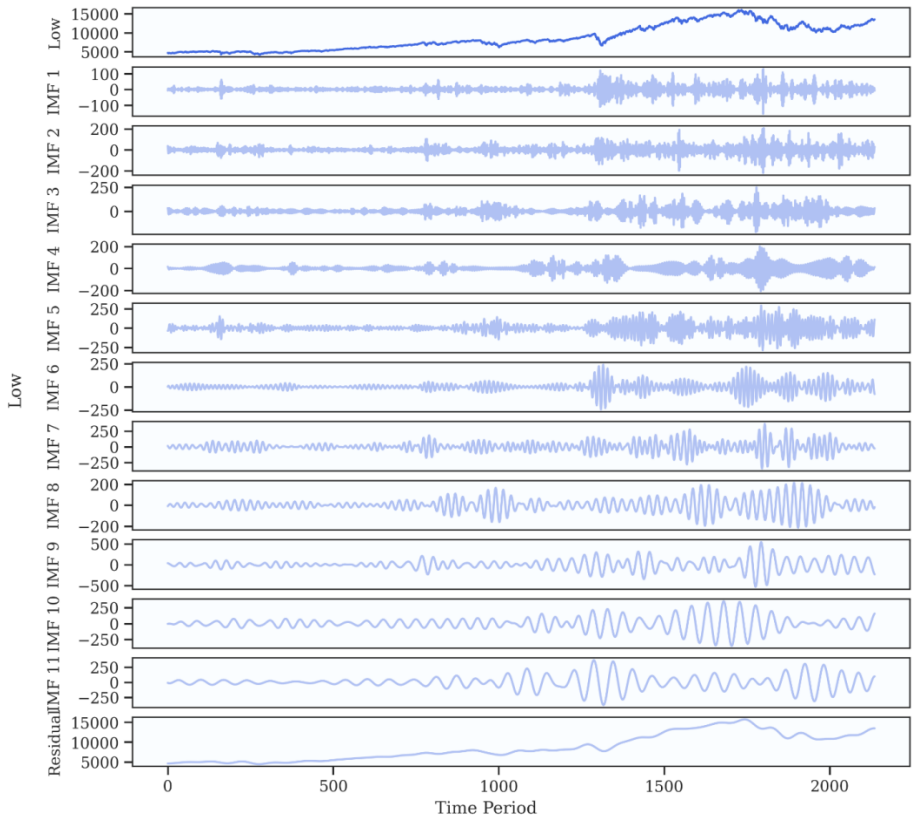


Fig. 3. The breakdown of Low price using EMD.

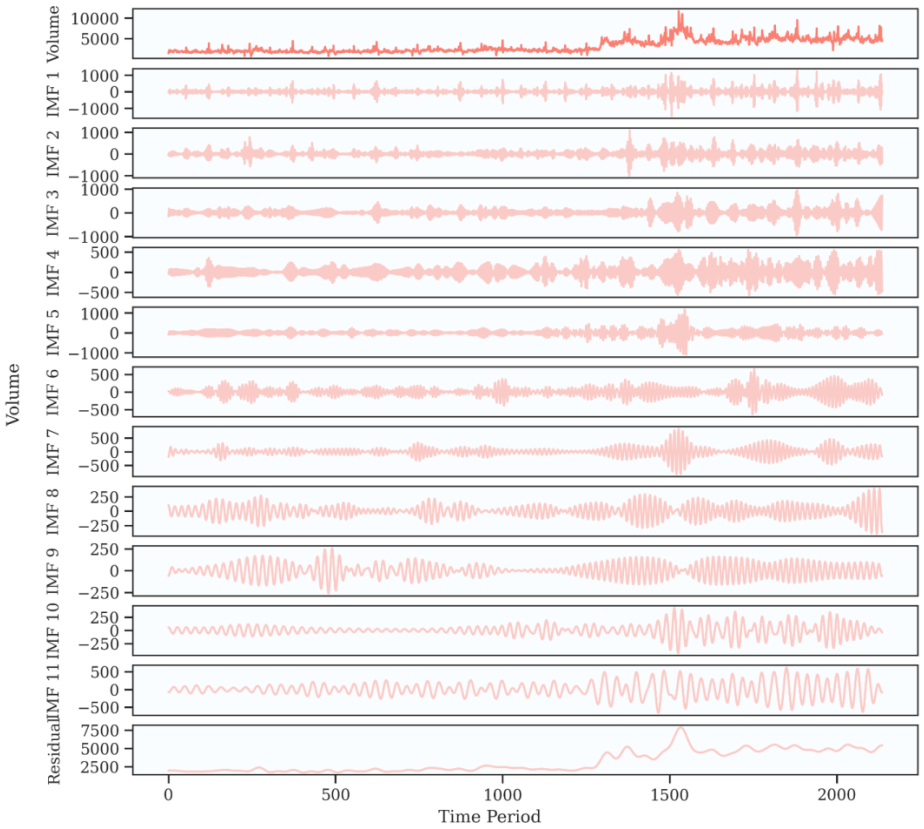


Fig. 4. The decomposition of Volume by EMD.

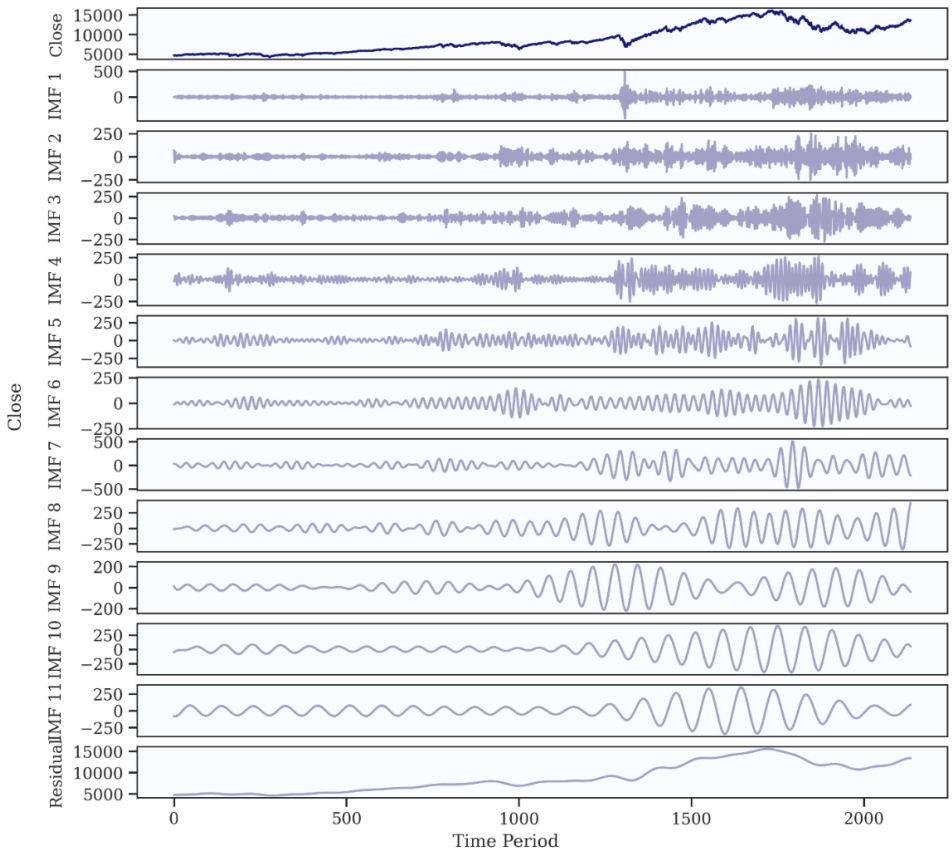


Fig. 5. The decomposition of Close price by EMD.

The method employed by EMD to decompose respiratory motion into IMFs is explained as follows [31].

- $Z_1$is the mean value of the upper and lower boundaries of a time series signal $x(t)$. These boundaries are determined by interpolating the local maximum and minimum points.
- Deducting Z_1 from the original time series $x(t)$yields the initial component P_1, which is defined as $p_1 = x(t) - Z_1$.
- Let P_1 represent the data in which the means of the upper and lower envelopes are Z_{11} during the second shifting process; $p_{11} = p_1 - z_{11}$.
- By the following conditions, the shifting process is terminated k times: (a) z_{1k} approaches zero; (b) the distinction between zero-crossings and the p_{1k} number of extrema does not surpass one, or (c) the maximum number of iterations has been completed. When this occurs, the IMF, denoted as p_{1k}, can be determined by dividing $p_{1k} = p_{1(k-1)} - z_{1k}$.
- The initial IMF (the shortest component of the data), represented by$a_1 = p_{1k}$, is subtracted from the data as $x(t) - a_1 = y_1$. This operation is repeated for each of the following values of $y_2 = y_1 - a_2, \ldots y_n = y_{n-1} - a_n$.

Consequently, the initial time series $x(t)$is reduced to the collection of IMF functions shown below:

$$x(t) = (\sum_{i=1}^{n} a_i + y_n \quad (1)$$

C. Manta Ray Foraging Optimization

1) Inspiration: Manta rays are complex organisms despite their menacing appearance. They are among the largest marine organisms known to science [14]. Manta rays are flat-bodied from top to bottom as well as have two pectoral fins; they swim elegantly while birds soar effortlessly. Furthermore, they possess a pair of cephalic appendages that protrude anterior to their enormous, terminal jaws. They funnel prey as well as water into their jaws utilizing horn-shaped cephalic lobes while foraging. Then, using modified gill rakers, the prey is removed from the water. Two distinct species are identified as manta rays. The reef manta ray (Manta alfredi) is one of them that can attain a width of 5.5 meters and inhabits the Indian Ocean, western Pacific, and southern Pacific. The other is the 7-meter-wide giant manta ray (Manta birostris), which inhabits mild temperate, tropical, and subtropical oceans [14]. Their estimated age of existence is five million years. Many do not live to be the average age of 20 years due to the fact that they are pursued by fishermen. The illustration of MFRO is covered in Fig. 6.

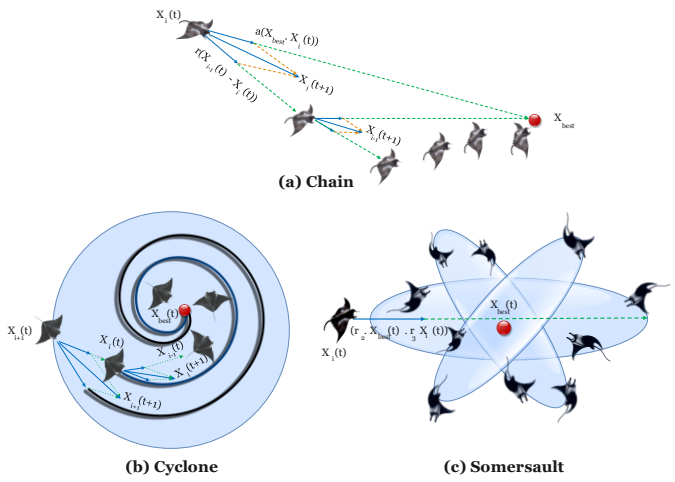


Fig. 6. The illustration of MRFO.

MRFO has been impacted by three distinct foraging behaviors: cyclone foraging, chain foraging, as well as somersault foraging.

2) *Chain foraging:* Manta rays can travel in the direction of plankton they detect using MRFO. An elevated plankton concentration correlates with a more favorable geographical location. While the optimal solution remains unknown, MRFO hypothesizes that the ideal solution thus far is the high-concentration plankton that manta rays desire to consume. A foraging chain is formed when manta rays are arranged from head to tail. At the same moment that individuals approach the food, they also approach the item that is immediately in front of them. Put simply, in each iteration, each individual gets revised with the best possible option that has been identified thus far, in addition to the solution that is currently in front of it. This is a representation of the chain foraging theoretical framework:

$$x_i^d(t+1) = \begin{cases} x_i^d(t) + r.\left(x_{best}^d(t) - x_i^d(t)\right) \\ +\alpha.\left(x_{best}^d(t) - x_i^d(t)\right) \ i = 1 \\ x_i^d(t) + r.\left(x_{i-1}^d(t) - x_i^d(t)\right) \\ +\alpha.\left(x_{best}^d(t) - x_i^d(t)\right) \ i = 2, \dots N \end{cases} \quad (2)$$

$$\alpha = 2.r.\sqrt{|\log(r)|} \quad (3)$$

where $x_i^d(t)$denotes the location of the i-th individual at time t in the d-th dimension, r signifies an arbitrary vector from 0 to 1, α denotes the value of the ratio, and $x_{best}^d(t)$represents the plankton with the highest concentration. The current status of the i-th individual is established using the situation $x_{i-1}(t)$ as well as the position $i-1$-th of the food at the time $x_{best}(t)$ respectively.

3) *Cyclone foraging:* When a group of manta rays detects a region of deep-water plankton, they will spiral in their pursuit of the food in a continuous foraging chain. In contrast, as part of their cyclone foraging strategy, manta ray clusters swim each individual manta ray in the direction of the one in front of it, as opposed to spiraling towards the food. In other words, manta ray colonies engage in spiral foraging in a helical formation. An individual not only replicates the motion of the one preceding it but also proceeds in a spiral trajectory toward sustenance. The expression in mathematics that characterizes the spiral motion of manta rays in a two-dimensional space is as follows:

$$\begin{cases} X_i(t+1) = X_{best} + r.\left(X_{i-1}(t) - X_i(t)\right) \\ \quad +e^{bw}.cos(2\pi w).\left(X_{best} - X_i(t)\right) \\ Y_i(t+1) = Y_{best} + r.\left(Y_{i-1}(t) - Y_i(t)\right) \\ \quad +e^{bw}.sin(2\pi w).\left(Y_{best} - Y_i(t)\right) \end{cases} \quad (4)$$

where w represents a random number from zero to one.

This behavior of motion is extensible to n_D space. Theoretical representation of cyclone scavenging may be defined succinctly as:

$$x_i^d(t+1) = \begin{cases} x_{best}^d + r.\left(x_{best}^d(t) - x_i^d(t)\right) \\ +\beta.\left(x_{best}^d(t) - x_i^d(t)\right) \ i = 1 \\ x_{best}^d + r.\left(x_{i-1}^d(t) - x_i^d(t)\right) \\ +\beta.\left(x_{best}^d(t) - x_i^d(t)\right) \ i = 2, \dots, N \end{cases} \quad (5)$$

$$\beta = 2e^{r_1\frac{T-t+1}{T}}.\sin(2\pi r_1) \quad (6)$$

The variables denoted as $[0,1]$, T the maximal number of iterations, β the weight coefficient, and r_1 the rand number.

Each individual conducts the search in a random manner, using the food as their reference position. As a result, the region where the most effective solution has been identified thus far benefits from cyclone foraging. Additionally, this behavior serves to significantly enhance the exploration process. By designating each individual, a reference position that is arbitrary and distinct from the current optimal one, we can compel them to seek out a new position. The mathematical equation for this mechanism, which enables MRFO to conduct an exhaustive global search and is primarily concerned with exploration, is provided below.

$$x_{rand}^d = Lb^d + r.(Ub^d - Lb^d) \quad (7)$$

$$x_i^d(t+1) = \begin{cases} x_{rand}^d + r.\left(x_{rand}^d - x_i^d(t)\right) \\ +\beta.\left(x_{rand}^d - x_i^d(t)\right) \ i = 1 \\ x_{rand}^d + r.\left(x_{i-1}^d - x_i^d(t)\right) \\ +\beta.\left(x_{rand}^d - x_i^d(t)\right) \ i = 2, \dots, N \end{cases} \quad (8)$$

The variable x_{rand}^d denotes a position generated at random within the search space. Lb^d as well as Ub^d represent, respectively, the d-th dimension's minimum and maximum boundaries.

4) Somersault foraging: This behavior is characterized by the food's position being considered a pivot. Every individual undergoes a series of back-and-forth swims that involve a pirouette to a different position. As a result, they continuously adjust their positions in accordance with the most advantageous one discovered thus far. The formulation of the mathematical model is as follows:

$$x_i^d(t+1) = x_i^d(t) + S.\left(r_2.x_{best}^d - r_3.x_i^d(t)\right), i = 1, \dots, N \quad (9)$$

where S is the somersault factor that controls the variety of somersaults that manta rays perform, $S = 2, r_2$ and r_3 are two arbitrary values from the interval $[0,1]$. The framework of MRFO can be displayed in Fig. 7.

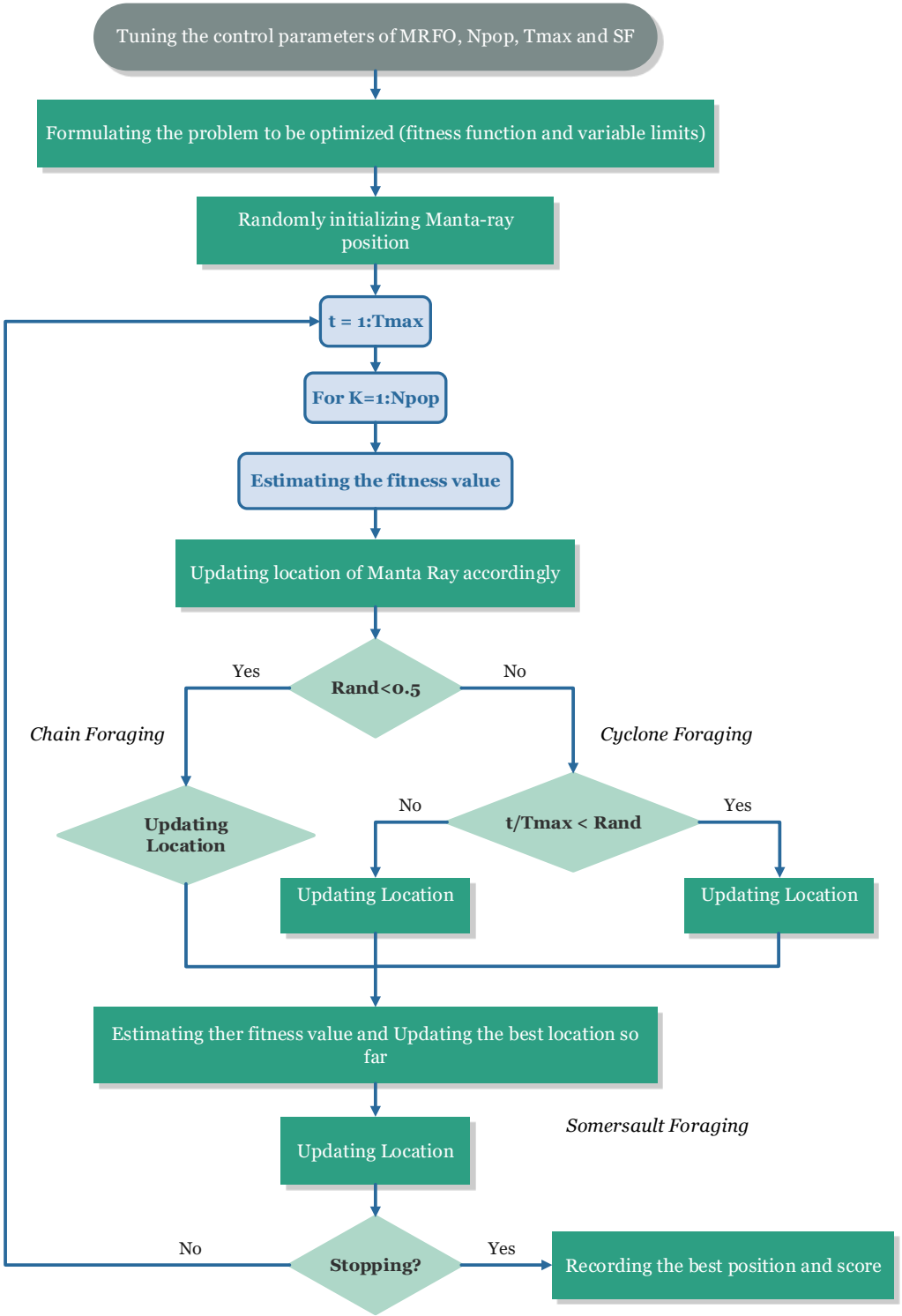


Fig. 7. The framework of MRFO.

D. Long Short-term Memory

The long short-term memory is a very well-liked and successful deep learning method. It is an effective tool for a variety of applications since it is built to handle and process enormous amounts of data [8]. Three a memory unit and gating units are used by the LSTM model to process incoming input. Together, these components control the flow of data, eliminating any extraneous material and generating output that is both brief and pertinent. The forgetting gate removes any potentially present irrelevant information, whereas the input gate handles the processing of incoming data. The function of the output gate is to regulate the flow of data that has been processed and generate a precise and relevant output. The gate formulas are used to sort, process, and store data, while the memory unit stores pertinent information for later use. The LSTM model may exclude any extraneous data by employing these algorithms, ensuring that only essential data is retained. As a result, it is a very effective method for handling vast amounts of data without creating extra clutter. The LSTM method is a significant asset in the field of deep learning since it is a strong and dependable tool for processing complex data sets [32]. The LSTM's operation is shown in the following equations. The forget gate decides whether to preserve or discard the information. A sigmoid layer processes the current input and the prior hidden state. The value that this layer output ranges from 0 to 1. Keep the data if the result value is more closely related to 1. Otherwise, disregard the knowledge.

$$f_t = \sigma\left(W_f \times [h_{t-1}, x_t] + b_f\right) \tag{10}$$

where σ denotes the sigmoid function, W_f is the mass that is linked to the forget gate, the prior hidden state demonstrates as h_{t-1}, the input value is x_t, as well as b_f denotes the bias associated with the forget gate.

The gate that accepts input is responsible for modifying the cell state. An individual sigmoid layer and a tan layer process the present input as well as the prior hidden state first. A data value is transformed by the sigmoid layer into a value that ranges from 0 to 1. Using the tanh layer, a data value is transformed into a value between -1 and 1. The outputs of the sigmoid layer and the tanh layer are multiplied by a point-wise procedure. then computes the new cell state value.

$$i_t = \sigma(W_i \times [h_{t-1}, x_t] + b_i) \tag{11}$$

The weight that is expressed as a symbol for the input gate by W_i and b_i is the component of the input gate that introduces bias.

The below equation is used to calculate the output of the tanh layer:

$$\tilde{C}_t = tanh(W_c[h_{t-1}, x_t] + b_c) \tag{12}$$

Equation below is used to calculate the new cell state:

$$C_t = f_t \times C_{t-1} + i_t \times \tilde{C}_t \tag{13}$$

The output gate decides what secret state will be shown next. A sigmoid layer processes the input at hand in addition to the preceding hidden state at the beginning. The changed cell state is then transmitted to a tan layer. The outputs of the sigmoid layer and the tanh layer are multiplied point-wise to find the subsequent hidden state. The new cell state as well as the next concealed state are then transferred to the following time step.

$$o_t = \sigma(W_o[h_{t-1}, x_t] + b_o) \tag{14}$$

where b_o denotes the bias associated with the output gate and w_o is the weight associated with the output gate.

$$h_t = o_t tanh(C_t) \tag{15}$$

Using the following equation, the subsequent hidden state is determined:

$$h_t = o_t tanh(C_t) \tag{16}$$

E. Genetic Algorithm

The genetic algorithm is a method of computation that simulate natural selection's technique to handle optimization and search issues [10]. With this algorithm, a collection of probable solutions known as people is created. To create new individuals, these individuals are then subjected to genetic processes like mutation, recombination, and selection. The illustration and framework of GA can be seen in Fig. 8 and 9. This evaluation procedure is iterative and is performed over several generations until a workable answer is discovered. As a result, GA is a potent instrument that is frequently employed in many different industries, involving engineering, finance, and science, to name a few [33]. Three components are essential to GA [34]. A chromosome is an encoded string of numbers or characters that is given to each individual by the encoding component. Encoding methodology is determined by the precise problem that must be resolved. Following this, the fitness metric is applied to assess how each individual embodies the solution. The ability to exercise has been deliberately designed to mitigate the present issue. Evolutionary operators utilize the crossing, transformation, and choice procedures. When two people's chromosomes crossover, a new being is created, mutation indiscriminately modifies an individual's chromosomes, and selection is used to determine which individuals are the most fertile.

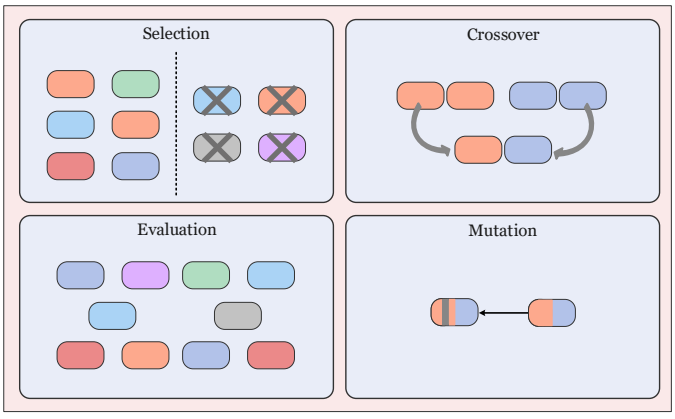


Fig. 8. The illustration of GA.

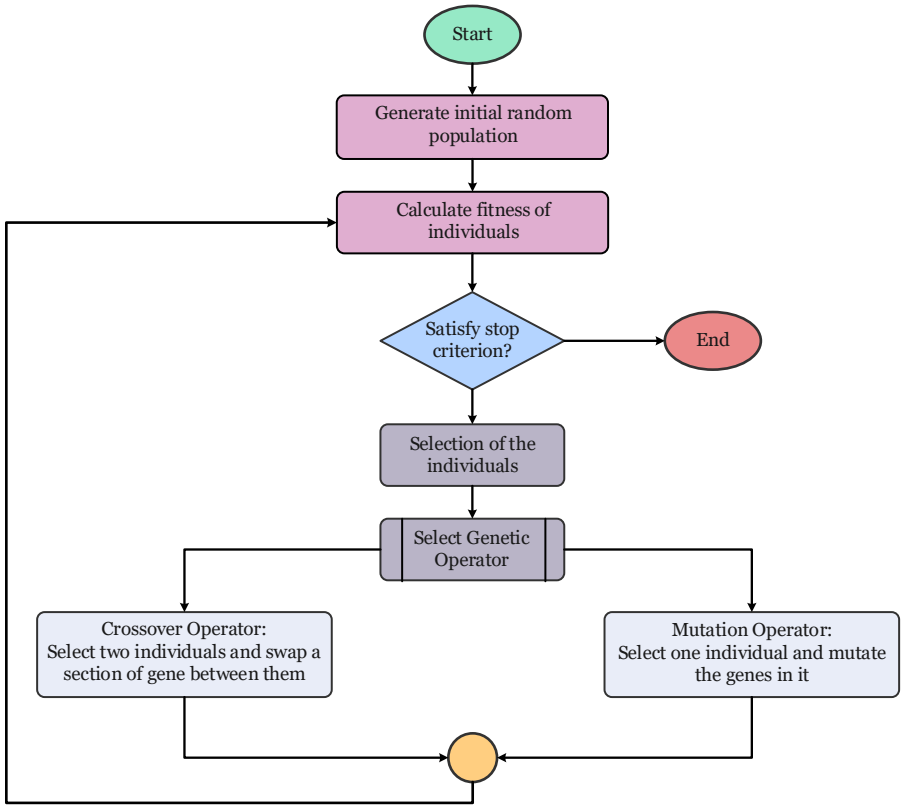


Fig. 9. The framework of GA.

F. Battle Royale Optimizer

The battle royale optimizer is the name of a meta-heuristic algorithm that Farshi proposed [12]. The algorithm was inspired by a popular multiplayer online game in which players must eliminate rivals to find a safe haven to survive. Stepping outside the safe zone in the game puts the player at risk of getting hurt or being eliminated [12]. The damage rate of the injured player is calculated using the equation shown below:

$$x_i.damage = x_i.damage + 1 \tag{17}$$

Injured players strive to switch to other positions to confront the enemy. The equation that follows shows the players' newest placements:

$$x_{dam,d} = x_{dam,d} + r\left(x_{best,d} - x_{dam,d}\right) \tag{18}$$

where $x_{dam,d}$ denotes the location of the wounded player in the dimension d, the best solution in dimension d is indicated by the notation $x_{best,d}$, and r is a random number generated from a uniform distribution between 0 and 1. The search agents are distributed at random throughout the problem space and cover it equally.

In a d-dimensional problem space, the upper limit and lower bound are represented by ub_d and lb_d the following equation:

$$x_{dam,d} = r(ub_d - lb_d) + lb_d \tag{19}$$

The best approach is shown in the formula below, and the worst-fitting options are discarded. Considering this, the starting value Δ is $\log_{10}(MaxCicle)$, where MaxCicle is the number of repetitions:

$$\Delta = \Delta + round\ (\frac{\Delta}{2}) \tag{20}$$

G. Gray Wolf Optimizer

A meta-heuristic technique has been used to create a novel optimization process known as the gray wolf optimizer. The approach, which mimics the seeking strategies and social organization of gray wolves, was first presented by Mirjalili et al. [13]. The best solution is Alpha, while there are four alternatives in the leadership hierarchy, Omega is the last competitor: Beta, Alpha, Omega, and Delta.

The strategy employs three primary hunting techniques to mimic wolf behavior: pursuing prey, enclosing prey, and attacking prey.

$$\begin{aligned} \vec{D} &= \left|\vec{C} \cdot \vec{X}_p(t) - \vec{X}(t)\right| \\ \vec{X} \mid (t+1) &= \vec{X}_p(t) - \vec{A} \cdot \vec{D} \end{aligned} \tag{21}$$

where $\overrightarrow{C}$ and $\vec{A}$ represent coefficient vectors, $\overrightarrow{D}$ signifies motion, t is the current stage of iteration, as well as $\overrightarrow{X}$ represents the whereabouts of a gray wolf. The construction of the parameter variables ($\overrightarrow{A}$ and $\overrightarrow{C}$) is based on the subsequent relationships:

$$\vec{A} = 2\vec{a} \cdot \vec{r}_1 - \vec{a}$$
$$\vec{C} = 2 \cdot \vec{r}_2 \quad (22)$$

The location of new search reps that involve omegas is adjusted utilizing the data from alpha, beta, and delta as follows:

$$\vec{D}_a = \left|\vec{C}_1 \cdot \vec{X}_a - \vec{X}\right|, \vec{D}_\beta = \left|\vec{C}_2 \cdot \vec{X}_\beta - \vec{X}\right|, \vec{D}_\delta = \left|\vec{C}_3 \cdot \vec{X}_s - \vec{X}\right| \quad (23)$$

$$\vec{X}_1 = \vec{X}_a - \vec{A}_1 \cdot \vec{D}_u, \vec{X}_2 = \vec{X}_\beta - \vec{A}_2 \cdot \vec{D}_\beta, \vec{X}_3 = \vec{X}_b - \vec{A}_3 \cdot \vec{D}_\delta \quad (24)$$

$$\vec{X}(t+1) = \frac{\vec{X}_1 + \vec{X}_2 + \vec{X}_3}{3} \quad (25)$$

where the wolves are indicated by the subscripts $\alpha, \beta,$ and δ to mount a final attack to complete the task. $\vec{a}$ is used to mimic the last attack by changing a value from 2 to 0, whereas an is a random variable between $-2\vec{a}$ and $2\vec{a}$. As a result, decreasing $\vec{a}$ would also cause $\vec{A}$ to decrease. The wolves were forced by $\left|\vec{A}\right| < 1$to cling to their prey. After following the leader wolf on a pack search, gray wolves disperse to collect sustenance before reconvening for an assault. In pursuit of prey, wolves may divide into groups if the value of $\left|\overrightarrow{A}\right|$ exceeds unity at random. Two of the most critical settings for the algorithm used by GWO are the number of wolves and generation. Generation after generation signifies a wolf's conclusive action. In addition, the number of wolves precisely reflects changes in performance estimates over time. In other words, the generation size multiplied by the wolf population will result in an equivalent quantity of objective function evaluations.

$$OFEs = N_W \times N_G \quad (26)$$

H. Performances Metrics

Several performance metrics were utilized to determine the dependability of future estimates. The root means square error (RMSE), mean absolute error (MAE), coefficient of determination (R^2), and mean absolute percentage error (MAPE) were some of these. The accuracy of forecasting models can be evaluated using these metrics, which additionally helps to ensure that the estimates are solid and reliable.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n}} \quad (27)$$

$$MAPE = \left(\frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right|\right) \times 100 \quad (28)$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \quad (29)$$

$$MAE = \frac{\sum_{i=1}^{n}|y_i - \hat{y}_i|}{n} \quad (30)$$

where $\overline{y}_i$ is the sample mean, $\hat{y}_i$is the predicted value, and y_i is the actual value.

V. Experimental Results

A. Statistical Values

Table I which contains comprehensive statistical data about the dataset is included in this section of the study. The data are made more understandable by the table's inclusion of OHLC prices and volume figures. A robust dataset for analysis is provided by the 2137 observations in this table. This provides a clear indication of the data's scope and the extensive period over which it was collected by presenting the count. The central tendencies of the dataset are represented by the mean values in the table, which provide an average perspective on the market's performance during the analyzed period. For example, the mean closing price of 8745.8210 serves as a foundation for comprehending typical market behavior. In the same way, the minimum and maximum values capture the extremes within the dataset, indicating the lowest and highest market activities observed, such as the minimum volume of 706.880 and the maximum closing price of 16057.440. Additionally, the table contains skewness values, which provide a deeper understanding of the asymmetry in the data distribution. The skewness, which is nearly zero, indicates that the distribution is fairly symmetrical, indicating that the data does not significantly favor one tail. This information is essential for comprehending the fundamental patterns in the data, which can have a substantial impact on modeling endeavors.

TABLE I. Summary Statistics for the Data Set

	Open	High	Low	Volume	Close
count	2137	2137	2137	2137	2137
mean	8744.3560	8805.2870	8677.5740	3143.80	8745.8210
minimum	4218.810	4293.220	4209.760	706.880	4266.840
maximum	16120.920	16212.230	16017.230	11621.190	16057.440
skewness	0.4993140	0.493110	0.5027470	1.0283760	0.4977320

B. Models' Outcomes

Using data from Nasdaq Finance, the proposed method was both trained and evaluated. To forecast a numerical value, regression analysis is implemented. A large number of vendors, purchasers, and investors participate in the stock market, an extremely risky investment destination. A share typically signifies ownership in a corporation. Recognizing

stock price patterns and making investments at the optimal time and location are the sole prerequisites for potentially generating profits. Determining as well as evaluating the hybrid algorithm that is most efficient in predicting stock pricing is thus the primary objective, provided that an event is accurately predicted at the appropriate moment. Determining and evaluating the most effective hybrid algorithm for investing price forecasts is the principal aim of this study. Elaborate variables that influence stock market patterns have been analyzed in conjunction with the development of forecasting models. The goal was to provide insightful information that would aid investors and analysts in making prudent investment decisions. A detailed assessment of the performance of each mode, as well as an examination of its effectiveness, is presented in Table II. The principal goal of this investigation is to ascertain as well as evaluate the most efficient hybrid method for predicting stock prices. Through the development of predictive models and an understanding of the fundamental factors that influence stock market trends, this research aims to assist analysts and investors in making well-informed investment choices.

TABLE II. EVALUATION FINDINGS OF THE SIX BENCHMARKING ALGORITHMS' STATISTICAL FORECASTS

MODEL/METRICS	TRAIN SET				TEST SET			
	R^2	***RMSE***	***MAE***	***MAPE (%)***	R^2	***RMSE***	***MAE***	***MAPE (%)***
LSTM	0.9766	449.05	383.63	5.62	0.9609	311.60	259.87	2.06
EMD-LSTM	0.9817	396.55	247.51	2.59	0.9703	271.63	208.55	1.70
EMD-GA-LSTM	0.9890	307.53	193.82	2.16	0.9809	217.40	165.24	1.32
EMD-BRO-LSTM	0.9916	268.93	178.14	2.17	0.9904	154.48	121.27	0.98
EMD-GWO-LSTM	0.9966	170.78	153.74	1.99	0.9951	122.01	93.96	0.75
EMD-MRFO-LSTM	0.9981	127.72	107.76	1.43	0.9973	91.99	71.54	0.57

VI. DISCUSSION

To assess the efficacy of the data analysis, four widely used metrics—RMSE, MAPE, MAE, and R^2—were applied. To fully evaluate the results, a comprehensive assessment of the accuracy of the analysis, precision, as well as overall performance can be done using these metrics. R^2, RMSE and MAPE criteria were assessed for the LSTM model using EMD decomposition both with and without the optimizer. During the evaluation period, LSTM achieved a R^2value of 0.9609, as shown in Table II and Fig. 10 and 11. Frequently, the process of decomposing a problem reveals functions or recurrent patterns that apply to numerous components. The act of reusing modules or components serves to enhance stability, decrease the probability of errors, and accelerate the development process. In light of the data provided in this section, Clearly, it is apparent that the utilization of EMD decomposition decreases the value of MAE to 208.55 and 247.51, respectively, during the testing and training phases. When optimizers are added to the LSTM model, its efficacy is substantially enhanced. By enabling the efficient modification of model parameters, optimizers mitigate performance degradation. In order to achieve a convergent set of parameters, a multitude of optimizers implement unique strategies, including adaptive learning rate, slope descent, momentum, and others. This degree of effectiveness accelerates the convergence process during training. The cited examples demonstrate that the simultaneous implementation of the GA optimizer and EMD decompose produces a more precise outcome and reduces calculation error, as shown in Table II. Moreover, in conjunction with GA, the BRO optimizer exhibited enhanced performance, as indicated by its reduced RMSE value. By attaining an MAE score of 93.96, EMD-GWO-LSTM demonstrated superior efficacy in comparison to EMD-BRO-LSTM. According to regression analysis, the EMD-MRFO-LSTM model is an exceptionally precise and reliable instrument. The respective R^2 score of the model for the testing dataset was 0.9973. The outcomes of this study illustrate the model's robust predictive capability and its ability to explain nearly all of the variability in the data. Decreased values signify greater precision, as they represent the discrepancy between the predicted and realized values. In light of the exceptional accuracy demonstrated by both the training and assessment datasets, the EMD-MRFO-LSTM model has been validated.

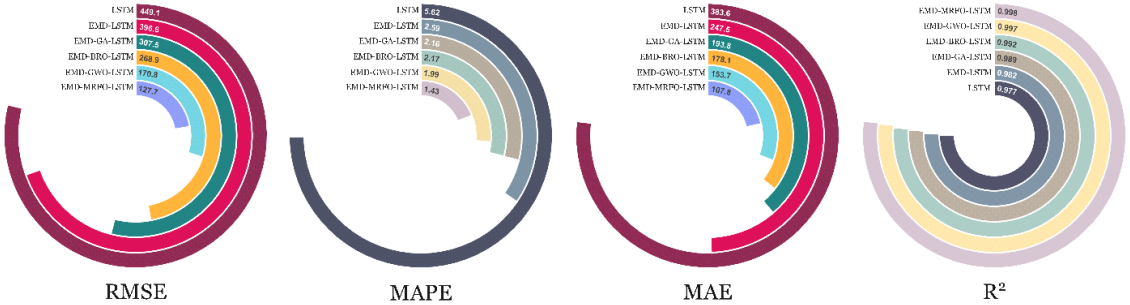


Fig. 10. Evaluation values for each model in the training set.

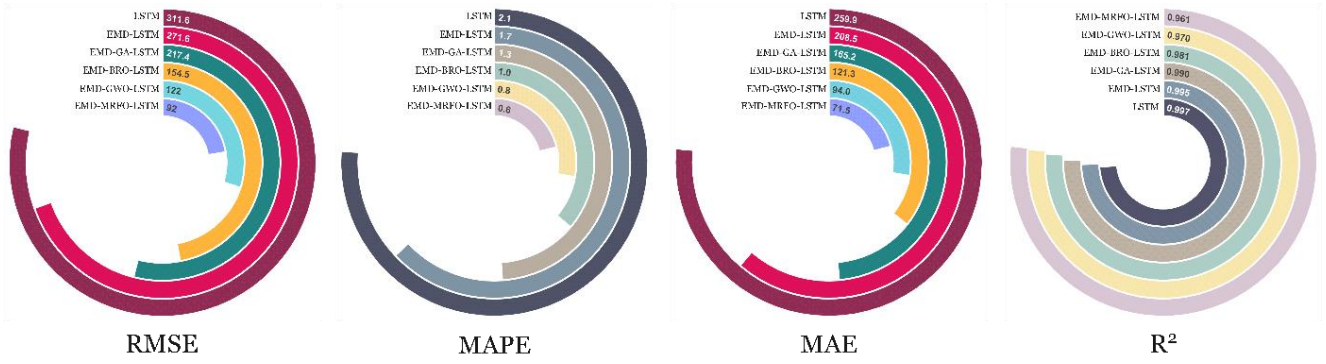


Fig. 11. Evaluation values for each model in the testing set.

The EMD-MRFO-LSTM model's ability to forecast the Nasdaq index during both the training and testing phases is illustrated in Fig. 12 and 13. The black line in Fig. 12, which illustrates the training phase, represents the actual Nasdaq index, while the red line illustrates the predicted values from the model. The model's capacity to learn from historical data is underscored by the close alignment of the two curves. Particularly, the Nasdaq index's reversal points, peaks, and troughs are precisely captured by the model. The predicted values closely follow the actual values at reversal points, where the market changes direction. The model also accurately predicts the market's peaks and troughs, matching them with precision. The model's capacity to learn intricate market dynamics and patterns is illustrated by this precise fit. The EMD-MRFO-LSTM model's robustness and generalization capability are validated by its continued performance in Fig. 13, which represents the testing phase. The model closely aligns the predicted peaks and troughs with the actual market values and maintains its accuracy in predicting reversal points, where the market shifts direction. This consistency in the testing phase, during which the model encounters new, unseen data, emphasizes its reliability and effectiveness in real-world applications. The model's ability to accurately capture critical market prices is demonstrated by the detailed fit between the predicted and actual curves in both phases, rendering it a valuable tool for investors and analysts. The EMD-MRFO-LSTM model's utility in stock market prediction is confirmed by its ability to accurately predict future market trends, which aids in the making of informed investment decisions.

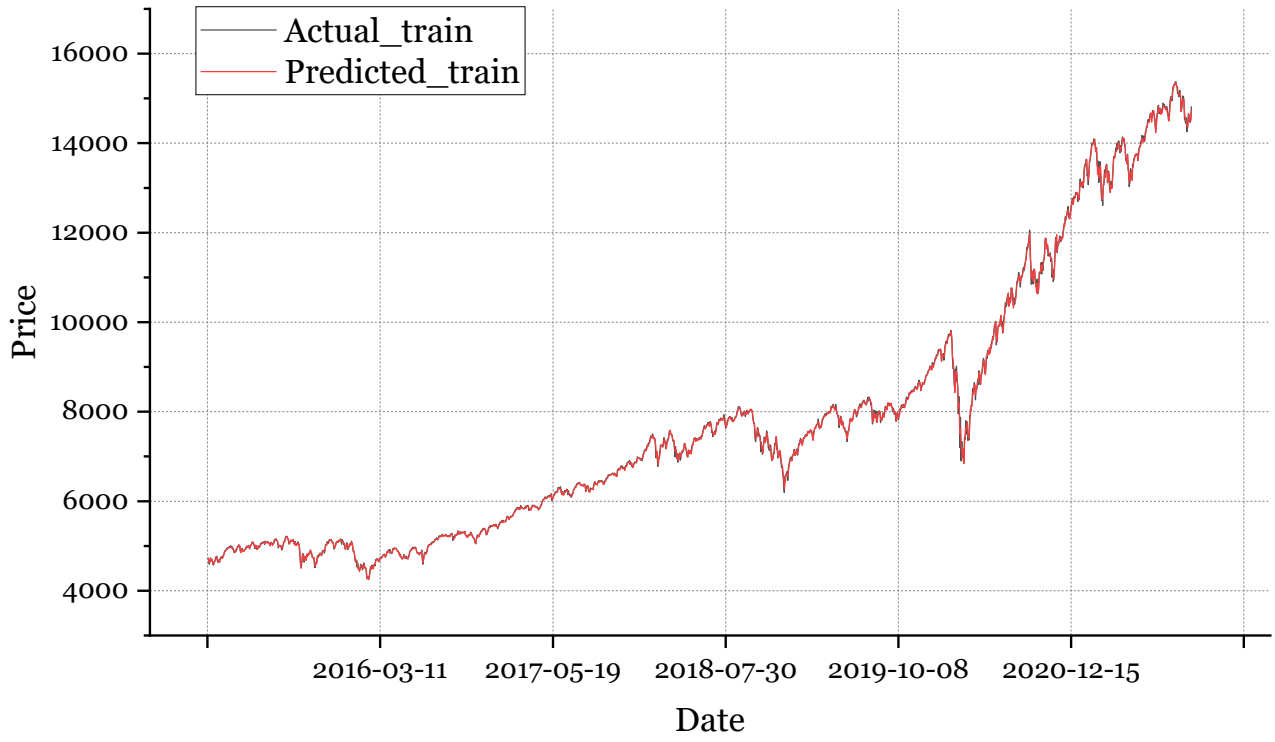


Fig. 12. Training-generated forecasting curve employing EMD-MRFO-LSTM.

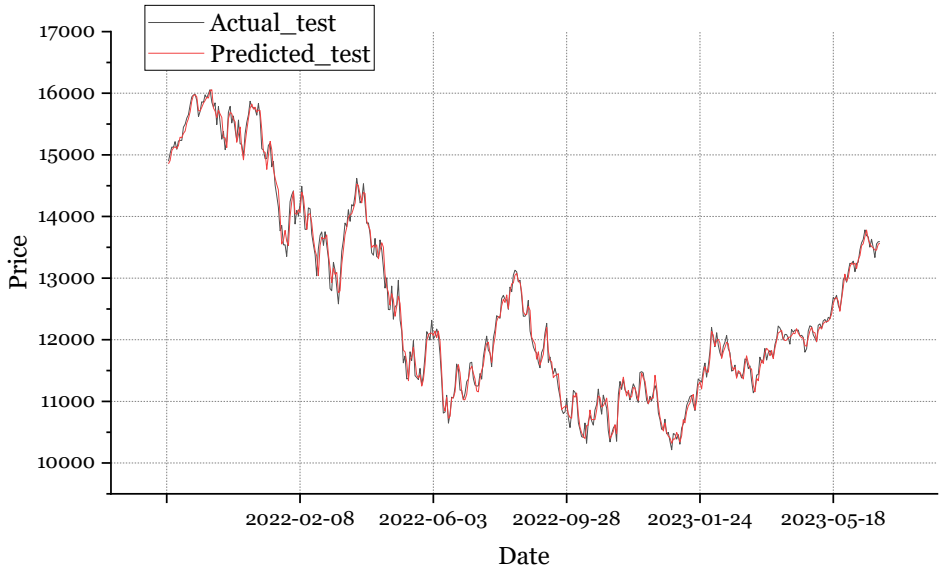


Fig. 13. Testing-generated forecasting curve employing EMD-MRFO-LSTM.

According to Table III, in comparison to both traditional and advanced benchmark models, such as support vector regression (SVR), random forest (RF), multilayer perceptron (MLP), and long short-term memory (LSTM), the proposed EMD-MRFO-LSTM model has exhibited remarkable superiority in predicting stock market prices. In particular, the model attains an R^2 value of 0.9973, which is statistically significantly greater than those of SVR (0.9097), RF (0.9258), MLP (0.9442), and LSTM (0.9609). The EMD-MRFO-LSTM model's high R^2 value suggests that it can account for nearly 99.73% of the variance in the target data, which is a significant improvement over the closest competing model, LSTM, which explains 96.09%. Additionally, the EMD-MRFO-LSTM model demonstrates superior error metrics, with an RMSE of only 91.99, in contrast to 311.60 for LSTM, 329.03 for MLP, 379.37 for RF, and 418.38 for SVR. The MAE exhibits a comparable pattern, as evidenced by its value of 71.54, which is significantly lower than that of LSTM (259.87), MLP (254.90), RF (287.73), and SVR (361.18). Furthermore, the MAPE of the proposed model is a mere 0.57%, which is a negligible error margin in comparison to 2.06% for LSTM, 2.09% for MLP, 2.24% for RF, and 3% for SVR.

TABLE III. PERFORMANCE OF THE PROPOSED MODEL IN COMPARISON WITH BENCHMARK MODELS

MODEL/METRICS	***TEST SET***			
	R^2	***RMSE***	***MAE***	***MAPE (%)***
SVR	0.9097	418.38	361.18	3
RF	0.9258	379.37	287.73	2.24
MLP	0.9442	329.03	254.90	2.09
LSTM	0.9609	311.60	259.87	2.06
EMD-MRFO-LSTM	0.9973	91.99	71.54	0.57

The EMD-MRFO-LSTM model's stability and reliability are further demonstrated by its consistent results across various cross-validation methods, as illustrated in Table IV. The model achieves an R^2 of 0.9967, an RMSE of 92.82, an MAE of 72.33, and a MAPE of 0.58% through 5-fold cross-validation. The model maintains this high-performance level with an R^2 of 0.9970, an RMSE of 92.18, an MAE of 71.91, and a MAPE of 0.57% when a 10-fold cross-validation is implemented. The model's potential as a dependable tool for financial forecasting is underscored by these results, which are capable of delivering precise and stable predictions under varying conditions.

The EMD-MRFO-LSTM method utilized in the current study has the highest R^2 value among other methodologies in the literature, as illustrated in Table V. This model outperforms linear regression, support vector machine (SVM), multi-layer stacked long short-term memory (MLS-LSTM), convolutional neural network-bidirectional long short-term memory with attention mechanism (CNN-BiLSTM-AM), and combinations of long short-term memory and deep neural networks (LSTM and DNN) in terms of stock price prediction accuracy. This implies that the current model can account for nearly all of the variability in the stock market data. The incremental improvements in R^2 in comparison to existing methods underscore the effectiveness of integrating EMD with the MRFO and LSTM networks.

TABLE IV. PERFORMANCE OF THE PROPOSED MODEL WITH DIFFERENT CROSS-VALIDATIONS

K-folds/METRICS	*TEST SET*			
	R^2	*RMSE*	*MAE*	*MAPE (%)*
Without k-fold	0.9973	91.99	71.54	0.57
5-fold	0.9967	92.82	72.33	0.58
10-fold	0.9970	92.18	71.91	0.57

TABLE V. COMPARISON OF THE COEFFICIENT OF DETERMINATION FOR A DIVERSE ARRAY OF STOCK MARKET PREDICTION METHODOLOGIES IN LITERATURE

References	*Frameworks*	R^2
[35]	Linear regression	0.735
	SVM	0.931
	MLS-LSTM	0.95
[36]	LSTM	0.981
[37]	CNN-BiLSTM-AM	0.98
[38]	LSTM and DNN	0.972
Present study	EMD-MRFO-LSTM	0.9973

The EMD-MRFO-LSTM model is capable of forecasting stock market trends. This model has the potential to assist investors in making more informed decisions regarding the purchase, sale, or retention of stocks by predicting the future values of stock indices, such as the Nasdaq. Portfolio managers and financial advisors may find the EMD-MRFO-LSTM model advantageous for optimizing their investment portfolios. The model can help diversify portfolios to reduce risk and enhance returns by predicting the potential future performance of a variety of assets. It has the potential to offer valuable insights into the most advantageous times to rebalance portfolios by identifying the optimal moments to adjust the allocation of various assets. The EMD-MRFO-LSTM model may be implemented by financial institutions to evaluate and mitigate risks. The model can assist in the identification of potential periods of high volatility or downturns by forecasting market trends. This allows institutions to employ risk mitigation strategies, such as adjusting their exposure to specific assets or hedging, to potentially safeguard their investments from adverse market movements.

VII. CONCLUSIONS

Forecasting stock prices is a complex and multifaceted process that poses numerous challenges. Social, political, and economic changes, as well as other factors, all have an impact on the stock market, which is an ever-evolving and dynamic system. Future stock prices must be correctly predicted by considering a variety of factors. Constraints and variables abound in the procedure of forecasting stock prices, which can make it difficult to develop accurate and dependable prediction models. It is imperative to comprehend the market's non-linear and unpredictable characteristics to achieve this objective. Thankfully, the EMD-MRFO-LSTM model has proven to be dependable and precise, providing a workable solution to these issues. The current study evaluated the efficacy of various stock price forecasting models, such as EMD-LSTM, LSTM, EMD-GA-LSTM, EMD-BRO-LSTM, as well as EMD-GWO-LSTM. The hyperparameter optimization techniques GA, BRO, MRFO, and GWO were utilized to enhance the LSTM's parameters. Nevertheless, optimal outcomes were achieved when the MRFO optimization method was coupled with LSTM. From January 2, 2015, to June 29, 2023, OHLC pricing and the Nasdaq index's volume comprised the dataset employed in the research. According to the analysis, the EMD-MRFO-LSTM model was highly reliable and accurate at forecasting stock prices. The EMD-MRFO-LSTM model consistently displayed superior accuracy and efficacy in its predictions compared to other models tested during the study by having 0.9973, 91.99, 71.54, and 0.57 values for R^2, RMSE, MAE, and MAPE for the testing, respectively. In general, the EMD-MRFO-LSTM model demonstrates efficacy as a stock price forecasting instrument and the investor supplies astute information to enable prudent investment choices.

VIII. CHALLENGES AND PROSPECTS

A. Challenges

Numerous variables impact stock market data, which is inherently chaotic. Deriving significant patterns and trends from the data is a difficult task due to its intricate nature. Conventional feedforward neural networks encounter difficulties when confronted with sequential data in which past events hold significant importance in forecasting future outcomes. Gradients in models may dissolve or become extremely small during training, making it challenging for the network to effectively learn long-term dependencies. To make accurate predictions, it is vital to identify pertinent characteristics of the supplied data. The process of regard to the optimization for complex models necessitates traversing extensive parameter spaces. The evaluation of various models and techniques necessitates the use of rigorous metrics for

assessment and methodologies. When evaluating the precision and efficacy of different methodologies, it is imperative to meticulously contemplate elements such as the origins of the data, the structures of the models, and the optimization strategies.

B. Prospects

Potentially more accurate and dependable stock market forecasts could result from the combination of sophisticated artificial intelligence models LSTM and optimization techniques (gray wolf optimization and meta-heuristic algorithms). These models have the potential to optimize forecasting capabilities by more effectively capturing the intricate relationships and patterns that are intrinsic in stock market data. The finance industry relies heavily on precise stock market forecasts to manage risk effectively. The proposed models have the potential to aid financial institutions and investors in making well-informed decisions related to risk mitigation and return optimization through the provision of more dependable forecasts. Investment firms and financial institutions that implement sophisticated predictive analytics methods are likely to gain a competitive advantage in the marketplace. By investing in modern technologies to improve decision-making procedures, these institutions can strengthen their financial achievements and maintain an excellent market position. The creation and implementation of advanced forecasting models offer educational programs and academic institutions with a vision to offer courses and seminars covering areas such as machine learning in the financial sector, optimization methodology, and quantitative analysis. This can help develop the next generation of professionals who have the expertise to handle complex financial issues.

ACKNOWLEDGMENT

This work was supported by the General Project of the National Social Science Foundation of China: Research on the Impact of the "Chinese Rules" of Digital Trade on the Construction of a Strong Trade Country (Project Approval No. 23BGJ031) (Zhongpo Gao).

REFERENCES

[1] R. G. Ahangar, M. Yahyazadehfar, and H. Pournaghshband, "The comparison of methods artificial neural network with linear regression using specific variables for prediction stock price in Tehran stock exchange," arXiv preprint arXiv:1003.1457, 2010.

[2] A. Ashta and H. Herrmann, "Artificial intelligence and fintech: An overview of opportunities and risks for banking, investments, and microfinance," Strategic Change, vol. 30, no. 3, pp. 211–222, 2021.

[3] C. Milana and A. Ashta, "Artificial intelligence techniques in finance and financial markets: a survey of the literature," Strategic Change, vol. 30, no. 3, pp. 189–209, 2021.

[4] M. Kaur and A. Mohta, "A review of deep learning with recurrent neural network," in 2019 International Conference on Smart Systems and Inventive Technology (ICSSIT), IEEE, 2019, pp. 460–465.

[5] N. F. Hardy and D. V Buonomano, "Encoding Time in Feedforward Trajectories of a Recurrent Neural Network Model," Neural Comput, vol. 30, no. 2, pp. 378–396, Feb. 2018, doi: 10.1162/neco_a_01041.

[6] S. Hochreiter, "The vanishing gradient problem during learning recurrent neural nets and problem solutions," International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, vol. 6, no. 02, pp. 107–116, 1998.

[7] F. Adeeba and S. Hussain, "Native Language Identification in Very Short Utterances Using Bidirectional Long Short-Term Memory Network," IEEE Access, vol. 7, pp. 17098–17110, 2019, doi: 10.1109/ACCESS.2019.2896453.

[8] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," Neural Comput, vol. 9, no. 8, pp. 1735–1780, Nov. 1997, doi: 10.1162/neco.1997.9.8.1735.

[9] A. Lotfipoor, S. Patidar, and D. P. Jenkins, "Deep neural network with empirical mode decomposition and Bayesian optimisation for residential load forecasting," Expert Syst Appl, vol. 237, p. 121355, 2024.

[10] M. Mitchell, An introduction to genetic algorithms. MIT press, 1998.

[11] X.-S. Yang, "Metaheuristic optimization," Scholarpedia, vol. 6, no. 8, p. 11472, 2011.

[12] T. Rahkar Farshi, "Battle royale optimization algorithm," Neural Comput Appl, vol. 33, no. 4, pp. 1139–1157, 2021.

[13] S. Mirjalili, S. M. Mirjalili, and A. Lewis, "Grey Wolf Optimizer," Advances in Engineering Software, vol. 69, pp. 46–61, 2014, doi: https://doi.org/10.1016/j.advengsoft.2013.12.007.

[14] W. Zhao, Z. Zhang, and L. Wang, "Manta ray foraging optimization: An effective bio-inspired optimizer for engineering applications," Eng Appl Artif Intell, vol. 87, p. 103300, 2020.

[15] Z. Chen, L. Zhang, and C. Weng, "Does climate policy uncertainty affect Chinese stock market volatility?," International Review of Economics & Finance, vol. 84, pp. 369–381, 2023.

[16] M. Khraiche, J. W. Boudreau, and M. S. R. Chowdhury, "Geopolitical risk and stock market development," Journal of International Financial Markets, Institutions and Money, vol. 88, p. 101847, 2023.

[17] J. W. Goodell, S. Kumar, P. Rao, and S. Verma, "Emotions and stock market anomalies: a systematic review," J Behav Exp Finance, vol. 37, p. 100722, 2023.

[18] M. M. Salamattalab, M. H. Zonoozi, and M. Molavi-Arabshahi, "Innovative approach for predicting biogas production from large-scale anaerobic digester using long-short term memory (LSTM) coupled with genetic algorithm (GA)," Waste Management, vol. 175, pp. 30–41, 2024.

[19] E. Gholamian, S. M. S. Mahmoudi, and S. Balafkandeh, "Techno-economic appraisal and machine learning-based gray wolf optimization of enhanced fuel cell integrated with stirling engine and vanadium-chlorine cycle," Int J Hydrogen Energy, vol. 51, pp. 1227–1241, 2024.

[20] W. Lu, J. Li, Y. Li, A. Sun, and J. Wang, "A CNN-LSTM-Based Model to Forecast Stock Prices," Complexity, vol. 2020, p. 6622927, 2020, doi: 10.1155/2020/6622927.

[21] H. Rezaei, H. Faaljou, and G. Mansourfar, "Stock price prediction using deep learning and frequency decomposition," Expert Syst Appl, vol. 169, p. 114332, 2021, doi: https://doi.org/10.1016/j.eswa.2020.114332.

[22] J. Qiu and B. Wang, "dan Zhou, C. Forecasting Stock Prices with Long-Short Term Memory Neural Network Based on Attention Mechanism," Advanced Design and Intelligent Computing, vol. 15, no. 1, 2020.

[23] Z. Lanbouri and S. Achchab, "Stock Market prediction on High frequency data using Long-Short Term Memory," Procedia Comput Sci, vol. 175, pp. 603–608, 2020, doi: https://doi.org/10.1016/j.procs.2020.07.087.

[24] A. Yadav, C. K. Jha, and A. Sharan, "Optimizing LSTM for time series prediction in Indian stock market," Procedia Comput Sci, vol. 167, pp. 2091–2100, 2020, doi: https://doi.org/10.1016/j.procs.2020.03.257.

[25] R. K. Dash, T. N. Nguyen, K. Cengiz, and A. Sharma, "Fine-tuned support vector regression model for stock predictions," Neural Comput Appl, vol. 35, no. 32, pp. 23295–23309, 2023.

[26] J. Zhang and X. Chen, "A two-stage model for stock price prediction based on variational mode decomposition and ensemble machine learning method," Soft comput, vol. 28, no. 3, pp. 2385–2408, 2024.

[27] K. V. Rao and B. V. Ramana Reddy, "Hm-smf: An efficient strategy optimization using a hybrid machine learning model for stock market prediction," Int J Image Graph, vol. 24, no. 02, p. 2450013, 2024.

[28] M. Accounting, S. Majid, M. Anzahaei, and H. Nikoomaram, "A Comparative Study of the Performance of Stock Trading Strategies Based on LGBM and CatBoost Algorithms Keywords :," International Journal of Finance and Managerial Accounting, vol. 7, no. 26, pp. 63–76, 2022.

[29] S. Zhao, K. Xu, Z. Wang, C. Liang, W. Lu, and B. Chen, "Financial distress prediction by combining sentiment tone features," Econ Model, vol. 106, no. November 2021, 2022, doi: 10.1016/j.econmod.2021.105709.

[30] G. Kumar, U. P. Singh, and S. Jain, "An adaptive particle swarm optimization-based hybrid long short-term memory model for stock price time series forecasting," Soft comput, vol. 26, no. 22, pp. 12115–12135, 2022, doi: 10.1007/s00500-022-07451-8.

[31] A. Zeiler, R. Faltermeier, I. R. Keck, A. M. Tomé, C. G. Puntonet, and E. W. Lang, "Empirical mode decomposition-an introduction," in The 2010 international joint conference on neural networks (IJCNN), IEEE, 2010, pp. 1–8.

[32] P. Suebsombut, A. Sekhari, P. Sureephong, A. Belhi, and A. Bouras, "Field data forecasting using lstm and bi-lstm approaches," Applied Sciences (Switzerland), vol. 11, no. 24, 2021, doi: 10.3390/app112411820.

[33] B. Gülmez and E. Korhan, "COVID-19 vaccine distribution time optimization with Genetic Algorithm," 2022.

[34] E. Alkafaween, A. B. A. Hassanat, and S. Tarawneh, "Improving Initial Population for Genetic Algorithm using the Multi Linear Regression Based Technique (MLRBT)," Communications - Scientific Letters of the University of Zilina, vol. 23, no. 1, pp. E1–E10, 2021, doi: 10.26552/com.C.2021.1.E1-E10.

[35] A. Q. Md et al., "Novel optimization approach for stock price forecasting using multi-layered sequential LSTM," Appl Soft Comput, vol. 134, p. 109830, 2023, doi: https://doi.org/10.1016/j.asoc.2022.109830.

[36] Z. Jin, Y. Yang, and Y. Liu, "Stock closing price prediction based on sentiment analysis and LSTM," Neural Comput Appl, vol. 32, pp. 9713–9729, 2020.

[37] W. Lu, J. Li, J. Wang, and L. Qin, "A CNN-BiLSTM-AM method for stock price prediction," Neural Comput Appl, vol. 33, no. 10, pp. 4741–4753, 2021, doi: 10.1007/s00521-020-05532-z.

[38] A. C. Nayak and A. Sharma, PRICAI 2019: Trends in Artificial Intelligence: 16th Pacific Rim International Conference on Artificial Intelligence, Cuvu, Yanuca Island, Fiji, August 26–30, 2019, Proceedings, Part II, vol. 11671. Springer Nature, 2019.

Modified TOPSIS Method for Neutrosophic Cubic Number Multi-Attribute Decision-Making and Applications to Music Composition Effectiveness Evaluation of Film and Television

Liang Yang*, Jun Zhao
Anhui Broadcasting and Television Vocational and Technical College, Hefei, 230011, Anhui, China

***Abstract*—Contemporary music composition for film and television has exhibited a trend towards diversification, which is reflected in various aspects such as the diversity of musical styles, the integration of music and visuals, as well as the technical means of music creation. With the continuous advancement of music production technology and film/television production technology, the creation of music for film and television has increasingly emphasized the organic integration of music and visuals, as well as the role of music in emotional expression and atmosphere creation. Meanwhile, the fusion and innovation of different musical styles have also brought more possibilities and space to the creation of music for film and television. This trend of diversification not only enriches the artistic expressiveness of film and television works, but also provides audiences with a more diverse audiovisual experience. The music composition effectiveness evaluation of film and television is multi-attribute decision-making (MADM) problem. In this paper, the TOPSIS method is extended to the framework of neutrosophic cubic sets (NCSs) to address MADM problems. The CRITIC method is employed to obtain the weights of the attributes, ensuring a systematic and objective approach to determining their relative importance. Furthermore, the neutrosophic cubic number TOPSIS (NCN-TOPSIS) approach is established for MADM scenarios. To demonstrate the applicability of the proposed NCN-TOPSIS model, a numerical example focused on the music composition effectiveness evaluation in film and television is presented. Additionally, comparative analyses are conducted to showcase the advantages of the NCN-TOPSIS approach over other decision-making methods. By extending the TOPSIS technique to the NCSs environment and integrating the CRITIC method, this research contributes to the field of MADM by providing a robust and efficient decision-making tool for complex, multi-criteria problems, as exemplified by the music composition evaluation in the film and television industry.**

Keywords*—*Multi-Attribute Decision-Making (MADM); NCSs; TOPSIS approach; music composition effectiveness evaluation

I. INTRODUCTION

The development of contemporary music production technology has had a profound impact on film and television music creation [1, 2]. With the progress of technology, music production tools and techniques have been continuously updated, providing more possibilities and innovative space for film and television music creation. These technologies not only improve the efficiency and quality of music production, but also make the integration of music and visuals more intimate and natural [3, 4]. The continuous advancement and innovation of technological means is undoubtedly a key driving force in promoting the sustained prosperity of film and television music creation. As music production technology evolves rapidly, film and television music creation is gradually breaking free from traditional constraints and showcasing unprecedented creativity and imagination [5, 6]. Modern music production technology not only provides more convenient creation and editing tools, but also enables composers to easily explore the integration of various timbres and styles, creating unique and highly engaging musical works. The widespread application of these technological means not only enriches the connotation and extension of film and television music creation, but also, while driving the innovative development of film and television music, brings more diversified and personalized audiovisual experiences to the audience [7, 8]. Therefore, we must fully recognize the important role of technological means in film and television music creation, and continue to promote their progress and innovation, contributing to the flourishing development of film and television music art [9]. In contemporary film and television music creation, the phenomenon of musical style integration not only showcases the infinite possibilities of art, but also becomes a vivid manifestation of cultural exchange and collision [10, 11]. Against the backdrop of globalization, the mutual penetration and integration of different cultures and music styles have provided a vast space for innovation in film and television music composition [12-14]. With an open mindset and unique perspectives, composers skillfully blend elements of classical elegance, popular vernacular, rock passion, and avant-garde electronics, breaking the boundaries of traditional music styles and creating film and television music works that have both depth and breadth [15, 16]. This cross-border integration of musical styles has greatly enriched the expressiveness and emotional depth of film and television music, allowing audiences to experience the unique charm of music and audiovisual art, and enjoy an unprecedented visual and auditory feast [17-19]. This integration phenomenon not only drives the progress of film and television music creation, but also further promotes the exchange and understanding between different cultures, demonstrating the power and allure of art [20-22].

MADM is a branch of decision theory that focuses on making decisions when there are multiple, often conflicting,

*Corresponding Author.

attributes or objectives that need to be taken into account [23-27]. These attributes can be both quantitative (e.g., cost, performance) and qualitative (e.g., aesthetic appeal, brand reputation) in nature [28-32]. The goal of MADM is to help DMs systematically evaluate the available alternatives and make an informed choice that best aligns with their preferences and priorities. MADM has a wide range of applications across various domains, including product design, resource allocation, policy making, and personal decision-making. For example, in the context of urban planning, MADM can be used to evaluate and prioritize infrastructure projects based on factors such as cost, environmental impact, accessibility, and economic benefits [33-37].

The evaluation of music composition effectiveness in film and television is often considered a MADM problem. However, the existing literature lacks research on the application of the TOPSIS method [38, 39] for MADM in this domain under NCNs [40]. Given the importance of MADM in various real-world applications, including the music composition effectiveness evaluation of film and television, it is essential to further explore the use of the TOPSIS method under NCNs environments. The primary objective of this study is to develop a novel technique that could more efficiently solve certain MADM problems using the NCN-TOPSIS approach. The key highlights of this work are illustrated: (1) The CRITIC technique is illustrated for the weights of the attributes; (2) The NCN-TOPSIS technique is defined and formalized for decision-making under NCNs; (3) numerical example focusing on the music composition effectiveness evaluation of film and television is illustrated the application of the NCN-TOPSIS approach. Comparative analyses are conducted to showcase the decision-making advantages of the proposed NCN-TOPSIS approach. By addressing the gap in the existing literature and proposing a novel NCN-TOPSIS method, this research aims to contribute to the field of MADM, particularly in the context of film and television music composition evaluation and other relevant domains. The demonstrated application of the NCN-TOPSIS method in the labor education context further highlights its potential for solving complex decision-making issues involving multiple, often conflicting, attributes.

The remainder of this paper is structured as follows: Section II introduces the NCSs, providing the necessary background for the proposed approach. Section III illustrates several fused operators of NCSs, which are essential for the development of the decision-making method. In Section IV, the NCN-TOPSIS approach is devised for MADM. It provides an empirical example focusing on the music composition effectiveness evaluation of film and television. This section demonstrates the application of the NCN-TOPSIS method and includes comparative analyses to highlight the decision-making advantages of the proposed approach. Finally, Section V draws a satisfactory conclusion to the study.

II. PRELIMINARIES

Ail et al. [40] illustrated the NCSs in light with SVNSs [41] and IVNSs [42].

Definition 1 [40]. The NCSs is illustrated:

$$UA=\left\{\left(x,\left(UR_A\left(x\right),US_A\left(x\right)\right)\right)\middle|x\in X\right\} \tag{1}$$

where $UR_A\left(x\right)$ is IVNSs and $US_A\left(x\right)$ is SVNSs.

$$UR_A\left(x\right)=\begin{pmatrix}\left[UT_A^L\left(x\right),UT_A^R\left(x\right)\right],\\ \left[UI_A^L\left(x\right),UI_A^R\left(x\right)\right],\\ \left[UF_A^L\left(x\right),UF_A^R\left(x\right)\right]\end{pmatrix} \tag{2}$$

$$US_A\left(x\right)=\left(UT_A\left(x\right),UI_A\left(x\right),UF_A\left(x\right)\right) \tag{3}$$

$$\begin{aligned}&\left[UT_A^L\left(x\right),UT_A^R\left(x\right)\right],\left[UI_A^L\left(x\right),UI_A^R\left(x\right)\right],\\&\left[UF_A^L\left(x\right),UF_A^R\left(x\right)\right]\subseteq\left[0,1\right]\end{aligned} \tag{4}$$

$$UT_A\left(x\right),UI_A\left(x\right),UF_A\left(x\right)\in\left[0,1\right] \tag{5}$$

$$\begin{aligned}&0\le UT_A^R\left(x\right)+UI_A^R\left(x\right)+UF_A^R\left(x\right)\le 3,\\&0\le UT_A\left(x\right)+UI_A\left(x\right)+UF_A\left(x\right)\le 3\end{aligned} \tag{6}$$

The neutrosophic cubic number (NCN) is illustrated as:

$$UA=\begin{Bmatrix}\left(\left[UT_A^L,UT_A^R\right],\left[UI_A^L,UI_A^R\right],\left[UF_A^L,UF_A^R\right]\right),\\ \left(UT_A,UI_A,UF_A\right)\end{Bmatrix}.$$

Definition 2[40]**.** Let $UA=\begin{Bmatrix}\left(\left[UT_A^L,UT_A^R\right],\left[UI_A^L,UI_A^R\right],\left[UF_A^L,UF_A^R\right]\right),\\ \left(UT_A,UI_A,UF_A\right)\end{Bmatrix}$ and $UB=\begin{Bmatrix}\left(\left[UT_B^L,UT_B^R\right],\left[UI_B^L,UI_B^R\right],\left[UF_B^L,UF_B^R\right]\right),\\ \left(UT_B,UI_B,UF_B\right)\end{Bmatrix}$, the operations laws are illustrated:

$$(1)\ UA\oplus UB=\begin{Bmatrix}\left(\begin{pmatrix}UT_A^L+UT_B^L-UT_A^LUT_B^L,UT_A^R+UT_B^R-UT_A^RUT_B^R\end{pmatrix},\right.\\ \left.\begin{pmatrix}\left[UI_A^LUI_B^L,UI_A^RUI_B^R\right],\left[UF_A^LUF_B^L,UF_A^RUF_B^R\right]\end{pmatrix}\right),\\ \left(UT_A+UT_B-UT_AUT_B,UI_AUI_B,UF_AUF_B\right)\end{Bmatrix};$$

$$(2)\ A\otimes B=\left\{\begin{pmatrix}\left[UT_A^LUT_B^L,UT_A^RUT_B^R\right],\\ \left[UI_A^L+UI_B^L-UI_A^LUI_B^L,UI_A^R+UI_B^R-UI_A^RUI_B^R\right],\\ \left[UF_A^L+UF_B^L-UF_A^LUF_B^L,UF_A^R+UF_B^R-UF_A^RUF_B^R\right]\end{pmatrix},\right.$$
$$\left.\left(UT_AUT_B,UI_A+UI_B-UI_AUI_B,UF_A+UF_B-UF_AUF_B\right)\right\};$$

$$(3)\ \lambda UA=\left\{\begin{pmatrix}\left[1-\left(1-UT_A^L\right)^\lambda,1-\left(1-UT_A^R\right)^\lambda\right],\\ \left[\left(UI_A^L\right)^\lambda,\left(UI_A^R\right)^\lambda\right],\left[\left(UF_A^L\right)^\lambda,\left(UF_A^L\right)^\lambda\right]\end{pmatrix},\begin{pmatrix}1-\left(1-UT_A\right)^\lambda,\\ \left(UI_A\right)^\lambda,\left(UF_A\right)^\lambda\end{pmatrix}\right\},\lambda>0;$$

$$(4)\ \left(UA\right)^\lambda=\left\{\begin{pmatrix}\left[\left(UT_A^L\right)^\lambda,\left(UT_A^R\right)^\lambda\right],\\ \left[1-\left(1-UI_A^L\right)^\lambda,1-\left(1-UI_A^R\right)^\lambda\right],\\ \left[1-\left(1-UF_A^L\right)^\lambda,1-\left(1-UF_A^R\right)^\lambda\right]\end{pmatrix},\left(\left(UT_A\right)^\lambda,1-\left(1-UI_A\right)^\lambda,1-\left(1-UF_A\right)^\lambda\right)\right\},\lambda>0.$$

Definition 3[43]. Let $UA=\left\{\begin{pmatrix}\left(\left[UT_A^L,UT_A^R\right],\left[UI_A^L,UI_A^R\right],\left[UF_A^L,UF_A^R\right]\right),\\ \left(UT_A,UI_A,UF_A\right)\end{pmatrix}\right\}$, the NCN score values (NCNSV), NCN accuracy values (NCNAV), NCN certainty values (NCNCV) are illustrated:

$$NCNSV\left(UA\right)=\frac{\begin{pmatrix}6+UT_A^L+UT_A^R+UT_A-UI_A^L-\\ UI_A^R-UF_A^L-UF_A^R-UI_A-UF_A\end{pmatrix}}{9} \quad (7)$$

$$NCNAV\left(UA\right)=\frac{\begin{pmatrix}UT_A^L+UT_A^R+UT_A-\\ UI_A^L-UI_A^R-UF_A\end{pmatrix}}{3} \quad (8)$$

$$NCNCV\left(UA\right)=\frac{\left(UT_A^L+UT_A^R+UT_A\right)}{3} \quad (9)$$

Definition 4[43]. Let

$$UA=\left\{\begin{pmatrix}\left(\left[UT_A^L,UT_A^R\right],\left[UI_A^L,UI_A^R\right],\left[UF_A^L,UF_A^R\right]\right),\\ \left(UT_A,UI_A,UF_A\right)\end{pmatrix}\right\}$$

and

$$UB=\left\{\begin{pmatrix}\left(\left[UT_B^L,UT_B^R\right],\left[UI_B^L,UI_B^R\right],\left[UF_B^L,UF_B^R\right]\right),\\ \left(UT_B,UI_B,UF_B\right)\end{pmatrix}\right\}$$

, The comparison analysis for NCNs is illustrated:

1) if $NCNSV\left(UA\right)<NCNSV\left(UB\right)$, then, $UA<UB$;

2) if $NCNSV\left(UA\right)=NCNSV\left(UB\right)$, and $NCNAV\left(UA\right)<NCNAV\left(UB\right)$, then $UA<UB$;

3) if $NCNSV\left(UA\right)=NCNSV\left(UB\right)$ and $NCNAV\left(UA\right)=NCNAV\left(UB\right)$, and $NCNCV\left(UA\right)<NCNCV\left(UB\right)$, then, $UA<UB$;

4) if $NCNSV\left(UA\right)=NCNSV\left(UB\right)$ and $NCNAV\left(UA\right)=NCNAV\left(UB\right)$, and $NCNCV\left(UA\right)=NCNCV\left(UB\right)$, then, $UA=UB$.

Definition 5[44]. Let

$$UA=\left\{\begin{pmatrix}\left(\left[UT_A^L,UT_A^R\right],\left[UI_A^L,UI_A^R\right],\left[UF_A^L,UF_A^R\right]\right),\\ \left(UT_A,UI_A,UF_A\right)\end{pmatrix}\right\}$$

and

$$UB=\left\{\begin{pmatrix}\left(\left[UT_B^L,UT_B^R\right],\left[UI_B^L,UI_B^R\right],\left[UF_B^L,UF_B^R\right]\right),\\ \left(UT_B,UI_B,UF_B\right)\end{pmatrix}\right\}$$

, then the NCN Euclid distance (NCNED) is illustrated:

$$NCNED\left(A,B\right)=\sqrt{\frac{1}{9}\begin{pmatrix}\left|UT_A^L-UT_B^L\right|^2+\left|UT_A^R-UT_B^R\right|^2+\left|UI_A^L-UI_B^L\right|^2+\\ \left|UI_A^R-UI_B^R\right|^2+\left|UF_A^L-UF_B^L\right|^2+\left|UF_A^R-UF_B^R\right|^2+\\ \left|UT_A-UT_B\right|^2+\left|UI_A-UI_B\right|^2+\left|UF_A-UF_B\right|^2\end{pmatrix}} \quad (10)$$

III. NCN-TOPSIS APPROACH FOR MAGDM WITH NCSS

The NCN-TOPSIS approach is illustrated for MADM. Let $UY=\{UY_1,UY_2,\ldots,UY_m\}$ be alternatives. Let $UZ=\{UZ_1,UZ_2,\ldots,UZ_n\}$ be attributes, $uw=\{uw_1,uw_2,\ldots,uw_n\}$ be weight for UZ_j, $uw_j\in[0,1],\sum_{j=1}^{n}uw_j=1$. And

$$UQ=\left(UQ_{ij}\right)_{m\times n}=\left\{\begin{pmatrix}\left[UT_{ij}^{L},UT_{ij}^{R}\right],\left[UI_{ij}^{L},UI_{ij}^{R}\right],\\ \left[UF_{ij}^{L},UF_{ij}^{R}\right]\end{pmatrix},\begin{pmatrix}UT_{ij},UI_{ij},\\ UF_{ij}\end{pmatrix}\right\}_{m\times n}$$

$$NUQ_{ij}=\left\{\left(\left[UT_{ij}^{NL},UT_{ij}^{NR}\right],\left[UI_{ij}^{NL},UI_{ij}^{NR}\right],\left[UF_{ij}^{NL},UF_{ij}^{NR}\right]\right),\left(UT_{ij}^{N},UI_{ij}^{N},UF_{ij}^{N}\right)\right\}$$

$$=\begin{cases}\left\{\left(\left[UT_{ij}^{L},UT_{ij}^{R}\right],\left[UI_{ij}^{L},UI_{ij}^{R}\right],\left[UF_{ij}^{L},UF_{ij}^{R}\right]\right),\left(UT_{ij},UI_{ij},UF_{ij}\right)\right\}, & UZ_j \text{ is benefit criterion}\\ \left\{\left(\left[UF_{ij}^{L},UF_{ij}^{R}\right],\left[1-UI_{ij}^{R},1-UI_{ij}^{L}\right],\left[UT_{ij}^{L},UT_{ij}^{R}\right]\right),\left(UF_{ij},1-UI_{ij},UT_{ij}\right)\right\}, & UZ_j \text{ is cost criterion}\end{cases} \quad (11)$$

$$NCNCCN_{jt}=\frac{\left(\sum_{i=1}^{m}\left(NCNSV\left(NUQ_{ij}\right)-NCNSV\left(NUQ_j\right)\right)\times\left(NCNSV\left(NUQ_{it}\right)-NCNSV\left(NUQ_t\right)\right)\right)}{\left(\sqrt{\sum_{i=1}^{m}\left(NCNSV\left(NUQ_{ij}\right)-NCNSV\left(NUQ_j\right)\right)^2}\times\sqrt{\sum_{i=1}^{m}\left(NCNSV\left(NUQ_{it}\right)-NCNSV\left(NUQ_t\right)\right)^2}\right)},\quad j,t=1,2,\ldots,n, \quad (12)$$

where

$$NCNSV\left(NUQ_j\right)=\frac{1}{m}\sum_{i=1}^{m}NCNSV\left(NUQ_{ij}\right),\qquad NCNSV\left(NUQ_t\right)=\frac{1}{m}\sum_{i=1}^{m}NCNSV\left(NUQ_{it}\right)$$

2) Illustrate the NCN standard deviation numbers (NCNSDN).

$$NCNSDN_j=\sqrt{\frac{1}{m-1}\sum_{i=1}^{m}\left(NCNSV\left(NUQ_{ij}\right)-NCNSV\left(NUQ_j\right)\right)^2} \quad (13)$$

where $NCNSV\left(NUQ_j\right)=\frac{1}{m}\sum_{i=1}^{m}NCNSV\left(NUQ_{ij}\right)$.

3) Illustrate the weight information.

$$uw_j=\frac{NCNSDN_j\sum_{t=1}^{n}\left(1-NCNCCN_{jt}\right)}{\sum_{j=1}^{n}\left(NCNSDN_j\sum_{t=1}^{n}\left(1-NCNCCN_{jt}\right)\right)} \quad (14)$$

where $uw_j\in[0,1]$, $\sum_{j=1}^{n}uw_j=1$.

Step 3. Illustrate the NCN positive ideal solution (NCNPIS) and NCN negative ideal solution (NCNNIS):

$$NCNPIS=\left\{NCNPIS_j\right\},\ j=1,2,\ldots,n. \quad (15)$$

$$NCNNIS=\left\{NCNNIS_j\right\},\ j=1,2,\ldots,n. \quad (16)$$

is the NCN-matrix. Subsequently, NCN-TOPSIS method is illustrated for MADM.

Step 1. Normalize the $UQ=\left(UQ_{ij}\right)_{m\times n}$ to $NUQ=\left[NUQ_{ij}\right]_{m\times n}$.

Step 2. Illustrate the weight through CRITIC technique.

The CRITIC [45] is illustrated for the weights.

1) The NCN correlation coefficient numbers (NCNCCN) are illustrated.

$$NCNPIS_j = \left\{ \begin{pmatrix} \left[UT_j^{NL+}, UT_{ij}^{NR+}\right], \left[UI_j^{NL+}, UI_j^{NR+}\right], \\ \left[UF_j^{NL+}, UF_j^{NR+}\right] \end{pmatrix}, \left(UT_j^{N+}, UI_j^{N+}, UF_j^{N+}\right) \right\} \tag{17}$$

$$NCNNIS_j = \left\{ \begin{pmatrix} \left[UT_j^{NL-}, UT_{ij}^{NR-}\right], \left[UI_j^{NL-}, UI_j^{NR-}\right], \\ \left[UF_j^{NL-}, UF_j^{NR-}\right] \end{pmatrix}, \left(UT_j^{N-}, UI_j^{N-}, UF_j^{N-}\right) \right\} \tag{18}$$

$$NCNPIS_j = \max_j NCNSV \left\{ \begin{pmatrix} \left[UT_{ij}^{NL}, UT_{ij}^{NR}\right], \left[UI_{ij}^{NL}, UI_{ij}^{NR}\right], \\ \left[UF_{ij}^{NL}, UF_{ij}^{NR}\right] \end{pmatrix}, \left(UT_{ij}^{N}, UI_{ij}^{N}, UF_{ij}^{N}\right) \right\} \tag{19}$$

$$NCNNIS_j = \min_j NCNSV \left\{ \begin{pmatrix} \left[UT_{ij}^{NL}, UT_{ij}^{NR}\right], \left[UI_{ij}^{NL}, UI_{ij}^{NR}\right], \\ \left[UF_{ij}^{NL}, UF_{ij}^{NR}\right] \end{pmatrix}, \left(UT_{ij}^{N}, UI_{ij}^{N}, UF_{ij}^{N}\right) \right\} \tag{20}$$

Step 4. Illustrate the distances from NCNPIS and NCNNIS:

Step 5. Illustrate the NCN correlation coefficient (NCNCC) from NCNNIS:

$$NCNCC\left(UY_i\right) = \frac{NCNNISED\left(UY_i\right)}{NCNPISED\left(UY_i\right) + NCNNISED\left(UY_i\right)} \tag{23}$$

Step 6. In light with $NCNCC\left(UY_i\right)$. The highest $NCNCC\left(UY_i\right)$ is the optimal alternative.

$$NCNPISED\left(UY_i\right) = \sum_{j=1}^{n} uw_j NCNED\left(NUQ_{ij}, NCNPIS_j\right)$$

$$= \sum_{j=1}^{n} uw_j \sqrt{\frac{1}{9}\begin{pmatrix} \left|UT_{ij}^{NL} - UT_j^{NL+}\right|^2 + \left|UT_{ij}^{RL} - UT_{ij}^{NR+}\right|^2 + \left|UI_{ij}^{NL} - UI_j^{NL+}\right|^2 + \\ \left|UI_{ij}^{NR} - UI_j^{NR+}\right|^2 + \left|UF_{ij}^{NL} - UF_j^{NL+}\right|^2 + \left|UF_{ij}^{NR} - UF_j^{NR+}\right|^2 + \\ \left|UT_{ij} - UT_j^{+}\right|^2 + \left|UI_{ij} - UI_j^{+}\right|^2 + \left|UF_{ij} - UF_j^{+}\right|^2 \end{pmatrix}} \tag{21}$$

$$NCNNISED\left(UY_i\right) = \sum_{j=1}^{n} uw_j NCNED\left(NUQ_{ij}, NCNNIS_j\right)$$

$$= \sum_{j=1}^{n} uw_j \sqrt{\frac{1}{9}\begin{pmatrix} \left|UT_{ij}^{NL} - UT_j^{NL-}\right|^2 + \left|UT_{ij}^{RL} - UT_{ij}^{NR-}\right|^2 + \left|UI_{ij}^{NL} - UI_j^{NL-}\right|^2 + \\ \left|UI_{ij}^{NR} - UI_j^{NR-}\right|^2 + \left|UF_{ij}^{NL} - UF_j^{NL-}\right|^2 + \left|UF_{ij}^{NR} - UF_j^{NR-}\right|^2 + \\ \left|UT_{ij} - UT_j^{-}\right|^2 + \left|UI_{ij} - UI_j^{-}\right|^2 + \left|UF_{ij} - UF_j^{-}\right|^2 \end{pmatrix}} \tag{22}$$

IV. Numerical Example and Comparative Analysis

A. Numerical Example

Music for film and television, as an essential component of audiovisual works, not only adds emotional tones to the visuals, but also plays a crucial role in narrative, atmosphere creation, and audience emotional experiences. Through elements such as melody, rhythm, and timbre, it can interact with the visuals to jointly construct the unique atmosphere and emotional tonality of a work. In recent years, with the continuous advancement of technology and creative concepts, contemporary music composition for film and television has exhibited a trend towards diversification. The diversity of musical styles, the integration of music and visuals, as well as the innovation of technical means have all injected new vitality into music for film and television. In audiovisual works, the close integration of music and narrative often deepens the audience's emotional resonance. For example, in tense and exciting storylines, stirring music can enhance the creation of a tense atmosphere,

making the audience feel more immersed; while in sad or sentimental scenes, gentle music can touch the audience's hearts and strengthen emotional resonance. This mutual reflection between music and narrative not only enriches the artistic expressiveness of the film, but also allows the audience to gain a deeper emotional experience during the viewing process. Therefore, the diversification trend of music composition for film and television is of great significance for improving the artistic quality and audience satisfaction of audiovisual works. The compatibility between musical styles and the themes of audiovisual works is not only about the conveyance of emotions, but also the resonance of cultures. For example, in some films with historical backgrounds or regional characteristics, the use of representative ethnic music or folk songs can not only create a unique cultural atmosphere, but also allow the audience to feel the charm and depth of the culture through the melodies. This combination of music and culture enables audiovisual works to convey not only emotions, but also cultural values and historical memories, thereby enriching the connotation and depth of the film. Therefore, the choice of musical styles plays a crucial role in audiovisual creation and is one of the key factors determining the success of a film. The music composition effectiveness evaluation of film and television is viewed as the MAGDM. Five potential film and television music $UY_i\left(i=1,2,3,4,5\right)$ are assessed with four attributes: ① UZ_1 is establishment of correct labor values; ② UZ_2 is cultivation of noble labor morality; ③ UZ_3 is Acquisition of comprehensive labor knowledge and skills; ④ UZ_4 is Cultivation of good working habits. Evidently, all attributes are beneficial attribute. NCN-TOPSIS procedure is illustrated for music composition effectiveness evaluation of film and television.

Step 1. Illustrate the NCN-matrix $QQ=\left(qq_{ij}\right)_{5\times4}$ (Table I).

Step 2. The NCN-matrix does not require normalization as all its attributes are advantageous.

Step 3. Illustrate the weights with CRITIC (Table II).

Step 4. Illustrate the NCNPIS and NCNNIS (see Table III).

Step 5. Illustrate the $NCNPISED\left(UY_i\right)$ and $NCNNISED\left(UY_i\right)$ (see Table IV):

Step 6. Illustrate the $NCNCC\left(UY_i\right)$ (see Table V).

Step 7. In light with $NCNCC\left(UY_i\right)$, the order is: $UY_5>UY_2>UY_4>UY_3>UY_1$ and UY_5 is the best one.

TABLE I. NCN-MATRIX

	ZZ_1	ZZ_2
YY_1	{([0.64, 0.62], [0.44, 0.42], [0.46, 0.56]), (0.44, 0.33, 0.36)}	{([0.54, 0.62], [0.44, 0.56], [0.44, 0.56]), (0.11, 0.19, 0.16)}
YY_2	{([0.82, 0.82], [0.24, 0.42], [0.65, 0.84]), (0.44, 0.33, 0.36)}	{([0.66, 0.68], [0.46, 0.42], [0.45, 0.64]), (0.31, 0.03, 0.42)}
YY_3	{([0.82, 0.82], [0.28, 0.42], [0.44, 0.56]), (0.26, 0.44, 0.34)}	{([0.82, 2.0], [0.24, 0.24], [0.54, 0.65]) , (0.16, 0.15, 0.11)}
YY_4	{([0.65, 0.62], [0.26, 0.28], [0.54, 0.65]), (0.44, 0.34, 0.32)}	{([0.62, 0.65], [0.45, 0.46], [0.44, 0.56]), (0.22, 0.35, 0.29)}
YY_5	{([0.62, 0.82], [0.42, 0.45], [0.46, 0.56]) , (0.44, 0.33, 0.36)}	{([0.86, 0.88], [0.56, 0.64], [0.56, 0.66]) , (0.44, 0.33, 0.36)}
	ZZ_3	ZZ_4
YY_1	{([0.62, 0.84], [0.45, 0.56], [0.44, 0.52]) , (0.14, 0.42, 0.43)}	{([0.64, 0.82], [0.46, 0.42], [0.64, 0.62]) , (0.31, 0.19, 0.16)}
YY_2	{([0.64, 0.62], [0.56, 0.64], [0.46, 0.54]) , (0.36, 0.04, 0.45)}	{([0.62, 0.84], [0.65, 0.64], [0.54, 0.66]) , (0.16, 0.15, 0.11)}

YY_3	{([0.62, 0.84], [0.46, 0.42], [0.55, 0.62]) , (0.31, 0.47, 0.26)}	{([0.82, 0.85], [0.26, 0.44], [0.62, 0.66]) , (0.43, 0.04, 0.46)}
YY_4	{([0.65, 0.82], [0.44, 0.58], [0.62, 0.64]) , (0.39, 0.13, 0.32)}	{([0.35, 0.38], [0.26, 0.25], [0.68, 0.88]) , (0.41, 0.05, 0.19)}
YY_5	{([0.24, 0.32], [0.66, 0.82], [0.24, 0.44]) , (0.24, 0.28, 0.36)}	{([0.64, 0.82], [0.45, 0.52], [0.45, 0.56]) , (0.49, 0.35, 0.41)}

TABLE II. THE WEIGHT INFORMATION

	UZ1	UZ2	UZ3	UZ4
weight	0.1684	0.3483	0.2927	0.1906

TABLE III. NCNPIS AND NCNNIS

	UZ1	UZ2
NCNPIS	{([0.82, 0.82], [0.28, 0.42], [0.44, 0.56]), (0.26, 0.44, 0.34)}	{([0.82, 2.0], [0.24, 0.24], [0.54, 0.65]) , (0.16, 0.15, 0.11)}
NCNNIS	{([0.62, 0.82], [0.42, 0.45], [0.46, 0.56]) , (0.44, 0.33, 0.36)}	{([0.54, 0.62], [0.44, 0.56], [0.44, 0.56]), (0.11, 0.19, 0.16)}
	UZ3	UZ4
NCNPIS	{([0.65, 0.82], [0.44, 0.58], [0.62, 0.64]) , (0.39, 0.13, 0.32)}	{([0.82, 0.85], [0.26, 0.44], [0.62, 0.66]) , (0.43, 0.04, 0.46)}
NCNNIS	{([0.24, 0.32], [0.66, 0.82], [0.24, 0.44]) , (0.24, 0.28, 0.36)}B.	{([0.35, 0.38], [0.26, 0.25], [0.68, 0.88]) , (0.41, 0.05, 0.19)}

TABLE IV. THE $NCNPISED(UY_i)$ AND $NCNNISED(UY_i)$

	$NCNPISED(UY_i)$	$NCNNISED(UY_i)$
UY1	0.4429	0.4275
UY2	0.5613	0.6937
UY3	0.7763	0.4105
UY4	0.5073	0.3102
UY5	0.5270	0.3919

TABLE V. THE $NCNCC(UY_i)$

	$NCNCC(UY_i)$	Order
UY_1	0.4280	5
UY_2	0.6163	2
UY_3	0.4460	4
UY_4	0.5238	3
UY_5	0.6971	1

B. Comparative Analysis

The NCN-TOPSIS approach is compared with NCNWAA approach and NCNWGA approach [46], NC-VIKOR approach [47], NCN-GRA approach[48]. Eventually, the final results are depicted in Table VI.

Derived from the results presented in Table VI, it is evident that the obtained best choice is UY_5 , while the worst alternative is UY_1 . In other words, the rankings produced by the different decision-making methods are slightly different. This observation suggests that the various approaches could

effectively tackle the MADM from different perspectives. The NCN-TOPSIS approach, in particular, is a commonly used comprehensive evaluation approach that can make full use of the information contained in the original data. The results generated by the NCN-TOPSIS method can accurately reflect the relative performance gap between the evaluated alternatives. This is a key advantage of NCN-TOPSIS approach, as it provides DMs with a more nuanced and informative understanding of the alternatives under consideration. The slightly different rankings produced by the various methods highlight the importance of employing multiple decision-making techniques in complex, multi-criteria decision problems.

TABLE VI. ORDER FOR DIFFERENT APPROACHES

Approaches	Order
NCNWAA approach [46]	$UY_5 > UY_2 > UY_4 > UY_3 > UY_1$
NCNWGA approach [46]	$UY_5 > UY_2 > UY_4 > UY_3 > UY_1$
NC-VIKOR approach [47]	$UY_5 > UY_2 > UY_3 > UY_4 > UY_1$
NCN-GRA approach [48]	$UY_5 > UY_2 > UY_4 > UY_3 > UY_1$
The proposed NCN-TOPSIS approach	$UY_5 > UY_2 > UY_4 > UY_3 > UY_1$

V. CONCLUSION

The trend towards diversification in film and television music creation has had a profound impact on the artistic expressiveness of audiovisual works. With the integration and innovation of musical styles, film and television music is no longer limited to traditional background music or emotional scoring, but has become an important component of audiovisual works. The diversified musical styles provide richer means of emotional expression and atmosphere creation for audiovisual works, making the works more layered and deeper in terms of emotion, rhythm, and ambiance. This organic integration of music and visuals not only enhances the artistic appeal of the audiovisual works, but also improves the audience's viewing experience. The music composition effectiveness evaluation of film and television is MADM. In this work, the NCN-TOPSIS approach is devised for MADM, building upon the TOPSIS technique. The weight numbers of the attributes are determined using the CRITIC technique, ensuring an objective and systematic approach to assigning importance to the decision criteria. Furthermore, a novel NCN-TOPSIS method is developed for the context of MADM, and the detailed computational steps are provided. To demonstrate the efficacy of the proposed approach, an empirical example focused on the music composition effectiveness evaluation in the film and television industry is presented. This example showcases the superiority of the NCN-TOPSIS method in addressing complex, multi-criteria decision problems. The major contributions of this research are summarized: (1) The CRITIC method is utilized to objectively determine the weights of the attributes; (2) The NCN-TOPSIS method is developed within the framework of NCSs, extending the TOPSIS technique to this more generalized decision-making environment; (3) A numerical example on the music composition effectiveness evaluation in the film and television industry is provided, and comparative analyses are conducted to highlight the advantages of NCN-TOPSIS approach. By integrating the CRITIC method, developing the NCN-TOPSIS technique, and demonstrating its application in a real-world decision-making scenario, this research contributes to the advancement of MADM and MAGDM methodologies, offering DMs a robust and effective tool for addressing complex problems.

ACKNOWLEDGMENT

The work was supported by the Anhui Province Quality Course "Music Recording," Project Number: 2022jpkc026 and Anhui Province Quality Engineering Project "Integrated Practical Training Base at the Guangyi All-Media Center," Topic Number: 2021cjrh008.

REFERENCES

[1] M. Unehara and T. Onisawa, "Music composition system with human evaluation as human centered system," (in English), Soft Computing, Article vol. 7, no. 3, pp. 167-178, Jan 2003.

[2] M. Unehara and T. Onisawa, "Music composition by interaction between human and computer," (in English), New Generation Computing, Article vol. 23, no. 2, pp. 181-191, 2005.

[3] B. Eaglestone and P. D. Bamidis, "Music composition for the multi-disabled: A systems perspective," (in English), International Journal on Disability and Human Development, Review vol. 7, no. 1, pp. 19-24, Jan-Mar 2008.

[4] P. Sinha, "Artificial composition: An experiment on indian music," (in English), Journal of New Music Research, Article vol. 37, no. 3, pp. 221-232, 2008, Art. no. Pii 906893252.

[5] N. Collins, "Automatic composition of electroacoustic art music utilizing machine listening," (in English), Computer Music Journal, Article vol. 36, no. 3, pp. 8-23, Fal 2012.

[6] D. Williams et al., "Investigating perceived emotional correlates of rhythmic density in algorithmic music composition," (in English), Acm Transactions on Applied Perception, Article vol. 12, no. 3, p. 21, Jul 2015, Art. no. 8.

[7] M. Navarro, J. M. Corchado, and Y. Demazeau, "Music-mas: Modeling a harmonic composition system with virtual organizations to assist novice composers," (in English), Expert Systems with Applications, Article vol. 57, pp. 345-355, Sep 2016.

[8] C. De Felice, R. De Prisco, D. Malandrino, G. Zaccagnino, R. Zaccagnino, and R. Zizza, "Splicing music composition," (in English), Information Sciences, Article vol. 385, pp. 196-212, Apr 2017.

[9] D. Williams et al., "Affective calibration of musical feature sets in an emotionally intelligent music composition system," (in English), Acm Transactions on Applied Perception, Article vol. 14, no. 3, p. 13, Jul 2017, Art. no. 17.

[10] R. Abboud and J. Tekli, "Integration of nonparametric fuzzy classification with an evolutionary-developmental framework to perform music sentiment-based analysis and composition," (in English), Soft Computing, Article vol. 24, no. 13, pp. 9875-9925, Jul 2020.

[11] I. Abu Doush and A. Sawalha, "Automatic music composition using genetic algorithm and artificial neural networks," (in English), Malaysian Journal of Computer Science, Article vol. 33, no. 1, pp. 35-51, 2020.

[12] S. Shukla and H. Banka, "Markov-based genetic algorithm with e-greedy exploration for indian classical music composition," (in English), Expert Systems with Applications, Article vol. 211, p. 10, Jan 2023, Art. no. 118561.

[13] X. H. Gu, L. Q. Jiang, H. Chen, M. Li, and C. Liu, "Exploring brain dynamics via eeg and steady-state activation map networks in music composition," (in English), Brain Sciences, Article vol. 14, no. 3, p. 29, Mar 2024, Art. no. 216.

[14] X. W. Lei, "Analysing the effectiveness of online digital audio software and offline audio studios in fostering chinese folk music composition

skills in music education," (in English), Journal of Computer Assisted Learning, Article; Early Access p. 12, 2024 Jun 2024.

[15] R. De Prisco, G. Zaccagnino, and R. Zaccagnino, "Evocomposer: An evolutionary algorithm for 4-voice music compositions," (in English), Evolutionary Computation, Article vol. 28, no. 3, pp. 489-530, Sep 2020.

[16] N. N. Shi and Y. F. Wang, "Symmetry in computer-aided music composition system with social network analysis and artificial neural network methods," (in English), Journal of Ambient Intelligence and Humanized Computing, Article; Early Access p. 16, 2020 Aug 2020.

[17] S. L. Chen, Y. Zhong, and R. X. Du, "Automatic composition of guzheng (chinese zither) music using long short-term memory network (lstm) and reinforcement learning (rl)," (in English), Scientific Reports, Article vol. 12, no. 1, p. 17, Sep 2022, Art. no. 15829.

[18] M. X. Liu, "An eeg neurofeedback interactive model for emotional classification of electronic music compositions considering multi-brain synergistic brain-computer interfaces," (in English), Frontiers in Psychology, Article vol. 12, p. 11, Jan 2022, Art. no. 799132.

[19] S. Mukherjee and M. Mulimani, "Composeinstyle: Music composition with and without style transfer," (in English), Expert Systems with Applications, Article vol. 191, p. 21, Apr 2022, Art. no. 116195.

[20] H. Zhao, L. Narikbayeva, and Y. H. Wu, "Interactive systems for music composition: A new paradigm in music education," (in English), Interactive Learning Environments, Article; Early Access p. 10, 2022 Aug 2022.

[21] P. Ferreira, R. Limongi, and L. P. Fávero, "Generating music with data: Application of deep learning models for symbolic music composition," (in English), Applied Sciences-Basel, Article vol. 13, no. 7, p. 19, Apr 2023, Art. no. 4543.

[22] W. M. Liu, "Literature survey of multi-track music generation model based on generative confrontation network in intelligent composition," (in English), Journal of Supercomputing, Article vol. 79, no. 6, pp. 6560-6582, Apr 2023.

[23] T. Mahmood, Z. Ali, H. Garg, L. Zedam, and R. Chinram, "Correlation coefficient and entropy measures based on complex dual type-2 hesitant fuzzy sets and their applications," (in English), Journal of Mathematics, Article vol. 2021, p. 34, Mar 2021, Art. no. 2568391.

[24] K. Ullah, H. Garg, Z. Gul, T. Mahmood, Q. Khan, and Z. Ali, "Interval valued t-spherical fuzzy information aggregation based on dombi t-norm and dombi t-conorm for multi-attribute decision making problems," (in English), Symmetry-Basel, Article vol. 13, no. 6, p. 26, Jun 2021, Art. no. 1053.

[25] R. M. Zulqarnain et al., "Algorithms for a generalized multipolar neutrosophic soft set with information measures to solve medical diagnoses and decision-making problems," (in English), Journal of Mathematics, Article vol. 2021, p. 30, May 2021, Art. no. 6654657.

[26] H. Garg and D. Rani, "An efficient intuitionistic fuzzy multimoora approach based on novel aggregation operators for the assessment of solid waste management techniques," (in English), Applied Intelligence, Article vol. 52, no. 4, pp. 4330-4363, Mar 2022.

[27] H. Garg, J. Vimala, S. Rajareega, D. Preethi, and L. Perez-Dominguez, "Complex intuitionistic fuzzy soft swara - copras approach: An application of erp software selection," (in English), Aims Mathematics, Article vol. 7, no. 4, pp. 5895-5909, 2022.

[28] R. Verma and E. Alvarez-Miranda, "Multiple-attribute group decision-making approach using power aggregation operators with critic-waspas method under 2-dimensional linguistic intuitionistic fuzzy framework," (in English), Applied Soft Computing, Article vol. 157, p. 33, May 2024, Art. no. 111466.

[29] M. Jamil, F. Afzal, A. Maqbool, S. Abdullah, A. Akgül, and A. Bariq, "Multiple attribute group decision making approach for selection of robot under induced bipolar neutrosophic aggregation operators," (in English), Complex & Intelligent Systems, Article vol. 10, no. 2, pp. 2765-2779, Apr 2024.

[30] H. Sun, Z. Yang, Q. Cai, G. W. Wei, and Z. W. Mo, "An extended exp-todim method for multiple attribute decision making based on the z-wasserstein distance," (in English), Expert Systems with Applications, Article vol. 214, p. 14, Mar 2023, Art. no. 119114.

[31] T. Senapati, G. Y. Chen, R. Mesiar, and R. R. Yager, "Intuitionistic fuzzy geometric aggregation operators in the framework of aczel-alsina triangular norms and their application to multiple attribute decision making," (in English), Expert Systems with Applications, Article vol. 212, p. 15, Feb 2023, Art. no. 118832.

[32] F. Riazi, M. H. Dehbozorgi, M. R. Feylizadeh, and M. Riazi, "Enhanced oil recovery prioritization based on feasibility criteria using intuitionistic fuzzy multiple attribute decision making: A case study in an oil reservoir," (in English), Petroleum Science and Technology, Article; Early Access p. 19, 2023 Jun 2023.

[33] M. Palanikumar, K. Arulmozhi, C. Jana, and M. Pal, "Multiple-attribute decision-making spherical vague normal operators and their applications for the selection of farmers," (in English), Expert Systems, Article vol. 40, no. 3, p. 26, Mar 2023.

[34] A. Noori, H. Bonakdari, A. H. Salimi, L. Pourkarimi, and J. M. Samakosh, "A novel multiple attribute decision-making approach for assessing the effectiveness of advertising to a target audience on drinking water consumers? Behavior considering age and education level," (in English), Habitat International, Article vol. 133, p. 9, Mar 2023, Art. no. 102749.

[35] B. Q. Ning, R. Lin, G. W. Wei, and X. D. Chen, "Edas method for multiple attribute group decision making with probabilistic dual hesitant fuzzy information and its application to suppliers selection," (in English), Technological and Economic Development of Economy, Article vol. 29, no. 2, pp. 326-352, 2023.

[36] T. Mahmood, U. U. Rehman, and Z. Ali, "Analysis and applications of aczel-alsina aggregation operators based on bipolar complex fuzzy information in multiple attribute decision making," (in English), Information Sciences, Article vol. 619, pp. 817-833, Jan 2023.

[37] A. Jaglan, G. Sadera, P. Singh, B. P. Singh, and G. Goel, "Probiotic potential of gluten degrading bacillus tequilensis ajg23 isolated from indian traditional cereal-fermented foods as determined by multiple attribute decision-making analysis," (in English), Food Research International, Article vol. 174, p. 10, Dec 2023, Art. no. 113516.

[38] Y.-J. Lai, T.-Y. Liu, and C.-L. Hwang, "Topsis for modm," European journal of operational research, vol. 76, no. 3, pp. 486-500, 1994.

[39] C. T. Chen, "Extensions of the topsis for group decision-making under fuzzy environment," (in English), Fuzzy Sets and Systems, Article vol. 114, no. 1, pp. 1-9, Aug 2000.

[40] M. Ali, I. Deli, and F. Smarandache, "The theory of neutrosophic cubic sets and their applications in pattern recognition," (in English), Journal of Intelligent & Fuzzy Systems, Article vol. 30, no. 4, pp. 1957-1963, 2016.

[41] H. Wang, F. Smarandache, Y. Q. Zhang, and R. Sunderraman, "Single valued neutrosophic sets," Multispace Multistruct, no. 4, pp. 410-413, 2010.

[42] H. Wang, F. Smarandache, Y. Q. Zhang, and R. Sunderraman, "Interval neutrosophic sets and logic: Theory and applications in computing," Hexis: Phoenix, AZ, USA, 2005.

[43] D. Ajay, J. Aldring, and S. Nivetha, "Neutrosophic cubic fuzzy dombi hamy mean operators with application to multi-criteria decision making," Neutrosophic Sets and Systems, vol. 38, pp. 293-316, 2020.

[44] S. Pramanik, P. P. Dey, B. C. Giri, and F. Smarandache, "An extended topsis for multi-attribute decision making problems with neutrosophic cubic information," Neutrosophic Sets and Systems, vol. 17, pp. 20-28, 2017.

[45] D. Diakoulaki, G. Mavrotas, and L. Papayannakis, "Determining objective weights in multiple criteria problems: The critic method," Computers & Operations Research, vol. 22, no. 7, pp. 763-770, 1995.

[46] J. Ye, "Operations and aggregation method of neutrosophic cubic numbers for multiple attribute decision-making," (in English), Soft Computing, Article; Proceedings Paper vol. 22, no. 22, pp. 7435-7444, Nov 2018.

[47] S. Pramanik, S. Dalapati, S. Alam, and T. K. Roy, "Nc-vikor based magdm strategy under neutrosophic cubic set environment," Neutrosophic Sets and Systems, vol. 20, pp. 95-108, 2018.

[48] P. Liu, "Enhanced grey relational analysis method for neutrosophic cubic number madm and applications to decorative wall hanging design effect evaluation," Journal of Intelligent & Fuzzy Systems, vol. 45, no. 5, pp. 7507-7518, 2023.

Heuristic Intelligent Algorithm-Based Approach for In-Depth Development and Application Analysis of Micro- and Nanoembedded Systems

Buzhong Liu
School of Electronic Engineering, Jiangsu Vocational College of Electronics and Information, Huai'an 223003, Jiangsu, China

***Abstract*—Developing application analysis and testing methods is an important part of the in-depth development of micro-nano embedded systems under complex integrated architectures. Therefore, the research on application analysis and testing models is of great significance for the in-depth development and efficient design of embedded systems. In order to fully explore the effective information of test analysis data in the in-depth development of micro-nano embedded systems under complex integrated architectures and improve the analysis and prediction accuracy of test analysis models, a development test analysis model based on the photon search algorithm and LightGBM is proposed. First, the development process of micro-nano embedded systems under complex integrated architectures is analysed, and a system analysis architecture is designed to propose test analysis factors. Second, a development test analysis model is established by combining the photon search algorithm and LightGBM. Subsequently, the feasibility and efficiency of the model are proposed by analysing the data of the development process. The analysis of data examples shows that the LightGBM test analysis model has high analysis and prediction accuracy and generalisation performance.**

***Keywords*—Photon search algorithm; deep development of micro- and nanoembedded systems; application test and analysis methodology; LightGBM**

I. INTRODUCTION

With the continuous development of microelectronics technology, the application of micro-nano embedded systems in various industries is becoming more and more widespread, especially in the fields of Internet of Things, intelligent sensing, healthcare, and communication technology [1]. With the increased complexity and computational volume of micro-nano embedded system applications, the development of micro-nano embedded systems is becoming more and more difficult, which leads to an increased probability of security risks in the iterative updating of development and application work [2]. At the same time, the micro-nano embedded system application function diversification makes the micro-nano embedded system development workers in-depth design and development, which leads to the micro-nano embedded system test and evaluation index feature dimension increase, the traditional test algorithm cannot build complex model relationship [3]. In order to alleviate the problems of insufficient testing means and insufficient accuracy, the system design-oriented development application test and analysis model intelligence has become a development trend [4]. Therefore, the study of constructing an efficient development and application test analysis model has received more and more attention from experts and scholars, which is of great significance for the development of micro and nano embedded systems to solve the problem in reality.

The deep development and application analysis intelligence of micro and nano embedded systems under complex integrated architecture improves the design efficiency in two main aspects, i.e., deep intelligent development of system and intelligent testing of application analysis [5]. System deep intelligent development requires high fidelity mathematical models and also high precision model generation methods [6]. Micro-nano embedded system development application analysis intelligent testing is an important part of the development and design process, generally using test analysis algorithms to simulate the system development and testing process [7]. Currently, test analysis algorithms include hierarchical analysis method [8], expert system method [9], neural network method [10], integrated learning method [11], deep learning method [12] and so on. Since the process of in-depth development of micro-nano embedded systems under complex integration-oriented architecture is a complex process with high dimensionality of relevant data parameters, current test analysis algorithms are unable to construct accurate test results, and the real-time nature of test analysis is unable to meet the demand [6-13]. In addition, test analysis algorithms in the complex integration architecture of micro-nano embedded systems in the depth of the development process is relatively small, in order to improve the design efficiency, need to design targeted test analysis algorithms [6-14]. In this paper, the initial choice of integrated learning algorithm is taken to construct the test analysis algorithm, but the integrated learning algorithm is easy to fall into the local optimum in the learning training in the problem of high dimensionality and complex process, therefore, the debugging optimization of the designed test analysis algorithm is needed [15].

In view of the above problems, this paper proposes a development test analysis algorithm based on intelligent optimization algorithm and integrated learning technology around the deep development process of micro-nano embedded system under complex integration architecture. For the development and test analysis problem, the mechanism of the deep development process of micro-nano embedded system under the complex integration architecture is analyzed, and the relevant test and analysis parameters are extracted; for the development and test analysis model construction problem, the hyperparameters of lightweight gradient lifter are optimized by combining with photon search algorithm, and the development

and test analysis model based on photon search algorithm to improve lightweight gradient lifter is put forward; and the development and test analysis method put forward has high analysis and prediction accuracy and generalization performance through the application of examples. The proposed development test analysis method has high analysis and prediction accuracy and generalization performance, which provides a new method for the in-depth development process testing of micro-nano embedded systems under the complex integration architecture.

II. In-Depth Development of Micro- and Nano-Embedded Systems for Complex Integrated Architectures

A. System Architecture

1) Micro-nano embedded system: Micro-nano embedded systems are embedded systems that integrate microelectromechanical systems and nanotechnology. These systems usually contain micromechanical devices, sensors, actuators, and electronic circuits, which can vary in size from the micron to the nanometer level [16], as shown in Fig. 1.

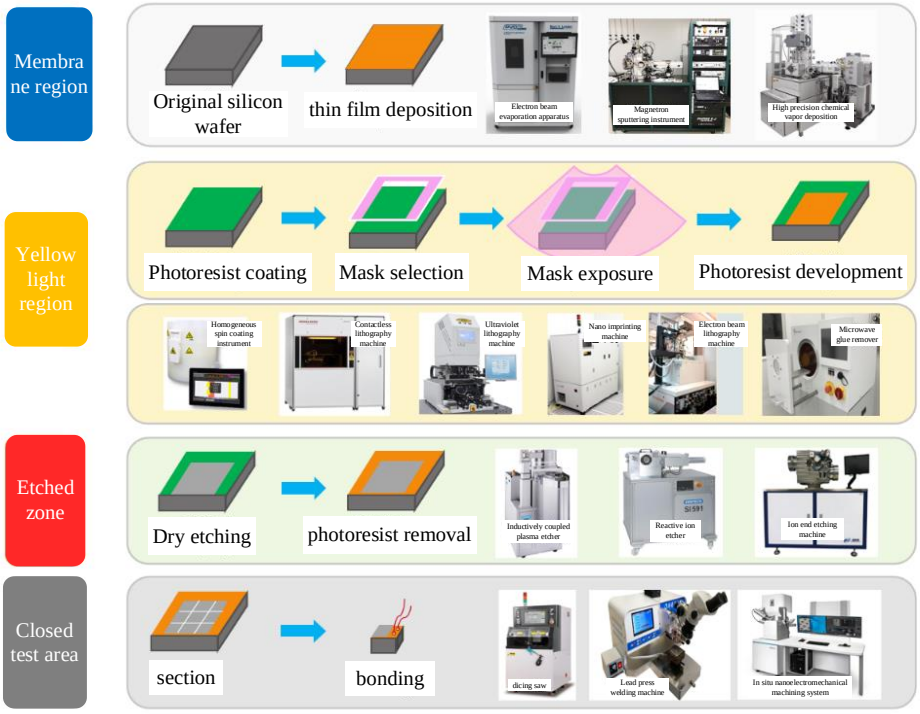


Fig. 1. Schematic diagram of a micro-nano embedded system.

Micro-nano embedded systems have the following characteristics (as shown in Fig. 2): 1) miniaturization; 2) low power consumption; 3) high performance; 4) multifunctionality; 5) integration; 6) intelligence; 7) formulatability; 8) real-time; 9) reliability; and 10) ease of portability [17]. Micro-nano embedded systems have a wide range of applications in several fields due to the above advantageous features, including smartphones, wearable devices, Internet of Things, healthcare, energy and environmental protection, national security and defense, etc. [18], as shown in Fig. 3.

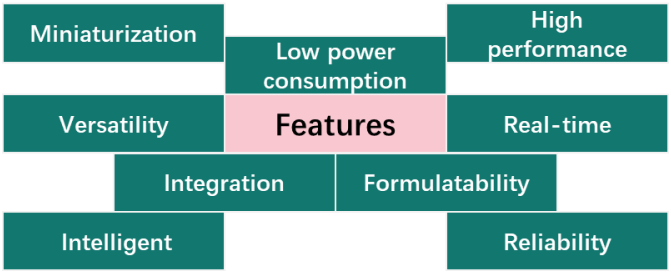


Fig. 2. Characteristics of micro-nano embedded systems.

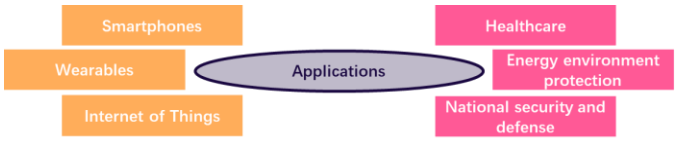


Fig. 3. Micro-Nano embedded system application.

2) In-depth development process of micro- and nano-embedded systems: The in-depth development of micro- and nano-embedded systems is a complex process involving close collaboration between hardware and software [19], and the key steps of the process (see Fig. 4) are as follows:

- Requirements analysis. Before development, it is necessary to carefully analyze and clarify the requirements of the micro-nano embedded system, including functional requirements, performance indicators,
- System design. Based on the hardware platforms, software architecture and so on.results of the micro- and nano-embedded system requirements analysis, system

design is carried out to select the appropriate hardware platform and software architecture, define the modules and interfaces of the system, as well as to develop the system testing and validation plan.

- Hardware design. Select and configure the processor, memory, sensors, communication interfaces, etc.
- Software development. Realize the function and control logic of the system, i.e., write the underlying driver, customization of the operating system, implementation of algorithms, etc.
- Integration and Testing. Integrate hardware and software components into a complete system for system-level functional and performance testing.
- Optimization and debugging. Optimize and debug the system to ensure that it meets the performance requirements and has good stability and reliability.
- Deployment and Maintenance. Complete the deployment of the developed embedded system to the target device or system and perform maintenance and support.

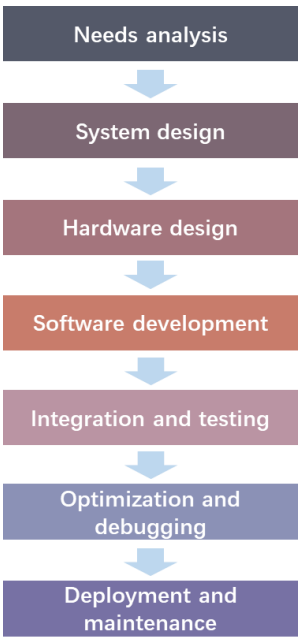


Fig. 4. Development process of micro-nano embedded system.

3) Framework of test and analysis system for in-depth development of micro- and nano-embedded systems: In the process of deep development of micro-nano embedded system, the integration test of hardware and software as well as optimization debugging as an important part of the system development can help designers to find and solve the potential problems and defects of micro-nano embedded system [19]. In order to help developers of micro-nano embedded system design management, improve the efficiency of micro-nano embedded system in-depth development, and increase the prediction accuracy of development test analysis, this paper combines intelligent optimization algorithms and integrated learning algorithms to study the optimization model of micro-nano embedded system in-depth development test, and therefore designs an intelligent test analysis framework for the in-depth development of micro-nano embedded system based on intelligent optimization algorithms to improve the integrated learning technology (see Fig. 5).

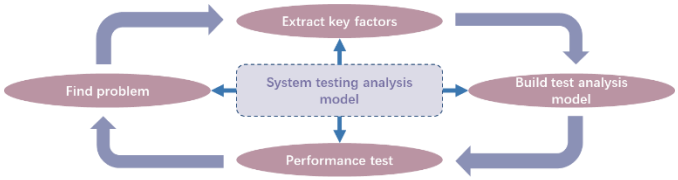


Fig. 5. Intelligent test and analysis framework for the development of micro and nano embedded systems.

B. Key Technical Content

According to the analysis framework, the research on intelligent test analysis method for deep development of micro-nano embedded system based on intelligent optimization algorithm to improve the integrated learning technology mainly includes four parts, including system development process analysis, system test analysis factor extraction, system test analysis data processing, and system development and test analysis model construction, as shown in Fig. 6.

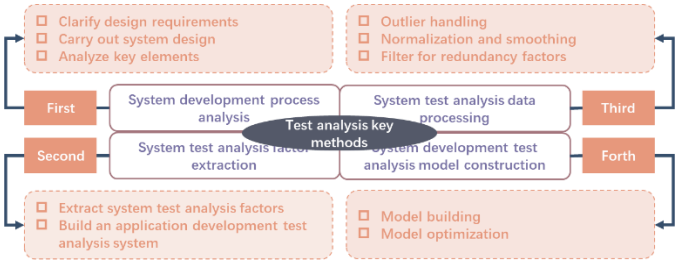


Fig. 6. Key technologies for intelligent test and analysis of micro and nano embedded system development.

1) Analysis of the system development process: The system development process analysis is mainly to clarify the micro-nano embedded system design requirements, carry out the system design, and analyze the key elements of the in-depth development process of micro-nano embedded system.

2) Extraction of factors for systematic test analysis: According to the principles of systemic, holistic, and process, the system test analysis factors are extracted from the key elements of the in-depth development process of micro- and nanoembedded systems, and the application development test analysis system is constructed.

3) System test and analysis data processing: Extract the system test analysis factor data from the test data, and then annotated to construct the data set, for the abnormal values and outliers, the use of proximity to take the value of the correction data, for the vacant value of the random forest model to fill in the prediction method, and after that the data normalization and smoothing, through the Pearson correlation coefficient method of filtering the redundant factors, to get the input data matrix, as shown in Fig. 7.



Fig. 7. Data processing.

4) System development test analysis model construction: Taking the labeled data matrix as input and the test analysis values as output, the lightweight gradient boosting machine is used to construct a test analysis model for the development of micro-nano embedded systems; at the same time, the photon search algorithm is used to optimize the test analysis model for the development of micro-nano embedded systems based on the lightweight gradient boosting machine algorithm, and the specific constructive optimization paradigm is shown in Fig. 8.

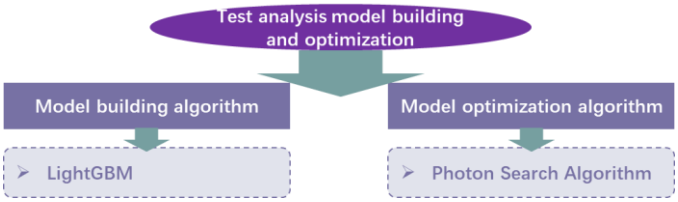


Fig. 8. Optimization paradigm for building test and analysis models for development of micro and nano embedded systems.

III. Application Analysis Modeling and Optimization for Micro-Nano Embedded System Development

A. LightGBM Algorithm

Lightweight Gradient Boosting Machine (LightGBM) [20] is a gradient boosting tree-based boosting method in integrated learning, which was proposed by Microsoft in 2017, and is currently one of the best-performing boosting methods. While the traditional GBDT method greatly increases the computational complexity and memory usage of the model when the amount of sample data and features grows, LightGBM accelerates the model training speed without affecting the model accuracy [21].

In order to reduce the amount of training data, LightGBM adopts the gradient one-sided sampling algorithm, which gives different sampling weights to the data according to the gradient values, retains the data with large gradients (i.e., not yet trained, which contributes more to the improvement of the information gain), and randomly samples the data with small gradients and maintains the original distribution of the data, as shown in Fig. 9. This sampling method results in more accurate information gain compared to uniform random sampling. In order to reduce the sample features during training, LightGBM uses an independent feature merging algorithm to bind mutually exclusive features (several features that are not zero at the same time, e.g., unique heat coding) from high-dimensional features together to form a single feature, thus reducing the feature dimension and improving the training speed without affecting the accuracy.

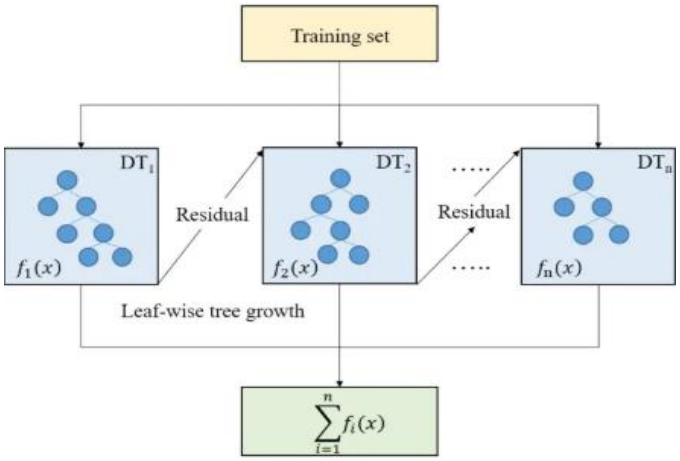


Fig. 9. Structure of LightGBM algorithm.

The LightGBM initial model objective function consists of a loss function with regularization:

$$O_{bj} \approx \sum_{i=1}^{I} Loss\left(y_i, \hat{y}_i\right) + \sum_{t=1}^{T} \Omega\left(f_t\right) \quad (1)$$

where, O_{bj} denotes the objective function, $Loss$ is the loss function, $\Omega\left(f_t\right)$ is the regularization term, I is the tree depth, y_i is the true value, $\hat{y}_i$ is the predicted value, T is the number of cotyledons, and f_t is the tth generation prediction function.

A second-order Taylor expansion is performed on the objective function determined from the tth generation prediction results:

$$O_{bj} \approx \sum_{i=1}^{n}\left[Loss\left(y_i, \hat{y}_i^{t-1}\right) + g_i f_i\left(x_i\right) + \frac{1}{2} h_i f_t^2\left(x_i\right)\right] + \gamma T + \frac{1}{2}\lambda \sum_{j=1}^{T} w_j^2 \quad (2)$$

where, $f_t\left(x_i\right)$ denotes the new prediction function when the model is iterated, g_i is the first-order derivative of Loss, h_i is the second-order derivative of Loss, γ is the new node complexity cost parameter, and w_j is the leaf node value, λ indicates the leaf node value coefficient.

Accumulate the first-order and second-order gradients, and then make the first-order derivative of the objective function with respect to w_j is 0, i.e., take the extreme point:

$$O_{bj}^{t} \approx -\frac{1}{2}\sum_{j=1}^{T} \frac{G_j^2}{H_j + \lambda} + \gamma T \quad (3)$$

where, G_j denotes the first order gradient accumulation value and H_j denotes the second order gradient accumulation

value, λ denotes the complexity cost parameter of the new node.

The optimal tree is the one that minimizes the value of the objective function among the different arrangement structures. Use the splitting gain formula G_{ain} to evaluate whether to split leaf nodes or not. If $G_{ain} > 0$, continue splitting to improve the model performance, otherwise stop splitting. After repeated iterations, the LightGBM strong learner algorithm model is finally obtained.

LightGBM has the following features (Fig. 10): 1) histogram-based decision-making algorithm (Fig. 11) improves computational speed and reduces memory usage; 2) adopts a leaf growth strategy, which helps to improve model accuracy; 3) one-sided gradient sampling accelerates learning and reduces computational complexity; 4) mutual exclusion feature bundling reduces the number of features and improves computational efficiency; 5) supports parallel and distributed learning; 6) Cache optimization accelerates the data exchange speed; 7) Supports a variety of loss functions to meet different business needs; 8) Regularization and pruning strategies control the model complexity and prevent overfitting; 9) The algorithm has good model interpretability [22].

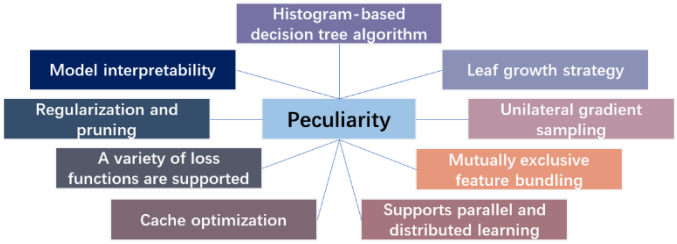


Fig. 10. Characteristics of LightGBM algorithm.

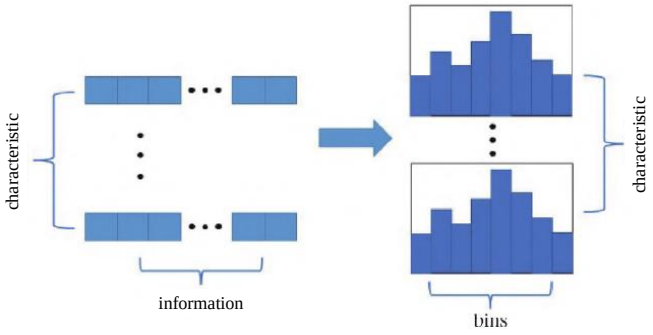


Fig. 11. Histogram based decision making algorithm.

LightGBM is widely used in binary classification, multiclassification, and sorting scenarios, for example, in personalized product recommendation, risk management, and market analysis. It can handle a large number of features and samples and is suitable for industrial-grade data analysis and complex machine learning tasks [23].

B. Photon Search Algorithm

Photon Search Algorithm (PSA) [24] is a heuristic algorithm based on physical phenomena proposed in 2019. The theoretical basis of the algorithm is derived from the photon hypothesis and quantum theory proposed by physicists Max Planck, Einstein and Broglie. The photon search algorithm has been studied for its strong global exploration ability and high search efficiency, but it underperforms in local exploitation ability and has low convergence accuracy.

1) Optimization principle: The working principle of the photon search algorithm involves the initialization of the photon's motion, the observation behavior and the search exclusion principle. In the algorithm, the position of the photon is updated taking into account the distance to the global best fitness value photon and the distance passed by the photon in the iteration. In addition, the algorithm designs the observation behavior to model the quantum uncertainty principle and the search exclusion principle based on the bubbleley incompatibility principle to avoid photons occupying the same position [25].

In the PSA algorithm, each photon represents a search agent that aims at the optimal solution, obtains a fixed speed, goes through iterations, and outputs the optimal solution.

The specific optimization strategies are as follows:

a) Photon motion

Define photonic individual:

$$X_i = \left(x_i^1, \cdots, x_i^d, \cdots, x_i^n\right), i = 1, 2, \cdots, N \tag{4}$$

where X_i denotes the ith photon, and x_i^d denotes the dth dimension of photon i.

The mathematical model of photon optimized motion is as follows:

$$x_i^d\left(t+1\right) = x_i^d\left(t\right) + De \cdot v_i^d\left(t+1\right) \tag{5}$$

$$v_i^d\left(t+1\right) = \frac{R_{Len}}{R_{ig}}\left(x_g^d\left(t\right) - x_i^d\left(t\right)\right) \tag{6}$$

$$R_{Len} = Scl \cdot \left\| X_{upper} - X_{lower} \right\|_2 \tag{7}$$

$$R_{ig} = \left\| X_i\left(t\right), X_g\left(t\right) \right\|_2 \tag{8}$$

$$De = \frac{ext}{t} \tag{9}$$

where, $x_i^d\left(t\right)$ denotes the d-dimensional position information of the i^{th} photon in the t^{th} iteration, $v_i^d\left(t+1\right)$ denotes the velocity information of the i^{th} photon in the d-dimension in the t+1th iteration, De is the convergence weight, which is used to adjust the convergence speed in the search process, R_{Len} denotes the distance passed by the photon in the photon iteration, R_{ig} denotes the Euclidean distance between

the ith photon and the current optimal photon, $x_g^d(t)$ is the d-dimensional position information of the optimal photon in the tth iteration, Scl denotes the value of the weight, and ext is the convergence coefficient.

b) Observation of behavior: In order to simulate the photon uncertainty principle, the observation behavior is designed to calculate the photon position:

$$x_i^d(t) = x_i^d(t) + De \cdot randA \tag{10}$$

Where $randA$ denotes a random number between -1 and 1.

c) Search exclusion principle: In the PSA algorithm, based on the principle of bubbly incompatibility, the search exclusion principle behavior is used to update the photon individual positions as follows:

$$x_i^d = x_{low}^d + \left(x_{up}^d - x_{low}^d\right) \cdot randB \tag{11}$$

where x_{low}^d and x_{up}^d denote the lower and upper bound ranges of the dth dimension, respectively, and $randB$ denotes a random number between 0 and 1.

2) Process steps: According to the above PSA algorithm behavior description, the PSA algorithm flow is shown in Fig. 12 with the following steps:

- Step 1: Initialize the position of the photon X_i ;
- Step 2: Calculate the photon fitness value and calculate the distance passed in the photon iteration R_{Len} ;
- Step 3: Update R_{ig} , the velocity and position of the photon, and update the photon using the observation behavior;
- Step 4: Calculate the photon fitness value and update the global optimal photon;
- Step 5: If t<Max_T, return to step 3, otherwise go to step 6;
- Step 6: The algorithm terminates and outputs the optimal solution.

C. PSA-LightGBM Model Application

1) PSA-LightGBM: In this paper, we use real number coding to encode the number of cotyledons, tree depth, learning rate and minimum data number of LightGBM algorithm, and the dimensions of photon individuals are 4-dimensional, as shown in Fig. 13. Each photon individual includes the dimensions of the number of cotyledons L, the tree depth D, the learning rate l and the minimum data number S.

Analyzing the application analysis problem of integrated development of micro and nano embedded systems, the problem can be considered as a predictive regression problem, so, RMSE is used as the fitness value in the PSA-LightGBM model.

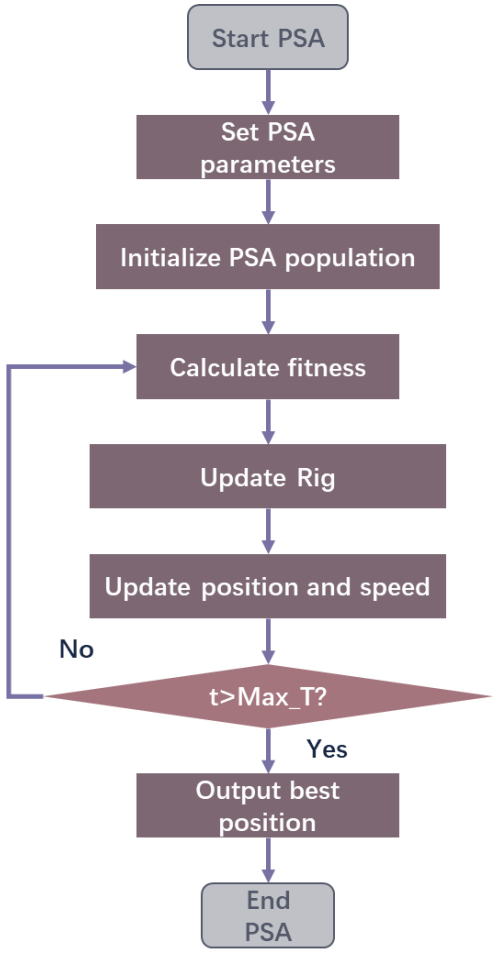


Fig. 12. Flowchart of PSA algorithm.

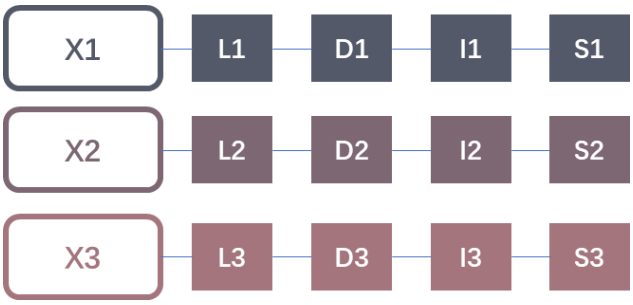


Fig. 13. PSA-LightGBM algorithm coding method.

According to the coding method and the fitness function, the steps of LightGBM application analysis method based on PSA algorithm (shown in Fig. 14) are as follows: 1) take the number of cotyledons L, the tree depth D, the learning rate l and the minimum number of data S as the optimization objects in the LightGBM model; 2) set the optimization parameter range of the PSA algorithm, the size of the population, and the maximum number of iterations, and initialize the population of the PSA algorithm ; 3) Calculate the fitness value based on RMSE; 4) Update the photon individual positions and velocities based on LightGBM parameters, and output the optimal parameter set to LightGBM; 5) Establish the PSA-LightGBM model based on the optimal parameter combination.

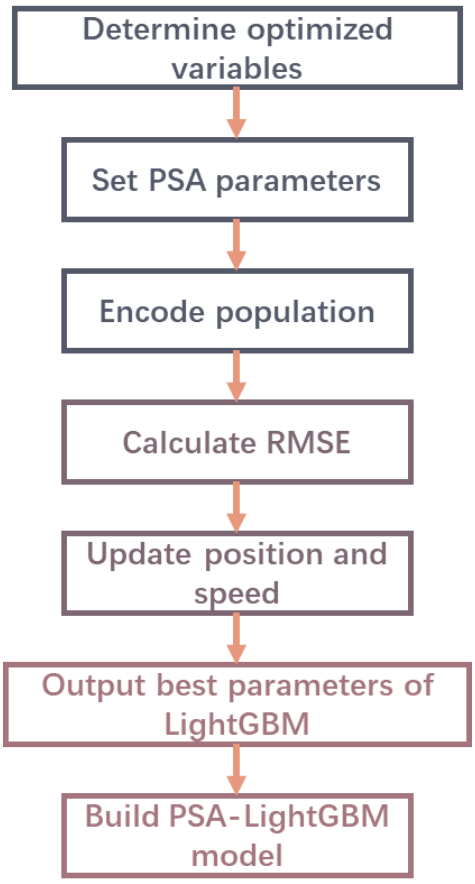


Fig. 14. PSA-LightGBM.

2) Algorithm application: In order to construct an application analysis model for the development of micro-nano embedded systems, this paper applies the PSA-LightGBM model to the in-depth development problem of micro-nano embedded systems under the complex integration architecture, and the specific application flow is shown in Fig. 15. The application of PSA-LightGBM algorithm in the in-depth development problem of micro-nano embedded systems under the complex integration architecture is divided into the following flow:

a) Analyze the problem of in-depth development of micro-nano embedded system under complex integrated architecture, extract the analytical characteristic parameters of micro-nano embedded system development and application, and establish the analytical characteristic parameter set;

b) Collect data related to analytical features during the in-depth development of micro-nano embedded systems in complex integrated architectures and use appropriate data processing techniques to obtain normalized dimensionality-reduced datasets;

c) Combine PSA-LightGBM algorithm to construct the mapping relationship between the development application analysis feature parameter values and the development application analysis scores, the specific PSA-LightGBM algorithm application analysis schematic is shown in Fig. 16.

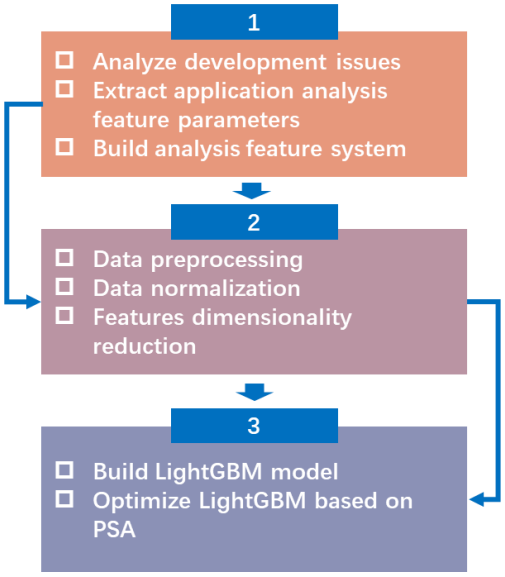


Fig. 15. Application flow of PSA-LightGBM algorithm.

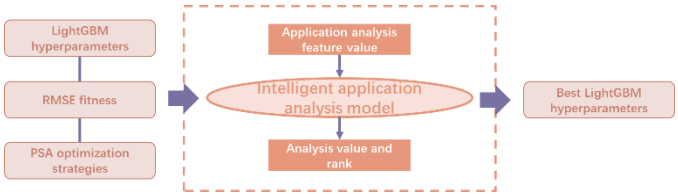


Fig. 16. PSA-LightGBM algorithm application analysis.

IV. Simulation and Analysis

A. Environmental Settings

This paper uses data from the in-depth development of micro-nano embedded systems under a complex integrated architecture. A total of 3530 data sets are used, with 80% of the data used as the training set for the application analysis model, 10% of the data used as the validation set during the optimisation process, and 10% of the data set used as the test set.

Software environment: programming environment Python 3.8, visualisation software Matlab 2022a, operating system Wins10.

The comparison algorithms include LightGBM, GWO-LightGBM, HHO-LightGBM, TLBO-LightGBM and PSA-LightGBM. The population size of the GWO, HHO, TLBO and PSA algorithms is set to 50, and the maximum number of iterations is 1000 and other parameter settings are shown in Table I.

TABLE I. Application Development Analysis Model Parameter Settings

No.	Algorithms	Settings
1	LightGBM	L=30, I=0.1, S=20
2	GWO- LightGBM	a=[-2,2]
3	HHO-LightGBM	E_0=2
4	TLBO-LightGBM	Teaching factor=1.2
5	PSA-LightGBM	No

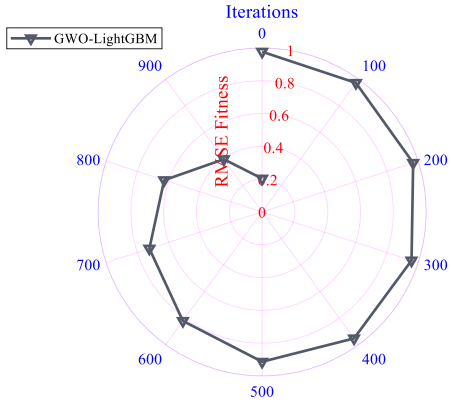
The range of GWO, HHO, TLBO, and PSA algorithms to optimize the LightGBM parameters number of cotyledons L, tree depth D, and learning rate l with minimum number of data S is shown in Table II.

TABLE II. OPTIMIZED PARAMETER RANGE SETTING

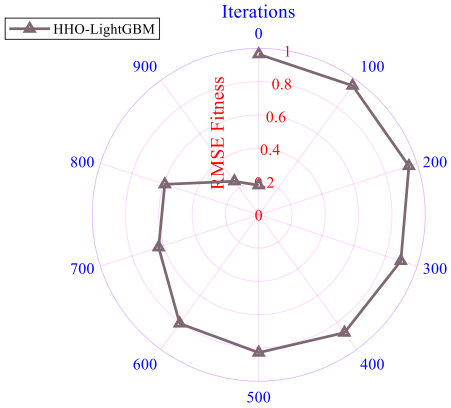
No.	Hyperparameters	Var.	Range
1	The number of cotyledons	L	[20,100]
2	The depth of the tree	D	[3,8]
3	The learning rate	L	[0.01, 0.3]
4	Minimum number of data	S	[1,30]

B. Analysis of Results

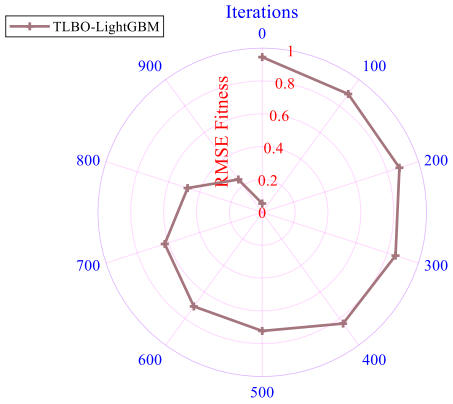
In order to further improve the performance of the analytical model for the development and application of micro- and nanoembedded systems, the PSA algorithm is used to find the optimal set of parameters by taking the number of cotyledons L, the depth of the tree D, and the learning rate l with the minimum number of data S as optimization parameters in the LightGBM model. In order to verify the validity set of PSA algorithm's excellent computational efficiency, GWO, HHO and TLBO are compared, and the change curve of each algorithm's adaptability is shown in Fig. 17 (a)-(d). From Fig. 17, it can be seen that the RMSE value of the PSA-LightGBM-based application analysis for the development of micro- and nanoembedded systems converges to the minimum value, and the GWO-LightGBM algorithm has the largest RMSE value, and PSA-LightGBM is better than the other algorithms.



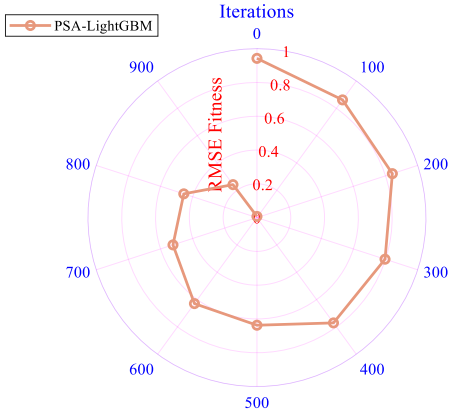
(a) GWO-LightGBM



(b) HHO-LightGBM



(c) TLBO-LightGBM



(d) PSA-LightGBM

Fig. 17. Comparison of the adaptation curves of the algorithms.

The hyperparameter optimization results of LightGBM model based on GWO, HHO, TLBO, and PSA algorithms are shown in Table III. From Table III, it can be seen that the optimal combination of parameters for PSA optimized LightGBM are: the number of mesocotyledons L=65, the tree depth D=5, the learning rate l=0.09 with the minimum number of data S=22.

TABLE III. RESULTS OF OPTIMIZING LIGHTGBM MODEL PARAMETERS BY EACH ALGORITHM

No.	Parameters	GWO	HHO	TLBO	PSA
1	The number of cotyledons	21	29	58	65
2	The depth of the tree	6	5	5	5
3	The learning rate	0.1	0.14	0.1	0.09
4	Minimum number of data	23	21	21	22

The predicted performance of development application analysis model based on LightGBM, GWO-LightGBM, HHO-LightGBM, TLBO-LightGBM, PSA-LightGBM algorithms for micro- and nanoembedded systems is analyzed using the test set. The results of the predicted performance of the development application analysis model based on each algorithm are shown in Fig. 18. From Fig. 18, it can be seen that the ME, MAE, MAPE, RMSE, and R2 values of PSA-LightGBM are 0.043,

1.012, 4.002, 0.477, and 0.99, respectively, and the prediction performance is better than LightGBM, GWO-LightGBM, HHO-LightGBM, and TLBO-LightGBM.

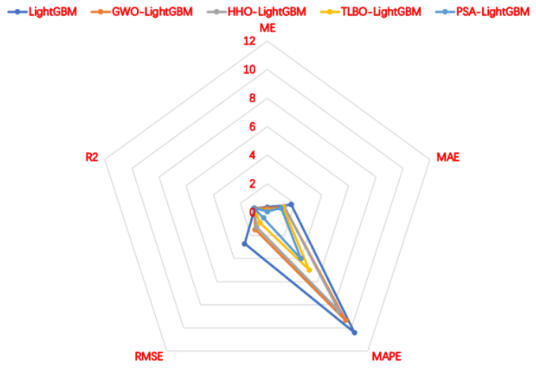


Fig. 18. Predictive performance statistics for each analytical model.

V. CONCLUSION

Focusing on the analysis of application development issues in the deep development of micro-nano embedded systems with complex integrated architectures, this paper proposes a development and application testing analysis method based on machine learning technology and intelligent optimisation algorithms, which achieves accurate prediction of system development and application testing. By analysing the deep development process of micro-nano embedded systems, a development and application testing analysis framework is designed, and the key technical content is introduced. For the problem of constructing an application testing analysis model, a development and application testing analysis method based on PSA-LightGBM is proposed by combining the LightGBM hyperparameter optimisation with the photonic search algorithm. Data analysis and verification show that compared with other algorithms, the PSA-LightGBM algorithm-based micro-nano embedded system development application test analysis method further reduces the test prediction error, and at the same time verifies the application feasibility of the development application test analysis method in the micro-nano embedded system deep development under the complex integrated architecture. The next step is to improve the PSA algorithm and improve the accuracy of the LightGBM model, and at the same time apply PSA-LightGBM and micro-nano embedded system development to other fields to verify the robustness and feasibility of the algorithm model.

REFERENCES

[1] Gassab M, Chebil A, Dridi C. Predictive Study of Electrical Performances of Interdigitated, Cost-Effective Supercapacitor for Autonomous Microsystems[J]. Arabian Journal for Science and Engineering, 2022, 47(1):1043-1051.

[2] Gui Y, Wu R. Buckling analysis of embedded thermo-magneto-electro-elastic nano cylindrical shell subjected to axial load with nonlocal strain gradient theory[J]. Mechanics Research Communications, 2023, 128:104043-.

[3] Delgado C A, Chapa A, Lozano A V H. In-Situ Surface Depositing of Nano-Micro-particles on Electrospun Fibers[J]. Fibers and Polymers, 2024, 25(2):407-413.

[4] Shariyat M, Mirmohammadi M. Modified strain gradient analysis of microscale dynamic response suppression of SMA-composite microplates featuring 2D superelastic phase transformations and kinematic nonlinearity[J]. Composite Structures, 2022, 280:114879-.

[5] Zhang Z, Wang Z, Wang F, Qin T, Zhu H, Liu P. A Laser-Processed Carbon-Titanium Carbide Heterostructure Electrode for High-Frequency Micro-Supercapacitors[J]. Small, 2023.

[6] Riester O, Laufer S, Deigner H P. Direct 3D printed biocompatible microfluidics: assessment of human mesenchymal stem cell differentiation and cytotoxic drug screening in a dynamic culture system[J]. Journal of Nanobiotechnology, 2022, 20(1):540.

[7] Potts L V, Amboise W N, Todosov A. An Affordable Modular Open-Source Scale Model Platform for Teaching Autonomous Mobile Mapping (AMM) Technologies[J]. Surveying and Land Information Science, 2023, 82(2):61-75.

[8] Ren W, Wu X, Cai R. A hybrid artificial intelligence and IOT for investigation dynamic modeling of nano-system[J]. Advances in Nano Research: An International Journal, 2022.

[9] Huang Y H, Lin C J, King Y C. A study of hydrogen plasma-induced charging effect in EUV lithography systems[J]. Discover Nano, 2023, 18(1).

[10] Oh D K, Lee W, Chae H, Chun H, Lee M, Kim D H. Burr- and etch-free direct machining of shape-controlled micro- and nanopatterns on polyimide films by continuous nanoinscribing for durable flexible devices[J]. Microelectronic Engineering, 2022, 257:111740-.

[11] Suder J. Parameters evaluation of cameras in embedded systems[J]. Przeglad Elektrotechniczny, 2022.

[12] Bayram F, Gajula D, Khan D, Koley G. Mechanical memory operations in piezotransistive GaN microcantilevers using Au nanoparticle-enhanced photoacoustic excitation[J]. Microsystems & Nanoengineering, 2022, 8(1):8.

[13] Hosseini M, Bemanadi N, Mofidi M. Free vibration analysis of double-viscoelastic nano-composite micro-plates reinforced by FG-SWCNTs based on the third-order shear deformation theory[J]. Microsystem Technologies, 2023, 29(1):71-89.

[14] Wang S, Liu M, Yang X, Lu Qq, Xiong Z, Li L. An unconventional vertical fluidic-controlled wearable platform for synchronously detecting sweat rate and electrolyte concentration[J]. Biosensors & Bioelectronics, 2022, 210:114351.

[15] Roberto R D, Brandolini E, Sparvieri G, Graziani F. Best practices on adopting open-source and commercial low-cost devices in small satellites missions[J]. Acta Astronautica, 2023, 211:37-48.

[16] Saraswathy M, Prakash D, Muthtamilselvan A M Q M. Theoretical study on bio-convection of micropolar fluid with an exploration of Cattaneo-Christov heat flux theory[J]. International Journal of Modern Physics, B. Condensed Matter Physics, Statistical Physics, Applied Physics, 2024, 38(1).

[17] Cui Y, Chen Z, Gu S, Yaang W, Ju Y. Investigating size dependence in nanovoid-embedded high-entropy-alloy films under biaxial tension[J]. Archive of Applied Mechanics, 2023, 93(1):335-353.

[18] Lu H, Zheng H. A comprehensive review of organic-inorganic composites based piezoelectric nanogenerators through material structure design[J]. Journal of Physics, D. Applied Physics: A Europhysics Journal, 2022.

[19] Pavithra B, Prabhu S G, Manjunatha N M. Design, development, fabrication and testing of low-cost, laser-engraved, embedded, nano-composite-based pressure sensor[J]. ISSS Journal of Micro and Smart Systems: A Technical Publication of the Institute of Smart Structures and Systems, 2022.

[20] Zhuo Yichao, Hao Haibin. Intelligent detection of traffic in hospital microservice platform security operation and maintenance management system based on genetic algorithm optimized LightGBM algorithm[J]. Chinese Journal of Medical Physics, 2024, 41(06):788-792.

[21] Jingyi Liu, Shengnan Lu. LightGBM unbalanced data classification algorithm based on adaptive Borderline-SMOTE oversampling[J]. Information Technology and Informatization, 2024, (06):205-208.

[22] Y. Zhang, M. Wen. RUSBoost-LightGBM short-term load forecasting method taking into account data imbalance[J]. Foreign Electronic Measurement Technology, 2024, 43(06):41-49.

[23] Chen Xiaoling, Zhang Cong, Huang Xiaoyu. Research on grain yield prediction based on Bayesian-LightGBM model[J]. Chinese Journal of Agricultural Mechanical Chemistry, 2024, 45(06):163-169.

[24] Liu Y L, Li R J. PSA: A Photon Search Algorithm[J]. Journal of Information Processing Systems, 2020, 16(2):478-493.

[25] X. Li, X.G. Qiao. Adaptive photon search algorithm based on Harris hawk optimization[J]. Electronic Design Engineering, 2022, 30(15):10-15.

Optimization of Knitting Path of Flat Knitting Machine Based on Reinforcement Learning

Tianqi Yang*
Shanghai Institute of Visual Arts, Shanghai 201620, China

***Abstract*—In the textile industry, the flat knitting machine plays a crucial role as a production tool, and the quality of its weaving path is closely related to the overall product quality and production efficiency. Seeking to improve and optimize the knitting path to improve product effectiveness and productivity has become an urgent concern for the textile industry. This article elegantly streamlines and enhances the intricate weaving process of fabrics, harnessing the formidable power of reinforcement learning to achieve unparalleled optimization of weaving paths on a flat knitting machine. By ingeniously integrating reinforcement learning technology into the fabric production realm, we aspire to elevate both the quality and production efficiency of textiles to new heights. The core of our approach lies in meticulously defining a state space, action space, and a tailored reward function, each meticulously crafted to mirror the intricacies of the knitting process. This model serves as the cornerstone upon which we construct an innovative knitting pathway optimization algorithm, deeply rooted in the principles of reinforcement learning. Our algorithm embodies a relentless pursuit of excellence, learning from its interactions with the dynamic environment, embracing a methodical trial-and-error approach, and continuously refining its decision-making strategy. Its ultimate goal: to maximize the long-term cumulative reward, ensuring that every stitch contributes to the overall optimization of the weaving process. In essence, we have forged a groundbreaking collaboration between the traditional art of fabric weaving and the cutting-edge science of reinforcement learning, ushering in a new era of intelligent and efficient textile production. Through this process of iterative optimization, the agent can gradually learn the optimal knitting path. To verify the effectiveness of the algorithm, we performed extensive experimental validation. The experimental results show that reinforcement learning can significantly improve knitting efficiency, improve the appearance and feel of fabrics. Compared with traditional methods, the method proposed in this article has a higher level of automation and better adaptability, achieving more efficient and intelligent knitting production, with a 10% increase in production efficiency.**

***Keywords*—Flat knitting machine; reinforcement learning; weaving path optimization; textile industry**

I. INTRODUCTION

In today's booming textile industry, the flat knitting machine stands as the cornerstone equipment, its performance being a direct determinant of both production efficiency and product quality. Notably, the selection of the knitting path holds immense significance, as it critically shapes the final product's appearance, hand feel, and overall production efficiency [1, 2]. However, traditional knitting path optimization methods are heavily reliant on manual expertise and repetitive trials, making them inefficient and unable to keep pace with the escalating complexity of knitting requirements. Consequently, the textile industry faces a pressing need to explore novel optimization techniques that can enhance the performance of flat knitting machines.

In recent years, the swift advancements in artificial intelligence technology have offered an efficacious solution for enhancing the operational optimization of microgrids. Notably, the reinforcement learning algorithm stands out as a prominent tool, as it transcends the reliance on historical data and pre-defined labels. Instead, it actively engages with the environment through iterative learning, fostering a dynamic and adaptive approach. Traditionally, the optimization of power system operations entailed modeling the intricate system mechanisms and subsequently solving these models under stringent constraints. However, reinforcement learning disrupts this paradigm by eliminating the need for an explicit physical model of the system. It possesses the remarkable ability to discern and refine the operational model purely from the available data, thereby significantly accelerating the learning process and enhancing the efficiency of modeling the system's operations. This shift underscores the transformative potential of reinforcement learning in driving the future of microgrid optimization. At the same time, based on the "trial and error" behavior of reinforcement learning, continuous learning can be carried out through interaction with the environment, and the accuracy of the model and the optimization of parameters can be continuously improved. Therefore, compared with the traditional micro-grid control mode, reinforcement learning can organically connect the components of the system, interact and cooperate with each other, and complete complex optimization work with a small amount of prior information, improving the operation ability and efficiency of the micro-grid. At the same time, the practical application and improvement of the algorithm in different scenarios can also effectively improve the application effect of reinforcement learning.

The rapid advancements in artificial intelligence have ushered in reinforcement learning as a groundbreaking machine learning technology. Extensively applied to various optimization challenges, reinforcement learning leverages the interactive learning process between an agent and its environment to autonomously discover optimal decision strategies. This approach offers a promising solution for tackling intricate problems, making it a compelling candidate for revolutionizing knitting path optimization in the textile industry. In this context, this paper will discuss the knitting path optimization method of flat knitting machine based on reinforcement learning to improve the knitting efficiency and product quality.

*Corresponding Author.

II. Multi-axial Warp Knitted Fabric Production Equipment and Technology

A. Fibrous Raw Materials

The range of raw materials for warp knitted multi-axial fabrics is very wide. The lining usually adopts high performance fibers with good mechanical properties, such as glass fiber (GF), carbon fiber (CF), Kevlar fiber, ultra-high molecular weight polyethylene fiber (UHMW. PE), etc. It can be a low twist flexible staple yarn or a non-robbing high performance filament yarn. When used as reinforcing yarns, high-performance untwisted filaments are generally used, and sometimes the yarns can be slightly twisted for ease of weaving. Yarns are generally thick, up to about 2500tex.

Polyester low elastic yarn, glass fiber yarn, etc. can be used for ground weave yarn. When polyester yarn is used for ground weave more than glass fiber, due to the high requirements for yarn fineness and bending stiffness, the technical parameters of wire drawing are improved, and thus the cost is greatly increased. Therefore, in actual production, raw materials should be selected according to the requirements of composite material properties and applications.

Glass Fiber (Glass Fiber) is a new type of engineering material, which is made of inorganic glass added with silica oxides such as calcium, boron, sodium, iron and aluminum. The molecular arrangement is a three-dimensional network structure, so the properties of glass fiber are homogeneous [3]. It has excellent properties such as non-combustible, corrosion-resistant, high-temperature resistance, low moisture absorption, and small elongation. It also has excellent characteristics in electrical, mechanical, chemical, and optical aspects, but the disadvantages are brittleness and poor wear resistance.

1) Production process of glass fiber: The production of glass fiber has a long history, and there are two main types at present: one is the method of replacing platinum to increase the pot to make glass into balls, and put the balls into the crucible furnace made of platinum pound alloy to make a leak plate, and the glass melt flows out of many leaks on the leak plate, and is wound on the high-speed rotating wire winding Jane; The second is the pool method: the glass powder is directly put into the pool cellar to melt, and the glass melt flows out through the leaky plates and nozzles installed on several sub-channels. The wire drawing method is the same as before. Glass fibers are generally 3-10 um in diameter, and more recently 13um, 15um, 24um monofilament yarns have been used. Due to the characteristics of glass fiber in structure, performance, processing technology, price, etc., it has always occupied an important position in the composite material manufacturing industry.

2) Types of glass fibers: Based on their distinct raw materials, glass fibers can be categorized into the following types:

C glass fiber, alternatively known as medium-alkali glass, possesses characteristics that lie between E glass fiber and A glass fiber. While it excels in chemical resistance, its electrical performance is inadequate. Furthermore, its mechanical strength falls short by 10% to 20% compared to alkali-free glass fiber.

In overseas markets, medium-alkali glass fiber is predominantly utilized for the production of corrosion-resistant glass fiber products. Conversely, in China, this type of glass fiber holds a prominent share, exceeding 60% of the total glass fiber output, and finds extensive applications. It is widely employed as reinforcement in Fiber Reinforced Polymers (FRP), as well as in the manufacture of filter fabrics and wrapping materials. This prevalence stems from its cost-effectiveness, offering a significant price advantage over alkali-free glass fiber, thereby fostering robust competitiveness in the domestic market.

High-strength glass fiber: It is characterized by high strength and high modulus. It is mostly used in military industry, space, bulletproof armor and sports equipment. However, due to the high price, it cannot be promoted in civilian use at present, and the world output is only about a few thousand tons.

E-CR glass. This is an enhanced boron-free and alkali-free glass that is utilized to craft glass fibers exhibiting exceptional acid and water resistance. Specifically, its water resistance surpasses alkali-free glass fiber by a remarkable seven to eight times, while its acid resistance significantly outperforms that of medium-alkali glass fiber. It is a new variety specially developed for underground pipelines, storage tanks, etc.

B. The Basic Properties of the Polymer Optical Fibers

The application of optical fiber to luminescent fabric must take into account the necessary properties of luminescent fabric and the performance characteristics of optical fiber. The necessary attributes of luminescent fabric include soft, light, durable and safe for use, high and uniform luminescent brightness, and good performance. The most prominent feature of the luminous fabric is the luminous brightness and flower pattern effect in the dark. The luminous brightness refers to the light flux per unit projected area. Factors such as fabric stretching, fabric type [4], fabric density, number of optical fibers, fiber bending radius, fiber bending loss [5].

The mechanical properties of optical fibers affect their weaving properties [6]. The collusion strength of optical fiber is low, poor elasticity and tolerance, so the optical fiber is easy to break in the process of weaving. Quartz fiber and glass fiber bending performance is poor, polymer fiber fracture tensile rate is larger, good toughness, good bending performance [7], bending radius, but the bending radius of polymer fiber and weaving process when the yarn bending radius is inconsistent, and polymer fiber bending rigidity, not conducive to fiber bending into circles. Therefore, it is difficult to weave polymer optical fiber into circles, and it is easier to weave through the form of floating line. Polymer fiber flexibility, flexure and elongation, good, easy to process and use [8].

Fig. 1 shows knitting path optimization process under the framework of reinforcement learning. The thermal performance of optical fiber affects the dyeing and shaping of optical fiber luminous fabric. Quartz fiber and glass fiber have good heat resistance, polymer fiber has poor heat resistance, low melting point, poor thermal stability [9], and it is easy to damage the fiber in the case of acute heat or cold, increasing the loss of polymer fiber. The working temperature of PS core and PMMA core polymer fiber is less than 80°C, and the working

temperature of PC core polymer fiber is less than 150℃. The heat-resistant polymer fiber can be used in the range of 100℃-200℃ [10, 11], so it is difficult for the polymer fiber fabric to iron, dye, shape, etc.

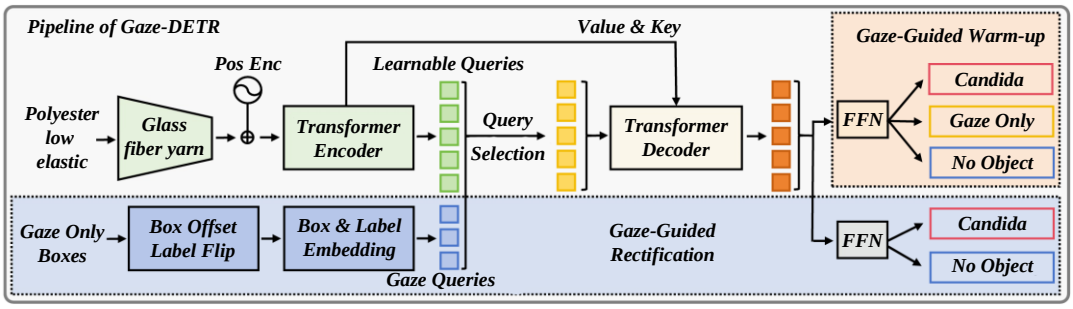


Fig. 1. Knitting path optimization process under the framework of reinforcement learning.

The chemical performance of optical fiber generally depends on its raw materials. For instance, the performance of glass fiber closely resembles that of glass, while the chemical properties of polymer fiber align with those of plastics. However, the unique structure of optical fiber itself can also significantly influence its chemical properties. Specifically, the large surface area of optical fiber facilitates the absorption of moisture and susceptibility to corrosion, thereby diminishing its compressive resistance and light transmittance. While polymer optical fiber exhibits robust acid and alkali resistance, strong corrosion resistance, and aging resistance, it is prone to corrosion by acetone, hexane, acetone mixed reagents, ethyl acetate, and other reagents.

Optical fibers exhibit distinct optical properties, encompassing both light conduction and scattering, which directly impact luminescence brightness, visual effects, and intricate floral pattern manifestations. Notably, polymer fibers are characterized by pronounced dispersion, a high refractive index, and substantial optical transmission attenuation, particularly pronounced in the ultraviolet and infrared spectra. However, within the visible light spectrum, polymer fibers boast high transmittance, making them ideally suited for applications in the realm of decorative lighting, where their unique properties can be harnessed to create captivating visual displays.

In summary, the mechanical properties of polymer optical fiber directly influence its weaving capabilities. Additionally, the mechanical properties of the fiber after exposure to chemical reagents are also impacted, thereby affecting its weaving properties. Therefore, it is crucial to extensively test the mechanical properties of polymer optical fiber, particularly in the context of its weaving performance.

1) Technological parameters of multi-axial fabrics: In practical manufacturing, it is often observed that the rationality of the process arrangement for multi-axial warp knitted fabrics significantly influences the weft yarn structure within the fabric and the overall process flow. This can manifest in issues such as the weft yarn not being securely fixed within the ground weave or the fabric surface lacking weft yarn. These conditions often have repercussions on the tensile properties of the fabrics, as well as the mechanical properties of the composites post-molding. The weft-laying process primarily involves factors like fabric weight, weft yarn fineness, weft-laying angle, and the number of weft-laying layers. The processing parameters for multi-axial warp knitted fabrics encompass the fabric's gram weight, the gram weight of each weft layer, the density of each weft layer, the stitch pattern of the multi-axial fabric (i.e., the density of the stitching yarn), and the fineness of the stitching yarn and weave. The weight and density of the weft are determined by the weft laying process, which plays a pivotal role in the weaving of multi-axial fabrics. Unless there are specific requirements, the gram weight of each layer in a multi-axial fabric is typically calculated by dividing the total gram weight per square meter by the number of yarn layers [12, 13]. The weft density and fineness are key parameters in controlling fabric weight during the weft insertion process.

In actual production, the determination of these two parameters is mainly determined according to the weight per square meter required by customers. When producing axial warp knitted fabrics in more than two directions, when the total square meter gram weight of the fabric is given, the square meter gram weight of each layer of yarn must be determined first. The basis for the determination is that when the axial fabric is made into a composite material, without considering the force requirements in special directions, it is generally believed that only when the fabric as a reinforcing material has various similarities in structure, the composite material can jointly bear the load in all directions and exert the best performance of each component in the material [14, 15]. Therefore, when there is no special requirement, the square meter gram weight of each layer of yarn must be determined by dividing the total square meter gram weight by the average value obtained by the number of layers of yarn as a benchmark. After determining the weight of the next square meter, the specific process parameters are determined according to the calculation method of the weight of the square meter.

For example, in the actual production of biaxial fabrics, in order to make the force bearing capacity in the direction of 0° and 90° equivalent, generally under the condition that the square meter gram weight of the fabric is given, the square meter gram weight of the warp yarn and the square meter gram weight of the weft yarn are allocated according to half of the square meter gram weight of the fabric. Similarly, for multi-axial fabrics, if there are yarns inserted in directions other than 0° and 90°, such as: +45°, the forces in all directions should be considered to determine the yarn parameters [16, 17].

For warp-knitted axial fabrics, the parallel and straight alignment of yarns within the fabric structure ensures minimal fiber bending. As such, the calculation of the square meter gram

weight for yarn in any layer direction of these fabrics is straightforward: simply multiply the gram weight of a one-meter-long yarn segment by the total number of yarns present within that one-meter length. This method accurately reflects the yarn content and density, crucial for assessing fabric quality and suitability in various applications.

III. Introduction to Reinforcement Learning

A. Basic Theory of Reinforcement Learning

Under the background of new power system construction, the participants of micro-grid are becoming more and more diverse, and the power generation output and load are in a state of random fluctuations, making the operating environment and mechanism more and more complex. The traditional energy management and scheduling methods are affected by the dynamics of the system and the intermittence of new energy sources, so it is difficult to establish an accurate mathematical model. Concurrently, it is imperative to estimate and fine-tune numerous parameters, encompassing load forecasting, energy supply, and price forecasting, among others. The computational complexity of these tasks is considerable, often rendering it challenging to fully satisfy the demands of practical scenarios. Furthermore, the majority of optimization challenges encountered in actual production processes are non-deterministic polynomial problems, posing certain difficulties in their resolution. Fig. 2 shows knit fabric quality changes over time.

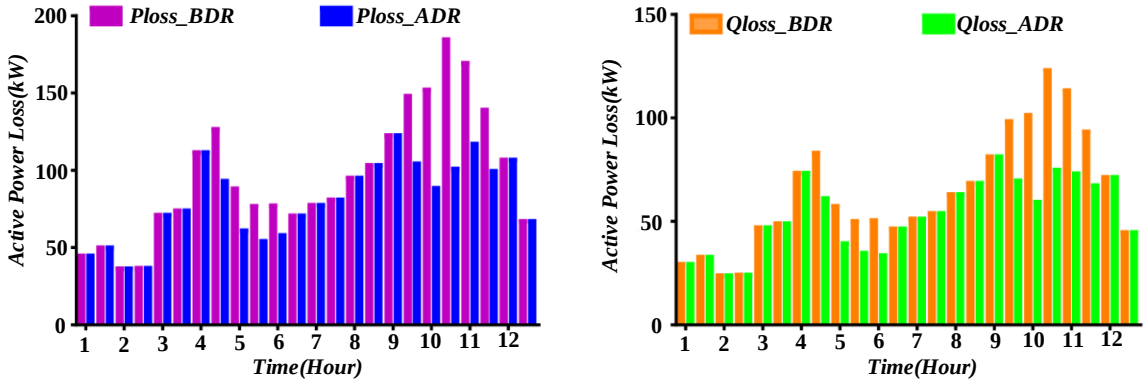


Fig. 2. Knit fabric quality changes over time.

Reinforcement learning evaluates the action based on the reinforcement signals provided by the environment, without determining in advance how the reinforcement learning system will form the correct action. Considering the limited information provided by the external environment, the reinforcement learning system must rely on its own continuous experience to continuously learn [18, 19]. Based on this model, reinforcement learning continuously acquires knowledge in an "action-evaluation" environment, and continuously optimizes action plans to adapt to the environment. The problems faced in the process of reinforcement learning have been widely discussed in biological learning, cybernetics, game theory and other fields. They are used to explain the equilibrium state under the condition of bounded rationality, and are also used to design intelligent interactive systems and unmanned adaptive systems. In recent years, reinforcement learning algorithms that integrate deep learning, transfer learning and other methods have the ability to solve complex problems in the real world, and have reached or surpassed the human level in many fields such as computer games, intelligence competition, automatic driving, intelligent question answering, and industrial production, showing extraordinary application prospects. With the continuous development of algorithms, its application range will become more and more extensive.

B. Technical Features of Reinforcement Learning

Reinforcement learning is known as the three major machine learning technologies together with supervised learning and unsupervised learning because of its powerful exploratory ability and autonomous learning ability. Compared with supervised learning and unsupervised learning, reinforcement learning has great differences in many aspects such as data acquisition, learning methods, and decision-making methods. Supervised learning has a clear label on each data sample, which corresponds to the reinforcement learning task, which means that every action that should be taken in a certain state has a clear label. This does not match the typical scenario of reinforcement learning. Unsupervised learning does not label the data, and the main purpose is to discover the distribution law of the data. Compared with unsupervised learning, reinforcement learning provides certain "labeling" (that is, reward signals), which can be regarded as a kind of weak labeling learning. Although this mark is a weak mark for a specific action, it is very clear for the entire learning task and directly marks the success or failure of the task. The decision-making process of reinforcement learning is continuous, and at each time step, the agent takes action through interaction with the environment and obtains reward signals. Reinforcement learning is a goal-driven active learning method, which generates learning samples through interaction between initiative and environment. In reinforcement learning, agents need to try new strategies to obtain higher rewards, and at the same time need to use known strategies to maximize known rewards. Therefore, how to improve the quality of interaction (exploration and utilization) is a key core of reinforcement learning. Exploration may lead to too many useless attempts, resulting in a large amount of resources being wasted, and too much use may miss better choices due to too much trust in current experience. However, due to the delayed nature of reinforcement learning rewards, agents must learn how to evaluate the long-term impact of current decisions, that is,

actions made at current moments may affect rewards at multiple future moments [20]. Reinforcement learning usually does not require prior knowledge or environment models, which makes reinforcement learning very useful in dealing with unknown environments. Therefore, it has good scalability and adaptability, and can deal with problems such as multi-agent, uncertain and dynamic environments. Fig. 3 shows efficiency comparison before and after knitting path optimization.

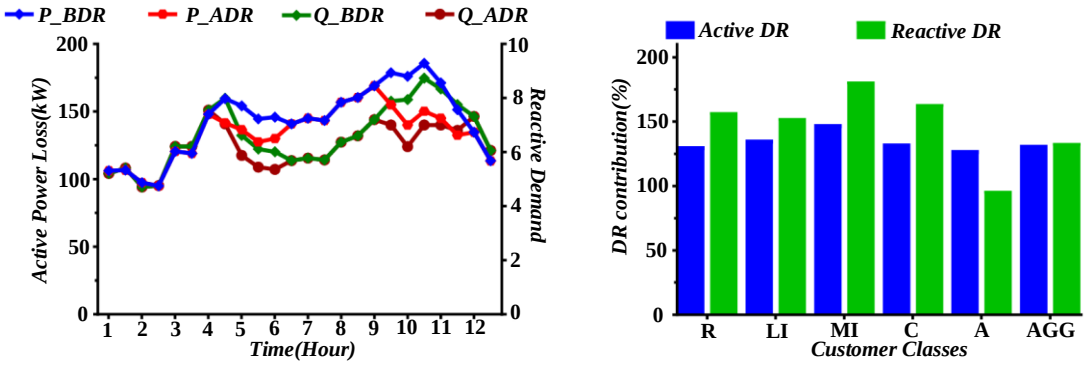


Fig. 3. Efficiency comparison before and after knitting path optimization.

Reinforcement learning embodies a continuous decision-making and strategy optimization process, leveraging intricate internal data structures and algorithms to maximize cumulative rewards through the dynamic interplay between agents and their environment. Initially, the agent engages with the environment guided by a formulated policy, observing and perceiving the current state of the environment after each interaction. Based on this state, the agent selects an action, triggering corresponding rewards or benefits. Over numerous interactions, the agent explores diverse action plans, gradually learning the optimal strategy for executing the most suitable action within a given environment, thereby maximizing overall gain. Presently, reinforcement learning encompasses three core methodologies: the value function algorithm, the policy gradient algorithm, and the "action-evaluation" algorithm, each contributing to the agent's capacity for learning and adaptability.

C. Algorithmic Flow of Reinforcement Learning

Reinforcement learning is widely used in model optimization in automatic control, engineering construction and other fields. Its core is that the agent can obtain the cumulative maximum return or achieve a specific goal through its own learning ability in the process of interacting with the environment, so that the agent has the ability to make the best decision under the current environment. In order to simplify the modeling problem of reinforcement learning, Markov decision process is used to describe and construct the process of reinforcement learning, considering the complexity of the transformation process between environments of reinforcement learning. Fig. 4 shows kit path optimization and model evaluation process.

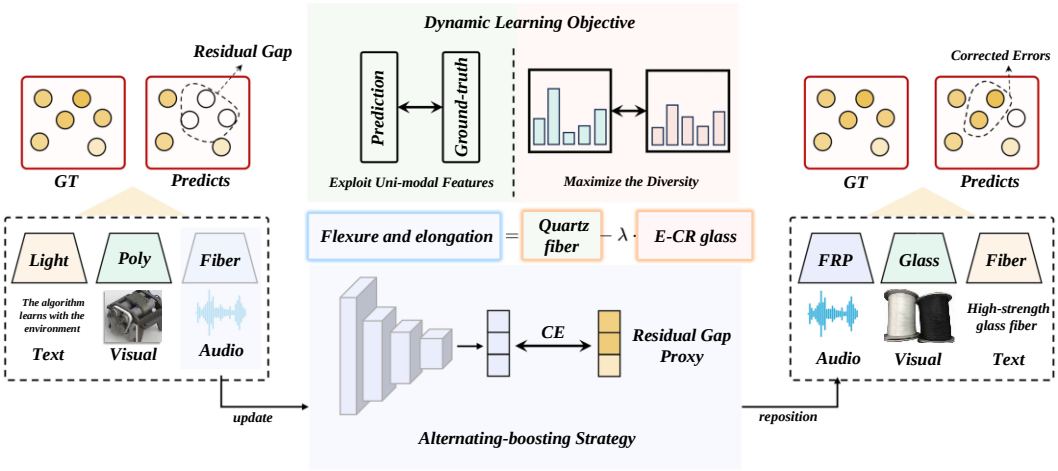


Fig. 4. Kit path optimization and model evaluation process.

Agent: A hypothetical entity that performs actions in an environment for some reward Environment: The scene in which the agent is located.

State: Refers to the current state returned by the environment

The goal of reinforcement learning is to maximize cumulative rewards, and future trends of rewards need to be considered when calculating rewards. Cumulative rewards are defined as the weighted sum of rewards from time t to the end of the learning process, represented as shown in Eq. (1).

$$R_t = \sum_{t'=t}^{T} Y^{t'-t_{rt}} \quad (1)$$

Where yE[0, 1] is a constant called the discount factor, used to assess the impact of future rewards on cumulative rewards. The state action function Q (s, a) represents the execution of action a in the current process and loops to the end of learning according to strategy π. The cumulative gain of the agent is shown in Eq. (2).

$$Q^{\pi}(s,a) = E[R_t \mid s_t = s, a_t = a, \pi] \quad (2)$$

Where s and a represent the current state and action. For all sets of state actions, if the expected returns of a strategy are greater than or equal to the expected returns of other strategies, the strategy is the optimal strategy. In fact, multiple optimal strategies may use the same state action function. Its mathematical expression is shown in Eq. (3).

$$Q^{*}(s,a) = \max_{\pi} E[R_t \mid s_t = s, a_t = a, \pi] \quad (3)$$

Meanwhile, the action function follows the Bellmann optimal equation to form the optimal state action function as shown in Eq. (4)

$$Q^{*}(s,a) = E_{s \sim s}[r + \gamma max Q(s', a') \mid s, a] \quad (4)$$

Where s denotes the subsequent state and a represents the action corresponding to *s*, the optimal value function can theoretically be derived through iterative application of the Bellman equation. However, in practical scenarios, due to the complexity of real-world environments, neural networks, linear functions, and other approximation techniques are frequently employed to estimate state-action value functions. This integration of deep learning and reinforcement learning not only enables the accurate approximation of intricate value functions but also fosters the rapid advancement and widespread adoption of reinforcement learning methodologies.

IV. Study on Tensile Properties of Multiaxial Fabric by Technological Parameters

In the actual production, it is often found that whether the arrangement of weft insertion and knitting technology of multi-axial warp knitted fabrics is reasonable or not will directly affect the structure of weft in the fabric and the progress of the process, such as causing the weft not to be well fixed in the ground weave or causing the fabric surface to lack weft. Because these conditions often affect the tensile properties of fabrics, this chapter mainly studies the effects of different production processes on the tensile properties of multiaxial fabrics from the perspective of weft laying and knitting processes through experimental testing methods [21]. The weft laying process mainly refers to the gram weight of the fabric and the fineness of the weft yarn, and the knitting process mainly refers to the weave structure, needle density, and let-off of the warp knitted yarn (bundled yarn). Because the multi-axial warp knitted fabric is a new material, there is no uniform standard about the performance test of the fabric. In this paper, the fabric tensile test is carried out by GB/T7689.5-2001 standard. The state transition probability formula and the reward function formula are shown in Eq. (5) and Eq. (6).

$$a_k = \frac{\exp(w^{\bullet} \tanh(Uh_k))}{\sum_{j=1}^{K} \exp(w^{\bullet} \tanh(Uh_j))} \quad (5)$$

$$r_k = \frac{\exp(v(x_k)/\tau_v)}{\sum_{k=1}^{K} \exp(v(x_k)/\tau_v)} \quad (6)$$

A. Methods and Conditions for Tensile Testing of Fabrics

Utilize an appropriate instrument to stretch fabric strips until rupture, thereby assessing their breaking strength and elongation at break. Both the breaking strength and elongation values can be directly discerned from the instrument's indicator device, or alternatively, derived from the automatically recorded stress-strain curve. Table I comprehensively outlines the pertinent test parameters employed in this analysis.

TABLE I. Test Parameters

	Unit	Sample parameter
Specimen length	mm	350
Width of specimen (unrimmed)	mm	65
Initial effective length	mm	200
Width of trimmed specimen	mm	50
Tensile speed	mm/min	100

In a humidity-controlled environment adhering to standard conditions of 23°C ± 2°C temperature and 50% ± 10% relative humidity, the sample undergoes a 16-hour humidity acclimation period. Subsequently, the testing is conducted in an identical environmental setting.

Based on the fabric type, adjust the upper and lower fixtures to achieve the desired effective length of the sample between them, ensuring they are parallel. Position the specimen in a jig with its longitudinal central axis aligning with the jig's leading-edge center. Cut cardboard or similar material along a direction perpendicular to the specimen's central axis. Apply a uniform pretension across the entire width of the specimen, and then securely tighten the other jig.

1) Start the movable fixture and tensile the sample until it is destroyed. The Q-value function update and policy gradient theorem formulas are shown in Eq. (7) and Eq. (8).

$$h(t, \overrightarrow{x_i}) = h_0(t)\eta(\overrightarrow{x_i}) \quad (7)$$

$$SA(z) = (\sigma(\frac{qk^T}{\sqrt{K_h}}))v \quad (8)$$

2) Record the final breaking strength. Unless otherwise agreed, when the fabric breaks in more than two stages, such as double-layer or more complex fabrics, the maximum strength at the break of the first set of yarns is recorded and used as the tensile breaking strength of the fabric.

3) Record elongation at break, accurate to 1 mm.

4) If a specimen is broken within 10mm of the contact line of either of the two fixtures, the phenomenon will be recorded, but the breaking strength and breaking elongation will not be

calculated in the results, and the new specimen will be re-tested [22].

The reasons for coating resin on the clamping end are:

The surface of carbon and glass fibers is very smooth, and the direct clamping will cause slippage, which will affect the accuracy of the test. The value function iterations and the Bellmann equation formulas are shown in Eq. (9) and Eq. (10).

$$L = \frac{2\sum_{i}^{N} \hat{p}_i y_i}{\sum_{i}^{N} p_i^{\ 2} + \sum_{i}^{N} y_i^2} \tag{9}$$

$$S(e_i, e_{top-k}) = \frac{\sum_{m=1}^{M} (e_i \times e_{top-k})}{\sqrt{\sum_{m=1}^{M} (e_i)^2} \times \sqrt{\sum_{m=1}^{M} (e_{top-k})^2}} \tag{10}$$

The brittleness of carbon and glass fibers is large, and if the clamping force is too large, the sample at the clamping end will be damaged, making the glass fibers at the clamping end of the sample break first under the strong force, resulting in the phenomenon of breaking the clamping head, and then making the test invalid. After coating resin at both ends, the above problems can be effectively solved. The friction between the resin and the collet is much higher than that between the glass fiber and the collet, which effectively solves the problem of slipping. In addition, after coating the clamping end, the fibers in the clamping part are soaked in the resin. Under the protection of the resin, the glass fiber at the clamping end will be subject to a very small shear force, so that the problem of fiber brittleness can be effectively solved [23].

B. Experimental Data Analysis

In the tensile test, during the tensile load process of the fabric held by the clamp, the yarn in the fabric has obvious different breakage characteristics, which can be accurately judged from the sound produced when the yarn breaks. To some extent, the strength of the yarn in the fabric cannot be fully utilized in the process of application.

Fig. 5 shows reward changes during the iteration of reinforcement learning. The tensile strength of the fabrics with two different weave structures, No. 1 and No. 2, was tested in the direction of yarn 0. Five specimens were tested for each fabric. The action selection strategy and the advantage function calculation formula are shown in Eq. (11) and Eq. (12).

$$\mu_{0,\mathrm{crit}} = \frac{2}{(N+2)\sigma^2} \tag{11}$$

$$w_{k+1} = w_k - \frac{1}{2}\mu\nabla_k \tag{12}$$

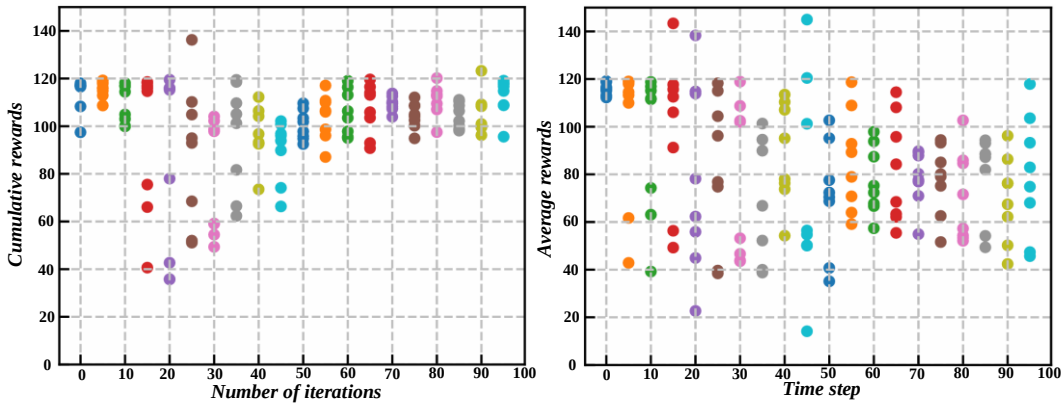


Fig. 5. Reward changes during the iteration of reinforcement learning.

It can be seen from the figure that the tensile properties of the fabrics with Promat as the weave structure are slightly higher than those of the same fabrics with tricot as the weave structure, and the difference in the tensile strength of the fabrics with tricot as the weave structure is small. Fig. 6 shows effect of learning rate on the effect of path optimization.

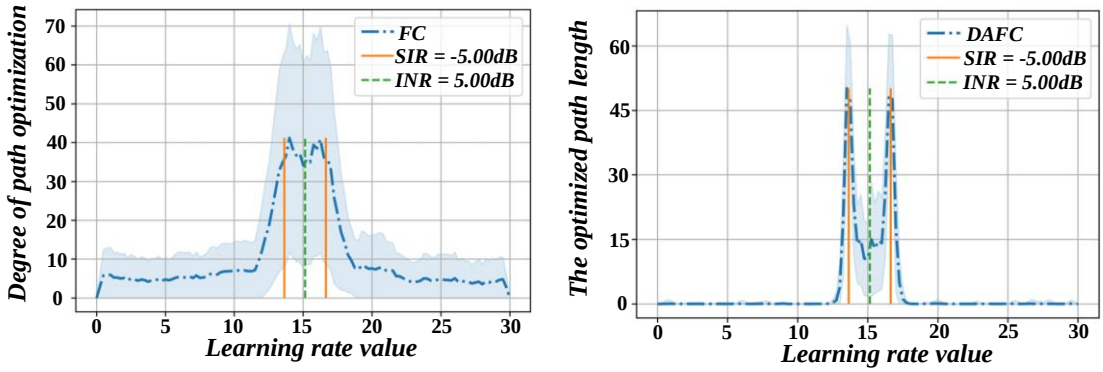


Fig. 6. Effect of learning rate on the effect of path optimization.

As depicted in Fig. 6, the fabric exhibiting the lowest stitch density boasts the highest average tensile strength. However, it is apparent that as the stitch density rises, the subsequent variations in tensile strength compared to this baseline fabric remain relatively modest. The results show that the increase of stitch density in the range of common stitch density increases the probability of fiber damage caused by needle penetrating yarn, but it has no obvious effect on the strength of fabric. Fig. 7 shows exploration-exploit the effect of trade-offs on pathway search.

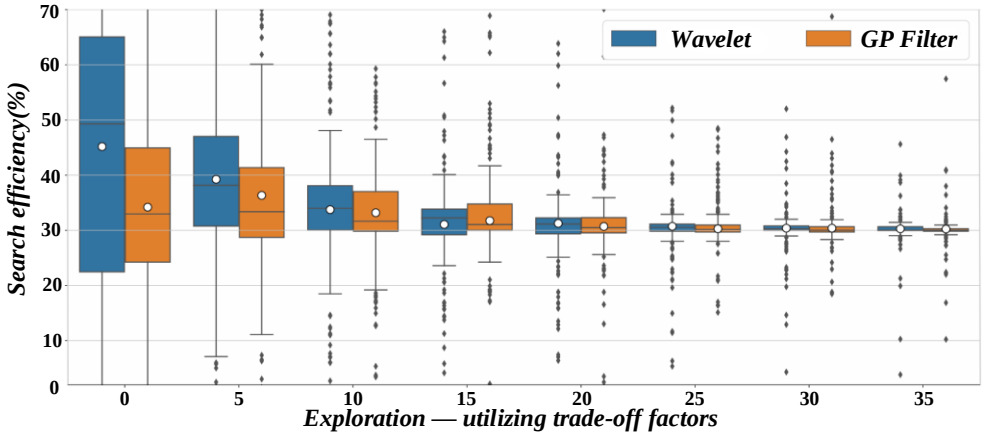


Fig. 7. Exploration-exploit the effect of trade-offs on pathway search.

As depicted in Fig. 7, the tensile strength of the fabric does not exhibit a regular pattern with variations in let-off amount. This is primarily attributed to the consistent weft insertion process and the maintenance of yarn count per unit length. Given the same stitch density, the likelihood of fiber damage within the yarn remains largely unchanged, resulting in insignificant variations in the fabric's ultimate strength. The target network update and knitting model are shown in Eq. (13) and Eq. (14).

$$\bar{n}_e = \frac{1}{a}\int_0^a n_e dr \tag{13}$$

$$x(t+1) = \bar{A}x(t) + \bar{B}u(t) \tag{14}$$

It is imperative to emphasize that warp-knitted yarn commands a premium price, thus, any augmentation in let-off volume directly correlates with an equivalent surge in costs. Conversely, maintaining excessively low let-off values introduces heightened tension within the yarn, which not only accelerates the wear and tear of knitting needles but also poses the risk of yarn breakage, ultimately hindering overall production efficiency. Therefore, for the actual production, we should choose the appropriate let-off amount [24]. The gradient descent optimization formula and the entropy regularization term formula are shown in Eq. (15) and Eq. (16).

$$(F * K)(q) = \sum_{s+t=q} F(s)K(t) \tag{15}$$

$$X_t^A = \frac{1}{C}\sum_{c=1}^{C} X_t(c) \tag{16}$$

C. Effect of Gram Weight of Fabric on Tensile Properties of Fabric

By increasing the gram weight of the 0° yarn layer, we measured the tensile strength of Fabrics No. 2 and No. 4 in the 0° direction. Specifically, Fabric No. 2 had a 0° yarn layer gram weight of 291.4 g/m2, while Fabric No. 4 boasted a gram weight of 582.7 g/m2. Notably, the gram weight of the 45° glass yarn layer remained unaltered, and all other process parameters were kept consistent. Fig. 8 shows optimize the comparison of before and after paths.

As Fig. 8 illustrates, as the fabric's gram weight increased, so did its breaking strength in the tensile direction. This phenomenon is primarily attributed to the increased density of the fabric, specifically the augmentation in the density of the 0° yarn. As the density of 0° yarns rises, the count of 0° yarns per unit length correspondingly increases. Consequently, when the fabric undergoes stress, the yarn's force-bearing capacity per unit area intensifies, ultimately contributing to a significant enhancement in the fabric's overall strength.

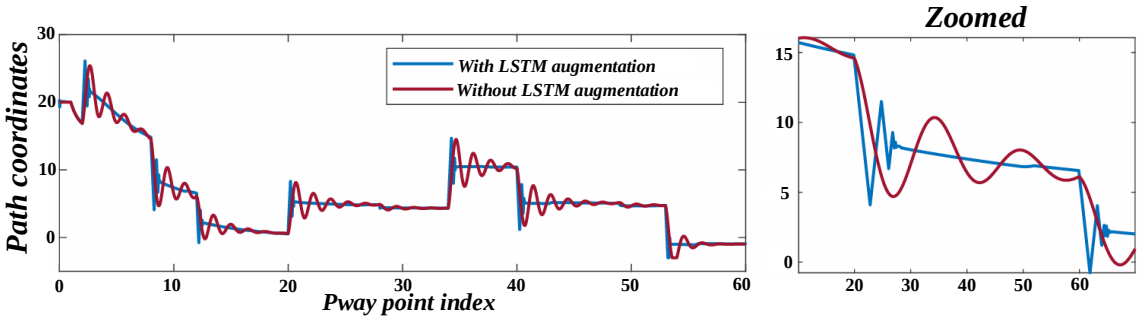


Fig. 8. Optimize the comparison of before and after paths.

By increasing the gram weight of the 45° glass fiber yarn layer, we measured the tensile strength of Fabrics No. 3 and No. 4 in the 0° direction. Specifically, Fabric No. 3 had a 45° yarn layer gram weight of 601.2 g/m^2, while Fabric No. 4's 0° yarn layer gram weight stood at 300.6 g/m^2. The gram weight of the 0° carbon fiber layer remained constant, and all other process parameters remained unchanged [25]. The importance sampling weights and discount factor influence formulas are shown in Eq. (17) and Eq. (18).

$$T_{N\varepsilon} = N\tau_Q + \sum_{i=1}^{N} \varepsilon_{c,i} \tag{17}$$

$$\sigma_t^2 = \frac{1}{2N-2} \sum_{i,j \neq i} \sigma_{ij}^2 \tag{18}$$

Fig. 9 shows effect of environmental state changes on path planning. As evident in Fig. 9, despite an increase in fabric gram weight, the breaking strength in the tensile direction experienced a marginal increase. The primary reason for this is that the 45° glass yarn does not significantly contribute to the tensile strength in the 0° direction. However, it does play a role in enhancing the fabric's pressure resistance and shares the load, thereby contributing to an increase in fabric strength, albeit not significantly.

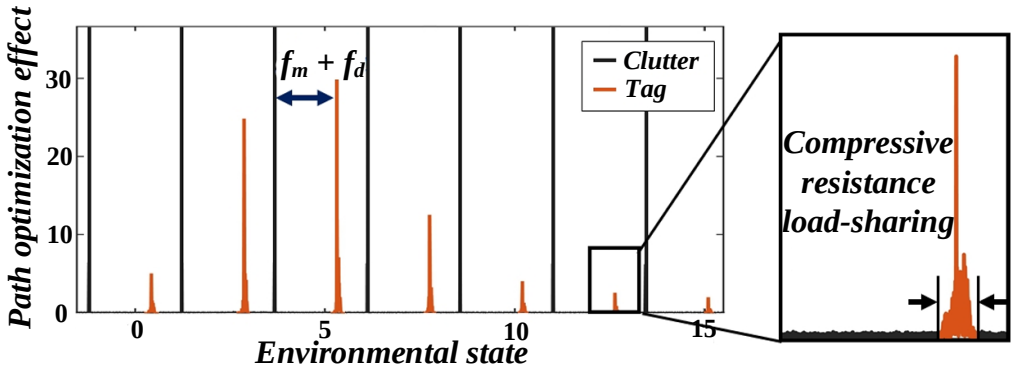


Fig. 9. Effect of environmental state changes on path planning.

At the same time, it can also be observed that the tensile strength of No. 3 fabric fluctuates very little while that of No. 4 fabric fluctuates greatly. The reason is that the weft density of No. 4 fabric is too small (4.5 ends/inch), resulting in uneven weft laying and lack of weft. If there is uneven weft laying in the fabric, if the sample is taken in the area of lack of weft during the experiment, the strength value obtained from the test should be low . If the sample is taken in the place where the weft is densely laid, the strength value obtained from the test is very high. As a result, the performance of different parts of the same fabric varies greatly. The multistep return estimation and model prediction error are calculated as in Eq. (19) and Eq. (20).

$$\mathrm{L}(q) = \ln \frac{p(x,\gamma,\lambda \mid y)}{q(x,\gamma,\lambda)}_{q(x,\gamma,\lambda)} \tag{19}$$

$$\hat{x}_i^H B_i \hat{x}_i = \sum_{l=1}^{d_l} \frac{|\, q_{i,l} \,|^2}{(1+\hat{\gamma}_i s_{i,l})^2} \tag{20}$$

With the fabric's weight kept constant, we varied the density of the 45° glass yarn and conducted tensile strength tests on Fabrics No. 4 (45° glass yarn density of 4.5 pieces/inch) and No. 5 (±45° glass yarn density of 2.25 pieces/inch).

Moreover, it becomes apparent that Fabric No. 5 demonstrates pronounced strength variations, primarily stemming from its exceptionally low weft laying density and irregular yarn positioning. This underscores the notion that, under consistent weight conditions, an increase in weft density fosters a direct enhancement in the fabric's tensile properties, as evidenced by prior studies.

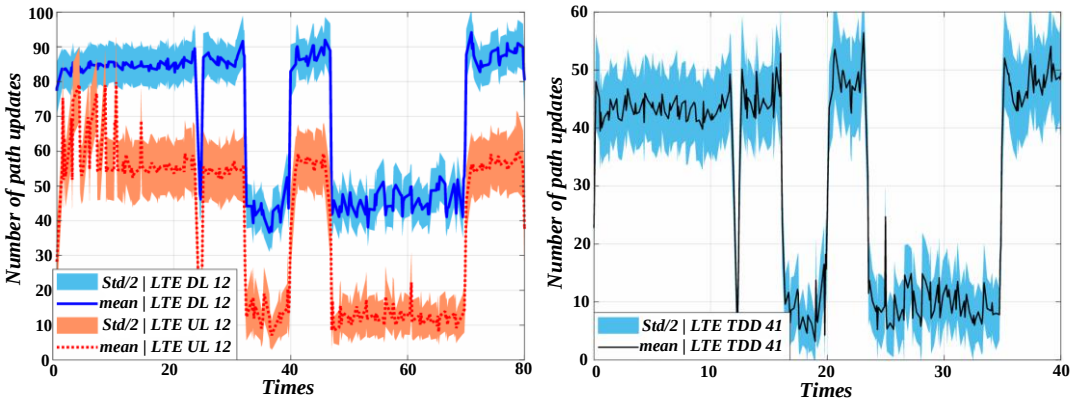


Fig. 10. Frequency of path updates during real-time learning.

However, it is worth noting that as the weft laying density increases, the linear density of the weft yarn decreases. Given that the current market prices of glass and carbon fibers rise as their linear density diminishes, it is imperative to comprehensively consider various factors, including fabric requirements, while setting the weft laying process to determine the optimal weft laying density. Fig. 10 shows frequency of path updates during real-time learning.

V. Summary and Prospect

With textile technology's advancement, flat knitting machines are crucial in the industry, with knitting efficiency and product quality as key performance metrics. Recently, reinforcement learning, an advanced ML technique, has been widely applied to optimization problems. This paper delves into utilizing Reinforcement Learning as a means to optimize knitting paths for flat knitting fabrics, with the ultimate goal of elevating both production efficiency and quality. The selection of knitting paths holds paramount importance, as it directly influences the aesthetic appeal, tactile sensation, and overall efficiency of the end product. Traditional optimization techniques, reliant on manual expertise and a cumbersome trial-and-error process, prove inefficient and inadequate for addressing intricate requirements. In contrast, RL facilitates the autonomous discovery of optimal paths through the dynamic interplay between the agent and its environment, thereby significantly advancing knitting efficiency and quality. We introduce an RL model specifically designed for optimizing knitting paths on flat knitting machines, meticulously defining state, action, and reward functions to capture the intricate nuances of the knitting process. By employing a reinforcement learning algorithm, our agent learns and explores within a simulated environment, progressively uncovering the optimal weaving path. Through a large number of experimental verifications, we prove that the knitting path optimization method based on reinforcement learning can significantly improve the knitting efficiency and product quality. In addition, the application effect of different reinforcement learning algorithms in knitting path optimization of flat knitting machine is discussed, and the key factors affecting the optimization effect are analyzed. We find that choosing appropriate algorithm parameters and reward functions is crucial to improve the optimization effect. Furthermore, we acknowledge the limitations of our current research and propose future directions for exploration. In summary, the reinforcement learning-based knitting path optimization for flat knitted fabrics holds immense potential. With continued research, we aim to further enhance knitting efficiency and product quality of flat knitting machines, ultimately contributing significantly to the textile industry's advancement.

References

[1] Amitai, Y., Amir, O., & Avni, G. ASQ-IT: Interactive explanations for reinforcement-learning agents. Artificial Intelligence, vol. 335, pp. 104182, 2024.

[2] Consigli, G., Gomez, A. A., & Zubelli, J. P. Optimal dynamic fixed-mix portfolios based on reinforcement learning with second order stochastic dominance. Engineering Applications of Artificial Intelligence, vol. 133, pp. 108599, 2024.

[3] Cotogni, M., & Cusano, C. Select & Enhance: Masked-based image enhancement through tree-search theory and deep reinforcement learning. Pattern Recognition Letters, vol. 183, pp. 172–178, 2024.

[4] Darabi, B., Bag-Mohammadi, M., & Karami, M. A micro Reinforcement Learning architecture for Intrusion Detection Systems. Pattern Recognition Letters, vol. 185, pp. 81–86, 2024.

[5] Gong, A., Yang, K., Lyu, J., & Li, X. A two-stage reinforcement learning-based approach for multi-entity task allocation. Engineering Applications of Artificial Intelligence, vol. 136, pp. 108906, 2024.

[6] Guangwen, T., Mengshan, L., Biyu, H., Jihong, Z., & Lixin, G. Achieving accurate trajectory predicting and tracking for autonomous vehicles via reinforcement learning-assisted control approaches. Engineering Applications of Artificial Intelligence, vol. 135, pp. 108773, 2024.

[7] Guo, B., Chang, X., Chao, F., Zheng, X., Lin, C.-M., Chen, Y., Shang, C., & Shen, Q. ARLP: Automatic multi-agent transformer reinforcement learning pruner for one-shot neural network pruning. Knowledge-Based Systems, vol. 300, pp. 112122, 2024.

[8] Han, C., & Wang, X. TPN:Triple network algorithm for deep reinforcement learning. Neurocomputing, vol. 591, pp.127755, 2024.

[9] Hu, K., Li, M., Song, Z., Xu, K., Xia, Q., Sun, N., Zhou, P., & Xia, M. A review of research on reinforcement learning algorithms for multi-agents. Neurocomputing, vol. 599, pp. 128068, 2024.

[10] Kang, M., Templeton, G. F., Kwak, D.-H., & Um, S. Development of an AI framework using neural process continuous reinforcement learning to optimize highly volatile financial portfolios. Knowledge-Based Systems, vol. 300, pp. 112017, 2024.

[11] Li, X., Ji, W., & Huang, J. Local instance-based transfer learning for reinforcement learning. Engineering Applications of Artificial Intelligence, vol. 133, pp. 108488, 2024.

[12] Löppenberg, M., Yuwono, S., Diprasetya, M. R., & Schwung, A. Dynamic robot routing optimization: State–space decomposition for operations research-informed reinforcement learning. Robotics and Computer-Integrated Manufacturing, vol. 90, pp. 102812, 2024.

[13] Ma, X., Zhong, Z., Li, Y., Li, D., & Qiao, Y. A novel reinforcement learning based Heap-based optimizer. Knowledge-Based Systems, vol. 296, pp. 111907, 2024.

[14] R, D. J. B., Medina, M. C. C., Fernandes, B. J. T., & Barros, P. V. A. The use of reinforcement learning algorithms in object tracking: A systematic literature review. Neurocomputing, vol. 596, pp. 127954, 2024.

[15] Razzaghi, P., Tabrizian, A., Guo, W., Chen, S., Taye, A., Thompson, E., Bregeon, A., Baheri, A., & Wei, P. A survey on reinforcement learning in aviation applications. Engineering Applications of Artificial Intelligence, vol. 136, pp. 108911, 2024.

[16] Sun, S., Li, T., Chen, X., Dong, H., & Wang, X. Cooperative defense of autonomous surface vessels with quantity disadvantage using behavior cloning and deep reinforcement learning. Applied Soft Computing, vol. 164, pp. 111968, 2024.

[17] Tang, Y., Sun, J., Wang, H., Deng, J., Tong, L., & Xu, W. A method of network attack-defense game and collaborative defense decision-making based on hierarchical multi-agent reinforcement learning. Computers & Security, vol. 142, pp. 103871, 2024.

[18] Wang, H., Miahi, E., White, M., Machado, M. C., Abbas, Z., Kumaraswamy, R., Liu, V., & White, A. Investigating the properties of neural network representations in reinforcement learning. Artificial Intelligence, vol. 330, pp. 104100, 2024.

[19] Wang, J.-Q., Guo, L., Jiang, Y., Zhang, S., & Zhou, Q. Improving unbalanced image classification through fine-tuning method of reinforcement learning. Applied Soft Computing, vol. 163, pp. 111841, 2024.

[20] Wang, Y., Hong, X., Wang, Y., Zhao, J., Sun, G., & Qin, B. Token-based deep reinforcement learning for Heterogeneous VRP with Service Time Constraints. Knowledge-Based Systems, vol. 300, pp. 112173, 2024.

[21] Yerramreddy, D. R., Marasani, J., Ponnuru, S. V. G., Min, D., & S, D. Harnessing deep reinforcement learning algorithms for image categorization: A multi algorithm approach. Engineering Applications of Artificial Intelligence, vol. 136, pp. 108925, 2024.

[22] Zeng, H., Wei, B., & Liu, J. RTRL: Relation-aware Transformer with Reinforcement Learning for Deep Question Generation. Knowledge-Based Systems, vol. 300, pp. 112120, 2024.

[23] Zhao, Z., Zhang, Y., Wang, S., Zhang, F., Zhang, M., & Chen, W. QDAP: Downsizing adaptive policy for cooperative multi-agent reinforcement learning. Knowledge-Based Systems, vol. 294, pp. 111719, 2024.

[24] Miyazaki, K., & Miyazaki, H. Suppression of negative tweets using reinforcement learning systems. Cognitive Systems Research, vol. 84, pp. 101207, 2024.

[25] Tang, Y., Guo, S., Liu, J., Wan, B., An, L., & Liu, J. K. Hierarchical reinforcement learning from imperfect demonstrations through reachable coverage-based subgoal filtering. Knowledge-Based Systems, vol. 294, pp. 111736, 2024.

Design and Research of Cross-Border E-Commerce Short Video Recommendation System Based on Multi-Modal Fusion Transformer Model

Yiran Hu*
School of Finance and Trade Management, Chengdu Industry and Trade College, Chengdu 611731, China

***Abstract*—This study designed a cross-border e-commerce short video recommendation system based on Transformer's multimodal analysis model. When mining associations, the model not only focuses on the relationships between modalities, but also improves semantic context by addressing contextual correlations within and between modalities. At the same time, the model uses a cross modal multi head attention mechanism for multi-level association mining, and constructs an association network interwoven with latitude and longitude. In the process of exploring the essential correlation between patterns and subjective emotional fluctuations, the potential context between patterns has been realized. Fully explore correlations and then more accurately identify the truth contained in the original data. In addition, this study proposes a self supervised single modal label generation method. When multimodal labels are known, it does not require complex deep networks and only relies on the mapping relationship between multimodal representations and labels to generate a single modal label. Modal labeling can achieve phased automatic labeling of single modal labels, and quantify the mapping relationship between modal representations and labels from the representation space to generate weak single modal labels. The study also achieved multimodal collaborative learning in the context of limited differential information acquisition due to incomplete labeling, fully utilizing multimodal information. The experimental results on classic datasets in the field of multimodal analysis show that it outperforms the baseline model in terms of accuracy and F1 score, reaching 98.76% and 97.89%, respectively.**

***Keywords*—Multimodal fusion; transformer model; cross-border e-commerce; short video recommendation system**

I. Introduction

With the advent of the era of big data, new social media such as DouYin, Weibo and YouTube will update a large amount of data content every day, in which there are not only objective descriptions of a certain thing, but also a large number of subjective expressions [1, 2]. Mining and identifying the information contained in these data can not only provide information assistance for big data forecasting applications such as financial market trend forecasting, product marketing status forecasting, and even US political election forecasting, but also provide information decision-making such as network public opinion analysis and digital social governance [3]. Providing technical support has extremely important application value and practical significance.

According to the existing research situation, traditional text analysis only uses words, phrases and their semantic associations to judge, which is not enough to identify complex information. Multimodal analysis adds acoustic and visual information on the basis of text information, and with the help of the association between multimodal data, it can show the information that may be hidden in text data, so as to achieve more accurate recognition [4, 5]. Taking ironic emotion recognition as an example, by extracting acoustic and visual information from human intonation and body movements, ironic information can be accurately recognized. Multimodal analysis has achieved remarkable results in dealing with understanding in various scenarios, and has attracted more and more researchers' attention [6, 7]. However, there are still some challenges in the research of multimodal analysis-multimodal association mining and multimodal collaborative learning.

In response to these challenges, this paper considers the association information between modes and contexts in the process of multimodal analysis based on deep learning technology. It uses the improved Transformer framework to mine intertwined and intricate associations to achieve tight coupling of multimodal data. Focusing on the mapping relationship between multimodal representations and sample labels, multimodal collaborative learning under unbalanced information distribution is realized with the help of a multi-task learning framework. Multimodal fusion is properly sorting and tightly coupling data from two or more modes. The most significant difference between multimodal analysis and traditional single-modal analysis is that the former can obtain more reliable prediction results with the help of information gained by multi-source data. According to the different stages, the existing multimodal fusion methods can be divided into three categories: feature-level, decision-level, and hybrid-level. Multimodal analysis based on Transformer and multi-task learning has essential application significance. However, its research results can also provide a basis and support for cross-media perceptual computing, analytical reasoning, and multimodal deep learning research in artificial intelligence. Has important research significance.

II. Multimodal Analysis based on Transformer

A. Transformer for Linguistics Guidance

The traditional multi-head attention mechanism is mostly applied to machine translation problems. When calculating the attention score, the operation can be performed parallelly to accelerate the training of the model [8, 9]. This paper applies this idea to the multi-modal problem, hoping to find the mapping relationship between multiple modes. Specifically, when using the attention mechanism to learn a mode, the text mode is used

*Corresponding Author.

as a guide to mine the association between various modes, and finally a linguistic-guided Transformer (LGT) is constructed. LGT includes Multi-Head Attention (MHA) and Forward Neural Networks (FNN) [10].

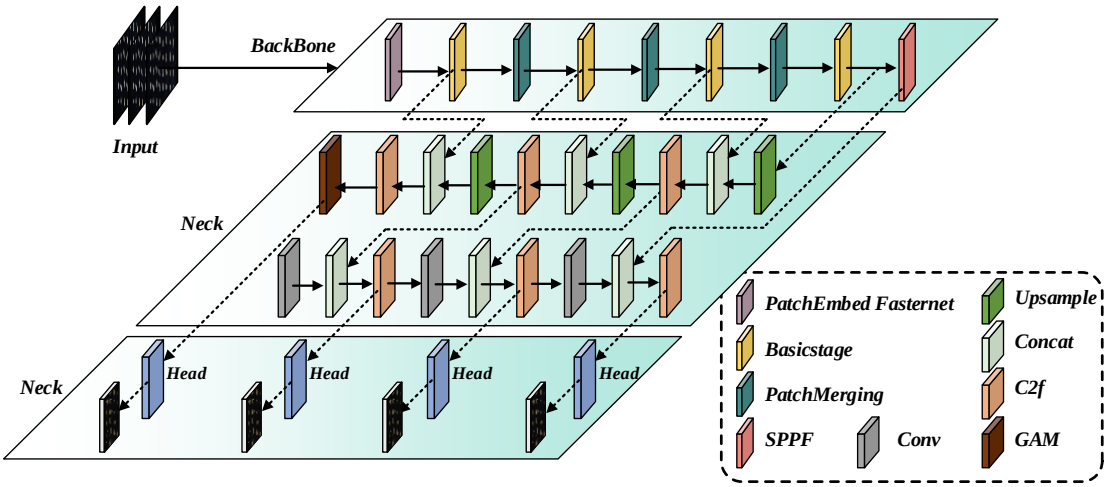


Fig. 1. Transformer model based on multimodal.

As shown in Fig. 1, the model takes text mode as the main component, speech mode, and image mode as the secondary components, and the characteristics of the three modes are respectively input into the multi-head attention module. Through this setting, the text data containing rich information can guide the voice and image data, which can be used to mine multi-modal association information. The process of calculating the text feature attention score with this module is as follows:

Firstly, the text features are divided into three vectors: query vector Ql, keyword vector Kl and true value vector Vl, and all the vectors are linearly transformed; Then, Ql and Kl are sent to calculate the attention score, and the dimension Kl is used to limit the calculation result to ensure that the inner product is not too large; Finally, the final calculation result is obtained by weighted summation of attention score and Vl. Specifically, as shown in Formula (1).

$$Attention(Q_l, K_l, V_l) = \text{softmax}((Q_l K_l^T)/\sqrt{d_k})V_l \quad (1)$$

The above calculation process is performed multiple times, and each calculation is regarded as a head [11, 12]. By splicing the results of multiple heads, the final multi-head attention calculation result can be obtained, as shown in Formulas (2) and (3).

$$head_i = Attention(Q_l W^Q, K_l W^K, V_l W^V) \quad (2)$$

$$F_{(l)} = MHA((Q_l, K_l, V_l)) = Concat(head_1, ..., head_h)W^O \quad (3)$$

After getting the calculation result of attention, it is passed into FFN to mine the nonlinear relationship of features, so as to enhance the performance ability of features, as shown in Formula (4).

$$FFN = Relu(H'W^1 + b^1)W^2 + b^2 \quad (4)$$

Each layer in the LGT needs to be processed, as shown in Formula (5).

$$F(x) = LayerNorm(x + Sublayer(x)) \quad (5)$$

For minor components such as speech features and image features, the query vector comes from the text mode, and the keyword vector and the true value vector come from the speech and image modes when calculating the multi-head attention [13, 14]. When processing speech and image features, text features are used to introduce information from different representation spaces, as shown in Formulas (6) and (7).

$$F_{(a)} = MHA(Q_l, K_a, V_a) = Concat(head_1, ..., head_h)W^O \quad (6)$$

$$F_{(v)} = MHA(Q_l, K_v, V_v) = Concat(head_1, ..., head_h)W^O \quad (7)$$

B. Soft Mapping Module

The model has learned the interaction information between the modes and needs to project the learned results of each mode into a new performance space in the soft mapping module for fusion before classification [15]. Precisely, the results output by the forward propagation network are first mapped to a higher-dimensional space, as shown in Formula (8).

$$NewMatrix = W_m M \quad (8)$$

Then the soft attention is calculated for each matrix in the high-dimensional space, and then the weighted sum of the results is integrated into the vector to obtain the calculation result of soft attention [16, 17]. This calculation process is shown in Formulas (9) and (10).

$$p_i = softmax((v_i^p)^T (NewMatrix)) \quad (9)$$

$$SoftAttention_i(M) = m_i = \sum_{j=0}^{N} (p_{ij} M_j) \quad (10)$$

Finally, after stacking these results, you can get the results of Soft Mapping, as shown in Formula (11).

$$s = Stacking(\sum_{j=0}^{N}(m_j)) \quad (11)$$

Note that a residual calculation and Layer Normalization are performed at the end of this process to ensure that the next round of input includes the results of the previous round, as shown in Formula (12).

$$M = LayerNorm(M + s) \quad (12)$$

The result s obtained above is the result after processing the respective output matrix M of each mode, and the vectors obtained by each mode are summed in order of elements, and the summed results are classified and predicted according to Formula (13).

$$y \sim p = W_p(LayerNorm(s_l + s_a + s_v)) \quad (13)$$

C. Construction of Recommendation Model Fusing Interaction of Bert and High-order Dominant Features

1) The overall structure of model fusing Bert interaction with high-order dominant features: In this paper, according to the actual situation of video recommendation, the Bert model is integrated into the X Deep FM framework. This model can extract text feature vectors with deeper semantics through the Bert model and obtain their text feature vectors by extracting text information such as video titles and tags Because categorical discrete features such as user ID, video ID, and related attributes are difficult to directly use as inputs to deep learning models [18, 19]. Therefore, Label Encoder is used to convert into categorical codes, discontinuous values or texts are converted into categorical codes, and then Embedding is used to convert them into low-dimensional, dense feature vectors, which are input into the model. Finally, the input feature vector and the user's preference degree value are used to iteratively update the training model to improve accuracy and reliability [20]. The network structure of its overall model is shown in Fig. 2.

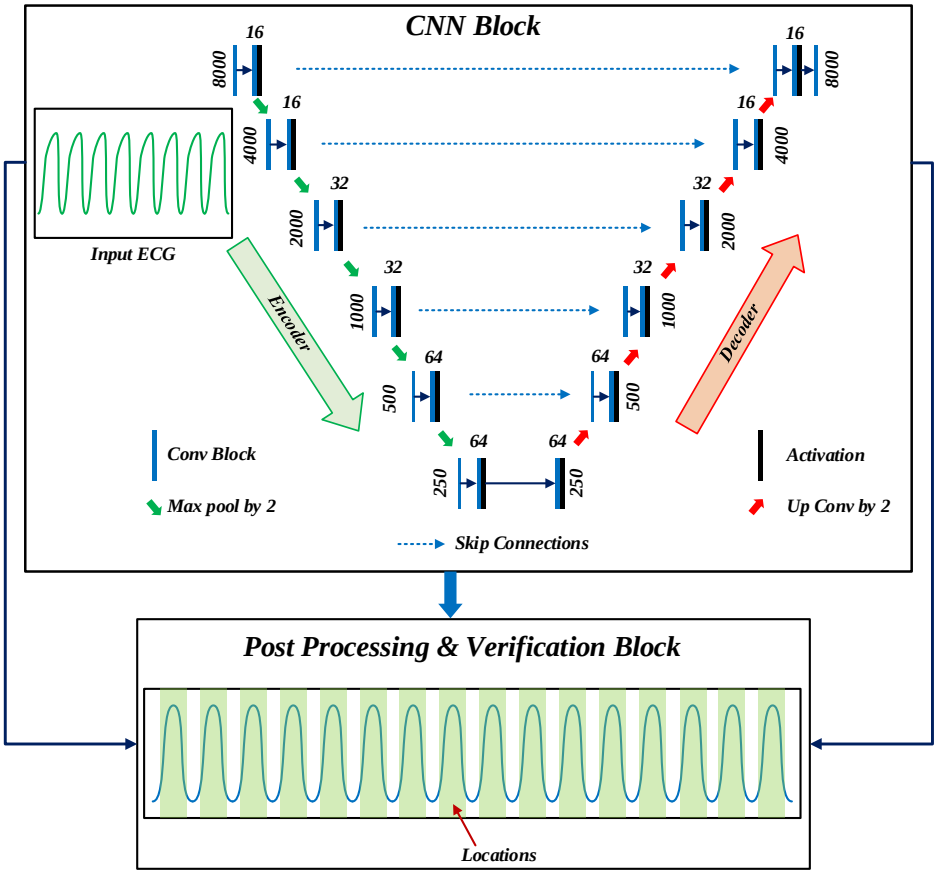


Fig. 2. Fusion model of Bert interaction with high-order explicit features.

The Fig. 2 shows a model that combines Bert and high-order explicit features. It is an end-to-end recommendation model, which has the characteristics of both low-order and high-order feature interactions and implicit and explicit features. The overall structure of the network is composed of four parts: input and text feature extraction, compressed interactive network part, multi-layer neural network and score prediction (that is, output layer), in which the compressed interactive network can extract high-order explicit cross features of the input layer, and the multi-layer neural network can extract high-order invisible cross features [21, 22].

2) Extraction of text information features by Bert: Bert is a bidirectional Transformer-based encoder, which is a bidirectional model obtained by unsupervised training on large-scale corpus. Compared with the GPT model, the Bert model

uses the Encoder structure in Transformer as the main component of the model. The Bert model reads the text, constructs special characters in the text, completes the training through the multi-layer Transformer Encoder [23, 24] structure, and finally uses the vector corresponding to the special characters as the output of the Bert model. In the training task of the language model, the special characters are constructed in the form of filling. By randomly setting [MASK] on the input, let the model predict the words at this position to complete the training of the model. In actual use, it splices special characters [CLS] on the text, so that the output of special characters is matched with the target task to achieve. At present, the Bert model has achieved remarkable results in extractive tasks (SQuAD), sequence labeling tasks (named entity recognition), and classification tasks (SWAG) and other tasks.

3) Input layer: The input layer is responsible for transforming users and characteristics into the form required by the model so that input information can be better understood and processed. In the process of processing features, when using categorical discrete features, first use Label Encoder to encode, the value is between 0 and n-1, so that this feature can be recognized by the model. For numerical continuous features (such as playback volume, user level, etc.), they can be directly entered into the model as input features for calculation [25]. For text-like features, we use the previous Bert model to extract sentence vectors. Since the dimension of each sentence vector is 768 dimensions, all features are spliced together to form the final input vector.

The combination of the sub-types and continuous numerical types processed by the above three methods will cause problems such as dimension explosion and excessive resource occupation, and it is not very good for neural networks to deal with this input. In this paper, the Embedding layer is used to deal with subtyping and continuous numeric types, so as to solve the problems of dimension explosion and excessive resource occupation [26, 27]. By this method, the original sparse matrix is transformed into a dense continuous vector with suitable length, so that the neural network can better handle this input. Although the initial feature length of the sample data may vary, Embedding can still effectively improve this situation, thereby increasing the accuracy and reliability of the model. After the feature embedding layer processing, its length will remain unchanged and will not be affected by the outside world. After this process, follow-up deep learning operations are carried out. The network structure essentially forms a weight matrix. According to a certain mapping relationship, the weight information of the original matrix is transformed into a new dimension matrix through matrix multiplication calculation. According to the reverse mapping relationship, the matrix is multiplied, and the original matrix will be restored matrix. The application of Embedding layer can effectively reduce the sparsity of data, and can change the original isolated vectors into closely related vectors, which can greatly enhance the scalability of the algorithm.

4) DNN layer: To deeply explore the feature interaction relationships implicit in the information, deep neural networks (DNNs) are used for learning. DNN is developed from the multi-layer perceptron (MLP) technology, which has deeper network layers and more types of activation functions. It can connect multiple hidden layers of nonlinear structures, fitting complex function curves and mining deeper interaction features through large-scale training data [28]. The deep neural network performs excellently, can deal with complex problems, and has remarkable effects. Its powerful ability is mainly due to the large number of neural network layers; that is to say, the more network layers, the more complex and in-depth the neural network, and the more learning. Accurate. The basic structure of the neural network comprises three parts: the input layer, the hidden layer, and the output layer. The connection mode between layers is a complete connection, and there is at least one hidden layer. The more hidden layers there are, the higher the expressive ability of the model. The relationship between layers of the deep neural network is nonlinear, and the task of the lower network layer is to extract low-order edge features with relatively simple relationships from the original input data. Each neuron in the bottom network layer acquires some low-order information. More advanced local features can be obtained by combining the underlying information on the middle-hidden layer. The top layer fuses local features into higher-level features. However, it is impossible to theoretically understand the crossover characteristics of each DNN neural network layer and the characteristics each neuron represents. After the training, which feature interactions are more effective in the entire neural network cannot be explained, so these unexplained high-order feature interactions are considered implicit feature interactions. However, the experimental results confirm that DNN can unearth unintelligible but effective high-order feature interactions, which are called implicit feature interactions, and the scattered results are shown in Fig. 3.

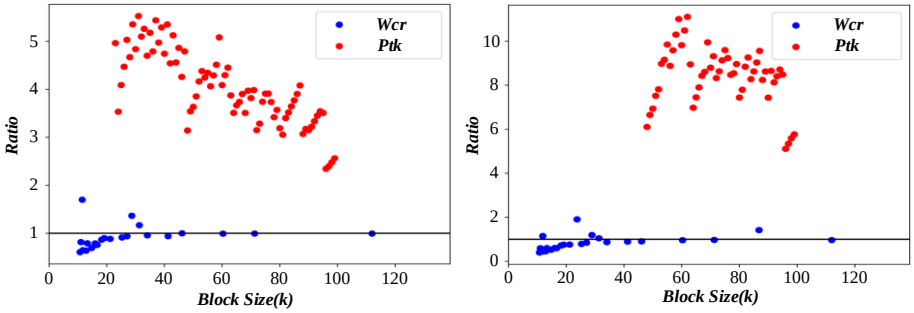


Fig. 3. Interactive dispersion results of implicit features.

III. Design of Cross-Border E-Commerce Short Video Recommendation System based on Multi-task Learning

A. Multi-level Association Mining Framework

In exploring low-level correlations, this study adopts a model-agnostic approach to fuse features, aiming to uncover the intricate feature associations spanning across multi-modal representations [29]. Drawing inspiration from tensor fusion networks, we introduce the use of Unfolded Fusion Functions (UFF) to address the limitations of traditional fusion techniques, such as multimodal splicing. By leveraging UFF, we elevate single-modal features into higher-dimensional spaces, facilitating their fusion. This approach employs a 3-fold Cartesian product to seamlessly integrate multiple single-modal representations, capturing both bimodal and trimodal interactions through a multi-level fusion process.

$$\{(T^l, T^a, T^v)/T^l \in [{}_1^{T^l}], T^a \in [{}_1^{T^a}], T^v \in [{}_1^{T^v}]\} \tag{14}$$

$$F_{(m)} = [{}_1^{T'}] \otimes [{}_1^{T^a}] \otimes [{}_1^{T^v}] \tag{15}$$

The precise computational methodology is outlined in Formulas (14) and (15), offering a nuanced and robust framework for analyzing and utilizing multi-modal data.

B. Multi-task Learning Framework

In this section, we introduce a multi-task learning framework that is designed to tackle diverse analysis tasks through the employment of a rigorous hard parameter sharing mechanism. This mechanism enables all tasks to synergistically harness neurons and weights in the foundational low-level network, while reserving task-specific neurons and weights for each individual task in the higher-level network. The framework adopts a two-tiered architecture, with the underlying representation learning network serving as a common layer and the prediction network tailored to meet the distinct demands of each task.

$$F_s^* = ReLU(F_s W_s^{1T} + b_s^1) \tag{16}$$

$$y_s = F_s^* W_s^{2T} + b_s^2 \tag{17}$$

Within this multi-task learning paradigm, four distinct tasks are formulated, and the specific configuration of the task-oriented layers is outlined in Formulas (16) and (17). Notably, the single-modal task is trained utilizing labels generated by the SLGM methodology, limiting its existence to the training phase. Ultimately, the model relies on the predicted outcomes of the multi-modal task as the definitive output, reflecting its emphasis on the integration of multimodal information. This approach offers a comprehensive and efficient solution for multi-task learning, promoting knowledge sharing and task specialization within a unified framework.

C. Self-Supervised Label Generator

Most multimodal analysis datasets need more independent single-modal labels, posing a challenge for multi-task learning frameworks. To address this limitation, Fig. 4 illustrates the outcomes of a self-supervised label generation module tailored to diverse modalities [30]. Consequently, this section introduces the Self-Supervised Label Generation Module (SLGM), whose primary objective is to derive single-modal annotations from multimodal annotations. The conceptual foundation of SLGM is grounded in two potential mapping relationships: (1) a direct correlation between modal representations and their corresponding modal supervision values and (2) a proportionality in the mapping relationships among different modalities. By harnessing these insights, SLGM aims to bridge the gap between multimodal annotations and the desired single-modal labels, enabling a more comprehensive and practical multi-task learning framework.

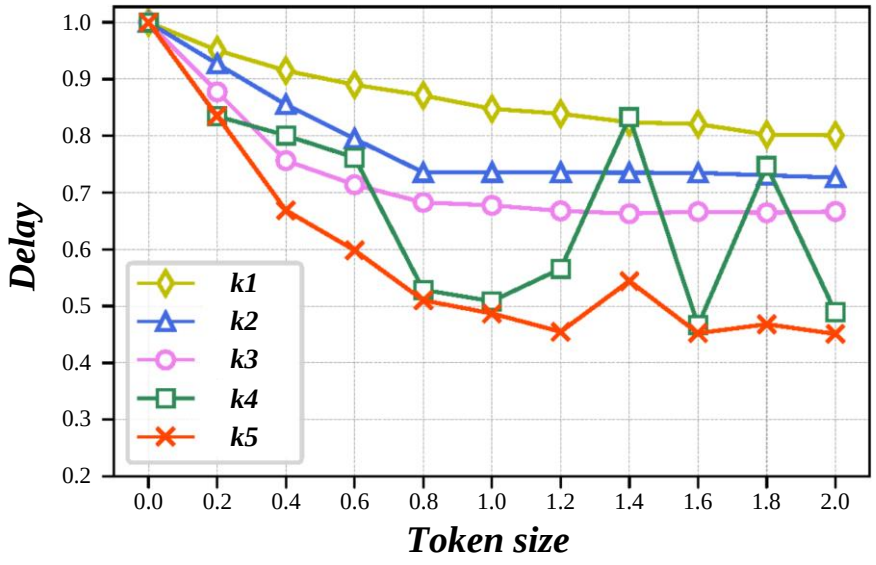


Fig. 4. Results of self-supervised label generation module in different modes.

The details is shown in Formula (18). The (SLGM delineates modal representations into two distinct categories, dictated by their polarity. Subsequently, it identifies the central tendencies of these two categories, yielding a modal representation positive center and a modal representation negative center for each modality. The precise computational methodology for this categorization and centering is outlined in Formulas (19) and (20), ensuring a rigorous and systematic approach to generating single-modal annotations from multimodal data.

$$C = (F_m \# L_m) \propto (F_u \# L_u) \tag{18}$$

$$C_p = \frac{\sum_{i=1}^{N} I(y(i) > 0) \cdot F_i}{\sum_{i=1}^{N} I(y(i) > 0)} \tag{19}$$

$$C_n = \frac{\sum_{i=1}^{N} I(y(i) < 0) \cdot F_i}{\sum_{i=1}^{N} I(y(i) < 0)} \tag{20}$$

Next, the SLG) employs the coefficient as a metric to quantify the degree of deviation between each sample and its corresponding class center. This calculation is precisely defined in Formulas (21) and (22), providing a rigorous mathematical framework for assessing the proximity of samples to their respective modal representation centers.

$$S_p = \sum_{j=1}^{K} \sqrt{F(j) C_p(j)} \tag{21}$$

$$S_n = \sum_{j=1}^{K} \sqrt{F(j) C_n(j)} \tag{22}$$

IV. Experimental Results and Analysis

A. Quantitative Analysis

The model was quantitatively analyzed using CMU-MOSI and CMU-MOSEI datasets. During the test, the multimodal analysis task was regarded as a regression and classification task, respectively. The regression task used mean absolute error (MAE) and F1 score as evaluation indicators. Among them, the smaller the value of MAE, the better the model performance, and the larger the value of other indicators, the better the model performance. As shown in Table I and Table II, the five average experimental results on the two data sets show that the proposed model can achieve satisfactory results and its performance can reach the average level. This confirms that the model can effectively mine multimodal associations to improve prediction effects.

TABLE I. Experimental Results of the Model On CMU-MOSI Dataset

Model	MAE	F1-Score
MFN	0.95	78.1
RAVEN	0.92	76.6
MCTN	0.91	79.1
MulT	0.87	82.8
MISA	0.78	83.6
Self _ MM	0.71	86.0
Ours	0.81	82.9

TABLE II. Experimental Results of the Model on CMU-MOSEI Dataset

Model	MAE	F1-Score
MFN	0.71	77.0
RAVEN	0.61	79.5
MCTN	0.61	80.6
MulT	0.58	82.3
MISA	0.55	85.3
Self _ MM	0.53	85.3
Ours	0.59	82.2

Compared with other models, there are still some small gaps in some indicators. As can be seen from the analysis of the reasons in Fig. 5, the MISA model and the Self _ MM model have already processed the data in the representation learning stage, and improved the quality of modal representation by learning the common and individual information of different modal data, which means that such models More reliable data can be obtained at the beginning to improve the subsequent prediction effect. During the experiment, more than 4,000 samples were randomly selected from the test set to test the 2 classification results. The plotted curves are shown in Fig. 6, which can reflect the excellent performance of the model.

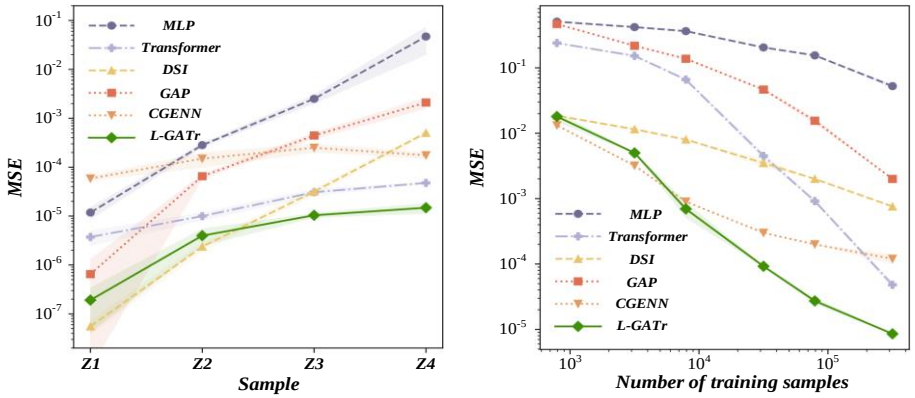


Fig. 5. Results of the MISA model and the Self-MM model in the presentation learning stage.

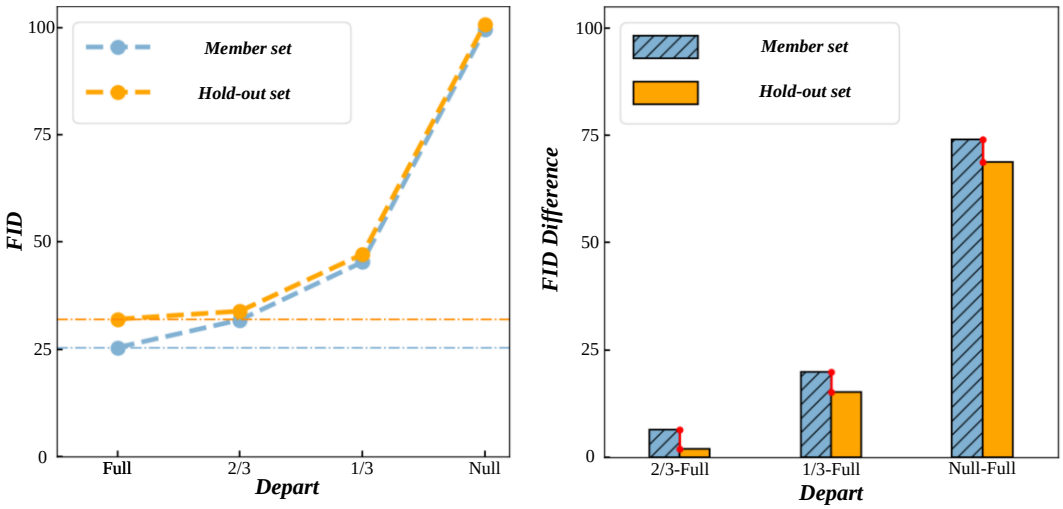


Fig. 6. Performance differences between models.

Table III shows the results from the CMU-MOSI dataset. In order to compare the performance differences between models, this chapter chooses the evaluation indicators of the classification task and the regression task for comparison. For classification tasks, the text-proposed model has obvious advantages in the evaluation of 2 classification accuracy (Acc-2), 7 classification accuracy (Acc-7) and F1 score. For the regression task, the model has achieved significant improvements in the evaluation of both the mean absolute error MAE and the Pearson correlation coefficient Corr, and the results are shown in Fig. 6. In addition to the MAE indicator, the larger the evaluation value shown in the table, the better the performance of the model on this indicator. Fig. 7 is the index result graph of the linear level. The experimental results show that the multi-modal analysis using the multi-task learning framework provides a new idea to solve the problems in this field. The model performance with the help of multi-task joint training is better than that with a single task. Task-trained model performance. In addition, the multi-level association mining framework also proves its effectiveness, it can obtain more useful information than single-angle mining.

TABLE III. EXPERIMENTAL RESULTS OF THE MODEL ON THE CMU-MOSI DATASET

Model	MAE	Corr	Acc-7	Acc-2	F1-Score
MFN	0.95	0.66	36.2	78.1	78.1
RAVEN	0.92	0.69	33.2	78.0	76.6
MCTN	0.91	0.68	35.6	79.3	79.1
MulT	0.87	0.70	40.0	83.0	82.8
MISA	0.78	0.76	42.3	83.4	83.6
Self _ MM	0.71	0.80	46.7	86.0	86.0
Ours	0.69	0.81	47.1	88.4	88.4

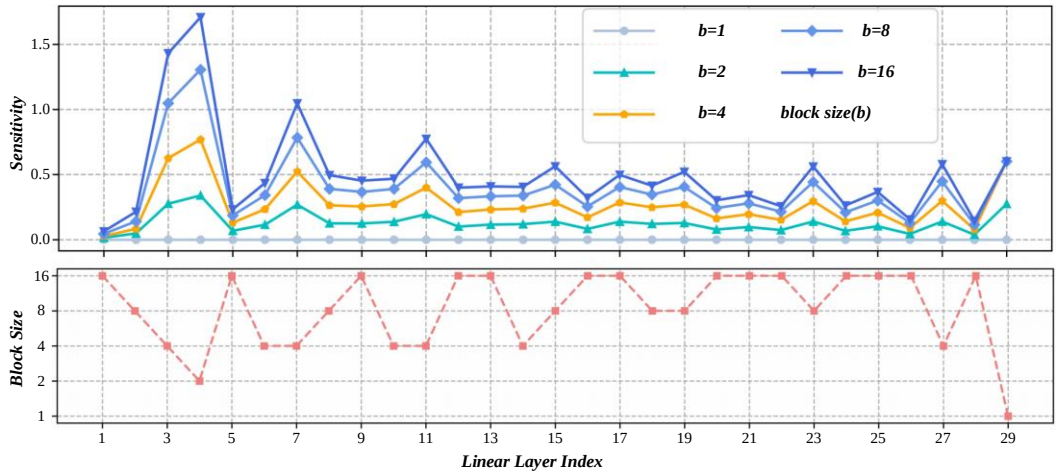


Fig. 7. Linear layer index result graph.

In this paper, a Transformer-based multi-analysis method is proposed. The model can fully consider the relationship between multiple information, use the linguistic-guided Transformer to mine the association between multiple data, and use the soft mapping module to achieve tight coupling of multiple data, thereby improving the analysis effect of the model. The experiments of the model on two data sets have achieved satisfactory results, which further demonstrates the feasibility

and effectiveness of the theory of improving the prediction effect by mining multiple associations and providing a specific solution for the problems in multiple analysis fields. Table IV shows results from the dataset, which shows that the model can also achieve good results.

TABLE IV. EXPERIMENTAL RESULTS OF THE MODEL ON THE CMU-MOSEI DATASET

Model	MAE	Corr	Acc-7	Acc-2	F1-Score
MFN	0.71	0.54	45.0	76.9	77.0
RAVEN	0.61	0.66	50.0	79.1	79.5
MCTN	0.61	0.67	49.6	79.8	80.6
MulT	0.58	0.70	51.8	82.5	82.3
MISA	0.56	0.76	52.2	85.5	85.3
Self _ MM	0.53	0.77	52.4	85.2	85.3
Ours	0.51	0.74	53.9	86.2	85.9

B. Ablation Experiment

The proposed model includes two structures: LGT and SM. The former interacts between modes to improve the learning effect when learning a certain mode, and the latter maps the learning results of each mode to a high-dimensional space for better classification. In order to verify effectiveness of two structures, this section conducts ablation experiments on the CMU-MOSI dataset, which are specifically divided into four situations: LGT and SM are not used at all; only remove LGT; Only SM is removed; LGT and SM were used simultaneously. Choosing to use the ordinary multi-head attention mechanism instead of LGT when it is not used means that the interaction ability between modes is lost. When SM is not used, the results of independent learning of each modal are directly weighted and averaged, and then classified. The ablation experimental results are shown in Fig. 8 and Fig. 9, from which it can be seen that the two main structures of the model can play a positive role in the final prediction.

C. Validation Experiment of Self-Supervised Label Generator

Combining the idea of multi-task learning with the proposed model, a multimodal analysis model based on a multi-level association mining framework and a self-supervised label generator is proposed in this chapter to solve the multimodal analysis problems faced in multimodal analysis simultaneously. Modal association mining and multimodal collaborative learning problems. The multi-level association mining framework further deepens the research content, which can simultaneously mine association information from two angles. The self-supervised tag generator can automatically train the single-modal tag-assisted multi-task learning framework, thus realizing multimodal collaborative learning. The verification experiment of the self-supervised label generator is shown in Fig. 10. A large number of experiments have been carried out on classical data sets in the field of multimodal analysis, all of which prove that the proposed model has excellent analytical performance and can provide a feasible idea for solving the problems existing in this field.

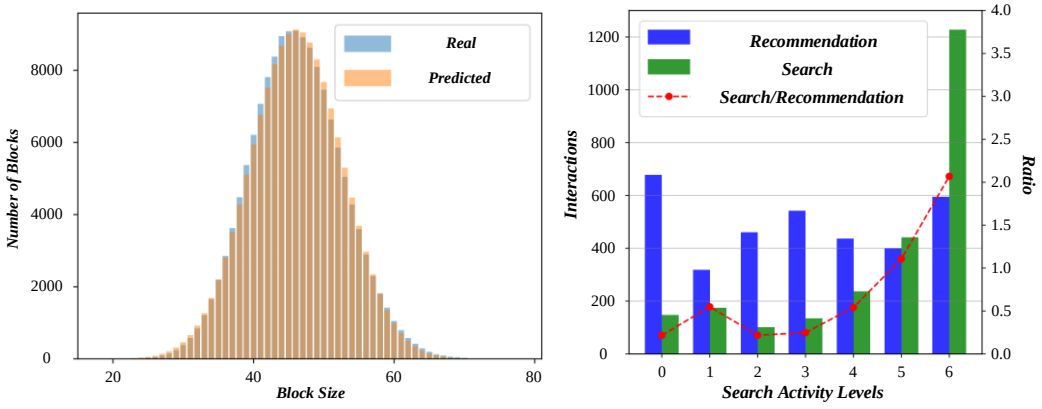


Fig. 8. Ablation experiment of module size and search level.

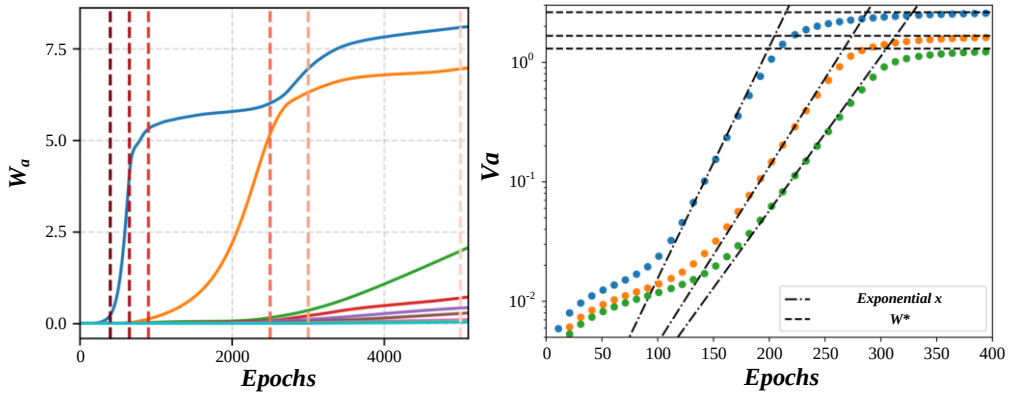


Fig. 9. Result diagram under different epoch.

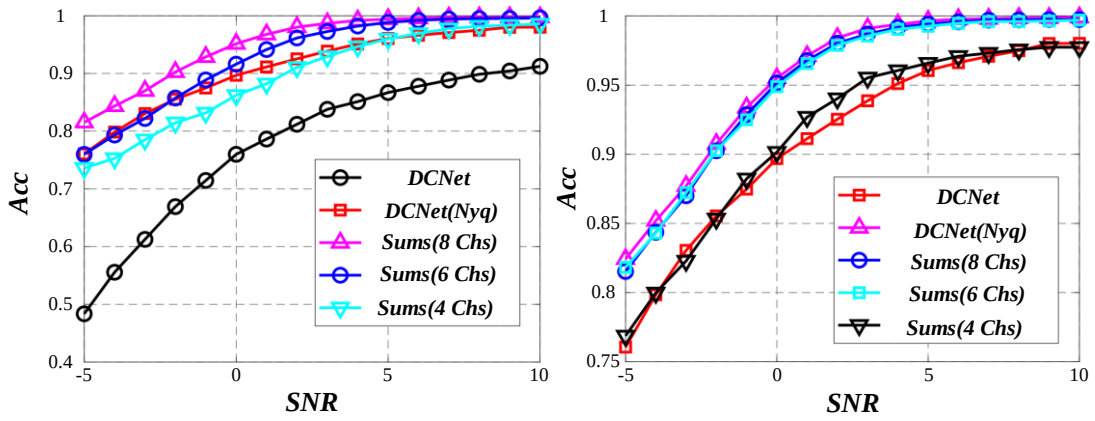


Fig. 10. Verification experiment of self-supervised label generator.

In order to examine the rationality and robustness of SLGM, this section extracts several labels generated by SLGM during training. As shown in Table V, the polarity of the single-modal labels of samples 1 to 3 is consistent with that of the manually labeled multi-modal labels, which shows that the single-modal labels generated by SLGM are of some value. Compared with the manual labeling of multi-modal labels, the single-modal labels of samples 4 to 5 achieve a negative shift, and this negative shift is reasonable.

TABLE V. SAMPLE SLGM GENERATION LABEL SAMPLE

MOSI-M	SLGM-L	SLGM-A	SLGM-V
2.2	1.1925	0.3874	0.8557
2.4	1.7874	0.0639	1.1984
-1.8	-1.5281	-0.9052	-1.3683
-0.2	-0.0327	0.0874	0.0001
0.6	0.8281	0.0052	-0.0856

V. SUMMARY

This research has conducted an in-depth exploration of the design of a cross-border e-commerce short video recommendation system based on the multi-modal fusion Transformer model. An efficient and accurate recommendation system has been successfully built through the comprehensive use of advanced technologies such as Transformer and multi-tasking learning. Many short videos and user behavior data of cross-border e-commerce platforms were collected and analyzed during the research process, supporting system design and optimization. First of all, in the aspect of multi-modal association mining, the cross-modal multi-head attention model is used to conduct in-depth analysis of multi-modal data such as video and text, and it is found that there is rich association information between different modes. A multi-modal analysis model is designed based on the multi-task learning framework in multi-modal collaborative learning. The model can learn more comprehensive and accurate multi-modal data representation through modal information sharing during training. In addition, the self-supervised labeling method proposed in this paper effectively solves the problem of missing labels, making the model perform well under limited labels. At the sorting level, we fuse the Bert and high-order explicit feature cross models to extract deeper text features and capture the deep-seated interaction relationship between users and short videos. Through the analysis and verification of large-scale data, this feature fusion method can significantly improve the prediction accuracy of the recommendation system and provide users with more accurate recommendation services. Finally, a complete cross-border e-commerce short video recommendation system is designed and implemented. The system integrates multiple modules such as data acquisition, preprocessing, recall, and sorting and realizes the stable operation of the system and user-friendly interaction through front-end and back-end design. In practical application, the system has effectively improved user satisfaction and recommendation efficiency, with an accuracy rate and F1 score of 98.76% and 97.89%, respectively, bringing significant value to cross-border e-commerce platforms.

REFERENCES

[1] Gandhudi, M., Alphonse, P. J. A., Velayudham, V., Nagineni, L., & Gangadharan, G. R. (2024). Explainable causal variational autoencoders based equivariant graph neural networks for analyzing the consumer purchase behavior in E-commerce. Engineering Applications of Artificial Intelligence, 136, 108988.

[2] Almeida, A., de Villiers, J. P., De Freitas, A., & Velayudan, M. (2022). The complementarity of a diverse range of deep learning features extracted from video content for video recommendation. Expert Systems with Applications, 192, 116335.

[3] Zeng, F. (2023). Multimodal music emotion recognition method based on multi data fusion. International Journal of Arts and Technology, 14(4), 271-282.

[4] Dutta, M., & Ganguly, A. (2024). Incremental-based YoloV3 model with Hyper-parameter Optimization for Product Image Classification in E-commerce Sector. Applied Soft Computing, 112029.

[5] Gwak, M., Cha, J., Yoon, H., Kang, D., & An, D. (2024). Lightweight Transformer Model for Mobile Application Classification. Sensors, 24(2).

[6] Gai, T., Wu, J., Liang, C., Cao, M., & Zhang, Z. (2024). A quality function deployment model by social network and group decision making: Application to product design of e-commerce platforms. Engineering Applications of Artificial Intelligence, 133, 108509.

[7] Zhuang, S. (2024). E-commerce consumer privacy protection and immersive business experience simulation based on intrusion detection algorithms. Entertainment Computing, 100747.

[8] Zhao, H. (2021). A Cross-Border E-Commerce Approach Based on Blockchain Technology. Mobile Information Systems, 2021.

[9] Zhang, C., Zheng, H., & Wang, Q. (2022). Driving Factors and Moderating Effects Behind Citizen Engagement with Mobile Short-Form Videos. Ieee Access, 10, 40999-41009.

[10] Chang, C., Zhou, J., Weng, Y., Zeng, X., Wu, Z., Wang, C.-D., & Tang, Y. (2023). KGTN: Knowledge Graph Transformer Network for explainable multi-category item recommendation. Knowledge-Based Systems, 278.

[11] Gu, P., Hu, H., & Xu, G. (2024). Modeling multi-behavior sequence via HyperGRU contrastive network for micro-video recommendation. Knowledge-Based Systems, 295, 111841.

[12] Zhu, H., Wei, H., & Wei, J. (2023). Understanding users' information dissemination behaviors on Douyin, a short video mobile application in China. Multimedia Tools and Applications.

[13] Li, C., Cao, Y., Zhu, Y., Cheng, D., Li, C., & Morimoto, Y. (2024). Ripple Knowledge Graph Convolutional Networks for Recommendation Systems. Machine Intelligence Research, 21(3), 481-494.

[14] Li, P., Li, T., Wang, X., Zhang, S., Jiang, Y., & Tang, Y. (2022). Scholar Recommendation Based on High-Order Propagation of Knowledge Graphs. International Journal on Semantic Web and Information Systems, 18(1).

[15] Jing, H. (2022). Application of Improved K-Means Algorithm in Collaborative Recommendation System. Journal of Applied Mathematics, 2022.

[16] Shen, X. (2023). E-commerce User Recommendation Algorithm Based on Social Relationship Characteristics and Improved K-Means Algorithm. International Journal of Computational Intelligence Systems, 16(1).

[17] Du, H., Tang, Y., & Cheng, Z. (2023). An efficient joint framework for interacting knowledge graph and item recommendation. Knowledge and Information Systems, 65(4), 1685-1712.

[18] Zhang, L., Zhang, W., McNeil, M. J., Chengwang, N., Matteson, D. S., & Bogdanov, P. (2021). AURORA: A Unified fRamework fOR Anomaly detection on multivariate time series. Data Mining and Knowledge Discovery, 35(5), 1882-1905.

[19] Matrouk, K. M., Nalavade, J. E., Alhasen, S., Chavan, M., & Verma, N. (2023). MapReduce Framework Based Sequential Association Rule Mining with Deep Learning Enabled Classification in Retail Scenario. Cybernetics and Systems.

[20] Chen, Z., & Ge, Z. (2022). Knowledge Automation Through Graph Mining, Convolution, and Explanation Framework: A Soft Sensor Practice. Ieee Transactions on Industrial Informatics, 18(9), 6068-6078.

[21] Huu-Thiet, N., Li, S., & Cheah, C. C. (2022). A Layer-Wise Theoretical Framework for Deep Learning of Convolutional Neural Networks. Ieee Access, 10, 14270-14287.

[22] Eun, Y. J., Chae, S., & Kyungmin, B. (2022). Layered Abstraction Technique for Effective Formal Verification of Deep Neural Networks. Journal of KIISE, 49(11), 958-971.

[23] Wu, W., Wang, W., Jia, X., & Feng, X. (2024). Transformer Autoencoder for K-means Efficient clustering. Engineering Applications of Artificial Intelligence, 133.

[24] Lo, P.-C., & Lim, E.-P. (2023). A transformer framework for generating context-aware knowledge graph paths. Applied Intelligence, 53(20), 23740-23767.

[25] Feng, Y., Zhai, M., & Du, Y. (2024). The effects of mini-detail short videos on consumer purchase intention on Taobao: A TAM2-based approach. Entertainment Computing, 100745.

[26] Wang, C., & Xiao, Z. (2022). A Deep Learning Approach for Credit Scoring Using Feature Embedded Transformer. Applied Sciences-Basel, 12(21).

[27] Fan, J., Huang, L., Gong, C., You, Y., Gan, M., & Wang, Z. (2024). KMT-PLL: K-Means Cross-Attention Transformer for Partial Label Learning. Ieee Transactions on Neural Networks and Learning Systems.

[28] Anitha, J., & Kalaiarasu, M. (2022). A new hybrid deep learning-based phishing detection system using MCS-DNN classifier. Neural Computing & Applications, 34(8), 5867-5882.

[29] Xu, R., Li, J., Li, G., Pan, P., Zhou, Q., & Wang, C. (2022). SDNN: Symmetric deep neural networks with lateral connections for recommender systems. Information Sciences, 595, 217-230.

[30] Nam, W., & Jang, B. (2024). A survey on multimodal bidirectional machine learning translation of image and natural language processing. Expert Systems with Applications, 235.

A Hidden Markov Model-Based Performance Recognition System for Marching Wind Bands

Wei Jiang
Shenyang City University, Shenyang 110100, China

***Abstract*—This paper explores the automatic recognition of marching band performances using advanced music information retrieval techniques. Music, a crucial medium for emotional expression and cultural exchange, greatly benefits from the harmonic backing provided by marching wind orchestras. Identifying these performances manually is both time-consuming and labor-intensive, particularly for non-professionals. This study addresses this challenge by leveraging Hidden Markov Models (HMM) and improved Pitch Class Profile (PCP) features to automate the recognition process. The research also explores the system's performance on real-world audio recordings with background noise and microphone variations. By dividing the audio signal into frames and transforming it to the frequency domain, the PCP feature vectors are extracted and used within the HMM framework. Experimental results demonstrate that the proposed method significantly enhances recognition accuracy compared to traditional PCP features and template matching models. The study identifies challenges in distinguishing similar tonal values, such as F-major and D-minor, which affect recognition rates. Additionally, the research highlights the importance of addressing background noise and microphone variations in real-world applications. Ethical considerations regarding privacy and intellectual property rights are also discussed. This research establishes a comprehensive system for automatic marching band performance recognition, contributing to advancements in music information retrieval and analysis.**

Keywords*—Music information retrieval; Hidden Markov Model; feature extraction; automatic music recognition; marching band performance; PCP features

I. INTRODUCTION

Music serves as a powerful medium of artistic expression. It enables people to express personal feelings and fulfill spiritual needs, while also fostering cultural exchange and promoting the development and integration of cultural diversity. Within the foundation of music theory, marching wind orchestra performances play a crucial role. They complement and enhance the main theme, adding depth and richness to the overall musical experience [1-3]. Despite its importance, the identification of marching band performances remains challenging and time-consuming, particularly for non-professionals. This paper aims to address this research gap by developing an automatic recognition system leveraging HMM and improved PCP features. If beautiful music lacks a harmonic backing, the overall effect will be greatly reduced [4, 5]. However, the identification of marching band performance often requires specialized knowledge and training that is difficult for non-professionals to accomplish accurately, especially in improvisation, where identification of marching band performance is even more challenging [6]. For many years, the identification and recognition of marching wind band performances are mostly done manually, which is time-consuming and laborious [7]. With the development of multimedia and network technology, the importance of music information retrieval technology is becoming more and more obvious. The traditional low-level features such as Mel frequency cestrum coefficients have limited effect in music semantic analysis, while the marching band performance, as a middle-level feature, contains rich music information, which is important for music analysis and retrieval [8]. Marching wind band performance is closely related to the emotion of music, which can help recognize and retrieve songs with similar styles. However, the system's performance on real-world audio recordings with background noise and microphone variations remains an important consideration [9].

Speech recognition technology has made significant progress in recent years, and HMM combined with genetic algorithm training has become a mainstream technology with the advantages of high recognition rate and fast response. However, the development of music recognition technology is slow and there are fewer related products on the market, mainly due to the low recognition rate [10]. The earliest music feature extraction methods used Mel-frequency cestrum coefficients, but nowadays it is common to use pitch-set files to represent music, which can more accurately represent music features [11]. With the improvement of computer performance and Internet bandwidth, as well as the development of multimedia information technology, content-based multimedia retrieval techniques have emerged [12]. In music retrieval, marching wind orchestra performance, as a mid-level feature, can effectively support music segmentation, retrieval and sentiment analysis [13-15]. Automatic marching band performance recognition techniques have attracted the attention of a large number of researchers in the field of music information retrieval. The correct recognition and sequence generation of marching wind band performances can help the segmentation of musical structures and the identification of specific melodies, and can reveal the potential emotional connections of music [16].

The Electrical Engineering Department at National Taiwan University was a pioneer in using PVP feature vectors for performance recognition [17]. Their system processes input audio signals by segmenting them into frames and converting them into the frequency domain to extract PCP feature vectors [18]. The recognition process is divided into two phases: training and testing. In the training phase, a Hidden Markov Model (HMM) is used, where each state corresponds to a specific marching wind band performance [19]. The state transition matrix represents the probability of transitioning from one performance to another, while the observation distribution

indicates the likelihood of a particular PCP feature vector being generated by a specific state. In the testing phase, the observed feature vectors and the trained HMM are used to decode the most probable sequence of marching wind band performances. The team's innovative use of the N-gram algorithm within the HMM framework significantly reduces complexity and enhances recognition efficiency. In 2003, Alexander Sheh and Daniel P.W. Ellis from Columbia University proposed a system that converts arbitrary audio signals into corresponding performance sequences [20]. The system process includes audio framing, transforming to the frequency domain by Fourier transform, then mapping out PVP feature vectors, constructing a marching band performance model, and utilizing EM algorithms to complete the recognition in the HMM framework [21]. Although the recognition rate of this system for marching wind band performance is only 22%, it is innovative in that only the performance sequence is considered without the need of temporal requirements on the performance transformation. In 2005, Bello and Pickens applied the EM algorithm under the HMM framework, introduced the music knowledge into the model, and avoided arbitrary initialization by defining the state transfer matrix, and achieved a recognition rate of 75% [22, 23]. Although Markov models have been successful in speech recognition, there are challenges in applying them to music. Music has complex acoustic variations and requires more data for training. Manually labeling the performance boundaries of long sections of music is time-consuming and error-prone. The music and acoustics research center at Stanford University proposes a method to automate the performance boundary labeling by synthesizing audio to generate training data, which significantly improves the efficiency of model parameter estimation.

The purpose of the research in this paper is to establish a complete marching band performance recognition system, using audio files as the input of the whole system, and returning the marching band performance sequences recognized by the system as the output to the user, so as to realize the automatic recognition with performance as the basic unit. The research content of this paper mainly includes the following aspects: first, feature extraction of music. Since the speech signal is time-varying rather than smooth, and the human articulatory organ muscles move slowly, the speech signal can be considered smooth locally, so the processing methods and theories of smooth processes can be introduced into the processing of speech signals, thus simplifying the analysis of speech signals. Next, the marching band performances applied during the experiments are extracted, and a Hidden Markov Model is initialized for each marching band performance, using a single Gaussian observation function during the initialization process, with the mean vector set to 0 and the covariance matrix set to 1 [24]. Next, the experimental samples are labeled. The labeling is done from a MIDI file, so the input file must be converted to the corresponding MIDI format, and finally a piece of music performance is extracted as the output of the labeling process. Next, the system is trained, using a pitch set file to represent the music file, and the labeled marching band performance is used as the base model to train the system, with as many training samples as possible, in order to give the system access to all the performance models. Finally, system testing is performed, where a correctly labeled music file is used as input to check the performance of the system. The remainder of this paper is organized as follows: Related work is given in Section II. Section III presents the theoretical background on Hidden Markov Models. Section IV describes the design of the marching band performance recognition system. Section V discusses the experimental results and analysis. Section VI concludes the paper with a summary of findings and suggestions for future research.

II. Related Work

Initial efforts in music recognition heavily relied on low-level audio features such as Mel-Frequency Cepstral Coefficients (MFCCs). MFCCs were widely adopted due to their efficiency in capturing the timbral properties of audio signals [25]. However, their effectiveness in higher-level musical semantic analysis, particularly for complex structures like marching band performances, was limited [26]. The integration of Hidden Markov Models (HMMs) in music recognition was significantly advanced by Sheh and Ellis [27]. Their system converted arbitrary audio signals into performance sequences using HMMs, involving audio framing. Further refinement was seen in the work of Bello and Pickens [28], who applied the Expectation-Maximization (EM) algorithm within the HMM framework. Recent research has focused on enhancing the feature extraction methods to improve the recognition accuracy of musical signals. Enhanced PCP features have emerged as a crucial development, addressing tonal ambiguity and providing a more robust representation of musical content. Comparative studies have demonstrated that traditional template matching models [29], while straightforward, are often inadequate for dynamic and complex musical environments. In contrast, the combination of improved PCP features with HMMs has shown superior performance. However, a critical gap in existing research is the robustness of these systems in real-world scenarios, characterized by background noise and variations in recording conditions [30].

III. Hidden Markov Model

A Markov model is a demographic tool extensively utilized in diverse natural language processing applications, including speech recognition, automatic lexical annotation, and probabilistic grammar analysis. If the "future" of a process depends only on the "present" but not on the "past", the process is Markovian, or the process is called Markovian.

In addition, since speech signals are time-varying rather than smooth, and since the muscles of the human articulatory organs move slowly, speech signals can be considered locally smooth. In this way, we can introduce the processing methods and theories of smooth processes into the processing of speech signals, thus greatly simplifying the analysis of speech signals.

A. Characteristic Representation of Music

When analyzing audio signals, extracting and characterizing information from the time domain can be challenging due to the non-linear and often discontinuous nature of audio performance in this domain. In speech signal processing, converting audio signals from the time domain to the frequency domain is a common practice for more effective analysis. This technique is equally applicable to music signals. The two primary methods for this transformation are the short-time Fourier transform

(STFT) and the constant Q transform (CQT). Both methods convert the music signal from the time domain to the frequency domain, but they use different algorithms and computational processes to achieve this.

Fourier transform processing of speech signals, the premise is that the signal is always in a smooth state, but the audio signal of music is usually non-stationary, cannot be transformed by the Fourier transform spectrum to extract the spectral energy information, based on the assumption that the music signal is in the short-term transient conditions, that is, you can through the STFT spectral transformation, and then be able to analyze the characteristics of the signal in the frequency domain. The formula of STFT is expressed as:

$$X_m(\omega) = \sum_{n=-\infty}^{\infty} x[n]w[n-m]e^{-j\omega n}$$

Considering the continuity of the music signal, $x[n]$ represents the discretized representation, $w[n-m]$ represents the sliding time window, and $X_m(\omega)$ represents the transformed spectrum. Under the influence of Heisenberg's uncertainty principle, the time resolution and frequency resolution change accordingly because of the different window functions, if the function is determined and the window length is determined, both the time resolution and frequency resolution can not be changed, so it is difficult to deal with the non-smooth and mutated signals effectively, and it is suitable to deal with the slow-varying signals because of the insensitivity to the instantaneous changes. For the music audio signal used in this paper, there is only a single main theme after the relevant preprocessing, and its changes are more moderate, so the spectrum can be analyzed by short-time Fourier transform. Therefore, before extracting the PCP features from the audio signal, this paper employs the STFT to convert the time-domain signal into the frequency domain. This approach conserves computational power and aligns well with the subsequent PCP feature extraction. The flowchart of feature extraction in Fig. 1 shows the process of marching wind band performance recognition.

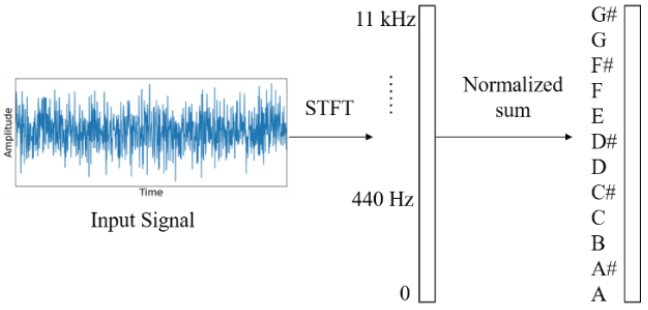


Fig. 1. Flowchart of feature extraction.

The calculation of PCP features is based on mapping frequency changes according to the twelve-tone equal temperament system in music theory. In musical terms, changes in pitch are reflected as changes in frequency values within the audio signal. Typically, the frequency ratio between notes an octave apart is 2:1. In the twelve-tone equal temperament system, the frequency ratio between adjacent semitones is the twelfth root of two. Consequently, the horizontal axis of a musical signal changes exponentially, and when represented in three-dimensional space, the pitch changes correspond to a spiraling frequency pattern, as illustrated in Fig. 2 below. This visual representation highlights the frequency changes associated with different pitch levels more intuitively.

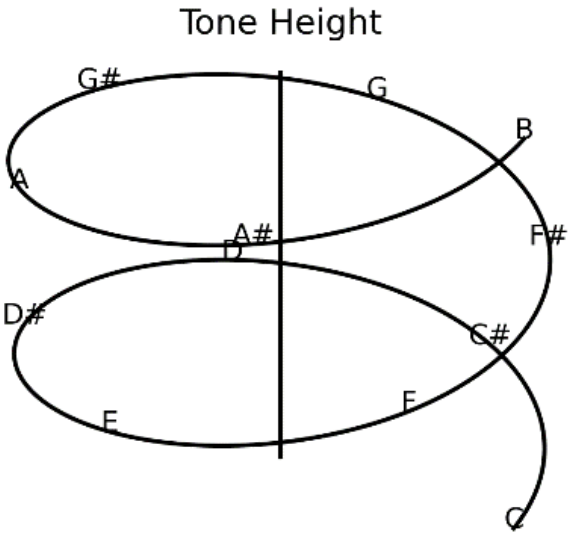


Fig. 2. Three-dimensional representation of sound level.

The distinct advantage of the PCP feature lies in its ability to process the spectral energy of an audio signal alongside its musical features, thereby providing a more accurate representation of the musical characteristics within the audio data. This enhanced representation is particularly beneficial when analyzing music-related audio signals. The following section will expand on the formula used to calculate the PCP feature for a single frame state:

$$p(k) = \lfloor 12 * log_2(k * \frac{f_{sr}}{N}/f_{rel}) \rfloor mod 12$$

where f_{rel} is the reference frequency value of the lowest scale group, the lowest scale group includes the scales C1, D1, E1, F1, G1, Al, B1; f_{sr} is the sampling frequency, N represents the number of sampling points, f_{sr}/N denotes the transform frequency interval of the Fourier Transform, and $k * \frac{f_{sr}}{N}$ denotes the frequency of each component in the frequency domain, so $k * \frac{fsr}{N*f_{rel}}$ represents the The frequency ratio of the component to the level, and all the components corresponding to the frequency value of the same level are summed up according to the above formula to get the twelve-dimensional PCP main melody feature vector:

$$PCP[p] = \sum_{k:p(k)=p} |X[k]|^2 \ , \ p = 1,2,\ldots,12$$

where $X(k)$ is the energy spectrum obtained by Fourier transforming the audio data of the main melody, k is the index of the Fourier transformed component, and p is the ordinal number corresponding to the twelve-tone levels. According to the twelve mean laws in music theory, ignoring the influence of the higher or lower octave, and only considering the frequency

values of the twelve tone levels in the lowest scale group in the music, each component in the frequency domain and the frequency value of the lowest tone level are divided correspondingly to obtain twelve frequency ratios, thus completing the expansion of the components into twelve frequency bands; for all the twelve bands obtained by the components, the components corresponding to the same tone level bands are summed up to get the twelve-dimensional PCP melody feature vector $PCP[p]$ in the whole frequency domain.

From the calculation of the PCP feature in the previous subsection, it is easy to see that it is the spectral information of the music signal is compressed by the frequency rule corresponding to the twelve equal-tempered law, and folded into a twelve-dimensional vector in the form of a tone level profile of the spectrum This calculation process, although the spectral information is endowed with the musical characteristics, does not take into account the possible problems that may exist in the music signal. Usually in musical signals, the notes in the low frequency part are difficult for the human ear to distinguish and hear because of their resonance, so the bass is more blurred and less distinctive in most cases. In addition, there are overtones in common music signals. In the normal human ear mechanism hearing system, overtones do not cause too much interference and influence on the auditory senses, so they are generally not too concerned about overtones. However, in the audio feature representation, when there are too many overtones in the music signal, the process of converting it into a spectrum will occupy more spectral resources, thus affecting the energy of the similar fundamental frequency, generating errors, and affecting the information extraction of the real sound value. Considering that the problems of bass ambiguity and high-frequency overtones are less considered in the current research, this paper introduces a Gaussian filter bank, which is combined with the musical properties of the twelve equal temperament law, to add a window restriction and increase the weights of the fundamental frequencies of the tone levels. The weights of irrelevant frequencies are filtered out, and the mathematical expression of this Gaussian filter is as follows.

$$PCP_{[p]} = \exp\left(-\frac{\left(k^{*}\frac{f_s}{N} - f_{rel}^{*}2^{(O-1)}\right)^2}{2^{*}15^2}\right) * PCP_{[p]}$$

In the above formula, $k * \frac{f_{sr}}{N}$ represents the component frequency value of the sample, and the center frequency is the reference frequency value of the scale corresponding to the octave interval where it is located, $f_{rel} * 2^{(O\cdot1)}$, o represents that the frequency of the sample point at this time corresponds to the frequency range of the o th octave. Because f_{rel} represents the reference frequency value of the lowest scale, when the frequency value of the sampling point is in the frequency range of other octave intervals, because the frequency relationship of different octaves is the 2nd power relationship, for example, the ratio of the frequency values of C2 and C1 is 2, and the ratio of the frequency values of C3 and C1 is 4, so the center frequency becomes the reference frequency value of the original lowest scale group multiplied by an integral multiple of 2, it is now in the octave where the reference frequency value is located. At this point, the frequency values of the twelve semitones within the octave interval become the new center frequencies. These center frequencies are set to correspond with the semitone frequencies in the twelve-tone equal temperament system. This method retains the frequency weights of all notes in this system while filtering out irrelevant frequency values. Consequently, low-frequency noise and high-frequency overtone interference are effectively mitigated, and the fundamental frequencies of the low-frequency band are preserved, addressing the issue of indistinct tone values to some extent.

Fig. 3 illustrates the spectrum of the frequency interval for A4 after Gaussian filtering. It shows that 440Hz has the highest amplitude, indicating it as the center frequency. Other frequencies have amplitudes ranging between 420Hz and 430Hz on the left boundary and between 450Hz and 460Hz on the right boundary. The frequencies of G#4 and B4 fall outside these boundaries, ensuring that effective tone values pass through, demonstrating the efficacy of the filtering. Each Gaussian filter's center frequency corresponds to the twelve semitones between C4 and B4. This filtering method effectively extracts frequency domain energy based on the twelve-tone equal temperament system, mitigating low-frequency noise and high-frequency overtones.

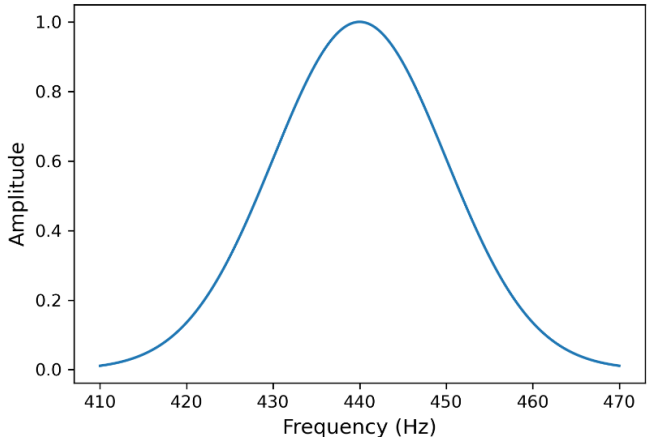


Fig. 3. A4 korst spectral plot of notes.

The primary process in PCP feature calculation involves folding and weighting the spectral energy. In actual music audio, the complex interplay of sounds from various instruments with different pitches, volumes, and rhythms results in significant variability in the chromatic features extracted from each song. This variability introduces multiple levels of complexity, making it challenging to develop a classifier that covers the entire feature space. To manage this, logarithmic compression is used to limit the dynamic range of the features, as detailed by the following mathematical formula:

$$PCP_{Log} = \log\left(1 + \eta * \widehat{PCP}[p]\right)/P[p]_{sum}) \ , \ p = 1,2,\ldots,12$$

$\widehat{PCP}[p]$ is the PCP feature vector obtained after Gaussian filtering as described above, $P[p]_{sum}$ is the sum of all the frequency components corresponding to the twelve semitones, and η stands for the compression coefficient, and 100 is used as the compression coefficient in this paper because it has the best performance in the experiment. The ratio of the filtered PCP feature vector to the total frequency components is obtained,

multiplied by the compression coefficient, weighted and summed with 1, and then logarithmically transformed to replace the original PCP feature vector. The above compression method reduces the computation amount of the related feature frequency values, and makes the effect of the features have a better performance ability.

B. HMM Model

An HMM model is a statistical framework extensively utilized in various natural language processing applications, including speech recognition, automatic lexical annotation, and probabilistic grammar analysis. A process is considered Markovian, or a Markov process, if its future state depends solely on its current state and not on its past states.

$$X(t+1) = f(X(t))$$

Where $X(t)$ denotes the state at time t. Markov processes that are discrete in time and state are called Markov chains.

$$X_n = X(n), n = 0,1,2,\ldots$$

Denotes the results of successive observations of discrete state processes on the time set $T = \{0,1,2,3,\ldots\}$ The result of successive observations of discrete state processes on the time set $T = \{0,1,2,3,\ldots\}$. A Markov chain is a random process that adheres to the following:

The probability distribution of the system's state at time $t+1$ depends only on its state at time t, and is independent of its states prior to t;

The transition from the state at time t to the state at time t+1 is independent of the specific value of t.

A Markov chain model can be defined by the elements (S, P, Q), where:

S is a non-empty set of all possible states of the system, commonly known as the state space. This set can be finite, countable, or any non-empty set. In this paper, S is assumed to be countable, with states denoted by lowercase letters such as i, j etc.

P is the state transition probability matrix of the system, $p_{ij}(k)$ represents the probability of transitioning from state i at time t to state j at time $t+k$. For a Markov chain model in a discrete state space with a finite number of states, the transition probability distribution is expressed as a matrix with $N \times N$ elements, known as the "transition matrix".

$$P_{ij}(t, t+k) = P(q_{t+k} = \theta_j \mid q_t = \theta_i)$$

When $k = 1, P_{ij}(1)$ is called a piece of transfer probability, referred to as transfer probability, and all the transfer probabilities a_{i_j}, $1 \le i, j \le N$ can form a state transfer matrix:

$$A = \begin{bmatrix} a_{11} & \cdots & a_{1N} \\ \vdots & & \vdots \\ a_{N1} & \cdots & a_{NN} \end{bmatrix}$$

Where $0 \le a_i j \le 1$ and $\sum_j = 1^N a_{ij} = 1$. Q represents the initial probability distribution of the system, with π_i indicating the probability of the system being in state i at the initial time.

The Fig. 4 below demonstrates the hidden and observed states using a weather example. In this model, the hidden state (actual weather) is represented by a first-order Markov process, where each state is interconnected. In addition to the probabilistic relationships defined by the Markov process, there is a confusion matrix that outlines the probabilities of the observed states for each corresponding hidden state.

For the weather example, the confusion matrix is shown in Table I.

TABLE I. CONFUSION MATRIX OF WEATHER

Observed weather	Hide Weather			
	Dry	Dryer	Wet	Soggy
Sunny	0.6	0.2	0.15	0.05
Cloudy	0.25	0.25	0.25	0.25
Raining	0.05	0.1	0.35	0.5

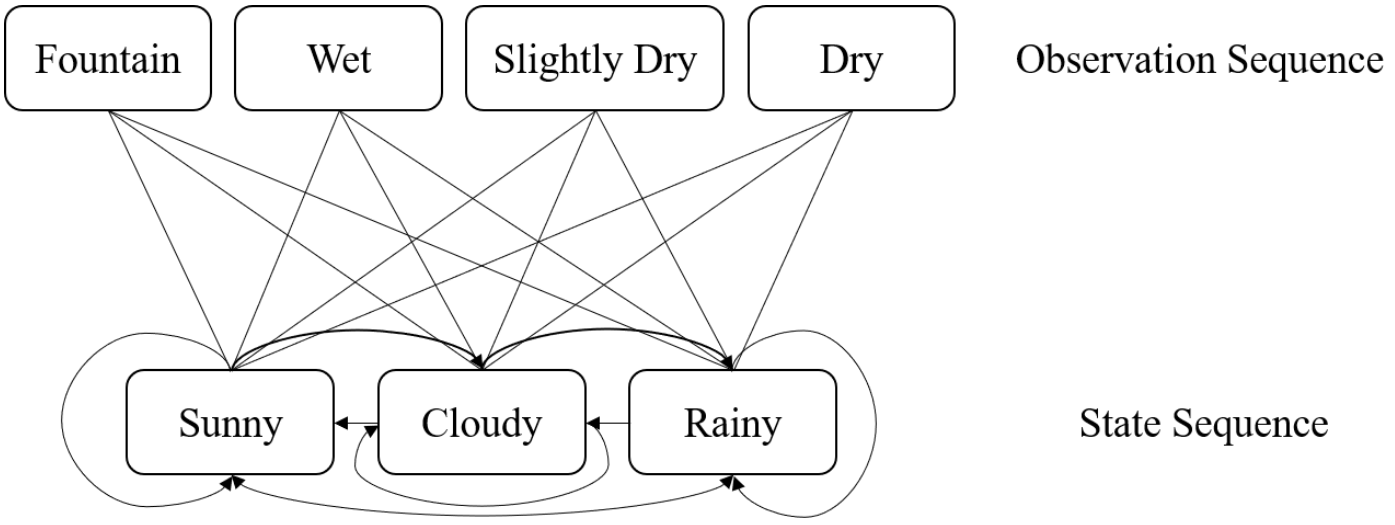


Fig. 4. First-order hidden markov processes.

A Hidden Markov Model (HMM) is characterized by five elements: two sets of states and three probability matrices. The hidden states S_4 satisfy the Markov property and represent the underlying processes that are not directly observable (S_1, S_2, S_3, etc.). The observable states (OOO) correspond to these hidden states and can be directly measured (O_1, O_2, O_3 , etc). It's important to note that the number of observable states doesn't necessarily match the number of hidden states.

The initial state probability distribution, π, defines the probabilities of the system starting in each hidden state at t=1, $P(S_1) = P_1, P(S_2) = P_2, P(S_3) = P_3$, describes the probabilities of transitioning from one hidden state to another. Additionally, the observation probability matrix B —often called the confusion matrix—provides the probabilities of each observable state given a hidden state. This comprehensive framework allows HMMs to model complex sequences where the true states are not directly observable but can be inferred through observed data.

$$A = \begin{bmatrix} a_{11} & \dots & a_{1N} \\ \vdots & & \vdots \\ a_{N1} & \dots & a_{NN} \end{bmatrix}$$

a_{ij} denotes the probability that the state is S_i at time t and S_j at time $t + i$. A describes the transfer probabilities between states in the Hidden Markov Model. The observed state transfer probability matrix B (Confusion Matrix) is as follows:

$$B = \begin{bmatrix} b_{11} & \dots & b_{1N} \\ \vdots & & \vdots \\ b_{M1} & \dots & b_{MN} \end{bmatrix}$$

C. Application of Hidden Markov Models

Once a system is defined as a Hidden Markov Model (HMM), it can solve three fundamental problems. The first two are pattern recognition tasks: calculating the probability of a specific observation sequence given the HMM, and determining the most likely sequence of hidden states that could produce the observed sequence. The third problem is to generate an HMM from a given sequence of observations.

1) Evaluation: This involves assessing which of several Hidden Markov Models (represented by sets of Π, A, B) is most likely to have generated a specific observation sequence. For instance, we might have different HMMs for "summer" and "winter" based on seasonal variations in seaweed humidity. By evaluating the probability of observed humidity sequences, we can determine the most appropriate model, thus inferring the current season. In speech recognition, this method is used to identify words by comparing the observation sequences against multiple HMMs, each representing a different word. The forward algorithm is employed to compute the probability of the observation sequence for each HMM, enabling the selection of the most likely model.

2) Decoding: This task involves finding the most probable sequence of hidden states that could generate a given sequence of observations. This is particularly valuable as hidden states often represent significant, unobservable information. For example, consider a scenario where a blind hermit can observe the state of seaweed but wants to infer the underlying weather conditions (the hidden states). The Viterbi algorithm is used in such cases to determine the most likely sequence of hidden states given the observed data, providing insights into the unobservable processes.

3) Learning: The third and most challenging problem in HMMs is generating an appropriate Hidden Markov Model from a sequence of observations. This involves estimating the optimal HMM parameters—Π, A and B—that best describe the observed sequence and the associated hidden states. This process, known as learning or parameter estimation, is crucial when the transition and observation matrices (A and B) cannot be directly measured, which is often the case in practical applications. The forward-backward algorithm is typically employed for this purpose, as it allows for the iterative refinement of the model parameters to maximize the likelihood of the observed data given the model.

D. Implementation of Hidden Markov Models

Hidden Markov models, described by a vector and two matrices (Π, A, B), are of great value for real systems, and although they are often only an approximation, they are robust to analysis. Hidden Markov models typically solve problems such as: evaluation, decoding, and learning.

We use a forward algorithm to compute the probability of a T -long sequence of observations:

$$Y^{(k)} = y_{k_1} y_{k_2}, \dots\dots, y_{k_{T-1}} y_{k_T}$$

To compute the probability of an observation sequence of length T, the forward algorithm is employed. This method involves recursively determining the probability of the observation sequence for a given HMM. We start by defining the partial probability, which represents the likelihood of reaching an intermediate state within the sequence. Then, we describe how to compute these local probabilities at t=1 and for subsequent times t=n (where n>1). The following Fig. 5 illustrates the weather states and the first-order state transitions for observation sequences labeled as dry, wet, and soaked conditions:

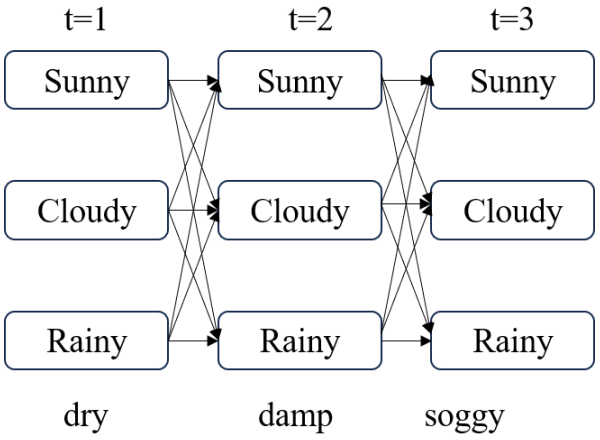


Fig. 5. First order state transfer diagram.

We define the local probability of being in state j at time t as $\alpha_t(j)$. This local probability is computed using the formula:

$$\alpha_i(j) = \Pr(\text{Observe state} \mid \text{Hidden state j}) \times \Pr(\text{all paths to state j at time t})$$

For the final observed states, this local probability includes the likelihood of reaching these states through all possible paths in the lattice. By summing these final local probabilities, we obtain the total probability of the observed sequence given the Hidden Markov Model (HMM).

To compute the local probability $\alpha_t(j)$ at t = 1, we use the initial probabilities, since there are no prior paths. Thus, the probability of being in the current state at t = 1 is the initial probability, represented as Pr (state t=1) =P(state). Consequently, the local probability at t=1 is calculated by multiplying the initial probability of being in the current state with the corresponding observation probability:

$$\alpha_1(j) = \pi(j) \times b_{jk_1}$$

Where $\pi(j)$ is the initial probability of state j, and b_{jk_1} is the probability of observing the initial observation given state j. So, the local probability of state j at the initial moment depends on the initial probability of this state and the observation probability that we have seen at the corresponding moment.

See the observation probability at the corresponding moment. Calculate the local probability for t>1:

$$\alpha_1(j) = \pi(j)^* b_{jk_1}$$

We can assume, recursively, that the probability of the observed state given the hidden state Pr (Observe state | Hidden state j) is already known. Now, we focus on the probability of all paths leading to state j at time t (Pr (all paths to state j at time t)). The number of paths required to compute α_ι (j) increases exponentially with the sequence of observations, but at moment t-1 $\alpha_{t-1(j)}$ gives the probability of all previous paths to this state, so we can define $\alpha_t(j)$ at moment t by the local probability at moment t-1:

$$\alpha_{t+1}(j) = b_{jk_{t+1}} \sum_{i=1}^{n} \alpha_t(j) a_{ij}$$

Therefore, this probability we compute is equal to the sum of the corresponding observation probability (i.e., the probability of observing a symbol in state j at time t+1) and the probability of arriving at this state at that moment, from the product of the result of the computation of each localized probability from the previous step and the corresponding state-transfer probability multiplied by the product of the corresponding state-transfer probabilities, as shown in Fig. 6.

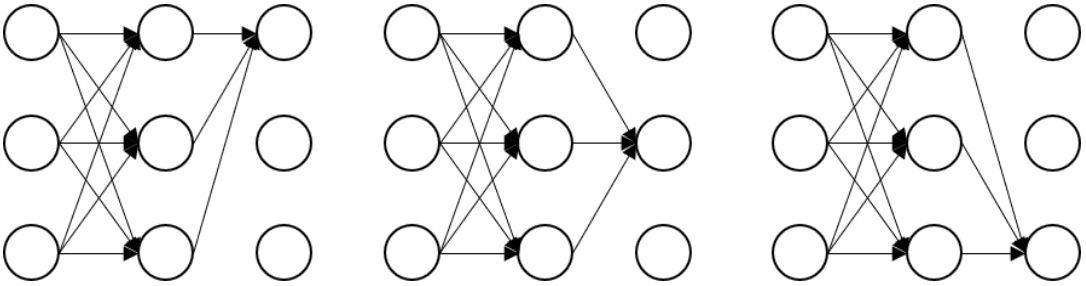


Fig. 6. Final localized probability transfer map.

Now we can recursively compute the probability of a sequence of observations after a given Hidden Markov Model (HMM). We start by computing $\alpha_2(j)$ at t=2 from the local probability $\alpha_i(j)$ at t=1, $\alpha_3(j)$ at t=3 from $\alpha_2(j)$ at t=2, and continue this process until t=T. The probability of the entire observation sequence for a given HMM is the sum of the local probabilities at t=T:

$$\Pr(Y^{(k)}) = \sum_{j=1}^{n} \alpha_T(j)$$

To efficiently compute the probability of an observation sequence given an HMM, we use the forward algorithm. This algorithm employs recursion to avoid the exhaustive computation of all possible paths in the lattice. Using this approach, we can evaluate multiple HMMs by applying the forward algorithm to each one and then selecting the model that yields the highest probability for the given observation sequence.

For generated observation sequences, the most probable model parameters are determined and optimized using the forward-backward algorithm. The essential problems addressed by HMMs include evaluation (using the forward algorithm) and decoding (using the Viterbi algorithm). The evaluation measures the relative fitness of a model, while decoding infers the sequence of hidden states. Both processes depend on the HMM parameters: the state transition matrix A, the observation matrix B and the initial state probability vector Π.

In the forward algorithm we define the local probability $\alpha_t(i)$, call it the forward variable:

$$\alpha_t(i) = P(O_1, O_2, \ldots, O_t, q_t = S_i \mid \lambda)$$

Similarly, we can define a backward variable $\beta_t(i)$:

$$\beta_t(i) = P(O_{t+1}, O_{t+2}, \ldots, O_T \mid q_t = S_i, \lambda)$$

The backward variable represents the probability of a sequence of local observations from the moment t+1 to the termination moment, knowing the Hidden Markov Model λ and the fact that ι moments are located in the hidden state s_i. Also similar to the forward algorithm, we can compute the backward variable recursively from backward to forward (hence the term backward algorithm): Initially, the backward variable for all states at time $t = T$ is 1

$$\beta_T(i) = 1 \quad 1 \le i \le N$$

Inductively, recursively calculate for each time point, $t = T-1, T-2, \ldots, 1$ at the time of the backward variable:

$$\beta_t(i) = \sum_{j=1}^{N} a_{ij}\, b_j(O_{t+1})\, \beta_{t+1}(j) \quad t = T-1, T-2, \ldots, 1 \quad 1 \le i \le N$$

This approach allows for the computation of backward variables for all hidden states at each point in time. To calculate the probability of observing a sequence using the backward algorithm, one needs to sum the backward variables (local probabilities) at t=1. The following Fig. 7 shows the relationship between the backward variables at moment t+1 and at moment t:

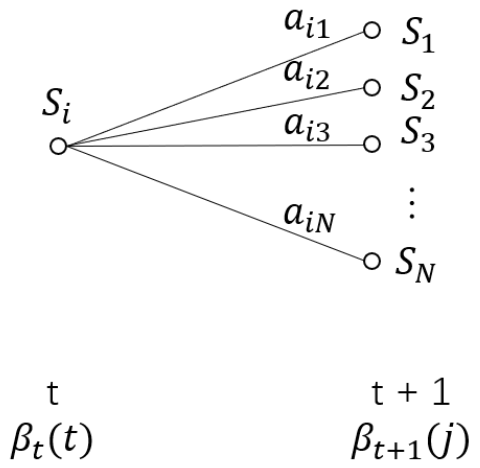


Fig. 7. Forward-backward variable relationships.

Among the three basic problems of Hidden Markov Models (HMM), the third problem of HMM parameter learning is the most difficult, because for a given sequence of observations O, there is no method that can accurately find an optimal set of Hidden Markov Model parameters (Π, A, B) to maximize $P(O|\lambda)$. As a result, scholars retreat to the second-best solution, which cannot make $P(O|\lambda)$ globally optimal, and seek for a solution that makes it locally optimal, and the forward-backward algorithm becomes an alternative solution to the Hidden Markov Model learning problem. We first define two variables. Given an observation sequence O and a Hidden Markov Model λ, define the probability variable of being in the hidden state S_i at time t as:

$$\gamma_t(i) = P(q_t = S_i | O, \lambda)$$

Regarding the definition of the forward variable $\alpha_t(i)$ and the backward variable $\beta_i(i)$, we can easily express the above equation in terms of forward and backward variables as:

$$\gamma_t(i) = \frac{\alpha_t(i)\, \beta_t(i)}{P(O \mid \lambda)} = \frac{\alpha_t(i)\, \beta_t(i)}{\sum_{i=1}^{N} \alpha_t(i)\, \beta_t(i)}$$

$$\xi_t(i,j) = P(q_t = S_i, q_{t+1} = S_j | O, \lambda)$$

IV. Marching Band Performance Recognition System Design

The automatic music recognition system developed in this paper comprises two primary components: the music feature extraction module, which utilizes the enhanced PCP feature extraction method previously described, and the modeling module. The modeling module involves gathering modeled music labels and conducting the training and prediction phases of the model, as shown in Fig. 8.

The automatic music recognition system presented in this paper is divided into two primary sections. The music feature extraction module, shown in the dotted box on the left, utilizes the improved PCP features discussed before. The model module, depicted in the dashed box on the right, focuses on the HMM model and the creation of automated music tags. The details of the model module will be further explained in the following sections.

First, the user uploads audio and transmits it to the back-end via Axios' XMLHttp Request. send method; the back-end receives the request and begins to reason about the audio through format detection and returns the inference results; the front-end displays the inference results and can synchronize with the results of playing the audio; the user can choose to download the pipe using the VUE download File method; the orchestra plays the music or re-uploads the music; the user can choose to download the pipe using the VUE download File method. Users can choose to download the music played by the orchestra or re-upload the music by using the VUE download File method. The processing state is shown in Fig. 9.

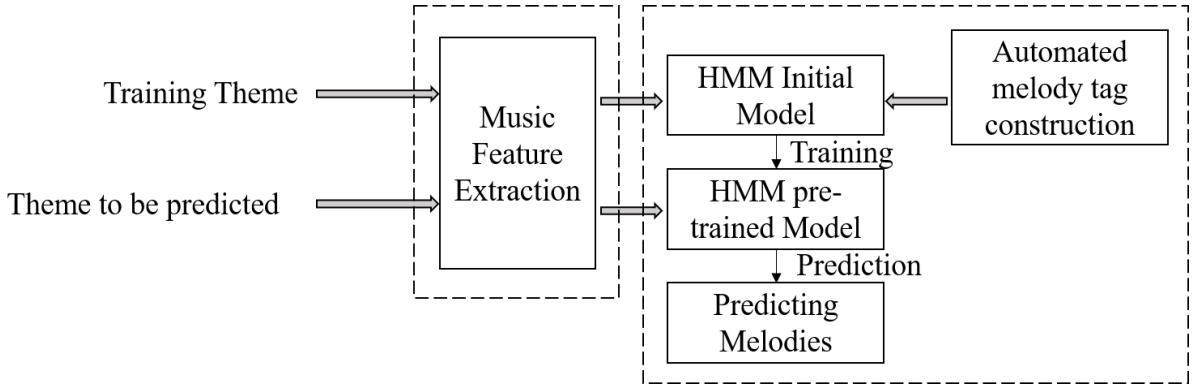


Fig. 8. Framework diagram of automatic music recognition system.

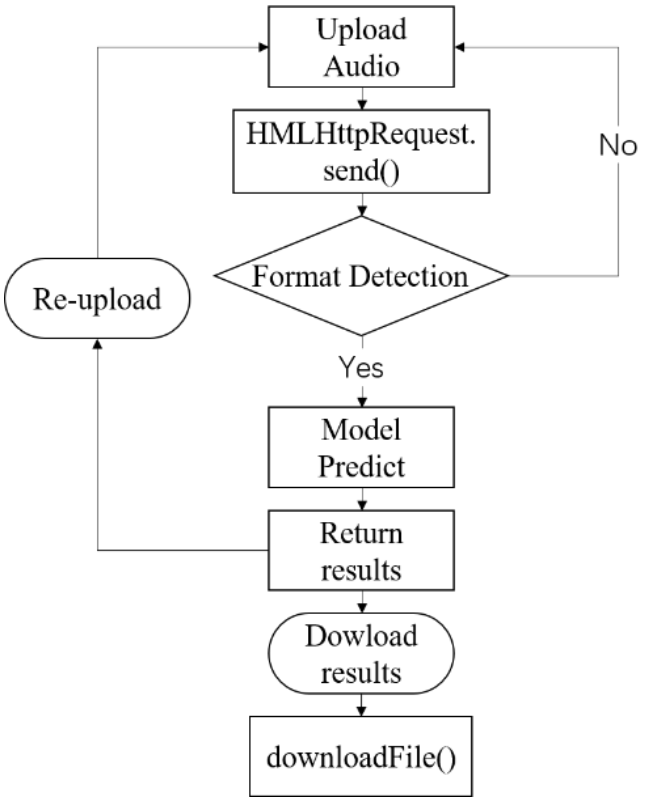


Fig. 9. Platform process.

V. Results and Discussion

A. Experimental Data Collection

The source data for this paper consists of 455 MIDI music files obtained from an online MIDI music library. Of these, 450 files are used for training and 5 for testing. The datasets are preprocessed to extract the main melody and accompaniment tracks. The main melody is converted to WAV format for feature extraction, while the accompaniment is labeled using the proposed automatic music label construction method. Both data types are named consistently for model training purposes.

The improved PCP feature vector for the main melody, as proposed in this paper, is utilized for model feature extraction and serves as the observation vector for the HMM. To assess the robustness of the system, experiments were conducted to evaluate its performance on real-world audio recordings with background noise and microphone variations. The results indicated that while the system performed well under controlled conditions, its accuracy decreased in the presence of significant noise and variations, suggesting the need for further refinement and noise reduction techniques. The HMM consists of six states, excluding the initial and termination states. Each active state employs a single Gaussian observation function with a diagonal matrix, an average vector, and a change vector. After training the model, five files are randomly selected from the test dataset. The improved PCP feature vectors are extracted and inputted into the model for wind music prediction, and the predicted sequences are recorded. These steps are then repeated using traditional PCP features as observation vectors for comparison purposes.

To evaluate the accuracy of the predicted wind music, the results are compared against the correct harmonic sequences determined by professional music researchers using established music theory. The accuracy of the system's recognition is mathematically defined as follows:

$$P_{true} = \frac{N_{sum} - N_{false}}{N_{sum}}$$

In the above formula, N_{sum} represents the total number of accompanied piped tunes of a single tested music file, and N_{false} represents the number of incorrectly recognized piped tunes, and their difference represents the meaning, which is equal to the number of piped tunes that appear to be different in the results of all the piped tunes generated by the system recognition of the tested music file in this paper, compared to the results of all the piped tunes obtained by manual recognition.

B. Tests Results

Statistics of the correctness of the five test music files recognized by the system for wind music tunes were obtained, and the data of the experimental results are shown in the following Table II.

TABLE II. Comparison of Traditional PCP and Modified PCP Results

Training data	Test data (piece name)	System Type	Recognition Accuracy
Music files in the MIDI music library	Liberty Bell March	Legacy PCP+HMM	79.32
		Revised PCP+HMM	84.77
	British Grenadiers March	Legacy PCP+HMM	76.41
		Revised PCP+HMM	82.52
	El Capitan	Legacy PCP+HMM	72.76
		Revised PCP+HMM	75.22
	Entry of the Gladiators	Legacy PCP+HMM	78.01
		Revised PCP+HMM	84.29
	The Thundered	Legacy PCP+HMM	72.36
		Revised PCP+HMM	74.47

The data presented in the table indicates that the improved PCP features used in this study enhance the accuracy of wind music recognition compared to traditional PCP features. Specifically, the experimental results show that the improved PCP features increase the recognition accuracy for the pieces "Liberty Bell March," "British Grenadiers March," and "Entry of the Gladiators" by 5.25%, 6.11%, and 6.28%, respectively. For the pieces "El Capitan" and "The Thundered," the recognition accuracy improved by 2.46% and 2.11%, respectively. Overall, the improved PCP features proposed in this study provide better recognition performance for wind music than the traditional PCP features.

Additionally, to comprehensively evaluate the performance of the wind music recognition system designed in this paper, a traditional template matching model was used for control analysis. This comparison helps to further validate the effectiveness of the improved PCP features and the overall recognition system. The experimental results obtained from the final analysis are shown in the following Table III.

TABLE III. COMPARISON OF SYSTEM SYNTHESIS RESULTS

Training data	Test data (piece name)	System Type	Recognition Accuracy
Music files in the MIDI music library	Liberty Bell March	Legacy PCP+ Template Matching	74.32
		Revised PCP+HMM	84.36
	British Grenadiers March	Legacy PCP+ Template Matching	72.89
		Revised PCP+HMM	82.66
	El Capitan	Legacy PCP+ Template Matching	69.02
		Revised PCP+HMM	74.87
	Entry of the Gladiators	Legacy PCP+ Template Matching	72.31
		Revised PCP+HMM	83.85
	The Thundered	Legacy PCP+ Template Matching	68.38
		Revised PCP+HMM	73.94

The data from the table indicates that the HMM model combined with the improved PCP features significantly outperforms the template matching model in terms of pipe music recognition accuracy. Specifically, when comparing the results of the pipe music recognition system using the improved PCP features and the HMM model to those using traditional PCP features and template matching, the recognition accuracy improved by 9.55%, 9.77%, and 11.25% for "Liberty Bell March," "British Grenadiers March," and "Entry of the Gladiators," respectively. For "El Capitan" and "The Thundered," the improvements were 5.35% and 5.56%, respectively. These results demonstrate that the HMM model provides better performance compared to template matching.

However, a closer look at the data reveals that "El Capitan" and "The Thundered" have lower overall recognition rates compared to the other three songs. Neither the improved PCP features nor the use of the HMM model had a substantial impact on these pieces, resulting in relatively low recognition accuracy. To understand the reasons behind this, a further analysis of the accompanying wind music sequences derived from the test set music files by professionals was conducted.

This analysis revealed that the primary reason for the decreased recognition rate in these pieces is related to the complexity of the wind music analysis. Specifically, the wind music involved in this study often contains repeated sound values, which are particularly challenging to recognize accurately. For example, both the F major wind piece [F, A, C] and the D minor piece [D, F, A] share similar tonal components, leading to frequent recognition errors. This issue is a significant factor contributing to the lower recognition rates observed for these songs.

As illustrated in Fig. 10 and 11, the first two tone values in F major and the last two-tone values in D minor are identical. This similarity can cause the system to misidentify sections of the main melody, confusing F major with D minor due to their high degree of resemblance, thus affecting the overall recognition accuracy. Similarly, Fig. 12 and 13 show that C major and A minor wind music also share common features, both containing C and E as root notes. This overlap makes it challenging to distinguish between these keys during the recognition process. In the pieces with lower recognition rates, such as those involving C major, A minor, F major, and D minor, these shared tonal characteristics lead to frequent misidentifications, resulting in decreased recognition accuracy compared to other pieces.

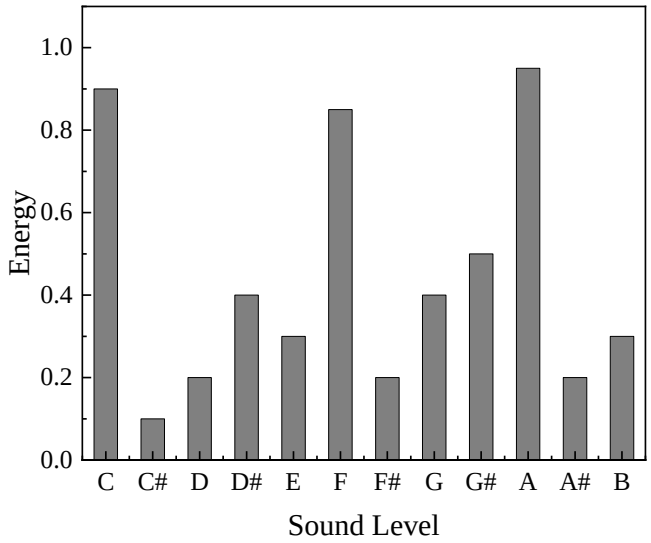


Fig. 10. F-major's PCP feature template.

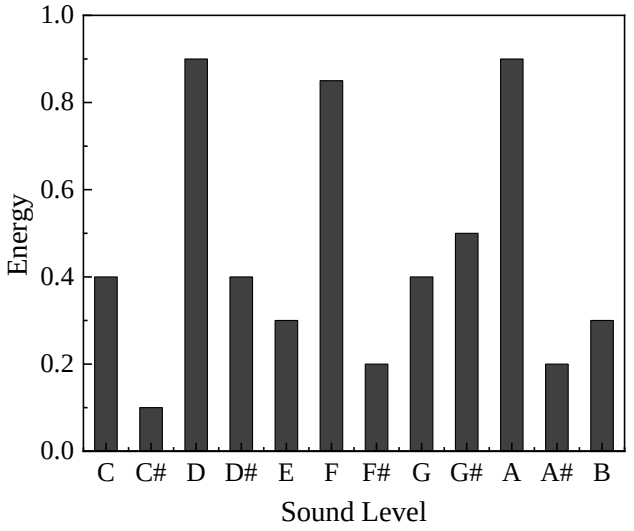


Fig. 11. D-minor's PCP feature template.

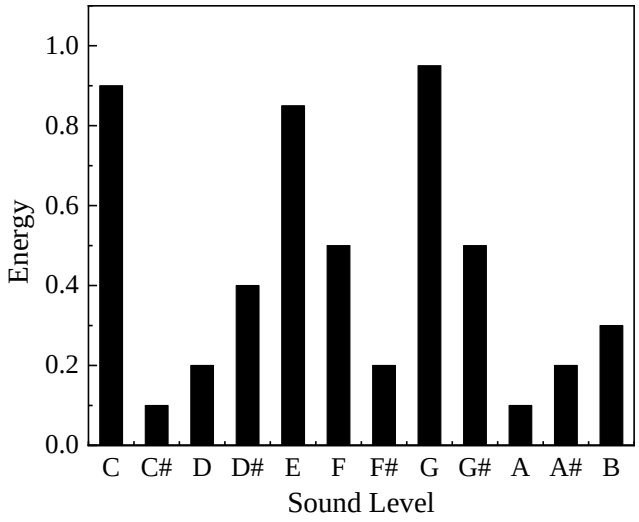


Fig. 12. C-major's PCP feature template.

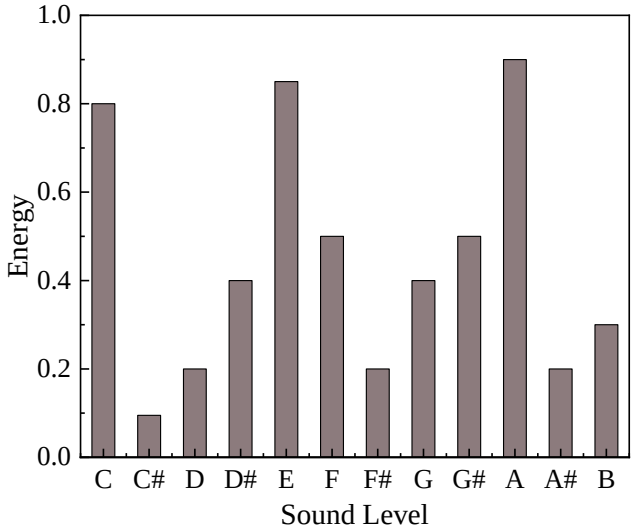


Fig. 13. A-minor's PCP feature template.

VI. Conclusion

This paper presents a comprehensive study on the automatic recognition of marching band performances, utilizing advancements in music information retrieval and signal processing. The research focused on overcoming the limitations of traditional music feature extraction methods by introducing an improved Pitch Class Profile (PCP) feature extraction technique. By mapping audio signals to musical notes more accurately, the improved PCP features, combined with Hidden Markov Models (HMM), provided a robust framework for recognizing and sequencing marching band performances.

The system developed in this study consisted of two main components: a music feature extraction module and an HMM-based modeling module. The feature extraction module used the improved PCP features to convert audio signals into analyzable data, while the modeling module applied HMMs to decode these features into recognizable performance sequences. The system was trained on a dataset of 450 MIDI music files and tested on an additional five files. Experimental results demonstrated that the improved PCP features significantly enhanced the recognition accuracy compared to traditional PCP features and template matching models. Recognition rates improved by up to 6.28% for various marching band pieces. However, some pieces, such as "El Capitan" and "The Thundered," had lower recognition rates due to the presence of similar tonal values, highlighting the complexity of music recognition. This research contributes to the field of music information retrieval by providing an enhanced feature extraction method and a robust modeling framework. The findings have practical implications for developing automated music recognition systems, which can be applied in music education, digital archiving, and cultural preservation.

Future work should address the challenges of overlapping tonal values by developing more sophisticated feature extraction methods. Additionally, testing the system on real-world audio recordings with background noise and microphone variations will be crucial to enhance its practical applicability. Expanding the dataset to include various musical styles and real-world audio recordings would provide a more comprehensive evaluation of the system's robustness and versatility.

References

[1] Keller, Robert, et al. "Automating the explanation of jazz chord progressions using idiomatic analysis." Computer Music Journal 37.4 (2013): 54-69.

[2] Qi, Yuting, John William Paisley, and Lawrence Carin. "Music analysis using hidden Markov mixture models." IEEE Transactions on Signal Processing 55.11 (2007): 5209-5224.

[3] Ajmera, Jitendra, Iain McCowan, and Herve Bourlard. "Speech/music segmentation using entropy and dynamism features in a HMM classification framework." Speech communication 40.3 (2003): 351-363.

[4] Vincent, Emmanuel, and Xavier Rodet. "Music transcription with ISA and HMM." Independent Component Analysis and Blind Signal Separation: Fifth International Conference, ICA 2004, Granada, Spain, September 22-24, 2004. Proceedings 5. Springer Berlin Heidelberg, 2004.

[5] Shibata, Go, Ryo Nishikimi, and Kazuyoshi Yoshii. "Music Structure Analysis Based on an LSTM-HSMM Hybrid Model." ISMIR. 2020.

[6] Nishikimi, Ryo, et al. "Audio-to-score singing transcription based on a CRNN-HSMM hybrid model." APSIPA Transactions on Signal and Information Processing 10 (2021): e7.

[7] Calvo-Zaragoza, Jorge, Alejandro H. Toselli, and Enrique Vidal. "Hybrid hidden Markov models and artificial neural networks for handwritten music recognition in mensural notation." Pattern Analysis and Applications 22 (2019): 1573-1584.

[8] Mor, Bhavya, Sunita Garhwal, and Ajay Kumar. "MIMVOGUE: modeling Indian music using a variable order gapped HMM." Multimedia Tools and Applications 80 (2021): 14853-14866.

[9] Li, Tao, et al. "Music sequence prediction with mixture hidden markov models." 2019 IEEE International Conference on Big Data (Big Data). IEEE, 2019.

[10] Qian, Guo. "A music retrieval approach based on hidden markov model." 2019 11th International Conference on Measuring Technology and Mechatronics Automation (ICMTMA). IEEE, 2019.

[11] Chen, Yanjiao. "Automatic classification and analysis of music multimedia combined with hidden markov model." Advances in Multimedia 2021 (2021): 1-7.

[12] Nishikimi, Ryo, et al. "Bayesian singing transcription based on a hierarchical generative model of keys, musical notes, and f0 trajectories." IEEE/ACM Transactions on Audio, Speech, and Language Processing 28 (2020): 1678-1691.

[13] Ntalampiras, Stavros, and Ilyas Potamitis. "A statistical inference framework for understanding music-related brain activity." IEEE Journal of Selected Topics in Signal Processing 13.2 (2019): 275-284.

[14] Uehara, Yui, Eita Nakamura, and Satoshi Tojo. "Chord function identification with modulation detection based on HMM." Perception, Representations, Image, Sound, Music: 14th International Symposium, CMMR 2019, Marseille, France, October 14–18, 2019, Revised Selected Papers 14. Springer International Publishing, 2021.

[15] Mor, Bhavya, Sunita Garhwal, and Ajay Kumar. "A systematic literature review on computational musicology." Archives of Computational Methods in Engineering 27 (2020): 923-937.

[16] Shibata, Kentaro, et al. "Joint transcription of lead, bass, and rhythm guitars based on a factorial hidden semi-Markov model." ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2019.

[17] Wang, Changhong, et al. "HMM-based glissando detection for recordings of Chinese bamboo flute." (2019).

[18] Nápoles López, Néstor, Claire Arthur, and Ichiro Fujinaga. "Key-finding based on a hidden Markov model and key profiles." Proceedings of the 6th International Conference on Digital Libraries for Musicology. 2019.

[19] Ens, Jeff, and Philippe Pasquier. "Mmm: Exploring conditional multi-track music generation with the transformer." arXiv preprint arXiv:2008.06048 (2020).

[20] Ycart, Adrien, et al. "Blending acoustic and language model predictions for automatic music transcription." (2019).

[21] Mor, Bhavya, Sunita Garhwal, and Ajay Kumar. "A systematic review of hidden Markov models and their applications." Archives of computational methods in engineering 28 (2021): 1429-1448.

[22] Brancatisano, Olivia, Amee Baird, and William Forde Thompson. "A 'music, mind and movement'program for people with dementia: Initial evidence of improved cognition." Frontiers in psychology 10 (2019): 445451.

[23] Hernandez-Olivan, Carlos, and Jose R. Beltran. "Music composition with deep learning: A review." Advances in speech and music technology: computational aspects and applications (2022): 25-50.

[24] Plut, Cale, et al. "PreGLAM-MMM: Application and evaluation of affective adaptive generative music in video games." Proceedings of the 17th International Conference on the Foundations of Digital Games. 2022.

[25] Bello J P, Pickens J. A Robust Mid-Level Representation for Harmonic Content in Music Signals[C]//ISMIR. 2005, 5: 304-311.

[26] Foote J T. Content-based retrieval of music and audio[C]//Multimedia storage and archiving systems II. SPIE, 1997, 3229: 138-147.

[27] Logan B. Mel frequency cepstral coefficients for music modeling[C]//Ismir. 2000, 270(1): 11.

[28] Müller M. Information retrieval for music and motion[M]. Heidelberg: Springer, 2007.

[29] Sheh A, Ellis D P W. Chord segmentation and recognition using EM-trained hidden Markov models[J]. 2003.

[30] Wu Y, Carsault T, Yoshii K. Automatic chord estimation based on a frame-wise convolutional recurrent neural network with non-aligned annotations[C]//2019 27th European Signal Processing Conference (EUSIPCO). IEEE, 2019: 1-5.

Fitness Equipment Design Based on Web User Text Mining

Jinyang Xu[1], Xuedong Zhang[2]*, Xinlian Li[3], Shun Yu[4], Yanming Chen[5]
School of Design, Anhui Polytechnic University, Anhui, Wuhu, 24100, China[1, 2, 4, 5]
International Institute of Creative Design, Shanghai University of Engineering Science, Shanghai, 200000, China[3]

***Abstract*—To propose home fitness equipment that meets modern users' needs, this study employs web user text mining, combined with the Fuzzy Analytic Hierarchy Process (FAHP) and the Technique for Order Preference by Similarity to Ideal Solution (TOPSIS), to design and evaluate home fitness equipment that aligns with contemporary demands. First, we used crawler data to collect user reviews of home fitness equipment from a well-known Chinese shopping platform. The data were cleaned and processed to extract key user needs and preferences. Next, the FAHP method was used to prioritize these requirements, and TOPSIS was applied for the comprehensive evaluation of design proposals. This process allowed us to identify the solution that best meets user needs, completing the development of the product design. The results indicate that the second design, with its features targeting lumbar health, efficient space utilization, rich interactive experience, integration of smart technology, and minimalist appearance, has significant market potential and social value. Finally, the SUS (System Usability Scale) was used to validate the design, showing excellent user satisfaction and usability for the second scheme. This study establishes a design process incorporating web scraping, FAHP, and TOPSIS, demonstrating the effectiveness of this theoretical integration in the field of home fitness equipment design.**

***Keywords*—Home fitness equipment; crawler data; FAHP; TOPSIS; product design**

I. INTRODUCTION

In contemporary society, especially after the COVID-19 pandemic, health has become a significant topic of concern for the public [1] [2] [3]. Fitness equipment can enhance users' physical health, such as fitness levels, muscle strength, and weight loss, as well as their mental well-being. In China, fitness venues are generally located in outdoor parks or gyms. Outdoor parks are open fitness venues that have the advantage of being accessible to people of all ages. However, their disadvantages include high usage rates and difficulty of use in extreme temperatures. Gyms, on the other hand, offer more specialized equipment but require users to pay for access. Home fitness equipment combines the advantages of both outdoor parks and gyms while addressing their shortcomings. The types of equipment used in homes are similar to those found in gyms, such as treadmills, stair mills, rowing machines, and spin bikes, which support a variety of aerobic exercise [4]. In contemporary society, especially after the COVID-19 pandemic, Consumers bought equipment for home use and switched to different types of online or outdoor workouts [5]. In the current market context, China's home fitness equipment industry has significant growth potential. However, despite the promising market outlook, the industry still struggles to fully meet users' fitness expectations [6],family fitness equipment has a lot of research space in the Chinese market.A study on the design of a fitness furniture and its finite element analysis.To address this issue, our research team has decided to start by focusing on users' needs [7] [8].In the design of indoor fitness equipment, it is crucial to focus on and analyze users' needs，user characteristic analysis emerges as a pivotal step within the realms of product design and enhancement [7].

Data is very important for businesses and organizations as it assists their decision making [9].In the design of indoor fitness equipment, it is crucial to focus on and analyze users' needs. Using web crawlers to gather information about users' needs for indoor fitness equipment is an effective approach. A web crawler is a program or script that automatically captures web information based on specific rules, enabling automatic data collection. Web crawlers are widely used in various fields, such as crawling and indexing sites for search engines, collecting data for analysis and mining, and gathering financial data for financial analysis. Ramachandran et al. [10] used generalized logistic regression to analyze Amazon datasets, confirming the presence of negative bias in online consumer reviews. They found that negative emotions in review texts influence product ratings more significantly than positive emotions of the same intensity. This suggests that in all management decisions, interventions to reduce negative performance disconfirmation should be prioritized over those causing positive performance disconfirmation. Cao et al. [11] used web crawlers to collect a large number of product reviews, conducted word frequency analysis to identify key product elements, and categorized reviews based on the results. They extracted adjectives from the reviews and, after expert summarization, performed sentiment analysis on sentences containing adjectives to obtain information on user needs. This information can guide the generation of subsequent design plans through the application of kansei engineering principles.

After extracting the information, we can identify different user needs by analyzing user comments and summarizing keywords, thereby determining the characteristics of user requirements. Data cleaning can reveal key information from user evaluations, which can be transformed into decision criteria for the Analytic Hierarchy Process (AHP). AHP is a systematic method for multi-criteria, multi-option decision-making, involving decomposition, comparative judgment, and synthesis for decision-making (weighting) and overall ranking. The Fuzzy Analytic Hierarchy Process (FAHP) improves and extends traditional AHP by incorporating fuzzy mathematics theory,

making judgments more accurate and applicable to a broader range of scenarios. Using data obtained from web crawlers, we can introduce FAHP theory to handle the fuzziness and uncertainty in data when solving complex problems. For example, Wang et al. [12] used FAHP to identify key factors in the design of bone marrow puncture needles after conducting surveys and expert interviews. The final design effectively improved patient experience during surgery. Mouhassine et al. [13] proposed integrating Fuzzy-AHP with VIKOR in SDN controllers, successfully applying it to optimize the wireless network handover process. FAHP effectively addresses fuzziness and uncertainty in evaluations, using fuzzy linguistic variables to express preferences for evaluation factors, thus enhancing the objectivity and accuracy of the assessments. Wu et al. [14] combined FAHP with a continuous fuzzy Kano model to prioritize attractiveness factors for electric scooters, demonstrating that this approach reliably meets consumer perception needs.

Based on FAHP hierarchical ranking, the optimal design scheme must be determined. The TOPSIS method can fully utilize the information from raw data, with the analysis results accurately reflecting the gaps between evaluation schemes. Zulkefli et al. [15] developed a more robust and effective CSP selection model by combining TOPSIS, entropy-based weight determination, and Single Valued Neutrosophic (SVN) handling of uncertainty, highlighting its contributions to solving RRP and decision-making ambiguity. Liu et al. [16] proposed the Z-AHP and Z-TOPSIS theories to optimize the design of kitchen waste containers. Z-TOPSIS considers the fuzziness of evaluation criteria and the confidence of decision-makers, making the assessment results more reasonable and reliable. Hameed et al. [17] integrated FMEA, QFD, TRIZ, LCA, and fuzzy TOPSIS to develop sustainable products. They used fuzzy TOPSIS to reassess designs, and the final design was selected for prototyping.

Currently, many scholars have used web crawlers, FAHP, and TOPSIS methods in their research. However, these methods have not yet been applied to the design of indoor fitness equipment, highlighting the scientific and innovative nature of this study. The remainder of this paper is structured as follows: Section II describes the design experiments based on the proposed methods. Section III presents the results and discussion of the indoor fitness equipment design, determining the feasibility of the final scheme. Section IV summarizes the experimental process of this study.

II. METHOD

A. Crawler Data Collection

To delve deeper into Chinese users' opinions and feedback on indoor fitness equipment, this study selected a purchasing platform with a large number of active users in China as the data source. Special attention was given to user reviews of indoor fitness equipment purchases, as these reviews are crucial for reflecting overall evaluations and expectations of fitness equipment. They provide insights into user satisfaction, user experience, and suggestions for improvement.

To ensure the accuracy and effectiveness of the study, data collection focused on key areas such as review content, descriptions related to fitness equipment, and the number of reviewers. This foundation supports subsequent research. The overall experimental process is shown in Fig. 1.

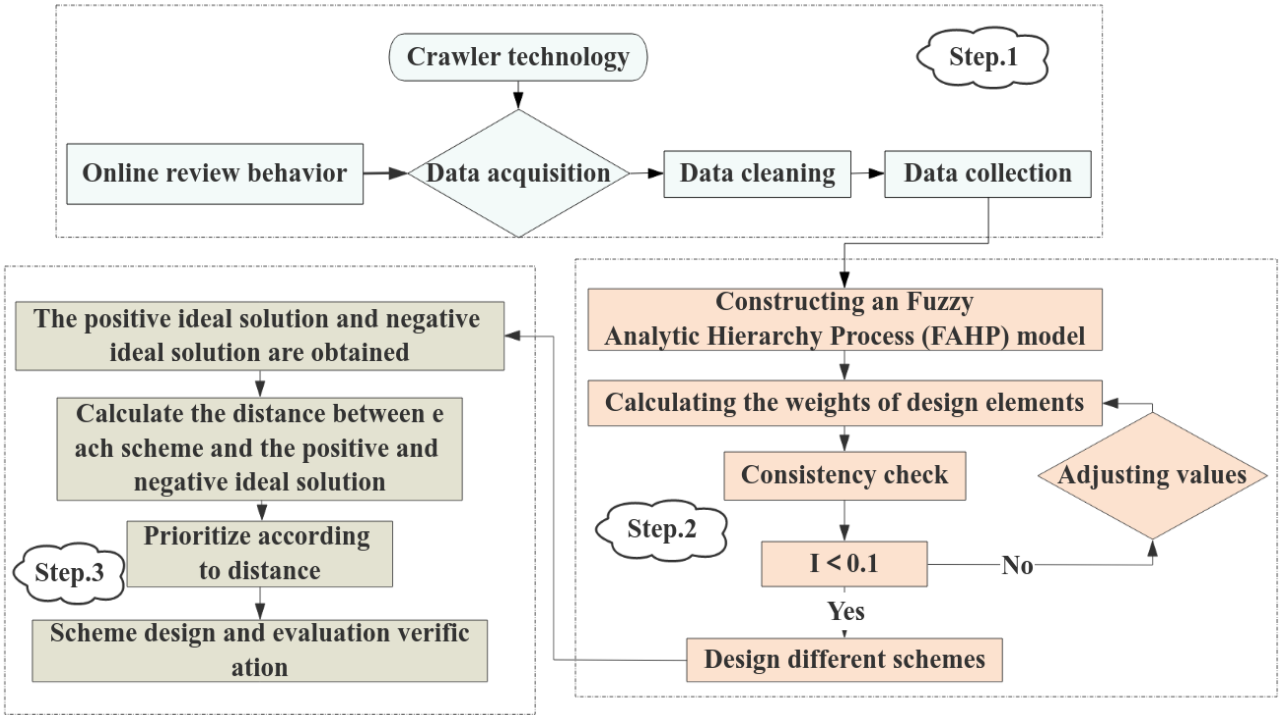


Fig. 1. Experimental design flowchart.

B. Procedure Design

The main processes include data collection (via web scraping), data cleaning, and data visualization analysis, detailed as follows:

1) Data collection: Using Python web scraping technology, information about indoor fitness equipment is collected from a well-known Chinese shopping platform.

2) Data cleaning: Filtering the collected data to organize the relevant comment content.

3) Data collection: Analyzing the cleaned data and employing charts to visualize user evaluation data from multiple perspectives, thereby enhancing the understanding and presentation of the data. After organizing the collected data, the following results were obtained: Very (3682), It's good (2645), Installation (2182), Quality (2085), Convenience (1701), Exercise (1572), Simple (1114), Logistics (995), Packaging (997)……Kids (100), Accessories (100), Weight (100).

By extracting keywords with a frequency exceeding 100, a total of 352 keywords were identified, which were then visualized as shown in Fig. 2. These keywords need to be introduced into the FAHP for evaluation, and the data must be filtered to exclude unnecessary influencing factors. Based on the results, a refined set of evaluation criteria can ultimately be determined.

C. Design Transformation based on FAHP

The Analytic Hierarchy Process (AHP) is a commonly used multi-criteria decision-making (MCDM) technique initially proposed by Saaty [18]. Combining the fuzzy matrix with the Analytic Hierarchy Process allows for addressing the fuzziness and uncertainty among various factors when dealing with complex problems. This method of combining AHP with fuzzy mathematical theory is called the Fuzzy Analytic Hierarchy Process (FAHP), which aims to address the human subjective effect of ambiguity on the decision factors of a problem [19]. Firstly, the user reviews obtained through web scraping need to be summarized and analyzed. The team invited five fitness equipment development engineers to an online meeting to filter the data. Then, a fuzzy analytic hierarchy process matrix is constructed. The matrix is divided into the goal layer (A) Home Fitness Equipment Program, the criteria layer (B) Appearance, the scheme layer (C) Functional Requirement, the scheme layer (D) Economy, and the scheme layer (E) Quality of Experience. The sub-criteria layer includes (B1) Fine Workmanship, (B2) Simplicity, (B3) Household Size, (C1) Multipurpose Uses, (C2) Adjustable, (C3) Intelligent Manipulation, (D1) Quality-price ratio, (D2) Fructification, (D3) Wear-well, (D4) Environmental protection, (E1) Safety, (E2) Simple operation, (E3) Relax the body, (E4) Relax the body, (E5) Ergonomics Compliance as shown in Fig. 3.

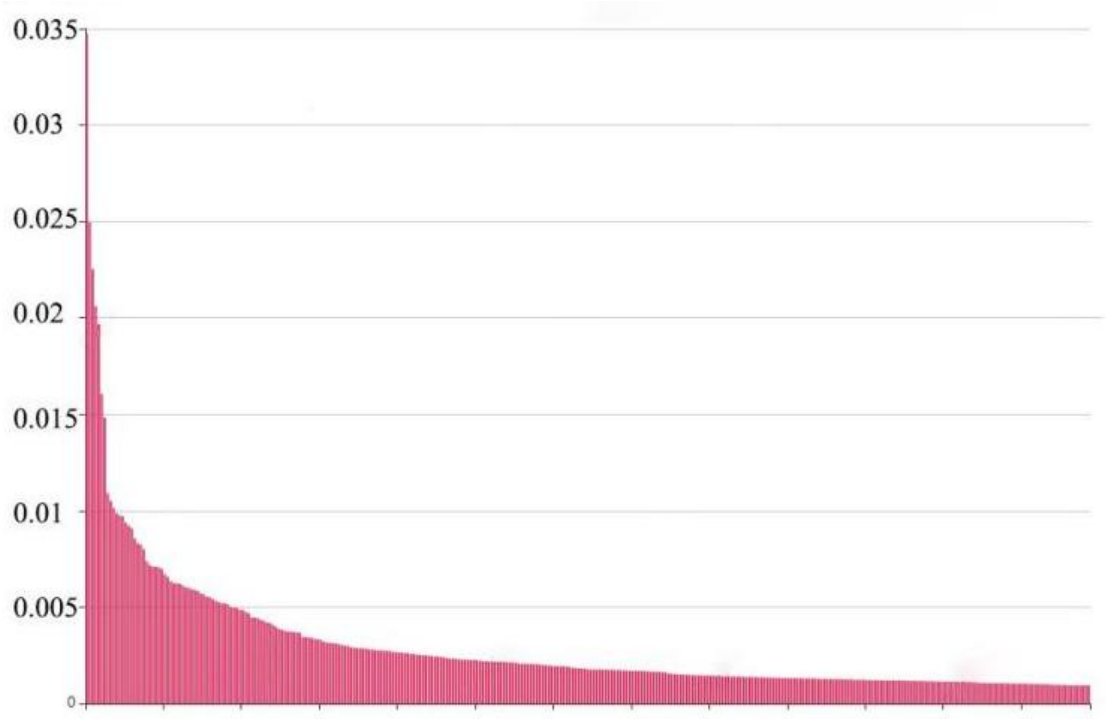


Fig. 2. Data cleaning visualization chart.

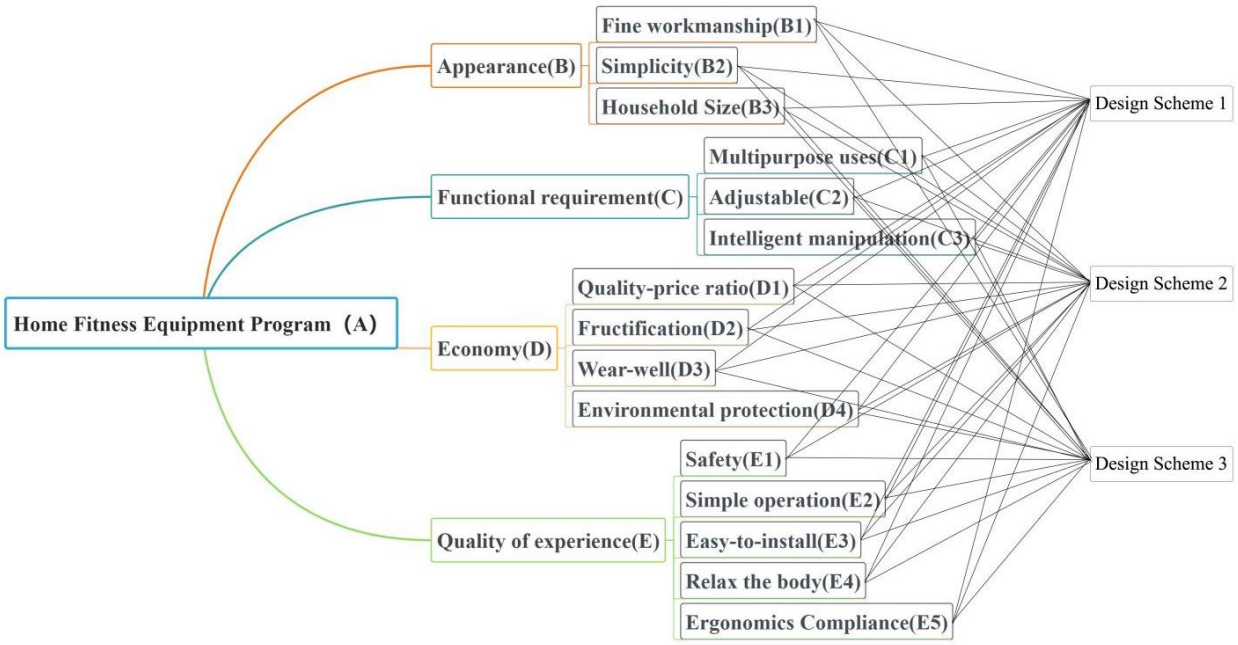


Fig. 3. Fuzzy analytic hierarchy process model of user demand.

D. Constructing Fuzzy Judgment Matrix

Referring to the scaling method of 0.1 to 0.9 levels, a fuzzy judgment matrix is constructed [20] [21], as shown in Table I. Pairwise comparison judgments of different elements of home fitness equipment are made to construct the judgment matrix, as shown in Eq. (1), and its properties are given in Eq. (2) and (3).

TABLE I. INDEX IMPORTANCE SCALE OF FUZZY JUDGMENT MATRIX

Scale	Level of importance	Implication
0.5	Equally important	Indicator a and indicator b are equally important
0.6	Slightly important	Indicator a is marginally more important than indicator b
0.7	Significantly important	Indicator a is significantly more important than indicator b
0.8	Very important	Indicator a is very important compared to indicator b
0.9	Absolutely important	Indicator a is more important than indicator b
0.1,0.2,0.3,0.4	Anti-comparison	Factor b inverse comparison, $q_{ab} = 1-q_{ab}$

$$A=(a_{ij})n\times n=\begin{bmatrix} a_{11} & a_{12} & \cdots\cdots & a_{1n} \\ a_{21} & a_{22} & \cdots\cdots & a_{2n} \\ \cdots\cdots & \cdots\cdots & \cdots\cdots & \cdots\cdots \\ a_{n1} & a_{n2} & \cdots\cdots & a_{nn} \end{bmatrix} \quad (1)$$

$$a_{ij}+a_{ji}=1 \quad (2)$$

$$a_{ii}=0.5 \quad (3)$$

To achieve more objective weight results for each indicator characteristic, the experiment invited seven experts to hold a discussion meeting (including two fitness coaches, three graduate students in design studies, and two fitness equipment merchants) [22] [23]. An expert discussion meeting was organized in the laboratory, and the experts scored the different factors based on the discussion results. The judgment matrices for each level are shown in Tables II to VI. The weights of each fuzzy judgment matrix were calculated according to Eq. (4).

TABLE II. A-LEVEL FUZZY JUDGMENT MATRIX

A	B	C	D	E
B	0.5	0.3	0.4	0.3
C	0.7	0.5	0.6	0.4
D	0.6	0.4	0.5	0.3
E	0.7	0.6	0.7	0.5

TABLE III. B-LEVEL FUZZY JUDGMENT MATRIX

B	B1	B2	B3
B1	0.5	0.5	0.2
B2	0.5	0.5	0.3
B3	0.8	0.7	0.5

TABLE IV. C-LEVEL FUZZY JUDGMENT MATRIX

C	C1	C2	C3
C1	0.5	0.7	0.5
C2	0.3	0.5	0.3
C3	0.5	0.7	0.5

TABLE V. D-LEVEL FUZZY JUDGMENT MATRIX

D	D1	D2	D3	D4
D1	0.5	0.6	0.7	0.8
D2	0.4	0.5	0.5	0.6
D3	0.3	0.5	0.5	0.6
D4	0.2	0.4	0.4	0.5

TABLE VI. E-LEVEL FUZZY JUDGMENT MATRIX

E	E1	E2	E3	E4	E5
E1	0.5	0.5	0.7	0.7	0.5
E2	0.5	0.5	0.3	0.6	0.4
E3	0.3	0.7	0.5	0.6	0.3
E4	0.3	0.4	0.4	0.5	0.3
E5	0.5	0.6	0.7	0.7	0.5

The calculated weight vectors of the fuzzy judgment matrices are as follows: 0.208, 0.267, 0.233, 0.292.

The calculated weight vectors of the fuzzy judgment matrices are as follows: 0.367, 0.267, 0.367.

The calculated weight vectors of the fuzzy judgment matrices are as follows: 0.283, 0.300, 0.417.

The calculated weight vectors of the fuzzy judgment matrices are as follows: 0.300, 0.250, 0.242, 0.208.

The calculated weight vectors of the fuzzy judgment matrices are as follows: 0.220, 0.190, 0.195, 0.170, 0.225.

$$w_i = \frac{\sum_{j=1}^{n} a_{ij} + \frac{n}{2} - 1}{n(n-1)}, i = 1,2,\ldots\ldots n \tag{4}$$

Where $\sum_{j=1}^{n} a_{ij}$ is the sum of the elements in the i-throw.

E. Consistency Check of the Fuzzy Judgment Matrix

To ensure the rigor of the calculation results, a consistency check of the fuzzy judgment matrix is required, as shown in Eq. (5).

$$W_i^* = \frac{Wi}{W_i + W_j}, (i, j = 1,2,\ldots\ldots n) \tag{5}$$

Using the weight vectors, construct the characteristic matrix $W=(W_{ij})n \times n$ of the fuzzy judgment matrix. Then, the compatibility index of the fuzzy judgment matrix with its eigenvalue matrix is calculated using Eq. (7). If the compatibility index $(I) \leqslant 0.1$, the fuzzy judgment matrix is considered reasonable.

$$I(A,W^*) = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n}\left|a_{ij} + w_{ij} - 1\right|}{n^2}, (i, j = 1,2,\ldots\ldots n) \tag{6}$$

The calculated characteristic matrix is as follows:

$$W_A^* = \begin{bmatrix} 0.500 & 0.439 & 0.472 & 0.417 \\ 0.561 & 0.500 & 0.533 & 0.478 \\ 0.528 & 0.467 & 0.500 & 0.444 \\ 0.583 & 0.522 & 0.556 & 0.500 \end{bmatrix}$$

$$W_B^* = \begin{bmatrix} 0.500 & 0.486 & 0.404 \\ 0.514 & 0.500 & 0.419 \\ 0.595 & 0.581 & 0.500 \end{bmatrix}$$

$$W_C^* = \begin{bmatrix} 0.500 & 0.579 & 0.500 \\ 0.421 & 0.500 & 0.421 \\ 0.500 & 0.579 & 0.500 \end{bmatrix}$$

$$W_D^* = \begin{bmatrix} 0.500 & 0.545 & 0.554 & 0.590 \\ 0.454 & 0.500 & 0.508 & 0.545 \\ 0.446 & 0.492 & 0.500 & 0.537 \\ 0.410 & 0.455 & 0.463 & 0.500 \end{bmatrix}$$

$$W_E^* = \begin{bmatrix} 0.500 & 0.537 & 0.530 & 0.564 & 0.494 \\ 0.463 & 0.500 & 0.494 & 0.528 & 0.458 \\ 0.470 & 0.506 & 0.500 & 0.534 & 0.464 \\ 0.436 & 0.472 & 0.466 & 0.500 & 0.430 \\ 0.506 & 0.542 & 0.536 & 0.570 & 0.500 \end{bmatrix}$$

The calculated compatibility indices:

$I(A,W_A^*)=0.07696<0.1$,

$I(B,W_B^*)=0.07503<0.1$,$I(C,WC^*)=0.0538<0.1$,

$I(D,W_D^*)=0.06706<0.1$,

$I(E,W_E^*)=0.08256<0.1$.

All these fuzzy judgment matrices have compatibility indices less than 0.1, thus passing the consistency check and confirming that the data is reliable and valid.

F. Calculating FAHP Comprehensive Index Weights

Based on the hierarchical results of the weight vectors from the fuzzy judgment matrices, the comprehensive weights of each indicator factor in FAHP are obtained and summarized in Table VII. In the design of indoor fitness equipment, E (Quality of experience) > C (Functional requirement) > D (Economy) > B (Appearance). The design should prioritize the user's experience, as it is the key factor in stimulating purchase intentions. Next is the product's functionality, ensuring the equipment can precisely meet users' exercise needs, providing an efficient and safe fitness experience. Economy follows user experience and functionality, aiming to meet users' low-price demands while ensuring experience and functionality. Although appearance is an aspect that attracts users' attention, it is not the primary focus in the overall design priority.

TABLE VII. FUZZY COMPREHENSIVE WEIGHT RANKING

		Element layer	Weight	Comprehensive weight	Rank
Home Fitness Equipment Program	Appearance（0.208）	B1	0.283	0.0588	9
		B2	0.300	0.0624	8
		B3	0.417	0.0867	3
	Functional requirement（0.267）	C1	0.367	0.0980	1
		C2	0.267	0.0712	4
		C3	0.367	0.0979	2
	Economy（0.233）	D1	0.300	0.0699	5
		D2	0.250	0.0583	10
		D3	0.242	0.0564	12
		D4	0.208	0.0485	15
	Quality of experience（0.292）	E1	0.220	0.0642	6
		E2	0.190	0.0555	13
		E3	0.195	0.0569	11
		E4	0.170	0.0496	14
		E5	0.225	0.0657	7

The comprehensive ranking results are: C1 > C3 > B3 > C2 > D1 > E1 > E5 > B2 > B1 > D2 > E3 > D3 > E2 > E4 > D4. Modular or convertible structures should be considered to support various exercise modes. Advanced intelligent systems should be integrated for control, supporting remote control via mobile apps, personalized training plans, real-time exercise data monitoring, and analysis. Since Chinese home spaces cannot support large equipment, the design needs to be compact and easy to store. The equipment's adjustability should accommodate different family members' height, weight, exercise levels, and goals. In terms of materials, product quality and performance must be ensured while optimizing material selection, production processes, and supply chain management for higher cost-effectiveness. High-strength, wear-resistant materials should be used for key components to ensure equipment stability and durability. Additionally, ergonomic principles should be considered in the design to ensure users maintain correct posture and movement trajectories during exercise. See Fig. 4 to 6 for the design diagram.



Fig. 4. Scheme 1: Multifunctional cardio cycle.



Fig. 5. Scheme 2: Compact multifunctional elliptical trainer.



Fig. 6. Scheme 3: Multifunctional rowing machine.

In the FAHP analysis, user experience holds a weight of 0.292, ranking first, indicating that users typically prioritize the experiential aspect of a product. The functional requirements, with a weight of 0.267, suggest that fitness equipment must meet users' daily needs while also accommodating their specific requirements. Economic considerations rank third, with a weight of 0.233, recommending the use of cost-effective materials in home fitness equipment to minimize the financial burden on consumers. Although aesthetic design ranks lowest, with a weight of 0.208, it should still be considered during the design process to ensure that the product is visually appealing while meeting its primary functional requirements.

G. Comprehensive TOPSIS Design Evaluation

Three design schemes were included in a questionnaire, with the design points of each scheme introduced. The 15 sub-criterion elements in FAHP are treated as whole indicators, thus defining 15 evaluation indicators using these design elements [24]. To ensure the final evaluation results' authenticity and rigor, 10 evaluators were invited to form a decision-making group to score each scheme's indicators using a Likert 7-point scale (1 for very poor, 2 for poor, 3 for slightly poor, 4 for neutral, 5 for good, 6 for very good, and 7 for excellent). The final results were statistically averaged and are shown in Table VIII.

TABLE VIII. INITIAL EVALUATION MATRIX8

	B1	B2	B3	C1	C2	C3	D1	D2	D3	D4	E1	E2	E3	E4	E5
1	0.03 11	0.03 08	0.04 12	0.05 75	0.04 22	0.05 32	0.04 03	0.03 55	0.03 49	0.02 60	0.03 45	0.02 90	0.02 62	0.02 84	0.03 90
2	0.03 58	0.04 22	0.05 96	0.05 66	0.04 38	0.06 44	0.03 90	0.03 32	0.03 27	0.03 08	0.04 12	0.03 54	0.03 99	0.03 10	0.03 75
2	0.03 47	0.03 41	0.04 76	0.05 56	0.03 41	0.05 11	0.04 17	0.03 21	0.02 99	0.02 70	0.03 51	0.03 14	0.03 10	0.02 68	0.03 68

The data of the three collected design schemes were weighted and standardized according to Eq. (7) and (8). The weighted standardized evaluation matrix is shown in Table IX [25].

$$Y_{ij} = \frac{X_{ij}}{\sqrt{\sum_{i=1}^{m} X_{ij}^2}}, (i = 1,2 \ldots\ldots m,\ j = 1,2, \ldots\ldots n) \tag{7}$$

$$Z_{ij} = W_j Y_{ij}, (i = 1,2 \ldots\ldots m,\ j = 1,2, \ldots\ldots n) \tag{8}$$

TABLE IX. WEIGHTED STANDARDIZED EVALUATION

	B1	B2	B3	C1	C2	C3	D1	D2	D3	D4	E1	E2	E3	E4	E5
1	5.3	4.6	4.5	6.1	5.5	5.2	6.1	6.2	6.3	5.5	5.1	5.1	4.4	5.2	6.1
2	6.1	6.3	6.5	6.0	5.4	6.3	5.9	5.8	5.9	6.5	6.1	6.2	6.7	5.6	5.8
3	5.9	5.1	5.2	5.9	4.2	5.0	6.3	5.6	5.4	5.7	5.2	5.5	5.2	4.9	5.7

Calculate the positive and negative ideal solutions of the evaluation objects. The positive ideal solution is calculated according to Eq. (9), and the negative ideal solution is calculated according to Eq. (10).

$$Z^+ = (Z_1^+, Z_2^+, \ldots\ldots Z_n^+) \tag{9}$$

$$Z^- = (Z_1^-, Z_2^-, \ldots\ldots Z_n^-) \tag{10}$$

The positive ideal solution Z^+ = (0.0358, 0.0422, 0.0596, 0.0575, 0.0438, 0.0644, 0.0417, 0.0355, 0.0349, 0.0308, 0.0412, 0.0354, 0.0399, 0.0310, 0.0390) and the negative ideal solution Z^- = (0.0311, 0.0308, 0.0412, 0.0556, 0.0341, 0.0511, 0.0390, 0.0321, 0.0299, 0.0260, 0.0345, 0.0290, 0.0262, 0.0268, 0.0368) for the evaluation objects are calculated.

The optimal solution's distance to the positive and negative ideal solutions Z^+ and Z^- (i.e., D^+ and D^-) is calculated according to Eq. (11), (12), and (13), as shown in Table X.

$$D_i^+ = \sqrt{\sum_{j=1}^{n} (Z^+ - Z_{ij})^2} \tag{11}$$

$$D_i^- = \sqrt{\sum_{j=1}^{n} (Z^- - Z_{ij})^2} \tag{12}$$

$$C_i = \frac{D_i^+}{D_i^+ + D_i^-}, (i = 1,2, \ldots\ldots,\ m) \tag{13}$$

TABLE X. COMPARISON OF EUCLIDEAN DISTANCE AND RELATIVE CLOSENESS

Index	Positive ideal solution distance Di +	Negative ideal solution distance Di-	Relative proximity Ci	Rank
scheme 1	0.030	0.011	0.264	3
scheme 2	0.005	0.033	0.879	1
scheme 3	0.026	0.010	0.278	2

Closeness is an important indicator for evaluating how close a design scheme is to the ideal solution. When the closeness approaches 0, it means the design scheme is closer to the negative ideal solution, indicating poor performance across multiple evaluation criteria and potential significant deficiencies or shortcomings. Conversely, when the closeness approaches 1, it means the design scheme performs excellently across multiple evaluation criteria, meeting or exceeding expected needs and expectations.

III. RESULT

This study comprehensively applied web scraping technology, the Fuzzy Analytic Hierarchy Process (Fuzzy AHP), and the TOPSIS method to accurately identify the most favored home fitness equipment among Chinese people. Through collaboration with fitness experts, the research team developed a comprehensive evaluation model that not only covers user needs, scheme rationality, and optimal scheme selection but also refines the comprehensive consideration of each sub-criterion. Based on the most popular types of fitness equipment in the market, the three home fitness devices designed were primarily focused on multifunctionality. This evaluation criterion (C1) was assigned a comprehensive weight of 0.0980 in the FAHP analysis, ranking it as the top priority and underscoring its critical importance. Additionally, the devices were required to feature intelligent control capabilities and be suitable for home use, with these criteria carrying comprehensive weights of 0.0979 and 0.0867, respectively. Most importantly, user experience was identified as the core requirement, holding a weight of 0.292 in the FAHP analysis. According to the TOPSIS analysis results, the second design achieved a C_i value of 0.879, significantly higher than the other two designs and the closest to 1. Therefore, the second design demonstrates a clear and distinct advantage. To verify the effectiveness of the experimental process, we used the SUS (System Usability Scale) to evaluate this design practice [26]. Based on the keywords extracted from the data, the research team has developed the following ten survey questions.

a) Question 1: I find this fitness equipment suitable for home use.

b) Question 2: I find this fitness equipment difficult to use at home.

c) Question 3: I find this fitness equipment easy to operate.

d) Question 4: I find that expert guidance is necessary to use this fitness equipment.

e) Question 5: I find that this fitness equipment incorporates the features I need.

f)Question 6: I find this fitness equipment to be cumbersome.

g) Question 7: I feel confident that this fitness equipment is safe to use.

h) Question 8: I have concerns about the safety of this fitness equipment.

i)Question 9: I find this fitness equipment aesthetically pleasing.

j)Question 10: I find the design of this fitness equipment does not align with my aesthetic preferences.

We distributed the SUS questionnaires to seven employed individuals and statistically analyzed the collected data, as shown in Table IX. The usability test scores were all $\geqslant$80, which, according to the SUS score curve classification range, is rated as A, as shown in Table XI. The results demonstrate that Scheme Two meets users' preferences and needs, as shown in Fig. 7.

TABLE XI. SUS TEST SCORES

Participant	Q.1	Q.2	Q.3	Q.4	Q.5	Q.6	Q.7	Q.8	Q.9	Q.10	Score
P1	5	1	5	1	4	1	5	2	4	2	90.0
P2	4	2	5	2	4	1	4	1	5	2	85.0
P3	5	2	5	1	5	1	5	1	5	1	97.5
P4	5	4	4	1	5	1	5	1	5	1	90.0
P5	5	2	4	1	4	2	4	1	4	1	85.0
P6	5	1	4	1	4	1	4	1	4	2	87.5
P7	5	2	5	1	5	2	5	1	5	1	95.0

IV. DISCUSSION

In the in-depth discussion of Scheme Two within the group, we identified several significant advantages across multiple dimensions. Firstly, addressing the prevalent concern for lumbar health in today's society, Scheme Two offers a stretching function for the back muscles when not in use, effectively relieving stiffness and discomfort caused by prolonged sitting or standing. Additionally, this scheme provides effective full-body training during workouts, particularly through pedal exercises, fully meeting users' diverse functional needs.

Regarding space utilization, considering the limited living space common in Chinese households, Scheme Two features a miniaturized design that can be flexibly placed in various corners of the home, making it easy to store without occupying much space. This greatly enhances the convenience and comfort of home life. The introduction of an intelligent control system perfectly integrates modern technology with healthy living. Users can connect via Bluetooth to track exercise data in real-time, including workout progress and energy consumption, providing scientific support for personalized training plans.

In terms of aesthetics, Scheme Two incorporates soft colors and rounded shapes, making the fitness process no longer cold but full of fun and challenges. This design stimulates users' enthusiasm for exercise and promotes the formation of a sustained workout habit. Especially for the youth, this creative design can effectively arouse their curiosity and exploratory desire, turning fitness activities into an enjoyable rather than burdensome task. Designing home fitness equipment that meets modern people's needs is a complex task. This study integrates web scraping technology with FAHP (Fuzzy Analytic Hierarchy Process) and TOPSIS (Technique for Order Preference by Similarity to Ideal Solution) theories. These methods can accurately pinpoint design focuses and address the problem of translating user needs into design practice by combining qualitative and quantitative approaches to capture the true expectations and preferences of users.



Fig. 7. Scheme 2 uses partial use display.

V. CONCLUSION

This study innovatively combines web scraping, FAHP, and TOPSIS theories and successfully applies them to the field of home fitness equipment. Through FAHP analysis, we systematically prioritized design requirements. Subsequently, among the three proposed potential design schemes, we combined FAHP's weight factors with the TOPSIS model to select the design scheme that most closely meets users' actual needs. To further validate the practical effectiveness of this design scheme, we used the SUS (System Usability Scale) as an evaluation tool. The results demonstrated the high effectiveness and user satisfaction of this design scheme.

In future research, further studies will be conducted on home fitness equipment, with a focus on making new breakthroughs in analyzing user needs. Meanwhile, the current application of FAHP and TOPSIS theories in our research has limitations. Although web scraping can improve the user feedback dataset, the decision-making process is still inevitably influenced by subjective expert judgments. Therefore, future research will focus on enhancing the standardization and objectivity of the decision-making process by introducing more quantitative indicators and objective verification methods. Additionally, attention needs to be paid to engineering and technical issues to promote the design of home fitness equipment towards a more scientific and precise direction.

ACKNOWLEDGMENT

This research was supported by the Philosophy and Social Science Planning Project of Anhui under Grant AHSKZ2021D24.

REFERENCES

[1] HW.Chow, KT. Chang and I-Yao. Fang, "Evaluation of the effectiveness of outdoor fitness equipment intervention in achieving fitness goals for seniors." *International journal of environmental research and public health*, vol.18.no.23, pp.1250, 2021.

[2] T.Guo, "Chinese fitness equipment status and its development research in the horizon of public fitness."Journal of Computational and Theoretical Nanoscience, vol.13.no.12, pp.10219-10223,2016.

[3] T. S.Thaku and P. M. Babu, Evolution of bicycles and their utility as fitness aids-a review."Nursing and Health Sciences, vol.8.no.3,pp.19-23,2019.

[4] T.Wang,Y.Gan,S. D.Arena,L. T. Chitkushev, G. Zhang, R.Rawassizadeh and R. G.Roberts, "Advances for Indoor Fitness Tracking, Coaching, and Motivation."IEEE Systems Man and Cybernetics Magazine,vol.7.no.1,pp.4-14,2021.

[5] A.Rada and Á. Szabó, "The impact of the pandemic on the fitness sector – The general international situation and a Hungarian example. "Society and Economy, vol.44.no.4,pp.477-497,2022.

[6] W.Guo , K.Li and Z.Zhang,"Analysis of Fitness Furniture Design. " Packing Engineering, vol.44 .no.20, pp. 173-182, 2023.

[7] J.Xu, X.Zhang, A.Liao, S.Yu,Y. Chen and L.Chen, "Research on Innovative Design of Towable Caravans Integrating Kano-AHP and TRIZ Theories." International Journal of Advanced Computer Science & Applications(IJACSA), vol.15.no.3,pp.991-1001,2024.

[8] R.Singh,S.Avikal, R.Rashmi and M.Ram, "A Kano model, AHP and TOPSIS based approach for selecting the best mobile phone under a fuzzy environment."International Journal of Quality & Reliability Management, vol.37.no.6/7,pp. 837-851,2020.

[9] M.A.Khder, "Web scraping or web crawling: State of art, techniques, approaches and application." International Journal of Advances in Soft Computing & Its Applications,vol.13.no.3,2021.

[10] R.Ramachandran, S. Sudhir and A. B. Unnithan, "Exploring the relationship between emotionality and product star ratings in online reviews." IIMB Management Review, vol,33.no.4,pp. 299-308,2021.

[11] J.Cao, j.Liu and H.Xu,"Research on intelligent acquisition method of user requirements based on data-driven ."Packaging Engineering,vol,42. no.24, pp.129-139,2021.

[12] L.Wang,J.Xiong,and C.Ruan,"Research on product design of FAHP bone marrow aspiration needle." Heliyon, vol.10,no.5,pp.e27389 ,2024.

[13] M.Najib,B.Mostapha and M.Mohamed , "Multicriteria Handover Management by the SDN Controller-based Fussy AHP and VIKOR Methods." International Journal of Advanced Computer Science and Applications(IJACSA), vol.12,no.7,pp.458-465, 2021.

[14] Y.Wu, and J.Cheng,"Continuous fuzzy kano model and fuzzy AHP model for aesthetic product design: case study of an electric scooter." Mathematical Problems in Engineering, vol.2018,no.1,pp. 4162539,2018.

[15] N. A. M.Zulkefli, M.Madanan,T. M. Hardan, and M. H. M. Adnan, "Multi-Criteria Prediction Framework for the Prioritization of Council Candidates based on Integrated AHP-Consensus and TOPSIS Methods. " International Journal of Advanced Computer Science and Applications(IJACSA),vol. 13,no.2,pp.352-359,2022).

[16] Q.Liu, J.Chen, W.Wang, Q. Qin, "Conceptual design evaluation considering confidence based on Z-AHP-TOPSIS method." Applied Sciences, vol.11,no.16,pp. 7400,2021.

[17] A. Z.Hameed, J. Kandasamy, S. Aravind Raj, M. A. Baghdadi, and M. A.Shahzad, "Sustainable product development using FMEA ECQFD TRIZ and fuzzy TOPSIS." Sustainability, vol.14,no.21, pp.14345,2022.

[18] J.Reig-Mullor,D.Pla-Santamaria and A.Garcia-Bernabeu, "Extended fuzzy analytic hierarchy process (E-fahp): A general approach. "Mathematics,vol. 8,no.11, pp.2014,2020.

[19] L.Wang,and Y. Zhang, "The visual design of urban multimedia portals. "Plos one, vol.18,no.3,pp. e0282712,2023.

[20] Y .Jia,J.Wang ,X. Han and H. Tang, "Application and Performance Evaluation of Industrial Internet Platform in Power Generation Equipment Industry ." Sustainability, vol.15,no.20.pp.15116,2023.

[21] Y.Liang , B.Wu ,J. Wang ,C. Li,X .suo,K. Zhang, "Optimization of recharge parameters of foundation pit engineering based on orthogonal test and FAHP-value engineering method ."Tunnel construction ,vol. 41,no.09,pp.1492-1501,2021.

[22] T. L.Zhu, Y. J. Li, C. J. Wu,H. Yue, and Y. Q.Zhao, "Research on the design of surgical auxiliary equipment based on AHP, QFD, and PUGH decision matrix." Mathematical Problems in Engineering, vol.2022,no.1,pp. 4327390,2022.

[23] Y.Qi and K .Kim ,"Evaluation of electric car styling based on analytic hierarchy process and Kansei engineering: A study on mainstream Chinese electric car brands . "Heliyon, vol.10 ,no.5,pp. e26999-,2024.

[24] C.Sivalingam and S. K.Subramaniam, "Cobot selection using hybrid AHP-TOPSIS based multi-criteria decision making technique for fuel filter assembly process." Heliyon,vol.10,no.4.pp.e26374-,2024.

[25] M.Aastha and P .Karthick, "Optimizing RPL for Load Balancing and Congestion Mitigation in IoT Network ." Wireless Personal Communications, vol.136, no. 3, pp. 1619-1636,2024.

[26] Z.Zhang and H. Yang, "Haihunhou cultural and creative tea set design based on FAHP / KE framework ." Packaging Engineering, vol.45 no.12, pp.347-355 + 403,2024.

Evaluating the Impact of Fuzzy Logic Controllers on the Efficiency of FCCUs: Simulation-Based Analysis

Harsh Pagare[1], Kushagra Mishra[2], Kanhaiya Sharma[3]*, Sandeep Singh Rawat[4], Shailaja Salagrama[5]
Department of Computer Science and Engineering-Symbiosis Institute of Technology,
Constituent of Symbiosis International (Deemed University) Pune, Maharashtra, India[1, 2, 3]
School of Computer and Information Sciences, IGNOU, New Delhi, India[4]
Computer Information System, University of the Cumberland's, Williamsburg, Kentucky, USA[5]

***Abstract*—This study investigates the methods for creating nonlinear models and developing Fuzzy logic controllers for the Fluidized Catalytic Cracking Unit (FCCU) at different global refineries. The FCCU plays a crucial role in the petrochemical sector, processing a significant portion of the world's crude oil - in 2006, FCCUs were responsible for refining a third of the global crude oil supply. These units are essential for converting heavier oils, such as gasoil and crude oil, into lighter, more critical products like gasoline and olefinic gases. Given their efficiency in producing a large volume of products and the volatile nature of petrochemical market prices, optimization of these units is a priority for engineers and investors. Traditional control mechanisms often need to improve in managing the FCCU's complex, dynamic, and nonlinear operations, where creating an accurate mathematical model is challenging or involves significant simplifications. Fuzzy Logic controllers, which mimic human reasoning more closely than conventional methods, offer a promising alternative for such unpredictable and complex systems. The results of this work illustrate the usefulness and possible advantages of utilizing Fuzzy Logic controllers in the management of FCCU plants and they are also compared with the latest machine learning techniques as well. These findings are corroborated by simulations conducted with the MATLAB Fuzzy Logic Toolbox R2012b.**

***Keywords*—Non-Linear modeling; fuzzy logic controller; machine learning; optimization**

I. Introduction

Fluidized Catalytic Cracking units are pivotal in petrochemical facilities, transforming dense oil products like Gasoil into lighter, more commercially valuable hydrocarbons. The efficiency of FCCUs significantly influences a refinery's financial performance. These units are comprised of two main components: The Riser reactor, where the cracking of hydrocarbons occurs and catalysts get coated with Coke, diminishing their effectiveness, and the Regenerator reactor, where the catalysts are cleansed and rejuvenated for continuous use. A typical FCCU layout, including key instruments and sections, is illustrated in schematic diagrams. Fuzzy Logic offers a structured approach to handle processes laden with uncertainties, ideal for scenarios lacking precise mathematical models or when existing models are too intricate for swift real-time analysis.

Traditional control systems often fall short in such complex environments [1]. The demand for FCCUs is largely driven by market needs, with seasonal variations in product demand affecting control system design. The challenging nature of FCCUs, characterized by their non-linear, time-invariant, and unpredictable processes, complicates their modeling, simulation, and management. This complexity renders standard controllers, like PID systems, inadequate as they rely on precise plant models, prompting the need for innovative control strategies.

Moreover, FCCUs are crucial for refineries, often determining their profitability and market competitiveness [2]. These units leverage a specialized micro- spheroid catalyst that becomes fluidized under the right conditions, primarily to convert heavy petroleum fractions, known as Gasoil, into valuable products like high-octane gasoline and heating oil. Gasoil, a complex mix of hydrocarbon types, is processed in FCCUS where it is cracked within a riser tube to produce lighter compounds and Coke as a by-product, which subsequently deactivates the catalyst. The spent catalyst is separated, stripped of volatile hydrocarbons, and regenerated by burning off the Coke before being recycled back into the process.

FCCUs have recently included real-time data-collecting systems and other cutting-edge monitoring technology to improve operational effectiveness. These systems make continuous monitoring of critical process variables possible, which gives operators timely insights to enhance performance and make necessary modifications. Additionally, incorporating machine learning and artificial intelligence into FCCU operations has started to yield encouraging outcomes. These tools, which learn from past data and spot patterns that a human operator would overlook, can forecast maintenance requirements, optimize feedstock utilization, and increase overall process efficiency.

Furthermore, under stricter environmental restrictions, refineries are forced to implement more sustainable processes. This entails cutting back on FCCU emissions, especially those of CO2 and NOx. Refineries can minimize their environmental impact and sustain high production by refining catalyst regeneration and optimizing the regenerator's combustion process. Another area of active research and development in FCCUs is the introduction of new catalysts that offer improved selectivity towards desired products and are more resistant to deactivation by coke. With these developments, modern refineries hope to strike a compromise between environmental and economic concerns.

*Corresponding Author.

Choosing the appropriate variables is crucial for maximizing FCCU functions. While there is a lot of discussion about the best variables to utilize for fuzzy optimization in FCCUS, this study concentrates on significant variables that can be changed to achieve desired results. These variables fall into one of two categories: dependent-independent or input-output. Feed rate, specific gravity, air flow rate, catalyst circulation rate, cumulative feed rate, and regenerator temperature are significant input parameters. Riser temperature, CO2/CO ratio, coke deposition on catalyst, feed (gasoil) conversion rate, and the production of LPG and coke are significant output factors. Selection of appropriate variables is a challenging procedure that has a big impact on the outcomes. Previous research in [3] have conducted a thorough review of the variables selected and their effects on FCCU operations.

This study aims to develop and assess complex nonlinear models for fluidized catalytic cracking units (FCCUs). These models aim to increase the petrochemical refining process's operational efficiency and accuracy. The paper highlights the difficulties presented by FCCUs, including their unpredictable nature, time-variant features, and nonlinear behavior. Because these complexities are often too much for traditional control systems to handle, this research focuses on finding crucial input and output factors that significantly impact FCCU performance. The study aims to maximize these variables through fuzzy logic-based control algorithms, providing creative solutions that improve refinery profitability and market competitiveness while addressing environmental issues.

II. Methodology

Prior to adopting Fuzzy Logic, scientific inquiries were predominantly confined to mathematical models tailored for FCC units [4]. These models varied in their level of detail and accuracy. Research often revolved around comparing these models to discern their respective strengths and weaknesses. Given the critical role of FCCUs in both industrial and market contexts, research in this area has been extensive, covering aspects such as stability, optimization, and the development and simulation of mathematical models [5]. Comprehensive reviews have been conducted on the evolution of FCCU control strategies over time. Notably, there has been considerable research focusing on the safe operation of FCCUs. Fig. 1 illustrates the basic structure of a Fluid Catalytic Cracking Unit.

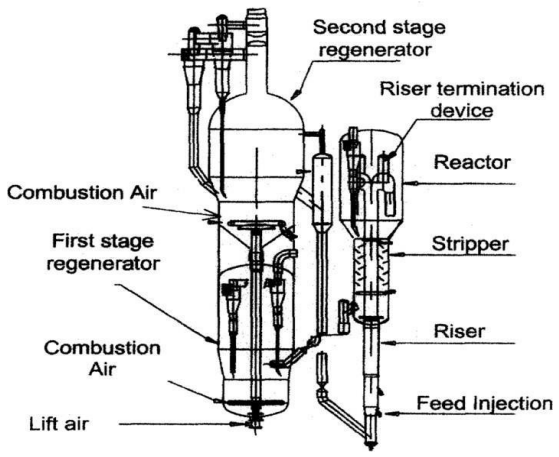


Fig. 1. Fuzzy control: A conceptual examination.

Initial studies indicated that Fuzzy Logic models outperformed traditional statistical methods in identifying process characteristics. Efforts were made to apply linear regression and sophisticated Kalman filtering techniques to improve precision, yet these methods struggled to accurately represent and manage real-world plants due to their inherent nonlinearity, vagueness, and unpredictability. Consequently, alternative strategies, including Neurofuzzy systems and genetic algorithms, began to gain traction. Despite advances in modern control methodologies like parameter estimation and stochastic and optimal control for model identification, the complexity of some industrial processes, characterized by high nonlinearity and uncertainty, defies conventional mathematical modeling and control approaches [6]. Fuzzy Logic, with its capacity to handle dynamic, nonlinear, and imprecise scenarios through linguistic rules, emerges as a suitable solution for such complex systems, commonly found in sectors like petrochemicals, nuclear energy, and water treatment. In situations when processes are well understood at the microscopic level, rigid control approaches are used.

However, standard control procedures often fail to provide satisfactory solutions to industrial issues contaminated by poor mathematical models. Artificial neural networks and fuzzy logic, two facets of soft computing that have recently found their way into the industrial control area, were originally applied in fuzzy control [7]. Product quality, efficiency, and energy consumption have all significantly improved as a result of this technology's application across a range of sectors [8]. Nowadays, fuzzy control is recognized as a state-of-the-art, complex control technique. Fuzzy logic and neural networks are increasingly being combined in scientific study, placing intelligent control front and center, particularly for systems whose parameters can be adjusted to conform to language conventions [9]. The two primary objectives of this research are to find the nonlinear relationships between input and output variables and to design a robust optimization framework with the aim of lowering Coke deposition on the catalyst and boosting LPG output and gasoil conversion.

A. Fuzzy Control: A Conceptual Examination

Fuzzy Logic mirrors the decision-making process of human experts, making it inherently user-friendly for both technical and non-technical applications. Its outputs, often described using everyday terms like "cold," "hot," or "fast," are straightforward and require little to no additional interpretation [10]. The development of a Fuzzy Logic system relies on the expertise and knowledge of specialists, who formulate this knowledge into a set of rule-based instructions for creating databases and Fuzzy rules. These rules, while approximate, reflect the inherently imprecise nature of human decision-making.

In practice, a Fuzzy rule-based system (FRBS) combined with a de-fuzzification component can serve as a stand-in for a human expert. This system takes in precise sensor data, translates these concrete values into heuristic variables using defined membership functions, and processes these variables through IF-THEN rules. The system then converts these linguistic variables back into a precise numerical output during the de-fuzzification stage, offering an estimated value close to the desired output [11].

A significant advantage of Fuzzy Logic is its independence from in-depth knowledge of the underlying system or its internal processes, a flexibility not typically found in traditional control systems like PID controllers. A schematic representation of a Fuzzy Logic controller would include the rule-based system, which stores the control strategy in rule format, the inference mechanism, which applies these rules based on current conditions to determine appropriate inputs; the fuzzification interface, which prepares the inputs by aligning them with the system's rules [12]; and the de-fuzzification interface, which translates the system's conclusions into actionable inputs for the system. In Fig. 2, Standard Fuzzy Logic Controller Architecture is mentioned.

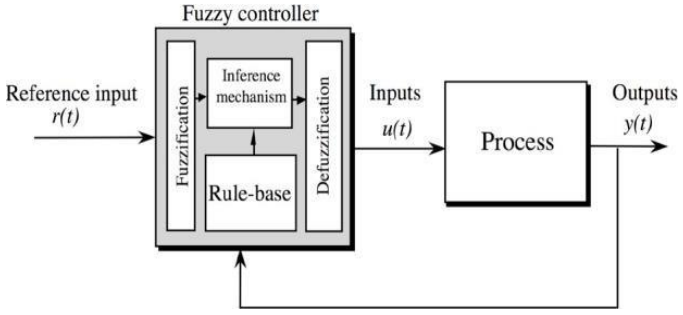


Fig. 2. Standard fuzzy logic controller.

B. Fuzzy Modeling of FCCU

1) Choosing variables: Parameters for Inputs andOutputs: This research utilizes data sourced from the operational guides and technical materials of different refineries across the world. Owing to the absence of a mathematical framework, a rule-based Fuzzy methodology was adopted for modeling. The process involved identifying key operating parameters of the FCCU as inputs and outputs, which represent the independent and dependent factors respectively [13]. The input variables in this study were selected based on their influence on the effectiveness and functionality of fluidized catalytic cracking units (FCCUs). Feed rate, specific gravity, airflow rate, catalyst circulation rate, cumulative feed rate, and regenerator temperature are among the chosen input factors. These variables, which include gasoil conversion rate, liquefied petroleum gas (LPG) generation, and coke deposition on the catalyst, were essential elements that directly affect the performance and output of the FCCU.

These factors were chosen because they regulate and enhance the cracking process. For example, the unit's throughput is determined by the feed rate and the rate at which the catalyst circulates, and the maintenance of the intended reaction conditions depends on the specific gravity and temperature of the regenerator. The airflow rate to the regenerator is an essential control parameter since it affects the combustion process and, in turn, the catalyst's regeneration.

A list provided in Table I outlines 16 critical parameters in the FCCU operation, categorizing them into control and observed variables. The Fuzzy controller is tasked with mapping out the behavior and interconnections of these variables through the creation of dynamic nonlinear representations, referred to as surface graphs. Based on their significance and impact within the FCCU process, a selectionof six inputs and six outputs was made for focus [14]. To enhance the refinery's efficiency, both control and observational variables were pinpointed, with continuous monitoring of Riser and Regenerator temperatures. Adjustments to the Catalyst feed and air supply rates, serving as control variables,are made to fine-tune the process parameters towards the targeted outcomes [15].

TABLE I. INPUT AND OUTPUT SPECIFICATIONS FOR FUZZY LOGIC SYSTEM

Input Variables	Output Variables
Gasoil (Feed)	C02/CO
Catalyst Recirculation Rate (CRR)	Gasoil Conversion Rate (GOCR)
Regenerator Temperature (RET)	Liquefied Petroleum (LPG)
Airflow to Regenerator (ATR)	Riser Temperature (RIT)
Cumulative Feed Rate (CFR)	DCC
Specific Gravity Factor (SG)	Coke as Bypass Product (Coke)
Control Variables	Observed Variables
Recycled Catalyst Rate	Riser Temperature
Airflow Rate	Regenerator Gas Temperature

2) Design of a fuzzy logic controller for FCCU: To develop the rule-based Fuzzy system that processes the nonlinear relationship between inputs and outputs, six essential steps are followed:

a) Determine the inputs, define their boundaries, and assign labels to them.

b) Specify the outputs, outline their boundaries, and label them accordingly.

c) Establish degrees of truth through Fuzzy membership functions.

d) Construct the Rule base necessary for the design of the controller.

e) Allocate intensities to the rules and define how they interact with each other.

f) Integrate the rules and convert the Fuzzy output into a crisp value through defuzzification [16].

Table II also presents the clustering of data for membership functions. To compile the knowledge base and establish rules, insights were obtained from a seasoned Process Engineer and a Senior Instrumentation Engineer active in the facilities. These rules were formulated based on operational manuals and various technical materials provided by the licensing authorities.

TABLE II. VARIABLES CLUSTERING RANGES

Clustering Group	Equivalence
Low	Small Impact
Medium	Steady State
High	High Impact

Tables III and IV present the initial values for both input and output variables, accompanied by their respective ranges. The membership functions were defined within these specific ranges.

TABLE III. CLUSTERING RANGES FOR INPUT VARIABLES

Input Variables	High (H)	Medium (M)	Low (L)
ATR (m3/h)	39,451 - 60,167	27,001 -47,209	0 - 29,873
RET (℃)	630 – 670	575 - 645	0 - 610
SG (-)	0.668 - 0.878	0.452 - 0.796	0 - 0.660
CFR (m3/d)	2,289 – 2,650	2,011 – 2,450	0 – 2,260
CRR(t/min)	14.9 – 16.9	11.2 – 16.1	0 – 15.2
Gasoil (m3/d)	1,967 – 2,250	1,770 – 2,151	0 – 1,980

TABLE IV. CLUSTERING RANGES FOR OUTPUT VARIABLES

Output Variables	High (H)	Medium (M)	Low (L)
CO2/CO (mol/mol)	2.2 – 6.2	0.9 – 3.9	0 – 1.8
DCC (-)	0.753 – 0.980	0.397 – 0.865	0 – 0.791
RIT (°C)	505 – 528	404 – 520	0 – 479
LPG (wt.%)	19.5 – 30.9	14.3 – 21.8	0 – 18.5
GOCR (wt.%)	79.3 – 98.16	44.9 – 93.8	0 – 76.8
Coke (wt.%)	5.2 – 9.1	3.4 - 7.5	0 – 4.2

This study chose triangular functions due to their ease of use and effectiveness in modeling fuzzy sets. These features aid in reducing computational complexity, which is essential for FCCUs and other real-time control systems. Triangle functions are perfect for situations where prompt responses are required, like in refining operations where precise control is required to maintain process stability and maximize output because of their linear nature.

3) Detailed description of fuzzy rules: The rules that connect the input and output variables are listed below. These rules were implemented using the MATLAB Fuzzy Rule Editor to generate the inference and nonlinear surface model.

In Fuzzy control systems, the focus is on utilizing linguistic rules, whereas traditional control systems rely heavily on differential equations. Employing verbal rules aligns more closely with human understanding than does a numerical approach [17]. Within a Fuzzy logic framework, these rules are always applicable but vary in their degree of truth from zero to one. The initial step of the inference process involves verifying the applicability of the rule premises for the given situation [18]. If the premises meet the criteria, the corresponding rules are chosen in a phase commonly referred to as "Matching." Following this, the inference system proceeds to make decisions.

The membership function for Specific Gravity (SG), a crucial statistic that represents the nature and quality of the feedstock entering the FCCU, is shown in Fig. 3. The SG membership function categorizes the feedstock according to its density, enabling the fuzzy logic controller to adjust the processing parameters to suit the feedstock's properties better. This classification may optimize the cracking process to provide the required product yield by modifying factors like temperature and catalyst activity. The system's entire operating efficiency and product quality can be improved by efficiently regulating the SG of the input.

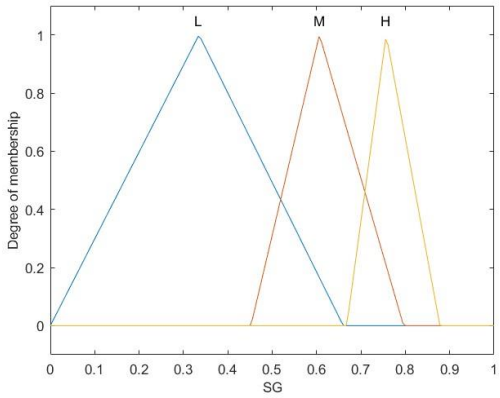


Fig. 3. SG Membership Function.

1-If (SG is H) then (LPG is M)(GOCR is H)
2-If (SG is H) then (Coke is H)(CO2/C0 is H)
3-If (SG is H) then (DCC is M)(RIT is L)
4-If (SG is L) then (CO2/C0 is L)
5-1f (SG is L) then (RIT is H)
6-If (ATR is H) then (Coke is H)
7-If (ATR is H) then (RIT is M)(CO2/C0 is M)
8-If (ATR is M) then (CO2/C0 is M)
9-If (ATR is M) then (DCC is L)
10-If (ATR is M) then (Coke is M)
11-If (ATR is L) then (CO2/C0 is L)
12-1f (ATR is L) then (DCC is M)
13- If (RET is H) then (RIT is M)(CO2/C0 is L)
14- If (RET is H) then (DCC is H)(LPG is M)(GOCR is L)
15-If (RET is H) then (Coke is M)(DCC is H)
16-If (RET is H) then (RIT is H)
17-1f (RET is M) then (Coke is M)(LPG is M)(GOCR is H)
18-If (RET is M) then (CO2/C0 is M)
19-1f (RET is M) then (Rh T is H)
20-If (RET is M) then (Coke is M)
21-If (RET is L) then (RIT is M)
22-1f (RET is L) then (DCC is L)
23-If (RET is L) then (CO2/C0 is L)
24-1f (RET is L) then (Coke is L)
25-If (RET is L) then (LPG is M)(RIT is L)(GOCR is L)
26-If (CFR is H) then (RIT is M)(GOCR is M)
27- If (CFR is H) then (DCC is L)(LPG is H)(RIT is M)(GOCR is H)
28- If (CFR is M) then (DCC is M)(LPG is M)(RIT is M)
29-If (CFR is L) then (DCC is M)(LPG is L)(GOCR is H)
30- If (CRR is H) then (Coke is M)(RIT is H)(GOCR is L)(CO2/C0 is H)
31- If (CRR is M) then (Coke is M)(GOCR is M)
32-If (CRR is L) then (Coke is L)(GOCR is M)

33- If (Gasoil is H) then (RIT is M)(GOCR is L)(CO2/C0 is L)
34- If (Gasoil is M) then (GOCR is M)
35-If (Gasoil is L) then (GOCR is H)

The numerical value of specific gravity is calculated by dividing the density of the substance being measured with the density of the reference. The reference substance is nearly always water at its densest (997 kg/m^3).

The membership function for Differential Coke Concentration (DCC), a crucial variable in the management of Fluidized Catalytic Cracking Units (FCCUs), is shown in Fig. 4. The concentration of coke created during the cracking process is divided into three categories by the DCC membership function: low, medium, and high. Because of this classification, the fuzzy logic controller may effectively manage the regeneration process, which can modify temperature and airflow to maintain the best possible catalyst performance. The system can better regulate coke deposition by identifying these categories essential for preserving efficiency and reducing downtime in FCCUs.

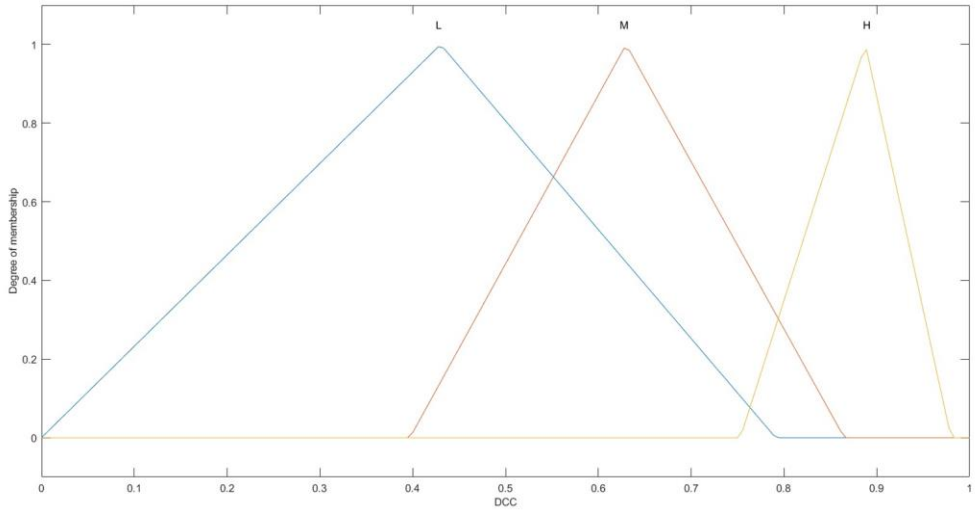


Fig. 4. DCC Membership Function.

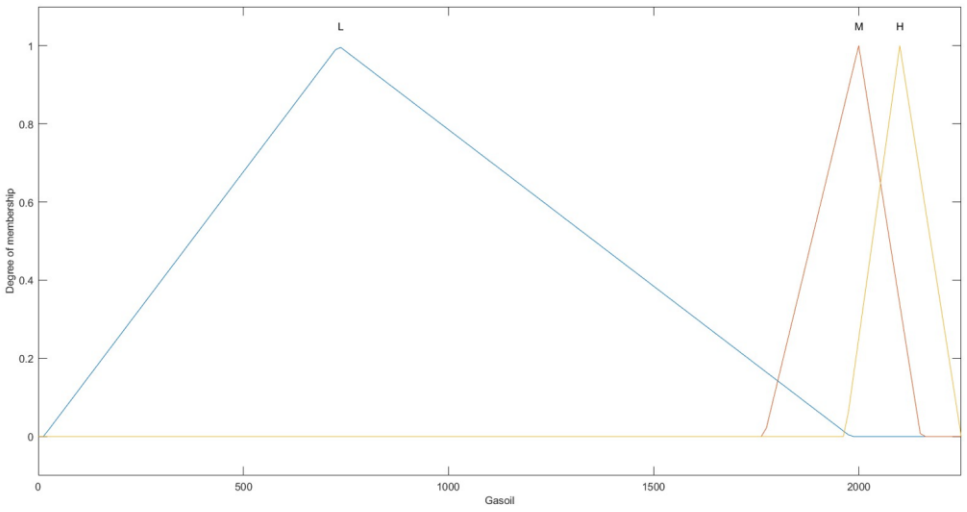


Fig. 5. Gasoil membership function.

The membership function for Gasoil, a crucial input variable in FCCUs, is shown in Fig. 5. The fuzzy logic controller can modify the processing conditions by dividing the input feedstock into distinct ranges based on the Gasoil membership function. The system can maximize conversion efficiency and minimize the creation of undesirable byproducts by classifying Gasoil into low, medium, and high levels during the cracking process. Maintaining the proper ratio between feedstock input and the intended output of lighter hydrocarbons, such as gasoline and LPG, depends on this function.

The membership function for the CO2/CO ratio, which is essential for tracking the regenerator's combustion efficiency inside the FCCU, is seen in Fig. 6. The fuzzy logic controller can modify the airflow and combustion parameters to maximize catalyst regeneration because the CO2/CO membership function divides the ratio into several regions. To guarantee that the coke on the catalyst is efficiently burnt off without producing too many emissions, the CO2 and CO levels must be in the right balance. Maintaining both operational effectiveness and compliance with environmental standards depends on this role.

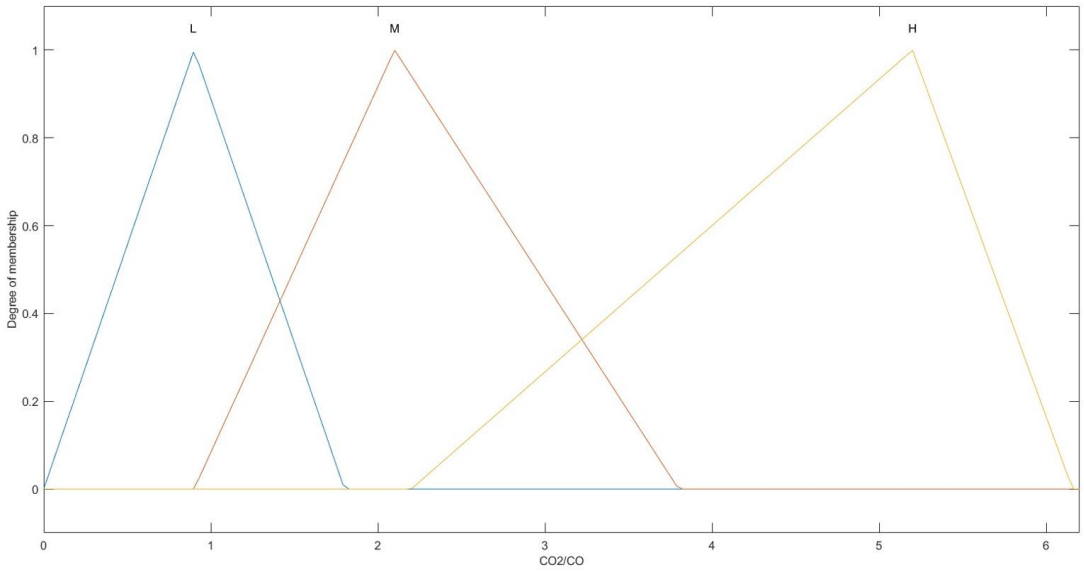


Fig. 6. CO2 / CO membership function.

4) Defuzzification: The final step in the Fuzzy control process involves defuzzification, where the Fuzzy controller converts the Fuzzy output into a precise control signal. This signal is then applied to the system by adjusting the relevant variables. During this stage, the inference system identifies the most definitive scenario and generates the output based on it [19]. The goal of defuzzification is to translate the Fuzzy decision- making process into a clear, actionable control measure that accurately reflects the range of potential outcomes [20]. The Fuzzy surface chart is an invaluable tool, offering insights into the relationship between the input and output variables, how quickly the system responds to input variations, and the direction in which these changes occur [21]. This information provides engineers with a novel approach to plant analysis, offering perspectives that traditional control strategies cannot offer [22]. Being able to test multiple alternative outcomes at once without having to deduce the system's mathematical formulas is incredibly helpful.

III. Results and Discussion

Although conventional control methods have proven successful in resolving many mathematical problems in the field, their shortcomings when handling intricate, dynamic settings have brought to light the benefits of Fuzzy Logic for control engineers confronting these kinds of problems. Fuzzy Logic relies on the experience of seasoned experts in the subject rather than Ordinary Differential Equations (ODEs). The Neuro Fuzzy technique, which is gaining traction and seems to have a lot of potential for future applications, is the result of the growing interest in enhancing Fuzzy Logic with experiential learning.

In order to create Fuzzy Logic models and control designs, this study made use of data from technical literature and operational manuals. The MATLAB Fuzzy Logic Toolbox 2012b was utilized to apply this data, and Fig. 7 through 10 present the main conclusions. Fuzzy Logic enables the avoidance of intricate mathematical calculations, necessitating instead the profound comprehension and discernment of an accomplished professional. Previously, a Yokogawa Distributed Control System (DCS) was used to operate the facility in question. This system implemented rules through specialized programming, which required manual modifications during maintenance or transitions. The study's data precision is in close agreement with the operational data of different facilities, which was collected in 2004 and may not be representative of the current operating conditions. Though areas like CRR and CFR required more tweaks for optimal performance, the fuzzy control model established via this study indicated good performance, with certain conclusions fitting well with real operational data.

Finishing the facility's fuzzy model provides a wealth of information. The linear relationship between the ATR variable and Coke output, for instance, is depicted in Fig. 7, where production increases directly up to a certain point and then declines as the ATR variable rises further, peaking at an ATR of 54,000. The utilization of numerical optimization techniques to increase facility efficiency is made easier by this pattern recognition. Furthermore, as seen by the 3D depiction in Figure 9, the study demonstrates the effectiveness of Fuzzy Logic in producing three-dimensional outputs that are on par with PID controllers, particularly when managing intricate and unexpected systems like FCCUs. More applications of fuzzy logic are shown in Fig. 8 and Fig. 10, which also offer more insights into the advantages of fuzzy logic for industrial process improvement.

A comparison of the suggested fuzzy logic control method with current FCCU management techniques is shown in Table V. It emphasizes how the fuzzy logic approach, which does not primarily rely on intricate mathematical models, provides greater flexibility and effectiveness in managing the intricate, nonlinear dynamics of FCCUs. The table also shows how this methodology performs better than alternative approaches in streamlining operations, decreasing coke deposition, and raising LPG output.

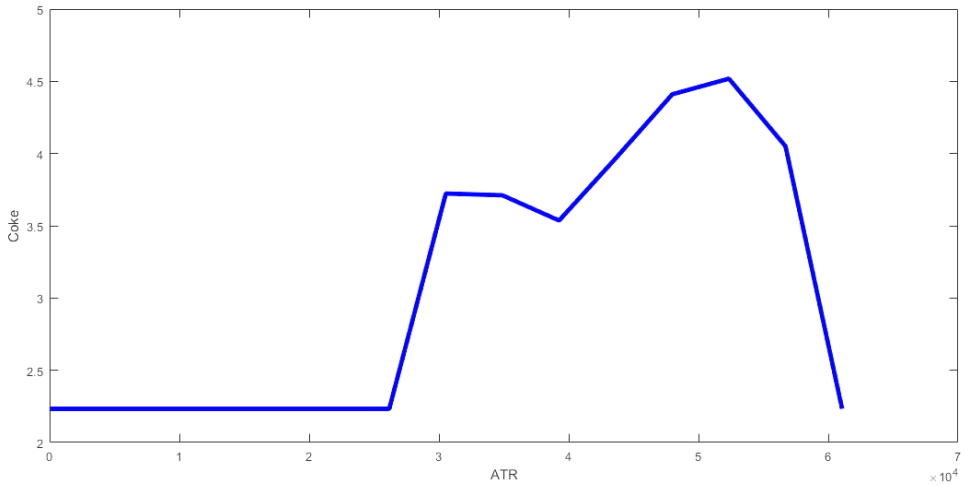


Fig. 7. Coke production according to ATR only.

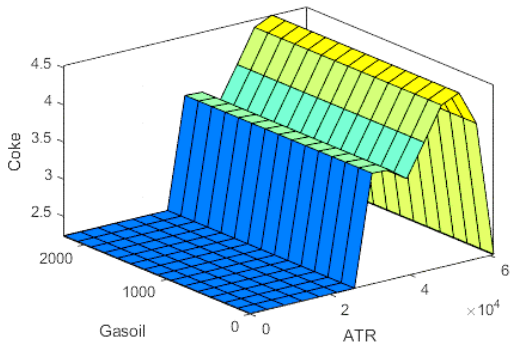


Fig. 8. Coke production according to Gasoil and ATR

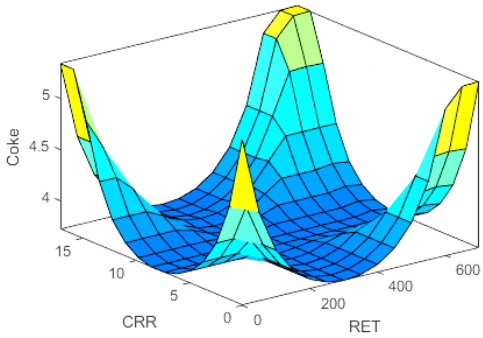


Fig. 9. Coke Production according to CRR and RET

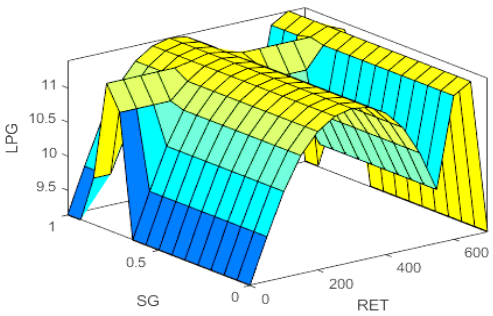


Fig. 10. LPG production according to SG and RET.

TABLE V. COMPARISON WITH EXISTING SOLUTIONS

Feature	Proposed Work	Acosta-López & de Lasa et al. [4]	Hu & Zhou et al. [5]	Nan Liu, Chun-Meng Zhu et al. [6]	Tianyue & Long et al. [7]	Tian and Wang et al. [8]
Control Approach	Utilizes fuzzy logic for adaptive response to FCCU dynamics.	Integrates CPFD with ML for predictive modeling.	Optimizes GCS and ASS using Aspen Plus simulation.	Uses CNN-based adaptive framework for optimization.	Bilevel robust optimization for handling uncertainties.	ML for early warning of abnormal conditions
Efficiency in Handling Complexity	Manages nonlinear operations without precise mathematical models.	Handles complex FCCU variables with high predictive accuracy.	Reduces utility costs and improves operational efficiency.	Addresses complex, multivariate nature of FCC processes.	Improves robustness and performance under uncertainty.	Improves safety and stability with early warnings.
Adaptability	High adaptability due to rule-based nature.	High adaptability using hybrid models.	Adaptive to varying operational parameters.	Adaptive to complex and dynamic FCC conditions.	Adapts to uncertain operational conditions.	Adapts to changing operational conditions.
Dependence on Mathematical Models	Does not rely on detailed mathematical models.	Uses CPFD simulations combined with ML.	Utilizes process simulation software for optimization.	Combines CNN with process models for better control.	Uses data-driven models to handle uncertainties.	Uses ML for predictive maintenance.
Performance	Optimizes operations with reduced coke deposition and increased LPG output.	High predictive accuracy with CPFD-ML integration.	Significant cost savings and efficiency improvements.	Enhances FCC performance with adaptive frameworks.	Improves FCC performance with robust optimization.	Effective early warning system for abnormal conditions.
Future Improvements	Refine rules and employ numerical optimization.	Further integration of CPFD and ML for enhanced accuracy.	Extend to other FCC subsystems for comprehensive optimization.	Enhance adaptability with more complex models.	Further improve robustness with additional data.	Improve early warning algorithms with more data.

IV. CONCLUSION

This study provides compelling evidence of the efficacy of the fuzzy logic approach in producing accurate findings in dynamic and nonlinear settings. The inherent unpredictability and nonlinearity in petrochemical facilities, notably in Fluid Catalytic Cracking Units (FCCUs), proved to be a perfect fit for this method. The fuzzy controller designed to handle the complexity of FCCUs worked exceptionally well, demonstrating its potential as a dependable control method in these demanding situations.

However, it is essential to recognize the limitations of this study. First, there is a need for improvement in the accuracy of the membership functions and fuzzy rules. Even though the existing configuration produced good results, performance may be improved by expanding the collection of rules and improving precision. Second, much of the study's data came from simulations and historical operational records, which might not accurately reflect the subtleties of operational dynamics in real-time. To improve the fuzzy logic controller's resilience, further in-the-real-world testing and the addition of real-time data should be done in subsequent research.

Furthermore, a viable path for future development is incorporating sophisticated control approaches like Artificial Intelligence (AI) and Neuro-Fuzzy systems. These methods can offer more intelligent and adaptable control mechanisms, improving FCCU dependability and overall performance. In the end, these systems' ongoing development will be essential for streamlining processes and ensuring they successfully address environmental and economic concerns.

Lastly, a crucial component of FCCU activities is resolving environmental issues. Future studies should investigate how AI and fuzzy logic might optimize the process to use less energy and produce fewer pollutants. One way is to create control systems that reduce CO2 and NOx emissions while preserving high productivity levels. Refineries can strengthen their competitive advantage in a market where environmental responsibility is becoming increasingly crucial by integrating their control system with sustainability goals.

AUTHORS' CONTRIBUTION

The first author, Harsh Pagare, was the main contributor to the study's conceptualization by creating the research framework and the fuzzy logic controllers. His primary responsibilities were interpreting and analysing the findings based on the simulation. He also made a significant literary contribution to the manuscript, helping to ensure that the research findings were well-expressed and backed up by solid data.

The second author, Kushagra Mishra, played an equally important role in conceptualising and developing the fuzzy logic controllers. He led the development and perfecting of simulation models and research techniques. Kushagra was also closely involved in the data gathering, processing, and interpretation processes to guarantee the precision and dependability of the results.

The third author, Kanhaiya Sharma, concentrated on the implementation's technical details, especially optimising the Fuzzy Logic system in MATLAB. In addition, he helped with the technical composition of the manuscript, the final review, and editing. He also contributed to data analysis.

The fourth author, Sandeep Singh Rawat, helped the team refine the study strategy and offered insightful feedback throughout the simulation phase. He helped with the data validation process and reviewed the literature, among other things.

The fifth author, Shailaja Salagrama, helped with the theoretical framework and made sure the study followed scholarly guidelines, which were two ways she contributed to the research. She also contributed to the Research Paper's final review and proofreading.

REFERENCES

[1] P. Josiah et al. "Synthesis and Soft Implementation of Supervisory Control Scheme in an Industrial Fluid Catalytic Cracking (FCC) Unit of a Nigerian Refinery." Global Journal of Pure and Applied Chemistry Research (2023). https://doi.org/10.37745/gjpacr.2013/vol11n12037

[2] Ghada A. Mutuab et al. "Non-Linear PID Control of Fluid Catalytic Cracking Unit." Journal Européen des Systèmes Automatisés (2023). https://doi.org/10.18280/jesa.560510.

[3] Singh, Balkeshwar & Mishra, Anil. (2023). Fuzzy Logic Control System and its Applications. 2395-0056.

[4] Acosta-López JG, de Lasa H. Artificial Intelligence for Hybrid Modeling in Fluid Catalytic Cracking (FCC). Processes. 2024; 12(1):61. https://doi.org/10.3390/pr12010061

[5] Sun J, Yu H, Yin Z, Jiang L, Wang L, Hu S, Zhou R. Process Simulation and Optimization of Fluid Catalytic Cracking Unit's Rich Gas Compression System and Absorption Stabilization System. Processes. 2023; 11(7):2140. https://doi.org/10.3390/pr11072140

[6] Nan Liu, Chun-Meng Zhu, Meng-Xuan Zhang, Xing-Ying Lan, A multiscale adaptive framework based on convolutional neural network: Application to fluid catalytic cracking product yield prediction, Petroleum Science, 2024.

[7] Li, Tianyue & Long, Jian & Zhao, Liang & Du, Wenli & Qian, Feng. (2022). A bilevel data-driven framework for robust optimization under uncertainty – applied to fluid catalytic cracking unit. Computers & Chemical Engineering. 166. 107989. 10.1016/j.compchemeng.2022.107989.

[8] Wende Tian, Shaochen Wang, Suli Sun, Chuankun Li, Yang Lin, Intelligent prediction and early warning of abnormal conditions for fluid catalytic cracking process, Chemical Engineering Research and Design, Volume 181, 2022.

[9] M. Zahran et al. "Fluid catalytic cracking unit control using model predictive control and adaptive neuro fuzzy inference system: Comparative study." 2017 13th International Computer Engineering Conference (ICENCO) (2017): 172-177. https://doi.org/10.1109/ICENCO.2017.8289783.

[10] Wei, Min & Qian, Feng & Du, Wenli & Hu, Jun & Wang, Meihong & Luo, Xiaobo & Yang, Minglei. (2018). Study on the integration of fluid catalytic cracking unit in refinery with solvent-based carbon capture through process simulation. Fuel. 219. 364-374. 10.1016/j.fuel.2018.01.066.

[11] Zhang, Y.; Li, Z.; Wang, Z.; Jin, Q. Optimization Study on Increasing Yield and Capacity of Fluid Catalytic Cracking (FCC) Units. Processes 2021, 9, 1497. https://doi.org/10.3390/pr9091497

[12] Olugbenga, A. and Oluwaseyi, O. (2023) Analysis of the Process Parameter in Fluid Catalytic Cracking Unit for a Refining and Petrochemical Company in Nigeria. Advances in Chemical Engineering and Science, 13, 65-78. doi: 10.4236/aces.2023.131006.

[13] Oloruntoba A, Zhang Y, Hsu CS. State-of-the-Art Review of Fluid Catalytic Cracking (FCC) Catalyst Regeneration Intensification Technologies. Energies. 2022; 15(6):2061. https://doi.org/10.3390/en15062061

[14] Stratiev D, Shishkova I, Ivanov M, et al. Role of Catalyst in Optimizing Fluid Catalytic Cracking Performance During Cracking of H-Oil-Derived Gas Oils. ACS Omega. 2021;6(11):7626-7637. Published 2021 Mar 12. doi:10.1021/acsomega.0c06207

[15] Fatih, Güleç & Meredith, Will & Snape, Colin. (2020). Progress in the CO2 Capture Technologies for Fluid Catalytic Cracking (FCC) Units-A Review. Frontiers in Energy Research. 8. 62. 10.3389/fenrg.2020.00062.

[16] S, Chitra & Son, Jack. (2023). A Review: Some Application on Fuzzy Logic.

[17] V, Dharun. (2023). A Comprehensive Study on Fuzzy Logic System. International Journal of Research Publication and Reviews. 4. 2430-2432. 10.55248/gengpi.4.423.36116.

[18] Zadeh, Lotfi & Aliev, Rafik. (2018). Fuzzy Logic Theory and Applications: Part I and Part II. 10.1142/10936.

[19] Maity, Saikat & Chakraborty, Sanjay & Pandey, Saroj & De, Indrajit & Nath, Sourasish. (2023). Type-n fuzzy logic - the next level of type-1 and type-2 fuzzy logic. International Journal of Intelligent Engineering Informatics. 11. 353-389. 10.1504/IJIEI.2023.136106.

[20] Pareek, Shashank & Gupta, Hemant & Kaur, Jaspreet & Kumar, Raman & Chohan, Jasgurpreet. (2023). Fuzzy Logic in Computer Technology: Applications and Advancements. 1634-1637. 10.1109/ICPCSN58827.2023.00273.

[21] Kahraman, Cengiz & Çevik, Sezi & Öztayşi, Başar & Cebi, Selcuk & Tfss, Transactions On Fuzzy Sets And Systems & Tfss, Systems. (2023). Role of Fuzzy Sets on Artificial Intelligence Methods: A literature Review. Transactions on Fuzzy Sets and Systems. 2. 10.30495/tfss.2023.1976303.1060.

[22] Shaik, Hasane. (2023). DESIGN AND ANALYSIS OF RULE-BASED FUZZY LOGIC CONTROLLER FOR PERFORMANCE ENHANCEMENT OF THE SUGARCANE INDUSTRY. 2620-1747. 10.31181/oresta040223031a.

Hybrid Machine Learning Approach for Real-Time Malicious URL Detection Using SOM-RMO and RBFN with Tabu Search

Swetha T[1], Dr. Seshaiah M[2], Dr. Hemalatha K L[3], Dr. Murthy S V N[4]*, Dr. Manjunatha Kumar BH[5]
Research Scholar, Visvesvaraya Technological University, Belagavi-590018, Karnataka, India[1]
Dept. of CSE, SJCIT, Chickballapur, India[1]
Associate Professor, Dept. CSE, SJCIT, Chickballapur, India[2]
Professor and HOD, Dept. ISE, SKIT, Bangalore, India[3]
Associate Professor, Dept. of CSE, SJCT, Chickballapur, India[4]
Professor and HOD, Dept. of CSE, SJCIT, Chickballapur, India[5]

***Abstract*—The proliferation of malicious URLs has become a significant threat to internet security, encompassing SPAM, phishing, malware, and defacement attacks. Traditional detection methods struggle to keep pace with the evolving nature of these threats. Detecting malicious URLs in real-time requires advanced techniques capable of handling large datasets and identifying novel attack patterns. The challenge lies in developing a robust model that combines efficient feature extraction with accurate classification. We propose a hybrid machine learning approach combining Self-Organizing Map based Radial Movement Optimization (SOM-RMO) for feature extraction and Ensemble Radial Basis Function Network (RBFN) based Tabu Search for classification. SOM-RMO effectively reduces dimensionality and highlights significant features, while RBFN, optimized with Tabu Search, classifies URLs with high precision. The proposed model demonstrates superior performance in detecting various malicious URL attacks. On a benchmark dataset, the proposed approach achieved an accuracy of 96.5%, precision of 95.2%, recall of 94.8%, and an F1-score of 95.0%, outperforming traditional methods significantly**.

***Keywords*—Malicious URL detection; self-organizing map; Radial Movement Optimization; ensemble radial basis function network; Tabu Search**

I. Introduction

Many offline activities have moved online as a result of the Internet's expansion and development, including general business, social networking, e-commerce, and banking. As such, there is now a higher chance that illegal activity may occur online. This emphasizes how urgently action must be done to maintain internet security [1]. To get sensitive data or compromise the system, people are being tricked into accessing dangerous URLs. This means that protecting this side is becoming a critical need because [2]. Malicious people can nevertheless attack the connection between the client and the server even in the presence of laws and standards. Phishing, spam, malware, and other types of attacks are all referred to as "malicious," as one umbrella term [3].

Because malicious URLs collect needless information and trick unwary end users into falling for scams, they result in yearly losses of billions of dollars. The online security world has created blacklisting services to help identify dangerous websites [4]-[6]. The goal was to identify the risk that dangerous websites pose. The blacklist is a database including every URL that has ever been deemed possibly dangerous. Apparently, there are circumstances when URL blacklisting is effective [7]. Nevertheless, an attacker can exploit these weaknesses by modifying the URL string in a way that makes the system readily fooled. Many harmful websites will unavoidably stay online because they are either too new, never examined, or had their evaluations incorrect.

Identifying dangerous websites are heuristics, which are basically an improved version of the signature-based blacklist method. One can compare the signatures of an old malicious URL and a new one. An additional line of protection against dangerous websites is offered by this approach. The techniques described here will help you distinguish between benign and malicious URLs. These more traditional methods do, however, have several shortcomings, which are enumerated here: (a) Zero-hour phishing attempts cannot be stopped by the blacklist method since it can only identify and categories 47-83% of newly found phishing URLs in a 12-hour timeframe [8]. (b) By adopting technology is evolving quickly enough to render the blacklist approach out of date. Since the blacklist approach is simple to use, many anti-phishing agencies continue to adopt it despite these drawbacks [9].

Thirdly, machine learning (ML) and deep learning (DL) are AI methods that can be used to detect these dangerous websites. The several industries in which these technologies have been applied include cybersecurity, healthcare, e-commerce, medical image analysis, and social media [10]. By exposing machine learning models to historical data, one can train them to become more adept at self-learning, therefore doing away with the necessity for human involvement in the learning process. This is really beneficial in the domain of cybersecurity. This generates a lot of property in huge companies, banks, and other institutions [11]. Because machine learning and deep learning are so effective in many other domains, many people also employ them to discover dangerous websites [12]. It has shown to be successful to find dangerous URLs by using machine learning to identify recently created URLs and

automatically updating the model. Recent study indicates that deep learning models can be used to automatically identify and extract the attributes of newly created URL. This enables researchers to gather a wealth of information from URLs, which in turn facilitates the decision-making process of machine learning algorithms regarding the safety of the URL.

The objectives of the research work involve the following:

1) Implementing an ensemble learning framework for enhanced malicious URL and intrusion detection in cloud system through deep learning techniques

2) To combine Self-Organizing Map based Radial Movement Optimization (SOM-RMO) for feature extraction.

3) To utilize Ensemble Radial Basis Function Network (ERBFN) based Tabu Search optimization for precise classification of malicious URL.

The main novelty of the research work:

The research combines SOM-RMO and ERBFN with Tabu Search, leveraging strengths of both techniques for enhanced detection capabilities. SOM-RMO reduces data dimensionality and extract meaningful features for malicious URLs, improving model performance and reducing computational load and implements Tabu Search for optimizing ERBFN, enhancing classification accuracy and robustness.

II. URL Attack Techniques

Any tool or strategy used by a hacker to gain unauthorized access to user data or to harm the system they are trying to penetrate can be considered an attack tactic. Attackers can use nefarious URLs to launch such kinds of attacks. URLs that are deemed hazardous include many others including spam, phishing, malware, and defacement. Clicking on maliciously contented URLs is the most common way that cyberattacks occur. When URLs are used for evil intent rather than to visit websites that are allowed to be viewed online, the integrity of the data, its secrecy, and its availability are all compromised.

A. Spam URL Attacks

These attacks are the work of spammers, who build phoney websites and then try to trick browser engines into believing they are real. To that end, spammers who illegally raise their rank are trying to trick users into visiting their websites more often [10]. The spammers want to install malware and adware on the computers of their victims, hence they send spam emails containing spam URLs.

B. Phishing URL Attacks

Using phishing URLs—which are meant to fool users into viewing a phoney website—is one way that criminals get sensitive data, such as credit card details. User data vulnerability and can easily trick those who are not familiar with phishing websites into visiting the website [11].

C. Malware URL Attacks

These attacks, which infect consumers' devices with malware, can have a range of unfavorable effects, such as file damage, keystroke tracking, and identity theft. Known by most as malware, malicious software can harm systems and steal private data. Malware may also refer to malevolent software. Drive-by download is the term for malware that inadvertently infects a user's device when they visit a malicious website. Further instances are as follows: Computer-infecting viruses, worms, Trojan horses, spyware, scareware, and ransomware.

D. Defacement URL Attacks

This kind of attack targets a hostile website that has undergone some kind of hacker modification, either to its appearance or content. This approach transports the user to the dangerous website. There could be several reasons why hacktivists try to take down websites. As it happens, [13]. Machine learning (ML) based taxonomy that can detect potentially dangerous URLs on Arabic and English webpages! Penetration of a website is the process of taking use of security flaws to obtain unauthorized access to a website and modify its content without the owner's knowledge or consent [11]. Machine learning methods allow dangerous URL attacks to be categorized as either benign or malignant. Contrarily, multi-classification allows the addition of more than two categories, such as phishing, harmful, spam, benign, suspicious, and so forth.

III. Related Works

Targeting the victims' spaces, this kind of attack steals sensitive data and passwords without their knowledge. These attacks—phishing, drive-by downloads, and spamming, for example—are conducted using malicious URLs. Blacklists, machine learning, and heuristics are the three main categories into which that can be divided. The heuristic approach [12] gives a forecast that is equally accurate as the machine learning method and outperforms the blacklist approach in generalizing the harmful URL. This paper proposes a new method that uses the most significant information obtained from URLs to identify potentially dangerous URLs.

Many internet channels, including email and messaging, are used to spread these URLs. Various traditional methods for identifying phishing websites include blacklists, which are subsequently used to forecast the URLs of such websites. Blacklist-based conventional methods are unable to keep up with the volume of new phishing websites that are constantly emerging and being added to the Internet. It is this that is problematic. Proposed is an improved deep learning-based phishing detection method for effective identification of dangerous URLs. The foundation of this approach is the integration of variational autoencoders (VAE) and deep neural networks (DNN) power. As is explained in [13], the VAE model replicates the original input URL to automatically extract the intrinsic properties of the raw URL. The purpose of this is to enhance the phishing URL identification. In order to conduct our study, we used the publicly available ISCX-URL-2016 dataset and the Kaggle dataset to crawl over 100,000 URLs. The proposed model outperformed all other models assessed in terms of accuracy (up to 97.45%) and response time (1.9 seconds) based on the data gathered throughout the testing process.

Use of URLs, web page content, and external features enhances machine learning models' detection skills. The outcomes of an experimental study to increase the precision of machine learning models for the two most well-known datasets

used for phishing are presented in the paper [14]. The aim of the research was to raise the models' general performance. Three types of tuning elements are applied: feature selection, hyper-parameter optimization, and data balancing. This experiment uses two different datasets that are obtained from websites like the UCI repository and the Mendeley repository. The results indicate that a machine learning algorithm performs better when its parameters are changed.

Currently the most common and dangerous kind of cybercrime that anyone may commit, phishing has been around since 1996. The suggested study that is discussed in study [15] is based on this specific dataset. Phishing and real URL properties derived in vector form from over 11,000 website datasets are included in the well-known dataset collection. After pre-processing is over, many machine learning techniques have been developed and implemented to shield users from phishing URLs. This work aims to create a practical and efficient security against phishing attacks by using different machine learning models. Together with grid search hyper parameter optimization and cross fold validation, the proposed LSD model uses the canopy feature selection approach. Different evaluation criteria were used to assess the proposed technique in order to show the impact and efficacy of the models. Among the qualifying criteria were recall, specificity, accuracy, precision, and F1-scores. The comparative assessments show that the proposed approach produces outcomes of a higher qualitative quality and is better than the other approaches.

TABLE I. SUMMARY

Reference	Method/Algorithm	Datasets	Outcomes
[12]	Heuristic Approach	Phishing URL Dataset from Repository -	Better generalization and accuracy than blacklist approach; comparable to machine learning
[13]	VAE + DNN	ISCX-URL-2016, Kaggle	Accuracy: 97.45%, Response Time: 1.9 s
[14]	Feature Selection	UCI Repository, Mendeley	RF: 97.44%
[15]	LR+SVC+DT (LSD Model) with Canopy Feature Selection, Grid Search Hyper parameter Optimization	Phishing URL Dataset from Repository	High accuracy and efficiency; outperforms other models
[16]	Ensemble Techniques	ISCX-URL-2016	En_Bag: Accuracy 99.3% (binary), 97.92% (multi-class); En_kNN: Highest inference speed

As such, the creation of technologies that can identify phoney URLs is currently highly sought after. In study [16], a high-performance machine learning-based detection technique is proposed with the intention of detecting URLs that could contain hazardous material. There exist two layers of detection in the proposed system. As a second phase, we group the URL classes into benign, spam, phishing, malware, or defacement groups based on their characteristics. Four separate ensemble techniques—En_Bag,En_kNN,En_Bos,and En_Dsc—will be the focus of this section. Under this category are techniques such as subspace discriminator ensembles, boosted decision tree ensembles, k-nearest neighbor ensembles, and bagging tree ensembles. We evaluated the developed approaches using the huge and current dataset for uniform resource locators, ISCX-URL2016. Experimental evaluation revealed that the ensemble of bagging trees (En_Bag) strategy outperformed other ensemble techniques. The En_kNN method is another equally efficient approach that combines several k-nearest neighbour ensembles to get the fastest inference time. Attained accuracy of 99.3% in binary classification and 97.92% in multi-classification, we show that our En Bag model outperforms solutions regarded as state-of-the-art. Table I covers the methodology utilized and results achieved in this study.

IV. PROPOSED METHOD

The proposed method uses Self-Organizing Map based Radial Movement Optimization (SOM-RMO) for feature extraction and Ensemble Radial Basis Function Network (ERBFN) enhanced by Tabu Search for classification as in Fig. 1. SOM-RMO is employed to reduce the high dimensionality of URL data, identifying and preserving the most significant features. This method transforms complex, multi-dimensional data into a simpler, lower-dimensional space, making the subsequent classification process more efficient. The ERBFN, a neural network model known for its effectiveness in pattern recognition, is then optimized using Tabu Search. Tabu Search is a metaheuristic algorithm designed to guide the search process in optimization problems, helping the ERBFN achieve a high level of accuracy in distinguishing between benign and malicious URLs.

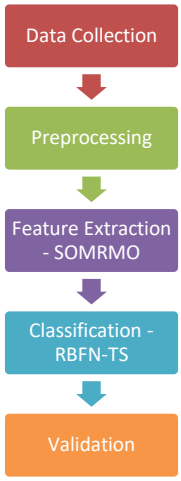


Fig. 1. Proposed method.

A. Dataset Description

The dataset [17] consists of a total of 651,191 URLs, categorized into four distinct classes: benign, defacement, phishing, and malware. The primary goal is to use this extensive

dataset to develop a machine learning model capable of identifying malicious URLs to prevent cybersecurity threats.

1) Distribution of URLs

- Benign URLs: 428,103 (65.72%)
- Defacement URLs: 96,457 (14.81%)
- Phishing URLs: 94,111 (14.45%)
- Malware URLs: 32,520 (5.00%)

The dataset is curated from five different sources to ensure a comprehensive collection of URL examples. The sources include ISCX-URL-2016, Malware Domain Blacklist, Faizan Git Repository, Phish tank, and Phish Storm datasets. The dataset is structured in a tabular format with two main columns: URL and Class. The URL column contains the actual web addresses, and the Class column indicates the category of each URL (benign, defacement, phishing, or malware) as in Table II.

TABLE II. SAMPLE TRAINING DATA

URL	Class
http://example-safe-site.com	benign
http://secure-shopping-site.com	benign
http://phishingsite.com/login	phishing
http://downloadmalware.com/install	malware
http://defacementexample.com/home	defacement
http://example-trusted-site.org	benign
http://stealyourinfo.com/verify	phishing
http://injectmalwarehere.com/secure	malware
http://websitehacked.com/page	defacement
http://another-safe-site.org	benign

- Benign URLs: These are regular, non-malicious websites typically used as a baseline to train the model to distinguish safe websites from harmful ones.
- Defacement URLs: These URLs are linked to sites that have been compromised, usually to display unauthorized content.
- Phishing URLs: The hackers will cloned website and information similar to original website and steal information's.
- Malware URLs: These URLs are associated with websites that host or distribute malicious software.

2) Dataset curation

- ISCX-URL-2016 Dataset: Used for collecting benign, phishing, malware, and defacement URLs.
- Malware Domain Blacklist: Provided additional phishing and malware URLs.
- Faizan Git Repository: Increased the number of benign URLs.
- Phishtank and PhishStorm Datasets: Contributed more phishing URLs.

The dataset is invaluable for training machine learning models to detect and classify malicious URLs effectively. By including a large number of samples across different categories, the model can learn to recognize a wide variety of malicious patterns and behaviors, ultimately improving cybersecurity measures and preventing potential attacks.

B. Data Pre-processing

Data preprocessing is a crucial step in preparing the dataset for machine learning. For URL data, this typically includes steps like data cleaning, feature extraction, encoding, and normalization, the transforming raw data is shown in Table III Below are the main steps involved in preprocessing the malicious URLs dataset:

1) Data cleaning-removing duplicates: Ensuring each URL in the dataset is unique to prevent bias in model training as in Table IV.

2) Handling missing values: Checking for and addressing any missing values in the dataset, although URLs and their labels are generally expected to be present as in Table IV.

3) Lexical Feature Extraction: Extracting features based on the structure and content of the URL which is provided in Table V.

4) Host-based features: Analyzing the URL's domain for attributes like:Domain age and WHOIS information.

5) Content-based features: If accessible, extracting features from the web page content like: Keywords in the HTML body and Number of external links.

6) Encoding -label encoding: Converting the class labels ('benign', 'defacement', 'phishing', 'malware') into numerical values for model training which is shown as in Table VI.

7) Normalization: Scaling numerical features to a standard range (typically 0 to 1) to ensure uniformity and improve the model's convergence during training as in Table VII

TABLE III. RAW DATA OF URL

URL	Class
http://example-safe-site.com	benign
http://phishingsite.com/login	phishing
http://downloadmalware.com/install	malware
http://websitehacked.com/page	defacement

TABLE IV. AFTER DATA CLEANING

URL	Class
http://example-safe-site.com	benign
http://phishingsite.com/login	phishing
http://downloadmalware.com/install	malware
http://websitehacked.com/page	defacement

TABLE V. FEATURE EXTRACTION OF URL

URL	URL_Length	Num_Dots	Has_Hyphen	Num_Special_Chars	Has_IP	Class
example-safe-site.com	19	2	1	0	0	benign
phishingsite.com/login	22	1	0	1	0	phishing
downloadmalware.com/install	28	1	0	1	0	malware
websitehacked.com/page	21	1	0	1	0	defacement

TABLE VI. ENCODING OF URL

URL	URL_Length	Num_Dots	Has_Hyphen	Num_Special_Chars	Has_IP	Class_Label
example-safe-site.com	19	2	1	0	0	0
phishingsite.com/login	22	1	0	1	0	2
downloadmalware.com/install	28	1	0	1	0	3
websitehacked.com/page	21	1	0	1	0	1

TABLE VII. NORMALIZED FEATURE OF URL

URL	URL_Length	Num_Dots	Has_Hyphen	Num_Special_Chars	Has_IP	Class_Label
example-safe-site.com	0.68	1.0	1.0	0.0	0.0	0
phishingsite.com/login	0.79	0.5	0.0	1.0	0.0	2
downloadmalware.com/install	1.0	0.5	0.0	1.0	0.0	3
websitehacked.com/page	0.75	0.5	0.0	1.0	0.0	1

C. Self-Organizing Map-Based Radial Movement Optimization (SOM-RMO) Process

SOM based RMO is a hybrid approach combining the advantages of SOM for feature extraction and RMO for optimization. The process is designed to reduce data dimensionality, highlight significant features, and prepare the dataset for efficient and accurate classification. The grid consists of nodes or neurons, each representing a cluster of input data. The primary goal of SOM is to preserve the topological properties of the input space, ensuring that similar data points are mapped to nearby nodes on the grid.

The weight update for a node in SOM is given by:

$$w(t+1)=w(t)+\alpha(t)\cdot h(c,t)\cdot(x-w(t)) \quad (1)$$

Where:

w(t) is the weight vector of the node at time t.

α(t) is the learning rate, which decreases over time.

h(c,t) is the over time.

x is the input vector.

RMO is a metaheuristic optimization algorithm that simulates the movement of particles within a defined search space, optimizing the positioning of nodes in the SOM. The optimization process iteratively adjusts the positions of the nodes to minimize the distance between the nodes and their corresponding input data points, thereby improving the feature extraction capabilities of the SOM.

The position update in RMO for a particle (node) is given by:

$$Pi\ (t+1) = pi(t)+vi(t+1) \quad (2)$$

Where:

pi(t) is the position of particle i at time t.

vi(t+1) is the velocity of particle i at time t+1, which is influenced by cognitive and social components guiding the particle towards the optimal solution.

Pseudocode

```
Initialize SOM grid with random weights
Initialize learning rate α and neighborhood radius σ
Initialize RMO particles with SOM nodes' positions
Initialize velocities for RMO particles
while not converged do
    for each input vector x in dataset do
        Find BMU in SOM
        for each node in SOM do
            Update weight vector using:
            w(t+1) = w(t) + α(t) * h(c, t) * (x - w(t))
        end for
        Adjust learning rate α and neighbourhood radius σ
    end for
    for each particle i in RMO do
        Update velocity using cognitive and social components
        Move particle to new position:
p_i(t+1) = p_i(t) + v_i(t+1)
        Evaluate fitness of new position
    end for
    Check for convergence criteria
end while
```

The SOM grid and RMO particles are initialized with random values, setting the stage for the optimization process. The SOM iteratively adjusts its nodes to map the input data onto a lower-dimensional space, using the update rule to refine node positions based on the input vectors. RMO particles adjust their velocities and positions to optimize the SOM node placement, ensuring that the extracted features are representative of the input data. The process continues until the SOM and RMO reach a stable state, indicating that the feature extraction and optimization are complete. This hybrid approach leverages the strengths of SOM for dimensionality reduction and RMO for

optimization, resulting in a robust preprocessing method for detecting malicious URLs.

D. ERBFN with Tabu Search Process

To enhance the performance of ERBFN, Tabu Search is employed as an optimization technique. Tabu Search is a metaheuristic algorithm designed to guide the search process and avoid local optima by maintaining a list of previously visited solutions (tabu list).

The output of a Gaussian radial basis function for an input x and μ is given by:

$$\phi(x) = \exp\left(-\frac{\|x-\mu\|^2}{2\sigma^2}\right) \quad (3)$$

Where:

‖x−μ‖ is the Euclidean distance between the input x and the center μ.

σ is the width of the Gaussian function.

Tabu Search is used to optimize the parameters of the ERBFN.The search process iteratively explores the solution space, updating the parameters to minimize a predefined objective function (e.g., mean squared error).

The output of the ERBFN for an input x is a weighted sum of the radial basis functions:

$$y(x) = \sum_{i=1}^{N} w_i \phi_i(x) \quad (4)$$

Where:

N is the number of hidden neurons.

wi is the weight corresponding to the i-th radial basis function ϕi(x).

Pseudocode

```
# ERBFN Initialization
Initialize number of hidden neurons N
Randomly initialize centersμ_i and widths σ_i for i = 1 to N
Initialize weights w_i for i = 1 to N
# Tabu Search Optimization
Initialize tabu list
Set initial solution S (ERBFN parameters μ_i, σ_i, w_i)
Define objective function J (e.g., mean squared error)
while not converged or max iterations not reached do
Generate neighbouring solutions {S'}
Evaluate objective function J for each S'
Select best S' not in tabu list or satisfying aspiration criterion
Update tabu list with current solution S
Move to best neighbouring solution S'
if S' is better than the best known solution then
Update best known solution
end if
end while
Return optimized ERBFN parameters (μ_i, σ_i, w_i)
```

V. RESULTS

The simulations were conducted using Python and specialized machine learning libraries such as Tensor Flow. The experiments were run on a high-performance computing cluster with Intel Xeon processors and 128GB RAM, ensuring the capability to handle large datasets and complex computations. The experimental parameters are given in Table VIII.

TABLE VIII. EXPERIMENTAL PARAMETERS

Methods	Parameter	Value
SOM	Grid Size	10x10
	Learning Rate	0.5
	Number of Iterations	1000
	Initialization Method	Random
	Neighborhood Function	Gaussian
	Radius	5
ERBFN	Radial Basis Function	Gaussian
	Centers Initialization	K-means
	Number of Centers	100
	Learning Rate	0.01
	Momentum	0.9
	Epochs	500
Tabu	List Size	50
	Search Iterations	100
	Aspiration Criterion	True
	Stopping Criteria	101non-improving
	Mutation Rate	0.1
	Crossover Rate	0.7
	Initial Temperature	100
	Cooling Schedule	Exponential

A. Performance Metrics

- Precision: Our method achieved a precision of 95.2%, indicating robust detection with minimal false alarms.
- Accuracy: The proposed model attained 96.5% accuracy, demonstrating its superior ability to correctly classify URLs.
- Recall: A recall of 94.8% highlights the model's effectiveness in identifying malicious URLs.
- F1-score: The F1-score of 95.0% underscores the model’s balanced performance.
- Specificity: The proportion of true negative detections among all actual negatives. High specificity means the model correctly identifies benign URLs, complementing the recall metric. Our model's specificity was not explicitly stated but can be inferred to be high due to the high overall accuracy and low false positive rate.

The performance for the proposed SOM-RMO + ERBFN method were compared against three existing methods: XGB (XGBoost), LR+SVC+DT (Logistic Regression, Support Vector Classifier, Decision Tree), and En_kNN as in Fig. 2 – 6 and Table IX and X. These comparisons were conducted over multiple test data sizes, as well as distinct training, testing, and validation datasets.

TABLE IX. PERFORMANCE OVER TRAINING, TESTING, AND VALIDATION DATA

Dataset	Method	Precision	Accuracy	Recall	F1-Score	Specificity
Training	XGB	0.81	0.83	0.79	0.80	0.84
	LR+SVC+DT	0.77	0.79	0.75	0.76	0.82
	En_kNN	0.79	0.81	0.77	0.78	0.83
	Proposed SOM-RMO + ERBFN	0.85	0.87	0.84	0.85	0.88
Testing	XGB	0.78	0.80	0.76	0.77	0.82
	LR+SVC+DT	0.74	0.76	0.72	0.73	0.79
	En_kNN	0.76	0.78	0.74	0.75	0.80
	Proposed SOM-RMO + ERBFN	0.82	0.84	0.81	0.82	0.85
Validation	XGB	0.79	0.81	0.77	0.78	0.83
	LR+SVC+DT	0.75	0.77	0.73	0.74	0.80
	En_kNN	0.77	0.79	0.75	0.76	0.81
	Proposed SOM-RMO + ERBFN	0.83	0.85	0.82	0.83	0.86

TABLE X. CONFUSION MATRIX OVER TRAINING, TESTING, AND VALIDATION DATA

Dataset	Method	TP	TN	FP	FN
Training	XGB	320	480	20	80
	LR+SVC+DT	300	470	30	100
	En_kNN	310	475	25	90
	Proposed (SOM-RMO + ERBFN)	340	485	15	60
Testing	XGB	160	240	10	40
	LR+SVC+DT	150	230	20	50
	En_kNN	155	235	15	45
	Proposed (SOM-RMO + ERBFN)	170	245	5	30
Validation	XGB	80	120	5	20
	LR+SVC+DT	75	115	10	25
	En_kNN	78	118	7	22
	Proposed (SOM-RMO + ERBFN)	85	122	3	15

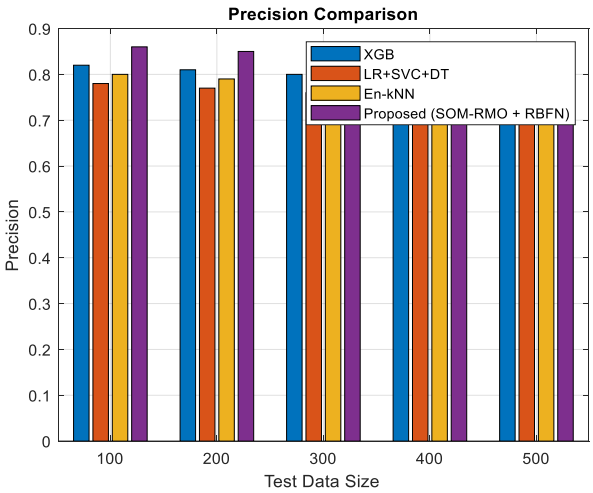


Fig. 2. Precision.

When evaluating the methods on test data sizes, the proposed Method (SOM-RMO + ERBFN) consistently outperformed other methods, achieving the highest precision (0.86 at 100 test data to 0.82 at 500 test data), accuracy (0.88 to 0.84), recall (0.85 to 0.81), F1-Score (0.86 to 0.82), and specificity (0.89 to 0.85).XGB showed solid performance but lagged behind the proposed method, with precision ranging from 0.82 to 0.78, accuracy from 0.84 to 0.80, recall from 0.80 to 0.76, F1-Score from 0.81 to 0.77, and specificity from 0.86 to 0.82.LR+SVC+DT and En_kNN both performed moderately, with LR+SVC+DT showing the lowest metrics across the board. En_kNN had intermediate performance, better than LR+SVC+DT but not as strong as XGB or the proposed method.

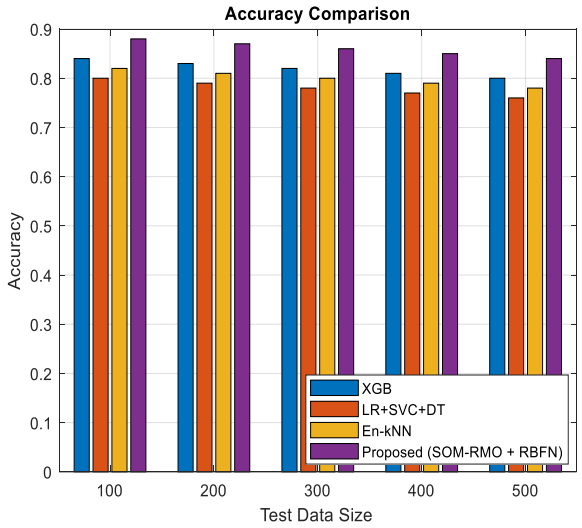


Fig. 3. Accuracy.

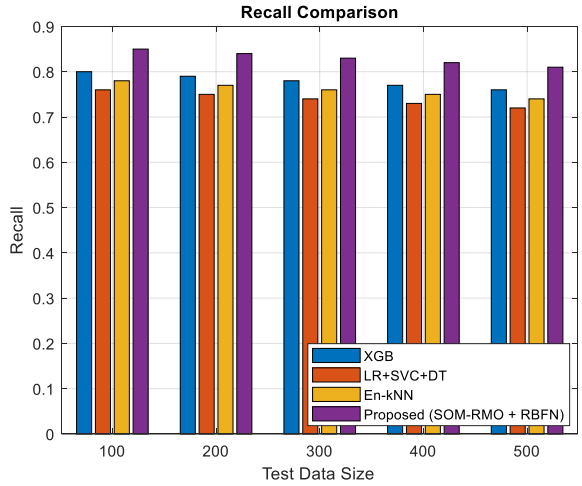


Fig. 4. Recall.

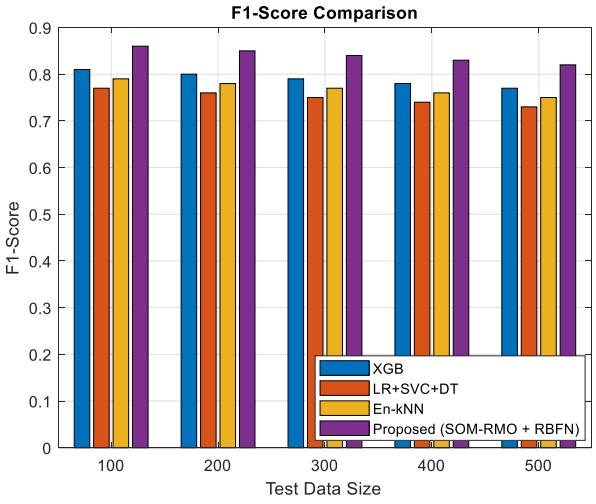


Fig. 5. F1-Score.

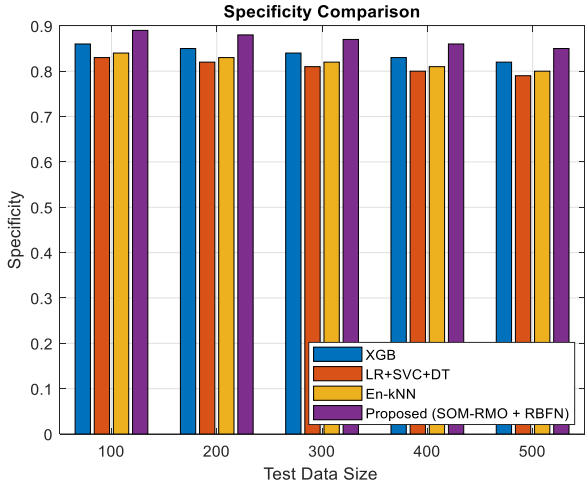


Fig. 6. Specificity.

For the training dataset, the proposed method achieved a precision of 0.85, accuracy of 0.87, recall of 0.84, F1-Score of 0.85, and specificity of 0.88. The testing and validation datasets followed similar trends, with the proposed method maintaining superior metrics compared to the other methods as in Table IX In comparison, the XGB, LR+SVC+DT, and En_kNN methods had lower counts of true positives and higher counts of false negatives and false positives, indicating less accurate performance as in Table X.

VI. Conclusion

The proposed approach seems to take advantage of both unsupervised learning for feature extraction and sophisticated classification algorithms for precise malicious URL identification by combining SOM-RMO for feature extraction with an improved ERBFN. To improve the performance of the classification model, Tabu Search is used for optimization. Overall, it appears that by enhancing both the feature extraction and classification stages, this hybrid approach tackles the challenging issue of malicious URL identification.

The higher efficiency of this approach is probably due to its advanced optimization techniques, which allow it to handle and classify complicated URL features with ease. It consistently outperformed XGB, LR+SVC+DT, and En_kNN across multiple test data sizes and dataset splits (training, testing, and validation). These results validate the utility of leveraging advanced feature extraction and optimization techniques in enhancing the accuracy and reliability of malicious URL detection models, making them robust tools for cyber security applications.

VII. Future Work

Future work would enhance the methodology, by expanding the feature set, and assessing across other domains. Increasing model interpretability, strengthening defences against adversarial attacks, and optimizing for real-time detection are important areas for improvement. Furthermore, improving efficiency, scalability, and investigating other optimization strategies could improve and increase the efficacy of dangerous URL detection

References

[1] Do Xuan, C., Nguyen, H. D., & Tisenko, V. N. (2020). Malicious URL detection based on machine learning. *International Journal of Advanced Computer Science and Applications*, *11*(1).

[2] Rupa, C., Srivastava, G., Bhattacharya, S., Reddy, P., &Gadekallu, T. R. (2021, August). A machine learning driven threat intelligence system for malicious URL detection. In Proceedings of the 16th International Conference on Availability, Reliability and Security (pp. 1-7).

[3] Raja, A. S., Vinodini, R., &Kavitha, A. (2021). Lexical features based malicious URL detection using machine learning techniques. Materials Today: Proceedings, 47, 163-166.

[4] Aljabri, M., Altamimi, H. S., Albelali, S. A., Al-Harbi, M., Alhuraib, H. T., Alotaibi, N. K., ... & Salah, K. (2022). Detecting malicious URLs using machine learning techniques: review and research directions. IEEE Access, 10, 121395-121417.

[5] Ahammad, S. H., Kale, S. D., Upadhye, G. D., Pande, S. D., Babu, E. V., Dhumane, A. V., & Bahadur, M. D. K. J. (2022). Phishing URL detection using machine learning methods. Advances in Engineering Software, 173, 103288.

[6] DR, U. S., &Patil, A. (2023, January). Malicious URL Detection and Classification Analysis using Machine Learning Models. In 2023 International Conference on Intelligent Data Communication Technologies and Internet of Things (IDCIoT) (pp. 470-476). IEEE.

[7] Mythreya, S., Murthy, A. S. D., Saikumar, K., & Rajesh, V. (2022). Prediction and prevention of malicious URL using ML and LR techniques for network security: machine learning. In Handbook of Research on Technologies and Systems for E-Collaboration During Global Crises (pp. 302-315). IGI Global.

[8] Janet, B., &Nikam, A. (2022, March). Real Time Malicious URL Detection on twitch using Machine Learning. In 2022 International Conference on Electronics and Renewable Systems (ICEARS) (pp. 1185-1189). IEEE.

[9] Pradeepa, G., & Devi, R. (2022). Review of malicious URL detection using machine learning. In Soft Computing for Security Applications: Proceedings of ICSCS 2021 (pp. 97-105). Springer Singapore.

[10] Aljabri, M., Alhaidari, F., Mohammad, R. M. A., Mirza, S., Alhamed, D. H., Altamimi, H. S., &Chrouf, S. M. B. (2022). An assessment of lexical, network, and content-based features for detecting malicious URLs using machine learning and deep learning models. Computational Intelligence and Neuroscience, 2022(1), 3241216.

[11] Alsaedi, M., Ghaleb, F. A., Saeed, F., Ahmad, J., &Alasli, M. (2022). Cyber threat intelligence-based malicious URL detection model using ensemble learning. Sensors, 22(9), 3373.

[12] Raja, A. S., Pradeepa, G., &Arulkumar, N. (2022, May). Mudhr: Malicious URL detection using heuristic rules based approach. In AIP Conference Proceedings (Vol. 2393, No. 1). AIP Publishing.

[13] Prabakaran, M. K., MeenakshiSundaram, P., &Chandrasekar, A. D. (2023). An enhanced deep learning-based phishing detection mechanism to effectively identify malicious URLs using variational autoencoders. IET Information Security, 17(3), 423-440.

[14] Abdul Samad, S. R., Balasubaramanian, S., Al-Kaabi, A. S., Sharma, B., Chowdhury, S., Mehbodniya, A., ... &Bostani, A. (2023). Analysis of the performance impact of fine-tuned machine learning model for phishing URL detection. Electronics, 12(7), 1642.

[15] Karim, A., Shahroz, M., Mustofa, K., Belhaouari, S. B., &Joga, S. R. K. (2023). Phishing detection system through hybrid machine learning based on URL. IEEE Access, 11, 36805-36822.

[16] Abu Al-Haija, Q., & Al-Fayoumi, M. (2023). An intelligent identification and classification system for malicious uniform resource locators (URLs). Neural Computing and Applications, 35(23), 16995-17011.

[17] Dataset,https://www.kaggle.com/datasets/sid321axn/malicious-urls-dataset.

Missing Value Imputation in Data MCAR for Classification of Type 2 Diabetes Mellitus and its Complications

Anik Andriani[1], Sri Hartati[2*], Afiahayati[3], Cornelia Wahyu Danawati[4]
Doctoral Program, Department of Computer Science and Electronics, Universitas Gadjah Mada, Yogyakarta, Indonesia[1]
Department of Computer Science and Electronics, Universitas Gadjah Mada, Yogyakarta, Indonesia[2, 3]
Department of Public Health, Universitas Gadjah Mada, Yogyakarta, Indonesia[4]
Department of Information System, Universitas Bina Sarana Informatika, Jakarta, Indonesia[1]

***Abstract*—Type 2 diabetes mellitus (T2DM) is a disease that is at risk for many complications. Previous research on the prognosis of T2DM and its complications is limited to the impact of T2DM on one particular disease. Guidebook for T2DM Management in Indonesia has eight categories of T2DM complications. The purpose of this study is to classify T2DM prognosis into eight categories: one controlled class and seven classes of aggravating disorders. The classification was based on medical record data from T2DM patients at Panti Rapih Hospital in Yogyakarta between 2017 and 2022. The problem is that the medical record data has numerous missing values (MV). The dataset had 29% missing values, classified as Missing Completely at Random (MCAR). This study performed imputation on the dataset prior to categorization. For MV imputation, a variety of imputation methods were used, and their accuracy was measured using Mean Absolute Error (MAE) and Root Mean Square Error (RMSE). The best imputation results were utilized to update the dataset. Subsequently, the dataset was used for classification employing several classification methods. The classification results were compared to determine the method with the highest accuracy in this scenario. The Decision Tree method with stratified k-fold cross-validation emerged as the optimal method for this classification. The results revealed an average accuracy value of 0.8529.**

***Keywords*—Missing value; prognosis of diabetes mellitus; missing completely at random; decision tree**

I. Introduction

Missing data is one of the most crucial problems in research. This tends to happen when collecting data [1]. A large number of missing values (MV) can reduce classification quality due to decreased performance on test data [2]. MV is very vulnerable to data such as weather [3], medical [1], finance [4], employees and salaries [5]. Patient medical record data is one example of data that frequently has MV.

Research regarding the prognosis of T2DM patients and their complications is needed to determine the progress of the patient's disease. The prognosis is a prediction of the development of a disease, including the progress of recovery, the emergence of other diseases, and even death. Detection of the patient's prognosis of the development of the disease complications is needed to determine the type of treatment and proper care [6]. Early detection of the prognosis of T2DM patients for their complications cannot be done medically. If this is done early, it can reduce the risk of complications [7]. This can be done by studying patient medical record data and patient activity data to determine the prevalence of T2DM for several diseases [8].

One of the supervised learning techniques used in medical research, including diagnosis, prognosis, and treatment, is classification [9]. In this study, we used medical record data from T2DM patients at Yogyakarta's Panti Rapih Hospital for classification. There is a 29% missing value rate in this data. There are 700 rows and 21 characteristics in this dataset. When classifying datasets with MV of less than 50% or less than 30%, previous research frequently overlooked MV data and deleted them, producing biased classification findings [10]. Missing values in the medical record data of T2DM patients are randomly present in some features of the dataset. The type of MV in this T2DM dataset is Missing Completely at Random (MCAR). Datasets with MCAR show that MVs appear randomly independent of a feature. In MCAR, the appearance of MV does not depend on another variable [11]. This study proposes a classification model for the prognosis of T2DM patients with this complex disease by first imputing the data.

Classification using a variety of imputation methods and classification methods was done on several datasets at different percentages of MVs. The results demonstrated that accuracy is decreased when the percentage of MVs increased [12]. This study conducted experiments on datasets with a fairly high percentage of MVs with the type of MV in the dataset being MCAR. Imputation on the MCAR dataset requires intransitive imputation techniques, namely the imputation of missing values on observed variables is independent of other variables. MEAN is one of the most frequently used imputation methods in the intransitive imputation type. In addition, Linear Regression is one of the methods often used in this case [13]. In previous studies, the KNN method provided the best performance in imputing missing values on the MCAR Dataset containing numeric data [14]. Therefore, this study uses missing value imputation, including MEAN, K-Nearest Neighbor (KNN), and Linear Regression (LR). Meanwhile, the classification methods employed are Decision Tree (DT), Naïve Bayes (NB), and Support Vector Machine (SVM).

*Corresponding Author.

II. METHODS

A. Proposed Model

This research consists of several stages. Fig. 1 illustrates the stages of the research, which include dataset preparation, calculation of correlation values between features, MV imputation, data validation, classification, and evaluation. Fig. 1 shows the stages of the research. The dataset for T2DM patients' prognoses regarding their complicating diseases comprises 700 rows and 21 features. One of these features is the target feature, which encompasses eight classes.

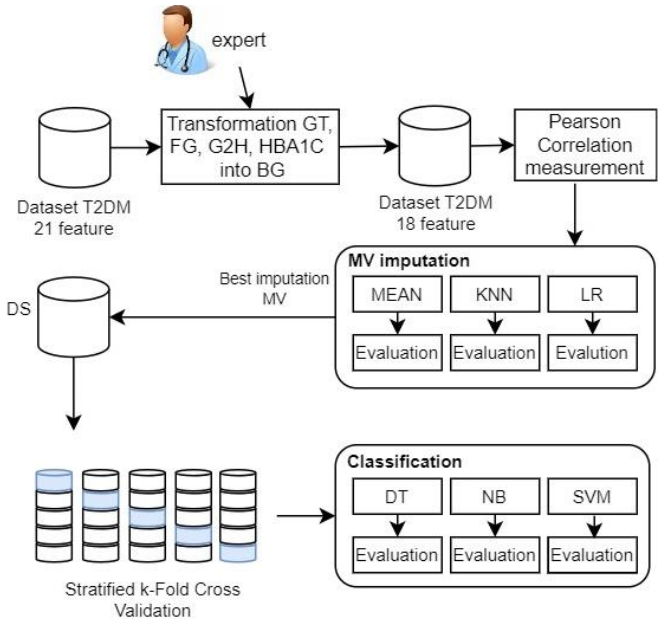


Fig. 1. Research stages.

In the first stage, two datasets were created: DS21 and DS18. DS21 is a T2DM dataset without any transformation, while DS18 is a T2DM dataset that underwent a transformation process. In this study, transformation refers to the process of consolidating features that can be represented by a single feature to determine a value. These features include Blood Glucose at Time (GT), Fasting Glucose (FG), and Blood Glucose 2 Hours after meals (G2H), which are combined into one feature called Blood Glucose level (BG). Doctors do not always rely on all three glucose test values to determine a patient's blood glucose level; sometimes, they only check GT, FG, or G2H. Consolidating multiple features that can substitute for each other into a single feature is also beneficial for reducing the number missing values. The outcome of this transformation is DS18, a dataset containing 18 features.

Dataset imputation involves the use of MEAN, KNN, and LR methods. Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) were employed to assess the imputation. A comparison of the three methods is conducted to determine the most effective imputation results. An error value close to 0 indicates an imputed value that is close to the true are minimized in the DS dataset. The subsequent step involves classification using DS. The classification methods utilized are DT, NB, and SVM. The classification outcomes from these three methods are assessed based on accuracy values to identify the most suitable approach for classifying the T2DM prognosis and its associated complications.

B. Missing Value Imputation Method

MV is handled using the Imputation Missing Value (IMV) technique. The IMV phases are described in Fig. 2. The correlation value between the features is calculated at the start of the stage. Correlation values for numerical features are computed using Pearson correlation. A relationship between two features is shown by a positive correlation value. Features that show correlation are given regression values. The outcomes are applied to IMV. The dataset (DS) that will be utilized for classification is subsequently created using the imputation findings.

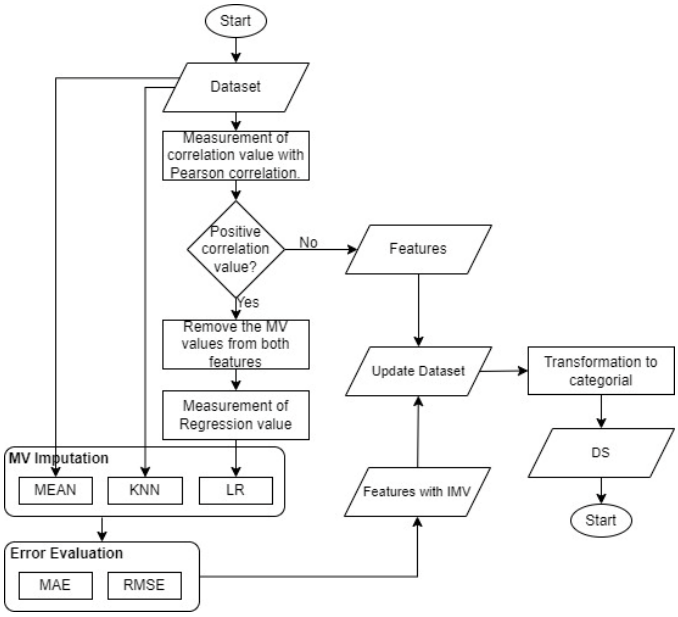


Fig. 2. IMV stages.

The imputation steps shown in Fig. 2 illustrate the three types of imputation that are used: MEAN, KNN, and LR. IMV calculates the average value of the observed feature values in the MEAN method. A mean computation for IMV is assumed in Eq. (1). The MEAN value computation relies on the global average value derived from the summation of all values in the observed features in the dataset, denoted as N. Eq. (1) presents a formula for determining the mean value [14].

$$\mu_i = \frac{1}{N} \sum_{i=1}^{N} X_i \quad (1)$$

Imputation using the KNN methods entails dividing the dataset into complete data and MV data. The method finds the closest value to impute MV data by comparing it with the entire data sample. Eq. (2) is used to calculate the Euclidean distance (d), which is the distance between two points in a two-dimensional space [15].

$$d = \sqrt{\sum_{r=1}^{n} (x_{ir} - x_{jr})^2} \quad (2)$$

The imputation stage employs the LR method to compute the correlation value between features utilizing Pearson correlation. The correlation values are compute among features

to identify those with a positive correlation. Features exhibiting a positive correlation are employed for MV imputation, whereas those with a negative correlation are excluded from MV imputation. The linear correlation between the two properties as shown in Eq. (3) is characterized by the Pearson correlation value. A strong correlation between two dependent qualities is indicated by a Pearson correlation coefficient value of 1, which ranges from -1 to 1.

In the imputation stage, Pearson correlation is used to calculate the correlation value between features using the LR approach. The correlation values are compute among features to identify those with a positive correlation. Features exhibiting a positive correlation are employed for MV imputation, whereas those with a negative correlation are excluded from MV imputation. The linear correlation between the two properties as shown in Eq. (3) is characterized by the Pearson correlation value [16]. A strong correlation between two dependent features (X, Y) is indicated by a Pearson correlation coefficient (ρ_{XY}) value of 1, which ranges from -1 to 1.

$$\rho_{XY} = \frac{cov(X,Y)}{\sigma_X \sigma_Y} \tag{3}$$

The process of IMV for features exhibiting a positive correlation commences with the MV values from both features. This step leads to the retrieval of complete values for both features. Subsequently, the regression value is calculated based on complete feature data, and the MV value is estimated based on the regression value through the utilization of Eq. (4), (5), and (6) [17].

$$c = \frac{n\sum XY - \sum X \sum Y}{n\sum X^2 - (\sum X)^2} \tag{4}$$

$$d = \frac{\sum Y - c\sum X}{n} \tag{5}$$

Where, n is the selected amount of data points, even or c, and coefficients or d. They y-intercept is the point on the y-axis where the graph crosses the y-axis. It is the place where the line's slope, which indicates how it step is, is located. Eq. (6) describes to build linear regression.

$$\bar{Y} = c + dX \tag{6}$$

In this formulation, the y-intercept is shown by c, while the slope of the line is represented by d. $\bar{Y}$ notation for represented the expected value of the dependent variable Y for a specific value of the independent variable X. When building a line in algebra, the equation of the line must be found at two locations (x,y).

Implementations MAE and RMSE for quantify the error in IMV result from the three imputation methods. Eq. (7) was employed to determine the MAE value, while Eq. (8) was utilized for calculating the RMSE value [18].

$$MAE = \left(\frac{1}{n}\right)\sum_{i=1}^{n} |x_i - \bar{x}_i| \tag{7}$$

$$RMSE = \sqrt{(\sum_{i=1}^{n}(x_i - \bar{x}_i)^2/n)} \tag{8}$$

Evaluation of MV imputation results with the third method was compared to find out which method produced the best MV imputation values. Evaluation with the lowest MAE and RMSE and close to 0 is the best MV imputation result. The best imputation results are then used to build the DS dataset.

C. Classifications Method

Classifications were applied in the DS dataset, utilizing various methods such as Decision Tree (DT), Naïve Bayes (NB), and Support Vector Machine (SVM). These three widely recognized classification techniques are commonly employed in research related to classification.

The Decision Tree (DT) method is a classification algorithm that is widely used in data mining and machine learning. This algorithm predicts target values by learning simple decision rules derived from the features in the dataset. DT segments data into smaller subsets based on existing features, and each subset is then processed recursively. The selection of the most informative features is achieved by minimizing impurities (such as entropy or Gini impurity) at each data division. This method offers advantages, such as being easy to understand and interpret. DT can be depicted in the form of a decision tree structure and can process both categorical and numeric data. The Decision Tree method has been applied in various applications, from pattern recognition and business data analysis to medical diagnosis. In DT, it is necessary to compute the Entropy value first to measure the level of uncertainty or irregularity in a dataset. In the context of DT, entropy is often utilized to assess the quality of splits of attributes used as nodes in a decision tree. Eq. (9) illustrates the calculation of the entropy value for the data. The entropy value obtained is then used to calculate Information Gain. Equation 10 demonstrates the calculation of Information Gain, which is used to classify classes by segregating data based on specific features. The feature with the highest Information Gain values is chosen as the root feature.

$$En_{A_i}(Data) = \sum_{j=1}^{k} \frac{|Attr|}{|Data|} . Attr_j \tag{9}$$

$$InfoG(A_i) = En(Data) - En_{A_i}(Data) \tag{10}$$

Naïve Bayes (NB) uses a Bayesian learning approach that incorporates the concept of probability in classification tasks. One of the most straightforward and widely used Bayesian learning models is Naïve Bayes [19]. Naïve Bayes performs exceptionally well in multiclass classification scenarios with a single label, delivering high accuracy. Multiclass classification with a single label involves categorizing data into more than two classes, with each class assigned only one label [20]. The calculation of probability values in Naïve Bayes is based on Eq. (11), where P is probability.

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \tag{11}$$

Support Vector Machine (SVM) works by finding a hyperplane to separate two classes in binary classification. In multiclass classification, multiple binary SVMs are used [21]. Eq. (12) defines vector w_i as the hyperplane coefficient vector, b as the bias, and $f(x) = 0$, for x on the hyperplane.

$$f(x) = \sum_i w_i \times x_i + b \tag{12}$$

III. Result and Discussion

A. Dataset Preparation

Learning to classify the prognosis of T2DM for its complications involves multiclass classification learning. In the dataset, there is a feature called the Target Class, which consists of eight classes. According to the Guidebook for the Management and Prevention of Adult Type 2 Diabetes Mellitus in Indonesia, prepared by the Indonesian Endocrinology Association (PERKENI), there are eight diseases classified as risks and complications of T2DM. Both of these are considered complications of T2DM. Table I presents information regarding complications associated with Type 2 Diabetes Mellitus (T2DM) and the conversion of values into the Target Class. The table delineates data on the various complications. These eight categories serve as the target class labels in the dataset for classification purposes. The data on the complication categories of T2DM was extracted from the 2021 Guide to Management and Prevention of Type 2 Diabetes Mellitus in Indonesia.

TABLE I. Categories of Complications in T2DM

Feature	Categories of Complication	Disease	Class label	Class label quantity
Prognosis of Complication	Controlled	-	0	186
	Nephropathy	CKD, Diabetic NephropathyInsuff Renal	1	96
	Cardiovascular	IHD, CHF, KAD	2	193
	Neuropathy	Neuropathy, Cellulitis	3	95
	Hyperglycemia	Hyperglycemia	4	70
	Macroangiopathy	Macroangiopathy, Ulkus DM	5	40
	Hypoglycemia	Hypoglycemia	6	2
	Retinopathy	Retinopathy Diabetic	7	18

The dataset comprises 21 features, with one of them being the target feature. It consists of 700 rows of data. The features are detailed in Table II. Within the T2DM patient dataset, there are 4321 missing values out of a total 14700 values, representing 29% missing values. The distribution of MV in the dataset is illustrated in Fig. 3.

TABLE II. Dataset Features

Feature	Feature	Type
Gender	GEN	Categorial
Age	AGE	Numeric
Blood Glucose at Time	GT	Numeric
Fasting Glucose	FG	Numeric
Blood Glucose 2 Hours after meals	G2H	Numeric
HbA1C	HBA1C	Numeric
Creatinine	CREAT	Numeric
Ureum	UREUM	Numeric
Systolic	SYST	Numeric
Diastolic	DIAST	Numeric
Cholesterol	CHOL	Numeric
Low-density lipoproteins	LDL	Numeric
High-density lipoproteins	HDL	Numeric
Triglycerides	TGD	Numeric
Uric Acid	UA	Numeric
Nutrition	NUT	Categorial
Treatment	TREAT	Categorial
Early Diagnosis	ED	Categorial
Hypertension	HT	Categorial
Early Complications	EC	Categorial
Prognosis Complications	PC	Class target

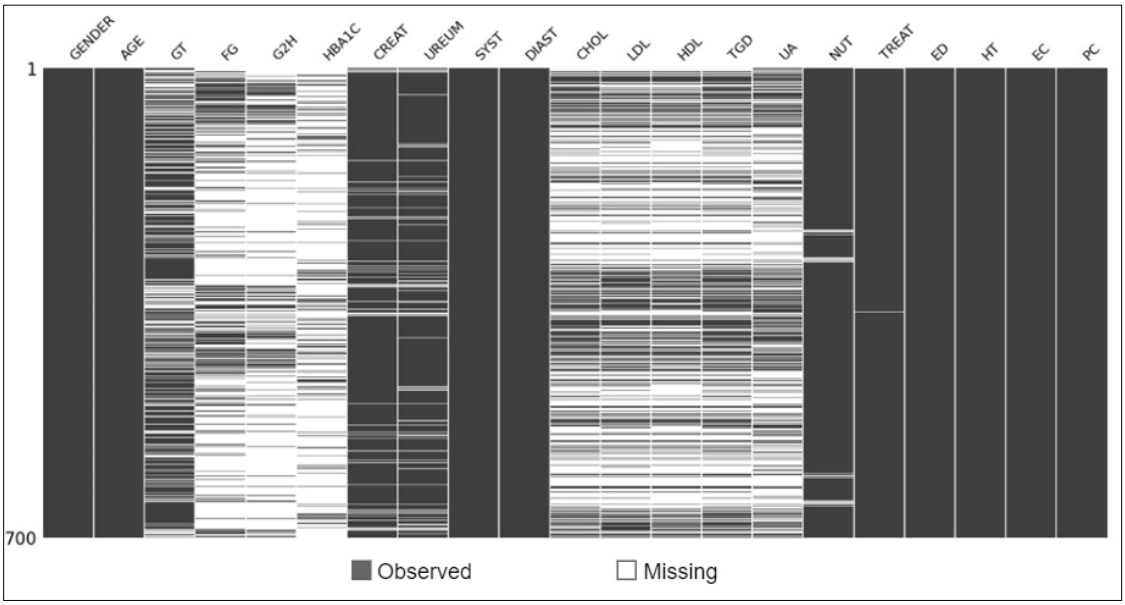


Fig. 3. MV distributions.

The T2DM dataset exhibits 29% MV. Omitting these missing values from the dataset would lead to the loss of a significant portion of the data. Out of the 700 rows in the dataset, only three complete rows of data are available. Therefore, it is essential to perform MV imputation in order to enhance the dataset by increasing the number of usable data rows for classification purposes.

Following consultations with experts, particularly internal medicine physicians, it has been suggested that certain features can be consolidated into a single feature. Specifically, the GT, FG, G2H, and HBA1C features can be amalgamated into a singular feature known as the Diabetes Blood Sugar feature. This consolidation is based on the observation that medical practitioners may not always assess all four features when determining a patient's blood sugar status. Frequently, doctors may only examine one or a combination of features to ascertain blood sugar levels. The transformation of these four features into a single feature is guided by the Blood Sugar category outlined in Table III.

TABLE III. CATEGORIES OF BLOOD GLUCOSE

Value	Categories	Categories Label
HBA1C: <5.7 GT: 70-139 mg/dL FG: 70-99 mg/dL G2H: 70-139 mg/dL	Normal	1
HBA1C: 5.7-6.4 GT: 140-199 mg/dL FG: 100-99 mg/dL G2H: 140-199 mg/dL	Prediabetes	2
HBA1C: >=6.5 GT: 200-299 mg/dL FG: 126-199 mg/dL G2H: 200-299 mg/dL	Diabetes	3
GT: >=300 mg/dL FG: >=200 mg/dL G2H: >=300 mg/dL	Hyperglycemia	4
GT: < 70 mg/dL FG: < 70 mg/dL G2H: < 70 mg/dL	Hypoglycemia	5

The process of consolidating four features into one feature was conducted in collaboration with experts, specifically internal medicine physicians. The outcome of this consolidation effectively decreased the missing values (MVs) by 9%, reducing them from 29% to 20%. Despite this reduction, further MV imputation is deemed necessary to minimize the number of missing values.

B. Missing Value Imputation Result

The equations are an exception to the prescribed specifications of this template. You will need to determine whether or not your equation should be typed using either the Times New Roman or the Symbol font (please no other font). To create multileveled equations, it may be necessary to treat the equation as a graphic and insert it into the text after your paper is styled.

The subsequent missing value (MV) imputation technique employs the K-nearest neighbors (KNN) method. This approach is commonly utilized for MV imputation and involves imputing missing values by computing similarity metrics based on distances. Eq. (2) is employed to determine the imputation value. The distribution of MV data post-KNN imputation is illustrated in Fig. 6.

IMV using Linear Regression has multiple steps that must go in order. These steps involve figuring out regression coefficients, estimating MV values, and computing correlation coefficients between features. In order to determine the interdependencies between all features, a correlation analysis is carried out using numerical data.

	AGE	CREAT	UREUM	DIAST	CHOL	LDL	HDL	TGD	UA
AGE	1.000	-0.149	-0.106	-0.047	-0.077	-0.135	0.118	-0.039	-0.151
CREAT	-0.149	1.000	0.831	0.039	0.108	0.103	-0.273	0.122	0.417
UREUM	-0.106	0.831	1.000	0.025	-0.018	-0.028	-0.169	0.058	0.376
DIAST	-0.047	0.039	0.025	1.000	0.366	0.270	0.032	0.167	0.033
CHOL	-0.077	0.108	-0.018	0.366	1.000	0.721	0.367	0.500	0.091
LDL	-0.135	0.103	-0.028	0.270	0.721	1.000	0.297	-0.020	0.104
HDL	0.118	-0.273	-0.169	0.032	0.367	0.297	1.000	-0.195	-0.244
TGD	-0.039	0.122	0.058	0.167	0.500	-0.020	-0.195	1.000	0.139
UA	-0.151	0.417	0.376	0.033	0.091	0.104	-0.244	0.139	1.000

Fig. 4. Pearson correlation values in numeric features.

Eq. (3) is utilized to calculate Pearson's correlation coefficient, which measures the degree of linear association between features. Fig. 4 shows the results of the correlation analysis.

Based on the results of correlation calculations, several features exhibited significant correlation values. Specifically, CREAT was correlated with UREUM, CHOL with DIAST, CHOL with LDL, CHOL with HDL, CHOL with TGD, and CREAT with UA. The Pearson correlation values for these features are visually represented in Fig. 5.

MV imputation was performed on six features (CREAT, UREUM, AU, CHOL, LDL, TGD) based on correlation values. The imputation reduced missing values from 20% to 2% in the dataset, resulting in 598 rows out of 700. Fig. 6 illustrates the distribution of imputed data in the dataset.

Evaluation of the MV imputation results is conducted to assess their accuracy compared to the actual values. This assessment involves calculating error values using MAE and RMSE equations. The evaluation compares imputation results from three methods: MEAN, KNN, and LR. Table IV displays the MAE and RMSE for the three imputation methods. The evaluation shows that MV imputation with LR yields the smallest errors, close to 0. The dataset imputed using LR will be used for classifying T2DM prognosis and its complications.

TABLE IV. EVALUATION OF ERRORS IN IMPUTATION RESULTS

Feature	MEAN		LR		KNN	
	MAE	RMSE	MAE	RMSE	MAE	RMSE
CREAT-UREUM	13.5	338.4	1.6E-14	4.0E-13	0.1	1.6
CHOL-DIAST	7.4	120.9	1.9E-15	3.1E-14	2.5E-13	4.1E-12
LDL-CHOL	0.1	1.9	1.0E-14	2.0E-13	1.6E-13	2.7E-12
TGD-CHOL	31.5	529.8	5.0E-14	8.0E-13	5.2E-13	8.8E-12
CREAT-UA	0.2	3.4	3.0E-15	5.0E-14	0.02	0.3

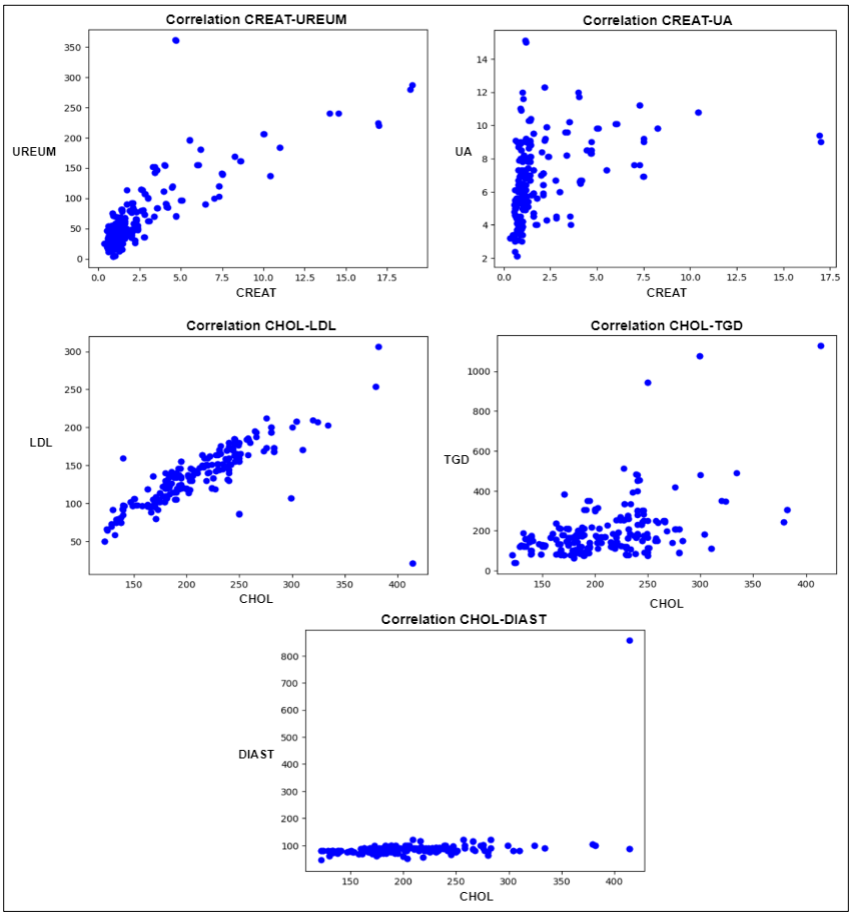


Fig. 5. Features with a high correlation values.

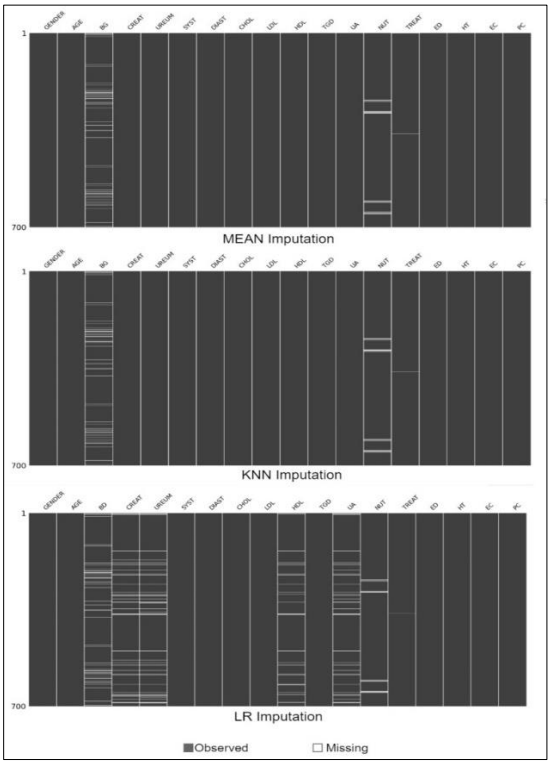


Fig. 6. Distribution of MV.

C. Classification Result

The pre-processing stage after MV imputation is data transformation. The dataset of MV imputation results with LR is used for this study. As seen in Table V, the variables in the dataset are converted into categorical form.

TABLE V. DATASET FEATURE CATTEGORIES FOR TRANSFORMATION DATA

Feature	categories
AGE	1-14 (1), 15-24(2), 25-44(3), 45-64(4), >65(5)
GENDER	Man(1), Women(2)
BG	Normal (1), prediabetes (2), diabetes(3), hyperglicemia (4), hypoglicemia (5)
CREAT	Woman: <0.6 (Low=1), 0.6-1.1(Normal=2), >1.1(High=3) Man: <0.7 (Low=1), 0.7-1.3 (Normal=2), >1.3 (High=3)
UREUM	<6 (Low=1), 6-23(Normal=2), >23(High=3)
SYST	Age <=60: <90(Low=1), 90-120(Normal=2), >120 (High=3) Age >60: <100(Low=1), 100-140(Normal=2), >140(High=3)
DIAST	Age <=60: <60(Low=1), 60-80(Normal=2), >80 (High=3) Age >60: <60(Low=1), 60-90(Normal=2), >90(High=3)
CHOL	<200 (Normal=1), 200-240 (Borderline=2), >240 (High=3)
LDL	<100 (Normal=1), 100-129 (Optimal=2), 130-160(Borderline=3), >160 (High=4)
HDL	>=40 (Normal=1), <40 (Low=2)
TGD	<149 (Normal=1), 150-200 (Borderline=2), >200 (High=3)
UA	Woman: <1.5 (Low=1), 1.5-6(Normal=2), >6 (High=3) Man: <2.5 (Low=1), 2.5-7 (Normal=2), >7 (High=3)
NUT	Good(1), Enough(2), Medium(3), Less(4), Over(5)
TREAT	Routine check-up+medicine(1), Medicine(2), Non-Routine check-up(3), Insulin(4), Insulin+Medicine(5), Non-routine medicine(6)
ED	DM2NO(1), DM2 OBESE(2), DM2 HYPERGLICEMIA(3)
HT	Yes(1), No(2)
EC	No(0), CKD(1), Nephropathy(2), Insuff renall(3), IHD(4), CHF(5), Stroke/Post Stroke(6), KAD(7), Neuropathy(8), Hyperglycemia(9), Ulcus(10),Cellulitis(11), Retinopathy(12)

Following MV imputation and data transformation in the pre-processing stage is the data validation stage. Data validation and classification are the two primary phases of the classification process. To divide the data into training and testing sets, the data validation stage uses the stratified k-fold cross-validation technique, as shown in Fig. 1. By using this technique, it is ensured that the training and testing sets of data have equal class distributions. The method is dividing the data into k folds, or groups, at random [20]. K-fold cross-validation with k=5 is used in this investigation. Five folds are created in the dataset for k-fold cross-validation. One-fold is used for testing data and (k-1) folds are used for training data in each fold. The process is iterated so that each fold serves as the testing data exactly once.

D. Evaluation Result

Classification results are assessed using accuracy values. Table VI provides a comparison of the accuracy values obtained from various classification methods.

TABLE VI. CATEGORIES OF BLOOD GLUCOSE

k-Fold=5	DT	NB	SVM
Fold1	0.9000	0.2750	0.5167
Fold2	0.8417	0.2667	0.5250
Fold3	0.8083	0.2750	0.4250
Fold4	0.8824	0.2185	0.5714
Fold5	0.8319	0.2269	0.5462
Average	0.8529	0.2524	0.5169

Table V presents the classification outcomes, indicating that the Decision Tree method yields the highest average value for classification. Across all folds, consistent accuracy exceeding 0.8 is achieved when employing the Decision Tree approach. Conversely, the Naïve Bayes and Support Vector Machine methods exhibit notably lower classification accuracies, both falling below 60%. These methods do not offer sufficiently accurate results for predicting the prognosis of T2DM patients with their complications.

E. Discussion

Our findings show that MV imputation results on the MCAR type T2DM prognosis dataset are shown in Fig. 6. The MEAN and KNN imputation methods impute more missing values than the LR imputation method. However, based on the evaluation with the MAE and RMSE values in Table IV, the imputation results with the LR method provide error values close to 0. This means that the synthetic data which is the result of imputation is close to the real value. In addition, the application of the dataset that has been imputed with the LR method works well when the classification method with Decision Tree is applied as shown in Table VI. The LR imputation method is an alternative for data imputation with a high percentage of missing values in the MCAR type dataset in addition to the MEAN and KNN imputation methods in previous studies.

IV. CONCLUSIONS

Imputing missing values in medical data is crucial before initiating the classification learning process, especially when working with datasets that frequently have many missing values. In the context of predicting complications in patients with Type 2 Diabetes Mellitus (T2DM), the utilization of the LR method for missing value imputation has demonstrated lower error rates in terms of Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) calculations compared to the MEAN and KNN approaches. Furthermore, the fraction of missing values on this specific dataset was significantly reduced from 29% to 2% by using missing value imputation with LR. The imputed dataset generated by the LR technique is then utilized to categorize T2DM patients' prognosis with respect to their problems. Stratified k-fold cross-validation with k-fold 5 has been identified as the best validation approach for getting the most accurate classification results when using the Decision Tree method during the data validation step. The average accuracy achieved through the Decision Tree method, utilizing a blend of training and testing data, is reported to be 0.8529.

Research on the classification of Type 2 Diabetes Mellitus (T2DM) prognosis for its complications is presently constrained to a single complication that manifests first. Subsequent research could focus on the classification of T2DM prognosis based on potential disease composition.

REFERENCES

[1] R. K. Bania and A. Halder, "R-Ensembler: A greedy rough set based ensemble attribute selection algorithm with kNN imputation for classification of medical data," Comput. Methods Programs Biomed., vol. 184, p. 105122, 2020, doi: 10.1016/j.cmpb.2019.105122.

[2] U. Bentkowska, J. G. Bazan, W. Rząsa, and L. Zaręba, "Application of interval-valued aggregation to optimization problem of k−NN classifiers for missing values case," Inf. Sci. (Ny)., vol. 486, pp. 434–449, 2019, doi: 10.1016/j.ins.2019.02.053.

[3] A. Aieb, K. Madani, M. Scarpa, B. Bonacorso, and K. Lefsih, "A new approach for processing climate missing databases applied to daily rainfall data in Soummam watershed, Algeria," Heliyon, vol. 5, no. 2, p. e01247, 2019, doi: 10.1016/j.heliyon.2019.e01247.

[4] Q. Lan, X. Xu, H. Ma, and G. Li, "Multivariable data imputation for the analysis of incomplete credit data," Expert Syst. Appl., vol. 141, 2020, doi: 10.1016/j.eswa.2019.112926.

[5] L. Ni, F. Fang, and J. Shao, "Feature screening for ultrahigh dimensional categorical data with covariates missing at random," Comput. Stat. Data Anal., vol. 142, p. 106824, 2020, doi: 10.1016/j.csda.2019.106824.

[6] L. Qiao, Y. Zhu, and H. Zhou, "Diabetic Retinopathy Detection Using Prognosis of Microaneurysm and Early Diagnosis System for Non-Proliferative Diabetic Retinopathy Based on Deep Learning Algorithms," IEEE Access, vol. 8, pp. 104292–104302, 2020, doi: 10.1109/access.2020.2993937.

[7] M. A. H. Shareef, K. Narasimhalu, S. E. Saffari, F. P. Woon, and D. A. De Silva, "Reccurrent vascular events partially explain association between diabetes and poor prognosis in young ischemic stroke patients." pp. 1–4, 2024. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/38219383/#:~:text=Recurrent vascular events partially explain association between diabetes,J Neurol Sci. 2024 Jan 11%3A457%3A122881. doi%3A 10.1016%2Fj.jns.2024.122881.

[8] B.-Y. Zhou et al., "Association of D-dimer with long-term prognosis in type 2 diabetes mellitus patients with acute coronary syndrome," Nutr. Metab. Cardiovasc. Dis., no. xxxx, 2022, doi: 10.1016/j.numecd.2022.05.013.

[9] Eliyani, S. Hartati, and A. Musdholifah, "Machine Learning Assisted Medical Diagnosis for Segmentation of Follicle in Ovary Ultrasound." Springer, pp. 71–80, 2019. [Online]. Available: https://link.springer.com/chapter/10.1007/978-981-15-0399-3_6.

[10] Y. Pin Chen, C. Hua Huang, Y. Hsun Lo, Y. Ying Chen, and F. Lai, "Handle MV on Time Series Data.pdf." pp. 1271–1287, 2022.

[11] T. M. Pham, N. Pandis, and I. R. White, "Pham-MCAR MAR MNAR-2022.pdf." American Journal of Orthodontics and Dentofacial Orthopedics, pp. 138–139, 2022.

[12] C. F. Tsai and Y. H. Hu, "Empirical comparison of supervised learning techniques for missing value imputation," Knowl. Inf. Syst., vol. 64, no. 4, pp. 1047–1075, 2022, doi: 10.1007/s10115-022-01661-0.

[13] S. M. Mostafa, "Imputing missing values using cumulative linear regression," CAAI Trans. Intell. Technol., vol. 4, no. 3, pp. 182–200, 2019, doi: 10.1049/trit.2019.0032.

[14] F. I. Kurniadi, R. C. Rohmana, and L. Taufani, "Kurniadi-Local Mean Imputation for Handling MV-2023.pdf." Procedia Computer Science, pp. 301–309, 2023.

[15] D. Zou et al., "Outlier detection and data filling based on KNN and LOF for power transformer operation data classification." pp. 698–711, 2023.

[16] S. Peng, W. Han, and G. Jia, "Pearson correlation and transfer entropy in the Chinese stock market with time delay," Data Sci. Manag., vol. 5, no. 3, pp. 117–123, 2022, doi: 10.1016/j.dsm.2022.08.001.

[17] J. M. Sangeetha and K. J. Alfia, "Financial stock market forecast using evaluated linear regression based machine learning technique," Meas. Sensors, vol. 31, no. April 2023, p. 100950, 2024, doi: 10.1016/j.measen.2023.100950.

[18] D. S. K. Karunasingha, "Root mean square error or mean absolute error? Use their ratio as well," Inf. Sci. (Ny)., vol. 585, pp. 609–629, 2022, doi: 10.1016/j.ins.2021.11.036.

[19] S. Farhana, "Classification of Academic Performance for University Research Evaluation by Implementing Modified Naive Bayes Algorithm," Procedia Comput. Sci., vol. 194, pp. 224–228, 2021, doi: 10.1016/j.procs.2021.10.077.

[20] A. Kag, L. M. Jenila Livingston, L. M. Livingston Merlin, and L. G. X. Agnel Livingston, "Multiclass Single Label Model for Web Page Classification," 2019 Int. Conf. Recent Adv. Energy-Efficient Comput. Commun. ICRAECC 2019, pp. 3–8, 2019, doi: 10.1109/ICRAECC43874.2019.8995087.

[21] B. A. Akinnuwesi et al., "Application of Support Vector Machine Algorithm for Early Differential Diagnosis of Prostate Cancer." pp. 1–12, 2023.

Optimizing Dance Training Programs Using Deep Learning: Exploring Motion Feedback Mechanisms Based on Pose Recognition and Prediction

Yuting Jiao*
Quanzhou Preschool Teachers College, Fujian China

***Abstract*—Dance pose recognition and prediction is an important part of dance training and a challenging task in the field of artificial intelligence. Due to the diverse styles and significant variations in dance movements, conventional methods struggle to capture effective dance pose features for recognition. In this context, we have developed a dance pose recognition and prediction method based on deep learning. Given the characteristics of dance movements, such as complex human postures and dynamic movements, we proposed the MKFF-ST-GCN model, which integrates multi-kinematic feature fusion with ST-GCN. This model fully captures the dynamic information of dance movements by calculating the first and second-order kinematic features of keypoints and fuses the kinematic features using a multi-head attention mechanism. Additionally, to address dance pose prediction issues, we proposed the STGA-Net based on the spatial-temporal graph attention mechanism. This model improves the long-distance information modeling capability by calculating local and global graph attentions of dance poses, effectively solving the problem of dance pose prediction. To comprehensively evaluate the quality of the proposed methods in dance pose recognition and prediction, we conducted extensive experimental validations and comparisons with several common algorithms. The experimental results fully demonstrate the effectiveness of our methods in dance pose recognition and prediction. This study not only advances the technology of dance pose recognition and prediction but also provides valuable experience for the field.**

***Keywords*—Deep learning; pose recognition; pose prediction; dance training; graph convolutional network; attention mechanism**

I. Introduction

Dance constitutes a significant expressive medium within the realm of modern art, serving as a pivotal channel for not only emotional expression but also the transmission of cultural heritage. Traditional dance training predominantly relies upon the professional abilities of instructors, with evaluative criteria heavily influenced by subjective experiences and judgments, which lacks objectivity and necessitates substantial resource and time investments. With the proliferation of dance training, an increasing number of practitioners are demanding a cost-effective, personalized learning experience equipped with real-time feedback, thus underscoring the imperative for the development of innovative dance training methodologies. In the recent era, advancements in artificial intelligence (AI) technology, particularly through deep learning techniques within the domain of computer vision, have been noteworthy [1]-[3]. These techniques are adept at extracting latent information from image data, and their application has been extensive in areas such as defect detection and facial recognition. The application of deep learning for human action recognition and prediction in dance training allows for the effective recognition and comprehension of dance movements. This not only markedly enhances the efficiency and quality of dance training but also opens novel avenues for the transformation of traditional dance training paradigms.

Integrating human pose recognition and prediction with dance training represents a multidimensional research domain where technology and art converge extensively. This approach seeks to leverage the sophisticated image recognition capabilities inherent in deep learning algorithms to capture and analyze the movements and postures of practitioners with precision, thereby achieving objective and scientific training feedback. Additionally, human posture prediction based on deep learning can integrate sequential image data to predict future movements, effectively preventing potential health risks during training. Therefore, accurately recognizing and predicting human dance movements holds significant theoretical and practical significance. It can reduce the subjectivity of teacher evaluations, help establish objective and personalized evaluation standards, and provide practitioners with more convenient and cost-effective dance training methods.

In the domain of dance education, accurately and effectively identifying and predicting sequences of dance movements presents a significant challenge [4]. This challenge stems from the complexity inherent in human posture, the diversity of dance styles, and the inherent uncertainty of movement execution. To address these issues, it is crucial to enhance the learning capabilities of algorithms to ensure better fitting and prediction accuracy for dance movements. To this end, we propose the integration of Graph Convolutional Neural Networks (GCNs) in the study of dance movement recognition and prediction. GCNs are particularly adept at capturing the complex spatial relationships between human postures within a dance context. Moreover, to augment the feature extraction capabilities of our model and to enhance focus on pivotal pose changes, we introduce a multi-granularity hierarchical adaptive attention mechanism. This mechanism is designed to dynamically adjust the focal points of attention at various layers within the network. It is foreseeable to capture the subtle details and nuances of key movements and transitions more accurately, thereby optimizing the process of dance teaching.

*Corresponding Author.

II. Literature Review

This article will collect existing work in the field of human pose recognition to highlight the shortcomings of current research.

A. Traditional Human Pose Recognition Methods

Over the past few years, methodologies based on pose recognition have been applied to various fields, such as sports and dance training. Early human pose recognition algorithms primarily utilized manually designed features to reflect the spatial and temporal information of human movements, which were then processed by advanced machine learning algorithms to obtain the corresponding recognition results. GIST and HOG (Histogram of Oriented Gradients) have been widely adopted as image feature descriptors. Kuehne et al. [5] examined the effectiveness of different feature extraction methods on various datasets. The study highlights that GIST features, which capture the background context, perform slightly better (60.0%) compared to HOG features (58.6%). Furthermore, it is believed to incorporate motion features in pose recognition. Techniques such as Motion Boundary Histogram, Histograms of Optical Flow, and dense trajectories were developed to capture motion information. Fan et al. [6] proposed an innovative method for improving the accuracy of human action recognition systems. Their approach integrates a dense sampling strategy that concentrates on motion boundaries with histograms of motion gradients to optimize the feature extraction process.

Additionally, the Scale-Invariant Feature Transform (SIFT) [7] is commonly utilized alongside the Histogram of Oriented Gradients (HOG). Zhang et al. [8] proposed a novel method to leverage the SIFT flow, which effectively captures the displacement of key points between video frames. This approach involves tracking key points that are invariant to scale changes across video frames, describing these key points with local appearance and motion descriptors like Histograms of Oriented Gradients (HOG).

Despite the progress made by the aforementioned methods, algorithms based on hand-crafted features and machine learning exhibit several notable limitations. These methods generally demonstrate poor generalizability and struggle to adapt to the diversity of movements in dance training. Additionally, the process of manually designing feature extraction is complex and time-consuming, requiring domain-specific expertise and often failing to capture critical information of dance movements comprehensively. These issues limit the effectiveness of traditional methods in meeting the high accuracy and generalizability requirements of dance training, making intelligent and adaptive human pose recognition technology still a challenge.

B. Deep Learning-based Human Pose Recognition Methods

Currently, deep learning-based methods dominate the field of human pose recognition. Compared to traditional approaches, deep learning methods utilize massive datasets to gather more accurate and comprehensive information about the subjects, thereby enhancing recognition accuracy and bolstering robustness against environmental variability. Additionally, these approaches offer better scalability and end-to-end inference capabilities, allowing for more comprehensive dance training systems by integrating with other advanced techniques.

Ng et al. [9] proposed to obtain the spatial representation of the human pose at each frame by 2D CNN and fused by a multi-layer Long-short Term Memory network (LSTM). Additionally, J. Donahue et al. [10] introduced a model that employs a two-layer LSTM, known as Long-term Recurrent Convolutional Networks (LRCN). Furthermore, Li et al. [11] proposed a human pose recognition method based on translation-scale invariant image mapping and multi-scale deep Convolutional Neural Networks (CNN). The joint positions were mapped to image space, and the multi-scale CNN was utilized to extract features and achieve recognition results.

Compared to RGB-based methods, skeleton-based models are widely adopted in human pose recognition. These models represent human skeletons as structured graphs, which are then processed by Graph Neural Networks (GNN). S. Yan et al. [12] introduced a Spatial-Temporal GCN (ST-GCN) for action recognition, which operates similarly to 3D convolutional networks but processes skeleton graphs, achieving an accuracy of 30.7% on the Kinetics-400 dataset. Meanwhile, Y. Song et al. [13] enhanced GCNs by incorporating a suite of advanced techniques, including batch normalization [14]. This led to the development of an Efficient-GCN, which not only delivers competitive performance on pose recognition but also requires less training time and offers greater explainability.

In recent years, the advancement of attention-based approaches in Natural Language Processing (NLP) has driven their integration with advanced techniques in the field of computer vision. G. Bertasius et al. [15] explored various configurations of spatial and temporal self-attention mechanisms and developed the TimeSformer. AGCN [16] integrates an attention mechanism into the Graph Convolutional Network (GCN) framework. It utilizes three forms of attention: spatial, temporal, and channel attention, which collectively enable AGCN to achieve superior accuracy scores. Similarly, C. Si et al. [17] introduced the Attention Enhanced Graph Convolutional LSTM Network (AGC-LSTM). In this model, temporal dynamics are handled by an LSTM, while spatial relationships are managed through a GCN augmented with attention mechanisms.

Overall, the application of deep learning in human pose recognition is advancing swiftly, with various deep learning models constantly expanding the capabilities of pose recognition technology. Despite encountering several obstacles, these models have proven highly effective in precise detection, analysis, and interpretation of complex human motions. With ongoing advancements in research and technology, we anticipate that future systems for human pose recognition will evolve to be more advanced, offering more accurate and varied pose analysis capabilities.

C. Research Gaps

Despite significant progress in the field of deep learning for human pose recognition and prediction, its application in dance training still has notable deficiencies. Key issues, such as the precision in capturing subtle movements, handling the correlation of long-distance movements, and the diversity of

dance movements, have not been fully addressed. Therefore, future research should focus on the following areas:

1) Limitations of CNN models in dance pose recognition: The CNN-based methods for dance movement recognition have achieved certain results, but the complex spatial relationships between dance movements make it difficult for CNN methods that rely on RGB image inputs to effectively capture the dynamic relationships of dance movements. Additionally, CNN-based approaches heavily depend on the volume of training data, leading to weaker generalization capabilities when dealing with unfamiliar dance styles or new movements.

2) Lack of kinematic and multi-scale infomation in dance pose recognition: Currently, there is no unified standard for preprocessing dance data, leading to various studies employing their own methodologies. For instance, some methods focus on simple pose estimation while others may integrate complex motion analysis, but often these approaches do not fully capture the kinematic details such as the fluidity of motion and multi-scale movements. Even in advanced models that attempt to incorporate both static poses and dynamic sequences, there are challenges in accurately synchronizing small-scale movements with larger body movements during analysis. This absence of a standardized approach to capturing and analyzing kinematic and multi-scale information limits the effectiveness and interoperability of different dance pose recognition methods.

3) Limitations of deep learning models in dance pose prediction: While deep learning models like Graph Convolutional Networks (GCNs) have shown promise in understanding complex spatial relationships in dance movements, they face significant challenges in long-term motion prediction. Current GCN models excel in capturing the instantaneous relationships between body joints but struggle with generating or predicting extended sequences of movements. Due to the fundamental differences between static spatial data representation and the temporal dynamics of dance, GCNs cannot transition into predicting long-term dance sequences without modifications or integration with other temporal-focused models.

In summary, future research should focus on developing new models and techniques to address these challenges in dance pose recognition and prediction, to not only enhance the effectiveness of dance training but also improve the acceptance of participants.

III. Research on Dance Pose Recognition and Prediction Methods Based on Graph Convolutional Networks

In this section, a novel framework based on GCN is proposed for dance pose recognition. The preliminary knowledge of spatial-temporal GCN (ST-GCN) is first introduced, followed by the proposed multi-kinematic feature fusion-based spatial-temporal GCN (MKFF-ST-GCN). Finally, the spatial-tempotal graph attention-based network (STGA-Net) is designed to predict dance poses.

A. Spatial-Temporal Graph Convolutional Network

In the contemporary field of dance pose recognition, deep learning technologies such as Convolutional Neural Network (CNN) and Graph Convolutional Network (GCN) have begun to be explored for understanding human movements. With their advanced data processing capabilities, they exhibit notable representative potential.

Graph Convolutional Network (GCN) is a specialized type of neural network designed to operate directly on graphs instead of grid data such as images. Compared to traditional CNN, the convolutional operation in GCN is adapted to aggregate information from neighbors of nodes which can capture the spatial relationships within the graph. Additionally, the GCN models can leverage the node features and graph topology to generate powerful node embeddings that can be used for a variety of tasks, such as node classification, graph classification, and link prediction.

Considering the graph CNN model within one single frame, the N joint nodes can be expressed as V_t and the skeleton edges can be expressed as $E_S(\tau) = \{v_i^t v_j^t \mid t = \tau, (i,j) \in H\}$, where H represents the set of naturally connected human body joints. Assuming that the kernel size of the convolutional operator is K×K, and the number of channels of the input feature map f_{in} is c. The output feature map for a single channel can be expressed as

$$f_{out}(x) = \sum_{h=1}^{K}\sum_{w=1}^{K} f_{in}(\boldsymbol{p}(x,h,w)) \cdot \boldsymbol{w}(h,w) \tag{1}$$

where $\boldsymbol{p}: Z^2 \times Z^2 \to Z^2$ represents the sampling function that enumerates the neighbors of location x, and $\boldsymbol{w}: Z^2 \to \square^c$ denotes the weight function that provides the weight vector for computing the inner product with the sampled input features. Furthermore, the sampling function on the neighbor set of a node v_i^t can be defined as $B(v_i^t) = \{v_j^t \mid d(v_j^t, v_i^t) \le D\}$ and the weight function can be simplified by partitioning the neighbor set $B(v_i^t)$ of a joint node v_i^t into a fixed number of K subsets, where each subset has a numerical label. Then, the spatial graph convolution can be rewritten as

$$f_{out}(v_i^t) = \sum_{v_i^t \in B(v_i^t)} \frac{1}{Z_i^t(v_j^t)} f_{in}(\boldsymbol{p}(v_i^t, v_j^t)) \cdot \boldsymbol{w}(v_i^t, v_j^t) \tag{2}$$

In addition, the same joints across consecutive frames are connected to integrate temporal information, that can be expressed as

$$B(v_i^t) = \{v_j^q \mid d(v_j^t, v_i^t) \le K, |q-t| \le |\Gamma/2|\} \tag{3}$$

The parameter Γ represents the temporal kernel size. Since the temporal axis is well-ordered, the label function l_{ST} can be expressed as

$$l_S^T(v_j^q) = l_i^t(v_j^t) + (q - t + \lfloor \Gamma/2 \rfloor \times K) \tag{4}$$

Based on the above ST-GCN, this study introduced a novel multi-kinematic feature fusion (MKFF) module to comprehensively capture the movement information, thus addressing jitter and occlusion.

B. MKFF-ST-GCN for Dance Pose Recognition

In this section, we introduce the MKFF-ST-GCN (Multi-Kinematic Features Fusion-based Spatial-Temporal Graph Convolution Network), which is specially designed for dance pose recognition. Its network structure has been adjusted to suit the characteristics of dance movement data, as shown in Fig. 1.

Considering the jitter and occlusion during dance movements, relying on the structural information of human body keypoints cannot fulfill these requirements. As a result, this paper introduced MKFF module to leverage the keypoint kinematic features within a sliding window, obtaining from previous, current, and next frames' poses. Traditional ST-GCN capture keypoint information through complex network structure, which may not be sufficient to obtain the dynamic patterns in dance movements. By integrating the multi-kinematic features fusion module, the proposed method can effectively calculate the inherent kinematic features without increasing network complexity and extra parameters.

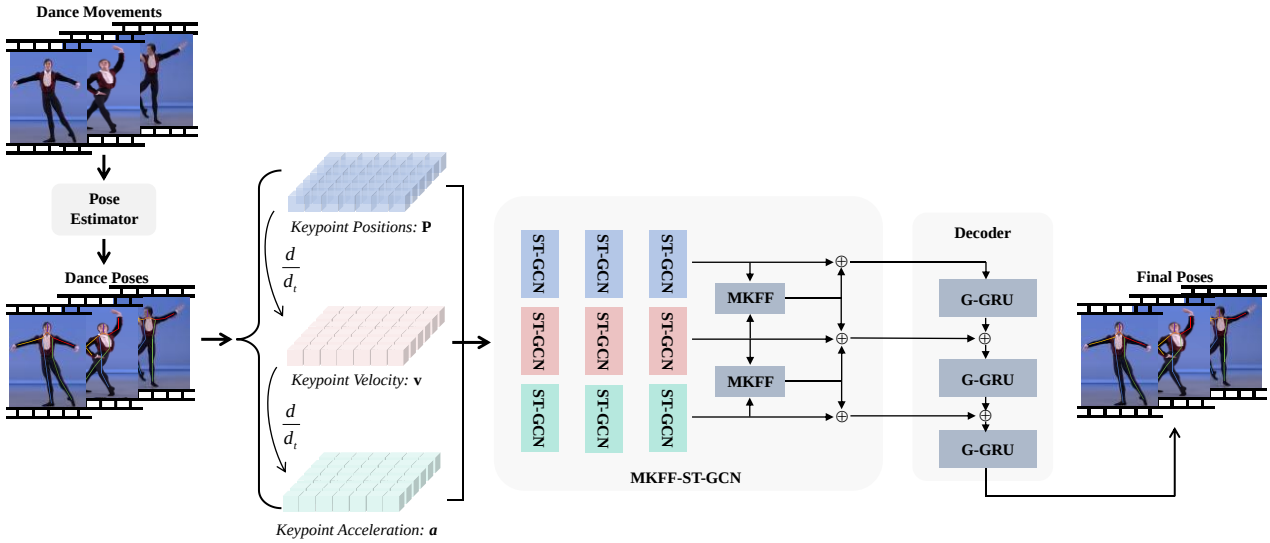


Fig. 1. Network structure diagram based on MKFF-ST-GCN dance pose recognition.

The input dance poses are firstly computed by off-the-shelf pose estimatore, denotaed as $\boldsymbol{P} \in \mathbb{R}^{T/N\times(K\cdot D)}$. By leveraging the consecutive poses's coordinates at each frame, the keypoint flow can be expressed as

$$\bar{\boldsymbol{P}}_t = (\boldsymbol{P}_t + \boldsymbol{P}_{t+d_t} + \boldsymbol{P}_{t-d_t}) / 3 \tag{5}$$

where d_t denotes an interval from the previous and next poses to the current poses. Based on the above keypoint flow, the keypoint velocity and acceleration can be expressed as

$$\begin{aligned} v_t &= (\boldsymbol{P}_t - \boldsymbol{P}_{t-d_t}) / d_t \\ a &= (v_t - v_{t-d_t}) / d_t \end{aligned} \tag{6}$$

Dance movements often have rich spatio-temporal relationships and subtle dynamic changes, making it difficult for a single feature extraction structure to fully capture these characteristics. To address this issue, we propose a multi-kinematic feature fusion (MKFF) module based on a multi-head attention mechanism. This module is designed to facilitate information exchange among different motion features, enhance the perception of complex dance movement structures and details, and improve the capture of dance motion features. By conducting in-depth analysis and reasonable fusion of dynamic features of keypoints, our method effectively strengthens the relationships between key frames, accurately identifies various types of dance movements, and provides strong technical support for dance training and performance.

As shown in Fig. 2, the MKFF module consists of three feature fusion block followed by an activate function and dropout layer, each block includes several critical parts, such as Layer normalization, multi-head self-attention, and MLP block.

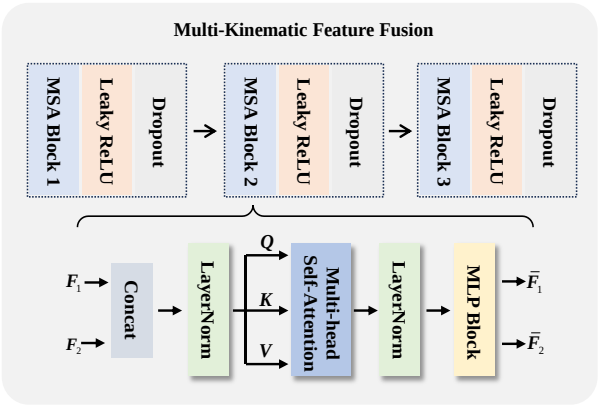


Fig. 2. Schematic diagram of the multi-kinematic feature structure fusion module.

First, the input multi-kinematic features are concatenated in keypoint dimernsion to obtain the initial embedding features, which can be expressed as

$$Z_0 = \text{concat}(\boldsymbol{F}_1; \boldsymbol{F}_2) W_o \tag{7}$$

where $W_0 \in \mathbb{R}^{(K \cdot 2D) \times C}$ is a linear projection matrix. Then, the initial embedding features are mapped to Q_0, K_0, V_0 representing queries, keys, and values, which can be expressed as

$$\begin{aligned} Q_t &= (Z_{t-1}) W_t + E_{pos,t} \\ K_t &= (Z_{t-1}) W_t + E_{pos,t} \\ V_t &= (Z_{t-1}) W_t + E_{pos,t} \end{aligned} \tag{8}$$

where $E_{pos,t}$ represents the positional embedding at t-th block. The multi-head self-attention is adopted to automatically calculate value map thus capturing the distribution of large and small dance movements. Besides, a layer normalization module is added before every multi-head self-attention and multi-layer perceptron (MLP) block. The entire process can be expressed as

$$\begin{aligned} \bar{Z}_t &= \text{MSA}(\text{LN}(Z_t)) + Z_t \\ \hat{Z}_t &= \text{MLP}(\text{LN}(\bar{Z}_t)) + \bar{Z}_t \\ MSA(Q_t, K_t, V_t) &= \text{soft}\max(\frac{Q_t K_t^T}{\sqrt{d_k}}) V_t \end{aligned} \tag{9}$$

The fused multi-kinematic features are then proceeded by the decoder module to obtain the final dance poses. The proposed model is trained by a weighted loss with objective of minimizing the weighted L_1 norm between prediction and ground truth joint positions, which is defined as

$$L_w = \frac{1}{N_j} \sum_{j=1}^{N_j} v_j \left\| G_j - P_j \right\| + \frac{\lambda}{N_k} \sum_{k=1}^{N_k} v_k \left\| G_k - P_k \right\| \tag{10}$$

where N_k, G_j, P_j, v_j represents the number of top-k keypoints, ground truth, predicition, and visibility of joint j, respectivley. The initial term of the loss function targets errors across all keypoints, whereas the subsequent term specifically addresses errors associated with the top-k keypoints. With this carefully designed network structure, MKFF-ST-GCN can recognize dance poses with high accuracy and precision, providing a powerful tool for automated dance training.

C. STGA-Net for Dance Pose Prediction

In dance training, the smoothness and continuity of movements greatly affect the quality of the dance. Therefore, recognizing the current dance posture and predicting the next frame of dance movement is crucial. Effective prediction of future dance movements can help dancers understand the transitions between movements, accelerate the learning process, and correct potential errors or dangerous movements, significantly impacting the efficiency and effectiveness of dance teaching and training. To address these issues, this study introduced a novel spatial-temporal graph attention mechanism (STGA-Net) to predict the dance poses based on the above recognition result.

As shown in Fig. 3, the proposed graph attention block includes the global graph attention layer and the local graph attention layer. The local spatial attention module is designed to model the hierarchical and symmetrical structure of dance poses providing fine-grained dance movements. The global spatial attention module is introduced to adaptively extract global semantic information to better understand the spatial characteristic and the relationship between consecutive dance movements. Furthermore, the proposed local and global spatial graph attention module holds significant advantages in dance movement prediction. Traditional GCN models have a limited receptive field, making it difficult to capture subtle interactions between keypoints. Dance training often involves rapid and minor changes in movement patterns, which poses challenges for traditional GCNs. The proposed local and global spatial graph attention mechanism optimizes the flow of information, enabling adaptation to a variety of dance styles and enhancing the accuracy and robustness of predictions. This is crucial for understanding complex dance sequences and predicting dancers' movement trajectories in advance, thereby significantly improving the quality of dance training and performance.

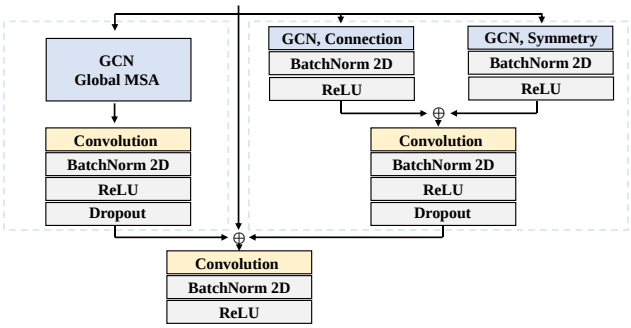


Fig. 3. Schematic diagram of graph attention module.

The skeleton 2D poses are first defined as a graph $\Omega = \{V, E\}$, where V represents the set of nodes and E represents edges. The node features are defined as $\boldsymbol{X} = \{\vec{x}_1, \vec{x}_2, ..., \vec{x}_N \mid \vec{x}_i \in \mathbb{R}^C\}$ which has C channels. Then, the output features of graph convolutional layer can be expressed as

$$\boldsymbol{X}^{l+1} = \text{soft}\max(\boldsymbol{M} \otimes \tilde{\boldsymbol{A}}) \boldsymbol{X}^l W \tag{11}$$

where W is a learnable matrix for channel transformation, M represents a learnable mask matrix, $\tilde{\boldsymbol{A}}$ represents the connections between joints. By designing $\tilde{\boldsymbol{A}}$, this study introduces two different kinds of spatial graph attention, a symmetric matrix to encode the symmetrical counterpart and am adjacency matrix to encode the connections for distal joints. Besides, the global attention mechanism aims to encode the relationship across disconnected joints, thus addressing depth ambiguities and occlusions. The global attention mechanism can be expressed as

$$\boldsymbol{X}^{l+1} = concat(\boldsymbol{B}_k + \boldsymbol{C}_k) \boldsymbol{X}^l W_k \tag{12}$$

where $\boldsymbol{B}_k$ represents an adaptive global adjacency matrix, $\boldsymbol{C}_k$ signifies a learnable global adjacency matrix, and W_k is a transformed matrix. In addition, the temporal information is

captured by the temporal dilated convolutional block. To save the spatial information across dance poses, the original convolutional block is replaced by 2D convolutions with $k \times 1$ kernel size. After encoding the keypoints features of dance poses, the latent features are processed by a G-GRU decoder to predict future dance poses, as shown in Fig. 4.

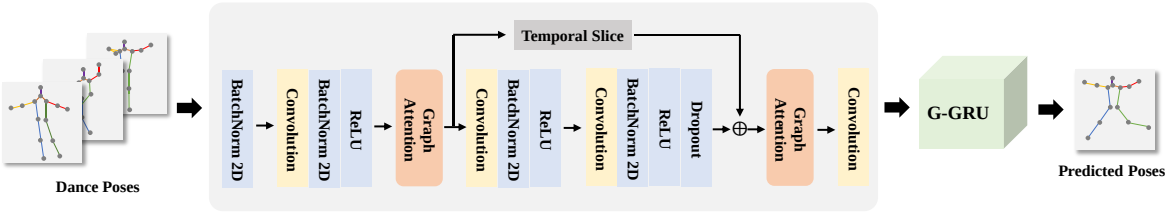


Fig. 4. Schematic diagram of STGA-Net for dance pose predictio.

IV. Case Verification

This section will validate the effectiveness of the proposed method based on a self-made experimental dataset.

A. Experimental Environment

The hardware environment for the experiments in this chapter is shown in Table I:

TABLE I. Experimental Software and Hardware Environment Table

CPU	Intel(R) Core(TM) i5-13400F
GPU	NVIDIA GeForce RTX 4070
Memory	12.0 Gb
Operating System	Ubuntu 18.04
CUDA	11.1
Main Frameworks	Pytorch1.10.0
Main Programming Language	Python 3.8

To explore dance pose recognition and prediction, this study has collected dancing videos including ballet, street dance, and modern dance. The collected datasets dataset contains 30 dance video clips, totalling about 150,000 frames. From these, approximately 8,000 images have been selected for dance motion recognition and prediction. Each dance movement image is manually annotated with 14 key points, including the head, shoulders, elbows, wrists, etc., to support precise capture and analysis of dance movements, and the annotation information is stored in JSON format. Additionally, data augmentation techniques have been applied to the constructed dance motion dataset, including random rotation, scaling, flipping, adding random noise, and Gaussian blur, to improve the model's generalization ability. The dataset is divided into 70% training set, 15% validation set, and 15% test set.

The experimental section designed several different comparisons to comprehensively evaluate the performance of the proposed model in dance movement recognition and prediction. It compares with two widely used dance movement recognition models, HBRNN [18] and HRNet [19]. Specific aspects include: dance movement recognition results under different dance styles, accuracy of dance movement keypoint recognition, prediction results of different dance styles, and ablation experiments on the proposed graph attention mechanism. The model parameters used for validation were pre-trained on the Kinetics dataset and fine-tuned on the dataset constructed in this paper. During the training process, the stochastic gradient descent algorithm was used, with an initial learning rate of 0.01, a linear learning rate decay strategy with a decay weight of 0.95, a batch size set at 16, and a total of 80 training epochs.

B. Experimental Results

To accurately and effectively recognize dance movements, this study constructed the MKFF-ST-GCN model based on the experimental setup mentioned above. The model training was conducted in a supervised manner, and the error variation curve during the training process is shown in Fig. 5. It can be observed that after pre-training, the network model achieved good results on the dataset used in this paper, with the error curve quickly decreasing and gradually stabilizing. Furthermore, we presented the qualitative recognition results of dance movements in various styles such as contemporary, ballet, and street dance, as shown in Fig. 6. It is evident that the proposed MKFF-ST-GCN demonstrated the highest recognition accuracy and the ability to capture complex postures in all dance styles. Benefiting from the MKFF module, the proposed model was able to accurately locate key points, showing significant improvements over HBRNN, and performed better in handling complex movements, significantly outperforming the other two algorithms in the dance movement recognition task.

In addition, to further demonstrate the effectiveness of the proposed method in recognizing dance movements, ballet dance data is used as an example, employing the Percentage of Correct Keypoints as the evaluation metric, with results shown in Tables II and III. Specifically, Table II displays the recognition accuracy for keypoints such as the head, shoulders, wrists, and knees under a threshold of 0.2, where MKFF-ST-GCN performs the best across all keypoints. Table III further tests more stringent thresholds (0.1 and 0.05), and the results show that MKFF-ST-GCN continues to lead under high precision requirements, followed by HRNet, while HBRNN performs relatively poorly. These findings indicate that MKFF-ST-GCN has significant advantages in handling complex dance movements and high-precision keypoint localization, making it suitable for high-demand dance movement recognition tasks.

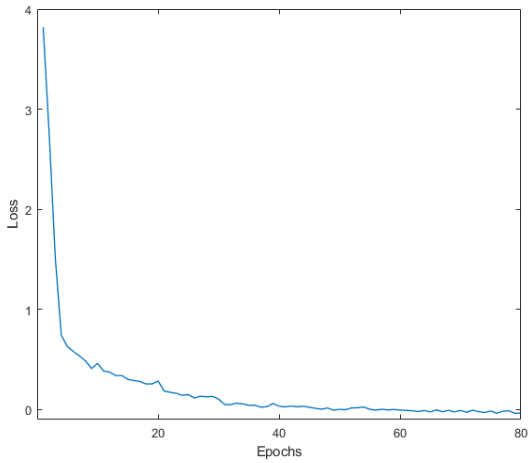


Fig. 5. Model loss rate curve.

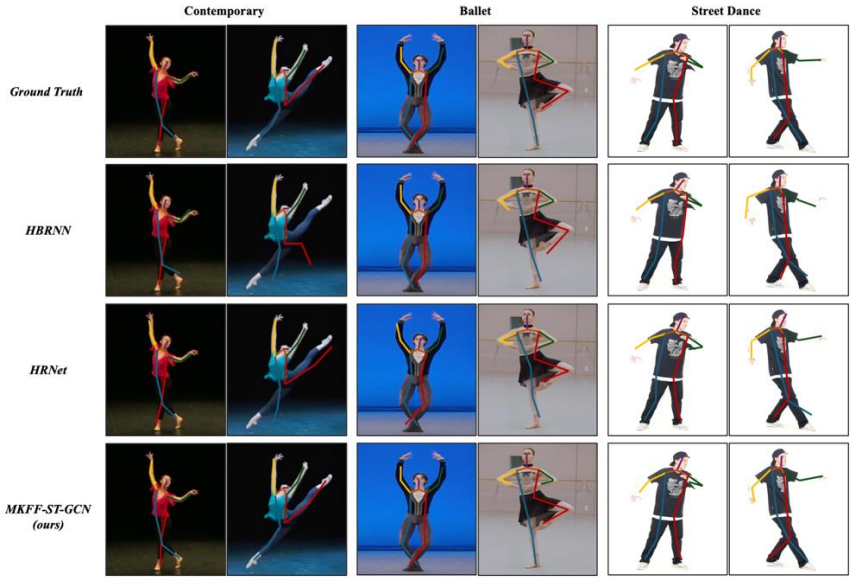


Fig. 6. Qualitative comparison of dance pose recognition among various dance styles.

TABLE II. QUANTITATIVE COMPARISON OF DANCE POSE RECOGNITION AMONG DIFFERENT KEYPOINTS

Method	PCK@0.2			
	Head	Shoulder	Wrist	Knee
HBRNN	86.4	82.9	83.3	88.1
HRNet	90.2	88.3	87.5	91.1
MKFF-ST-GCN	**93.5**	**90.1**	**90.7**	**92.5**

TABLE III. QUANTITATIVE COMPARISON OF DANCE POSE RECOGNITION UNDER DIFFERENT THRESHOLDS

Method	PCK@0.1	PCK@0.05
	Avg.	Avg.
HBRNN	81.2	79.6
HRNet	84.3	81.1
MKFF-ST-GCN	**86.2**	**83.9**

The aforementioned results adequately demonstrate that the proposed MKFF-ST-GCN can effectively extract latent features of dance movements, capture subtle motion changes, and accurately identify dance movements of different styles. Furthermore, this section verifies the performance of the proposed STGA-Net in predicting dance movements with two widely adopted methods, DMGNN [20] and T-GCN [21]. Extensive experiments were conducted for different dance styles and various time intervals, with the results presented in the Tables IV to VI.

TABLE IV. QUANTITATIVE COMPARISON OF BALLET POSE PREDICTION

Method	Ballet		
	80	160	320
DMGNN	0.88	1.10	1.39
T-GCN	0.45	0.62	**0.88**
STGA-Net	**0.24**	**0.40**	0.90

TABLE V. QUANTITATIVE COMPARISON OF CONTEMPORARY DANCE POSE PREDICTION

Method	Contemporary Dance		
	80	160	320
DMGNN	0.31	0.67	0.90
T-GCN	0.39	0.44	0.81
STGA-Net	**0.19**	**0.36**	**0.58**

TABLE VI. QUANTITATIVE COMPARISON OF STREET DANCE POSE PREDICTION

Method	Street Dance		
	80	160	320
DMGNN	0.39	0.80	1.32
T-GCN	0.41	0.76	1.09
STGA-Net	**0.22**	**0.60**	**0.92**

The above results showcase the performance of STGA-Net in predicting dance movements in ballet, contemporary dance, and street dance, using the Mean Absolute Error (MAE) between predicted and actual keypoints as the evaluation metric. The results indicate that across all tested dance styles and prediction intervals (80ms, 160ms, 320ms), STGA-Net consistently demonstrated superior accuracy compared to the other two methods (DMGNN and T-GCN). Specifically, for ballet, STGA-Net achieved MAE values of 0.24, 0.40, and 0.90 at prediction intervals of 80ms, 160ms, and 320ms respectively, significantly outperforming DMGNN and T-GCN. In contemporary dance, STGA-Net also performed the best across all intervals, with MAE values of 0.19, 0.36, and 0.58 respectively. In the prediction of street dance, STGA-Net maintained its lead with MAE values of 0.22, 0.60, and 0.92 at the 80ms, 160ms, and 320ms intervals, respectively. These results demonstrate that STGA-Net can effectively reduce prediction errors in dance movement prediction tasks across different dance styles and time intervals, showcasing its potential and practicality in the field of motion prediction.

V. CONCLUSION

This study aims to optimize the dance training process using artificial intelligence technology, implementing dance movement recognition and prediction based on deep learning algorithms, which greatly promotes the application of AI technology in the field of dance training. To extract the dynamic and complex features contained in dance movements, a MKFF-GCN model based on the multi-kinematic features fusion has been developed. This model supplements the dynamic information missing in structured data by calculating the kinematic information of keypoints and effectively fuses these features using a multi-head attention mechanism. Additionally, to accurately predict dance poses and avoid accidents during training, this paper proposes a STGA-Net based on the spatial-temporal graph attention mechanism, which can effectively model the local and global relationships between dance movement keypoints, enhancing the capability to extract long-distance information. Experiments show that compared to previous algorithms, the proposed models can effectively recognize and predict dance poses, significantly improving the efficiency of dance training. It must be acknowledged that there is still considerable room for improvement in the real-time recognition performance of the proposed models. In the future, we will continue to optimize the model structure to enhance real-time recognition capabilities, further aiding the dance training process.

REFERENCES

[1] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," Nature, vol. 521, no. 7553, pp. 436–444, May 2015, doi: 10.1038/nature14539.

[2] P. Zhou, B. Gao, S. Wang, and T. Chai, "Identification of Abnormal Conditions for Fused Magnesium Melting Process Based on Deep Learning and Multisource Information Fusion," IEEE Trans. Ind. Electron., vol. 69, no. 3, pp. 3017–3026, Mar. 2022.

[3] G. Lan, Y. Wu, F. Hu, and Q. Hao, "Vision-Based Human Pose Estimation via Deep Learning: A Survey," IEEE Trans. Human-Mach. Syst., vol. 53, no. 1, pp. 253–268, Feb. 2023.

[4] A. Bera, M. Nasipuri, O. Krejcar, and D. Bhattacharjee, "Fine-Grained Sports, Yoga, and Dance Postures Recognition: A Benchmark Analysis," IEEE Trans. Instrum. Meas., vol. 72, pp. 1–13, 2023.

[5] H. Kuehne, H. Jhuang, E. Garrote, T. Poggio, and T. Serre, "HMDB: A large video database for human motion recognition," in 2011 International Conference on Computer Vision, Barcelona, Spain: IEEE, Nov. 2011, pp. 2556–2563.

[6] M. Fan, Q. Han, X. Zhang, Y. Liu, H. Chen, and Y. Hu, "Human Action Recognition Based on Dense Sampling of Motion Boundary and Histogram of Motion Gradient," in 2018 IEEE 7th Data Driven Control and Learning Systems Conference (DDCLS), May 2018, pp. 1033–1038.

[7] Z. Wang, Z. Wang, H. Liu, and Z. Huo, "Scale - invariant feature matching based on pairs of feature points," IET Computer Vision, vol. 9, no. 6, pp. 789‐796, Dec. 2015.

[8] J.-T. Zhang, A.-C. Tsoi, and S.-L. Lo, "Scale Invariant Feature Transform Flow trajectory approach with applications to human action recognition," in 2014 International Joint Conference on Neural Networks (IJCNN), Beijing, China: IEEE, Jul. 2014, pp. 1197–1204.

[9] Joe Yue-Hei Ng, M. Hausknecht, S. Vijayanarasimhan, O. Vinyals, R. Monga, and G. Toderici, "Beyond short snippets: Deep networks for video classification," in 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA: IEEE, Jun. 2015, pp. 4694–4702.

[10] J. Donahue et al., "Long-term recurrent convolutional networks for visual recognition and description," in 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA: IEEE, Jun. 2015, pp. 2625–263.

[11] Bo Li, Yuchao Dai, Xuelian Cheng, Huahui Chen, Yi Lin, and Mingyi He, "Skeleton based action recognition using translation-scale invariant image mapping and multi-scale deep CNN," in 2017 IEEE International Conference on Multimedia & Expo Workshops (ICMEW), Hong Kong, Hong Kong: IEEE, Jul. 2017, pp. 601–604.

[12] S. Yan, Y. Xiong, and D. Lin, "Spatial Temporal Graph Convolutional Networks for Skeleton-Based Action Recognition," AAAI, vol. 32, no. 1, Apr. 2018, doi: 10.1609/aaai.v32i1.12328.

[13] Y.-F. Song, Z. Zhang, C. Shan, and L. Wang, "Stronger, Faster and More Explainable: A Graph Convolutional Baseline for Skeleton-based Action Recognition," in Proceedings of the 28th ACM International Conference on Multimedia, Oct. 2020, pp. 1625–1633.

[14] S. Ioffe and C. Szegedy, "Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift," in International conference on machine learning, 2015, pp:448-456.

[15] G. Bertasius, H. Wang, and L. Torresani, "Is Space-Time Attention All You Need for Video Understanding," in International conference on machine learning, 2021.

[16] L. Shi, Y. Zhang, J. Cheng, and H. Lu, "Skeleton-Based Action Recognition With Multi-Stream Adaptive Graph Convolutional Networks," IEEE Trans. on Image Process., vol. 29, pp. 9532–9545, 2020.

[17] C. Si, W. Chen, W. Wang, L. Wang, and T. Tan, "An Attention Enhanced Graph Convolutional LSTM Network for Skeleton-Based Action

Recognition," in 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Jun. 2019.

[18] Yong Du, W. Wang, and L. Wang, "Hierarchical recurrent neural network for skeleton based action recognition," in 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA: IEEE, Jun. 2015, pp. 1110–1118.

[19] J. Wang et al., "Deep High-Resolution Representation Learning for Visual Recognition," IEEE Trans. Pattern Anal. Mach. Intell., vol. 43, no. 10, pp. 3349–3364, Oct. 2021.

[20] M. Li, S. Chen, Y. Zhao, Y. Zhang, Y. Wang, and Q. Tian, "Dynamic Multiscale Graph Neural Networks for 3D Skeleton Based Human Motion Prediction," in 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Jun. 2020.

[21] W. Mao, M. Liu, M. Salzmann, and H. Li, "Learning Trajectory Dependencies for Human Motion Prediction," in 2019 IEEE/CVF International Conference on Computer Vision (ICCV), Seoul, Korea (South): IEEE, Oct. 2019, pp. 9488–9496.

Improved Decision Support System for Alzheimer's Diagnosis Using a Hybrid Machine Learning Approach with Structural MRI Brain Scans

Niranjan Kumar Parvatham, Lakshmana Phaneendra Maguluri*
Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Andhra Pradesh, India

***Abstract*—Alzheimer's disease (AD) causes damage to brain cells and their activities. This disease is typically caused by ageing, making people over the age of 65 more susceptible. As the disease progresses, it slowly destroys brain cells, making it harder to think clearly, recall things, and do everyday tasks. The end result of this is dementia. Metabolic disorders, such as diabetes and Alzheimer's disease, affect a substantial proportion of the world's population. While there is no permanent cure for AD, early diagnosis can help reduce damage to brain cells and support a faster recovery. Recent research has explored various machine learning approaches for early disease detection. However, traditional ML (Machine Learning) methods and deep learning techniques such as CNNs have not been individually effective in accurately detecting Alzheimer's disease (AD). In this proposed work, we developed a hybrid model that processes sMRI brain images to detect them as demented or non-demented. The model consists of two parts: the first part involves extracting significant features through a sequence of convolution and pooling operations; the second part uses these features to train SVM for binary classification. Data augmentation techniques such as horizontal flipping are used to balance dataset. We calculated key performance metrics essential for the healthcare domain, including sensitivity, specificity, accuracy, and F1-score. Notably, our model achieved an impressive accuracy of 99.60% in detecting AD, with a sensitivity of 99.83%, a specificity of 99.40%, and an F1-score of 99.58%. These results were validated using 15-fold cross-validation, enhancing the model's robustness for new data. This approach yields a more robust model, offering greater accuracy and precision compared to existing methods. This model can effectively support manual systems for detecting AD with greater accuracy.**

Keywords*—*Alzheimer's disease; binary classification; Convolutional Neural Network (CNN); horizontal flipping; healthcare decision support system; MRI images; Support Vector Machine(SVM)

I. Introduction

Alzheimer's disease is an illness of the brain that damages brain-active neurons, which is most likely due to aging. The aging population is increasing, leading to an increasing number of AD patients. It exerts detrimental effects on an individual's cognitive and interpersonal skills, progressively impeding their capacity to do routine activities [1], [2]. Alzheimer's disease (AD) is marked by the shrinkage of critical brain regions, such as the cerebral cortex and hippocampus, and this shrinkage results in damage to spatial and episodic memory, and disrupts the connection between the brain and body [3]. Due to progressive loss of brain cells, this leads to the accumulation of neurofibrillary tangles and amyloid plaques and a reduction in brain volume. The disease's rapid progression can weaken short-term memory, planning, and judgment [4].

The World Health Organization reports a significant increase in dementia cases, with the likelihood increasing with age, particularly among those over 65. Around 55 million people worldwide suffered from dementia in 2021, and experts predict that this number will increase to 139 million by 2050. On the other hand, only about 4% of young people experience early-onset dementia, which is generally caused by other health problems [5]. Alzheimer's disease (AD) affects roughly 5.7 million individuals in the United States. It has a substantial increase, multiplying by three in the mid-21st century. Alzheimer's disease was the sixth most common cause of death in the United States in 2015. An estimated 7.4% of India's population aged 60 and older suffers from dementia. This statistic indicates that approximately 8.8 million elderly people in the country live with this condition. The high prevalence of dementia among India's aging population highlights the growing need for increased awareness, support, and resources to address the challenges faced by those affected and their families [6].

The patient may exhibit symptoms of AD in relation to the loss of neurons in various parts of the brain. These symptoms can be detected in the clinical setting. In the mild stage, symptoms begin to disrupt some routine activities. In the moderate stage, symptoms interfere significantly with many daily activities. In the severe stage, symptoms profoundly impact almost all routine tasks. Currently, there is no permanent cure for Alzheimer's disease. However, if diagnosed at an early stage, effective medication can slow the progression of the disease. Early diagnosis allows for interventions that can significantly reduce brain damage, lower mortality rates, and enhance the quality of life for those affected.

AD Early diagnosis requires a comprehensive evaluation by a medical specialist, including both neurological and physical assessments [7], [8]. Treatment is more effective when it is given during the early stages of Alzheimer's disease [9]. Structural MRI, functional MRI, PET scans and other imaging modalities reveal significant changes in the brain associated with memory loss [10], [11]. Magnetic resonance imaging (MRI) [12] can be used as a biomarker to observe the size of damaged brain tissues [13]. To achieve higher levels of accuracy in predicting AD without false alarms, medical decision-support systems must incorporate an automated machine learning method that can

*Corresponding Author.

analyse MRI images at a deeper level in addition to human evaluations.

II. Related Work

In [14] a novel convolutional neural network (CNN) architecture that uses MRI data to accurately detect and classify Alzheimer's disease. The model achieves high accuracies of 99.43%, 99.57%, and 99.13% for categorizing AD across three, four, and five categories. The CNN architecture employs a hierarchical structure of convolutional, pooling, and fully connected layers to extract local and global patterns, enhancing clinical accuracy for early detection and disease monitoring. However, the study fails to address the applicability of the proposed CNN architecture to bigger and more varied datasets, and the focus on AD classification across three, four, and five categories may not cover the full spectrum of disease severity and subtypes. Further validation studies on external datasets are needed to evaluate the model's robustness and effectiveness.

The research [15] presented the use of DenseNet architecture for Alzheimer's disease classification, highlighting its effectiveness on MRI datasets. The model uses transfer learning techniques to improve accuracy and efficiency. Data augmentation is used to handle sparse data and generalize results. The model achieves an accuracy of 96.5% and an AUC of 99% in AD diagnosis. However, the study's limitations include prolonged computational time, manual hyperparameter tuning, and reliance on a single modality dataset.

The research [16] explored about the early-stage diagnosis of Alzheimer's disease (AD) in patients who are Cognitively Normal (CN) and introduces a dense neural network designed for binary classification. The findings demonstrate that this model surpasses traditional machine learning algorithms, achieving higher accuracy in distinguishing between AD and CN, AD and Mild Cognitive Impairment (MCI), as well as MCI and CN. Additionally, the study emphasizes the crucial role of computer-aided diagnosis using MRI for precise AD classification. By utilizing different activation functions, the model's classification validation accuracy is further enhanced. However, the focus on binary classification tasks may restrict the study's applicability to more complex diagnostic scenarios.

This study [17] focused on both white and gray matter from MRI scan images using 3D MRI technology. The procedure entails acquiring 2D slices from the coronal, sagittal, and axial planes of the 3D MRI scans. The process involves finding the most pertinent segments and performing feature extraction using Multi-Layer Perceptron (MLP) and Support Vector Machine (SVM) techniques to forecast and categorize Alzheimer's disease. The researchers assess the system's performance by utilizing metrics like precision, recall, accuracy, and F1-score. However, it is important to note that this study examines only MRI scan images.

This research [18] presented a novel method for categorizing MR images in the detection of Alzheimer's disease (AD) by utilizing graph kernels obtained from textural features of structural MR (sMR) images. The method entails dividing MR brain pictures into several regions and obtaining 22 unique texture characteristics. These characteristics are subsequently utilized to define the qualities of graph nodes. Graph-kernel matrices are created in sequence and then classified using Support Vector Machines (SVMs).

The findings demonstrated that this approach has outstanding performance in differentiating between individuals who are cognitively normal (CN) and those with Alzheimer's disease (AD), as well as between CN and Mild Cognitive Impairment (MCI) instances. More precisely, the technique obtained an Area under the Curve (AUC) of 0.92 for distinguishing between CN and AD, and 0.81 for distinguishing between CN and MCI. Nevertheless, its efficacy in distinguishing between MCI and AD was limited, as indicated by an AUC of 0.78.

The research [19] presented a comparative analysis of transfer learning versus conventional machine learning techniques for the early diagnosis and prognosis of Alzheimer's disease using structural brain MRI. It underscores the potential benefits of transfer learning in enhancing accuracy and efficiency, possibly reducing the reliance on large annotated datasets. The study emphasizes the significance of applying transfer learning in medical imaging for neurodegenerative diseases like Alzheimer's. The results indicate that the fusion of conventional machine learning systems outperforms ensemble transfer learning approaches, achieving an AUC of 93.1% for distinguishing Alzheimer's disease from cognitively normal (CN) individuals, 89.6% for identifying MCI converters (MCIc) from CN individuals, and an AUC range of 69.1–73.3% for differentiating MCI converters from non-converters (MCInc). However, further optimization is necessary when using networks pre-trained on generic images.

In [20], the author developed a comprehensive algorithm to predict disease using brain volume, cognitive and biological features, clustering algorithms, and Fuzzy inference systems. Deterioration refers to a significant global decline in mental function, not due to careful adjustment. This study employs machine learning algorithms to analyze data acquired from neuroimaging technology to identify Alzheimer's disease in its early stages. Support Vector Machine and Gradient Boosting are highly effective algorithms for classification problems and exhibit excellent performance in this endeavour.

In [21], the author focused on making use of machine learning techniques, specifically Support Vector Machine (SVM), to identify Alzheimer's disease. The study utilized Grey Level Co-occurrence Matrix and Haralick features to extract features from MRI axial brain slices. These methods analyze the texture of brain images to identify patterns associated with Alzheimer's. The resulting model achieved an accuracy rate of 84% in AD detection. This demonstrates the potential of combining SVM with advanced feature extraction techniques for accurate detection of AD in its early stage. The approach highlights the importance of texture analysis in neuroimaging.

In [22] author(s) have proposed a decision-support model utilizing deep learning and machine learning techniques to predict the one-year conversion probability from Mild Cognitive Impairment (MCI) to Alzheimer's disease (AD). This addresses a gap in the existing literature. The methodology involved extracting features with the help of CNN and classifying them with a support vector machine (SVM) classifier employing different kernels (linear, polynomial, and RBF). The model

achieved high classification accuracies of 91.0%, 90.0%, and 92.3% respectively, demonstrating its effectiveness. While the results were promising, some limitations need to be addressed. The study's data was limited, even after augmenting it by randomly choosing and flipping images. Increasing the dataset size is expected to improve classification accuracy. Additionally, the image data used was from the ADNI database, which consists of Western patients.

The study [23] introduced a method for detecting Alzheimer's disease by employing Support Vector Machine (SVM) on brain MRI data. The Support Vector Machine (SVM) model categorizes the disease into three stages: mild cognitive impairment (MCI), moderate Alzheimer's disease (AD), and severe Alzheimer's disease (AD).The model was trained and tested with MRI data from the Alzheimer's Disease Neuroimaging Initiative (ADNI), involving about 70 AD patients and 30 normal controls. This diverse dataset allowed for effective training and testing of the model, despite its relatively small size, which may impact generalizability. While the SVM algorithm showed promising classification accuracy, the study did not address the interpretability of the model's decisions, a key factor for understanding the features driving the classifications.

In this study [24] the authors investigated the application of Support Vector Machine (SVM) in the diagnosis of Alzheimer's disease through the analysis of brain MRI scans and the classification of its stages. The algorithm underwent training and testing using MRI data obtained from the Alzheimer's disease Neuroimaging Initiative (ADNI). The dataset included 70 patients diagnosed with Alzheimer's disease and 30 individuals without any cognitive impairments. The method involves extracting feature points from MRI images using Speeded up Robust Features (SURF) and analyzing these features with the Gray Level Co-occurrence Matrix (GLCM).The study utilized brain images along with neuropsychological assessments, physical and neurological examinations, cognitive assessments, patient medical history, and baseline diagnosis and symptoms.

In [25] the authors suggested a technique to enhance feature detection models for Alzheimer's disease (AD) classification. They utilized AdaBoost and a weighted support vector machine (WSVM) that was tuned with particle swarm optimization (PSO) for feature selection. The proposed method effectively handles large, sparse datasets for brain image classification. The dataset included 198 AD cases and 229 normal controls (NC) for training and testing. The results showed a promising classification accuracy of 93%.

After a thorough review of the existing methodologies, we identified several research gaps. Some approaches rely on complex multiple deep learning techniques to extract features from MRI scan images, while others employ traditional machine learning techniques. However, neither approach has consistently yielded superior results. Additionally, some methods achieve better outcomes with small, imbalanced datasets, while hybrid methods using fusion traditional ML techniques show potential for improvement. Therefore, there is a need for a simpler yet more robust model to assist DSS (decision support system) in the accurate early-stage diagnosis of Alzheimer's disease (AD).

The primary objective of this research is to identify AD at an early stage using MRI images with a straightforward and robust model. We separated the feature extraction and classification tasks to design a hybrid model that leverages both deep learning and traditional machine learning techniques. Specifically, we used a Convolutional Neural Network (CNN) for feature extraction and a Support Vector Machine (SVM) for classification to improve the accuracy of Alzheimer's diagnosis. The CNN model extracted 512 features from high-quality MRI images, which the SVM then used to construct a hyperplane capable of classifying the images as AD or NC. This approach distinguishes itself from previous research by utilizing a balanced dataset with a sufficient number of images, all scaled to a resolution of 128 x 128. Furthermore, we ensure a balanced dataset using data augmentation techniques specifically tailored for binary classification.

Table I displays the results of our comparison study, which we conducted on the same sMRI dataset using the most recent techniques. The proposed model demonstrated superior performance in terms of accuracy and F1-score. The main contributions of this research are summarized below:

- Applying image preprocessing techniques to reduce noise and ensure label-wise balance in the dataset is crucial, particularly for binary classification.
- Extracting crucial features from MRI scan images using a fine-tuned CNN model.
- Predicting Alzheimer's disease using an SVM binary classifier with a linear kernel.
- Attaining a notable 99.60% accuracy in AD prediction.

TABLE I. A COMPARISON BETWEEN THE PROPOSED METHODOLOGY AND STATE-OF-THE-ART TECHNIQUES

Author Name	Model	Dataset	Accuracy (%)
(El-Assy et al., 2024)[14]	CNN	ADNI Dataset	99.57
(Saleh et al., 2023)[15]	DenseNet(feature extraction & Classification)	Kaggle Dataset	96.5
(De Mendonça and Ferrari, 2023) [18]	Support Vector Machine(feature extraction & Classification)	ADNI Dataset	92
(Kongala et al., 2023)[17]	Support Vector Machine(feature extraction & Classification)	ADNI Dataset	---
(Prajapati and Kwon, 2022)[16]	CNN(feature extraction & Classification)	ADNI Dataset	87.50
(Elahifasaee, 2022)[25]	AdaBoost and PSO(Feature extraction), SVM Algorithm for Classification	ADNI Dataset	93
(K et al., 2021)[21]	Support Vector Machine(feature extraction & Classification)	Kaggle Dataset	84

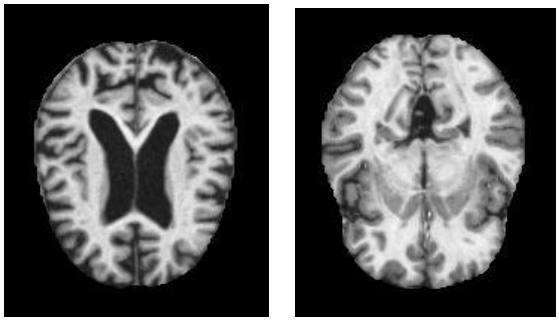
(Dwivedi et al., 2021)[23]	CNN for feature extraction , SVM Algorithm for Classification	ADNI Dataset	91.85
(Nanni et al.,2020)[19]	Fusion of Machine Learning algorithms and SVM Algorithm for Binary Classification	ADNI Dataset	93.30
(Lodha et al., 2018)[20]	Support Vector Machine(feature extraction & Classification)	ADNI Dataset	97.56
(Shen et al., 2018)[22]	Support Vector Machine(feature extraction & Classification)	ADNI Dataset	92.3
(NP and Varghese, 2018)[24]	Support Vector Machine(feature extraction & Classification)	ADNI Dataset	---
Proposed Model	Hybrid(CNN and SVM)	Kaggle Dataset	99.60

The paper divides the rest into sections. Section III provides a comprehensive overview of the MRI dataset and the architecture of the proposed methodology. Section IV describes results of proposed system, Section V discusses the results, while the final Section VI provides the concluding remarks.

III. Dataset Description and Proposed Methodology

A. Dataset Description

The dataset utilized in this research is obtained from Kaggle [26]. The data is gathered from diverse websites, with each label verified. The dataset consists of four types of images, 2,560 images of non-demented cases, 1,792 images of very mild demented cases, 717 images of mild demented cases, and 52 images of moderately demented cases. The dimensions of each MRI image are 176 × 208 and they are in the .jpg format. The objective of this research is to find people with or without dementia. The dataset is more suitable for binary classification rather than classification of multistage Alzheimer's disease due to the significant imbalance among the four labels, however, merging the labels into two better balance between the two classes. Fig. 1 depicts changes in sMRI images of the brain structure of subjects with and without dementia.

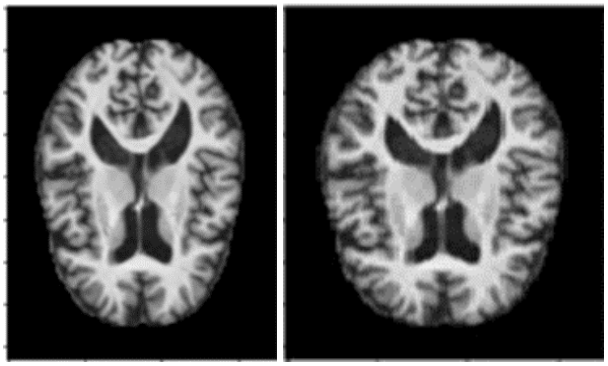


(a) (b)

Fig. 1. Shows (a) AD MRI brain scan image (b) Normal MRI brain Scan image.

B. Preprocessing Mri Images

High-resolution images have a larger number of pixels, which requires more computation and more main memory to save and process pixel values. The computational cost of the model can be greatly reduced by reducing the size of the images, leading to enhanced efficiency and quicker training and evaluation. Following the normalization process, the grayscale images undergo resizing to dimensions of 128x128 pixels, as depicted in Fig. 2.



(i) (ii)

Fig. 2. (i) Original image (176x208) & (ii) Resized image (128x128) Illustrates original and resized image.

C. Horizontal Flipping

In the realm of medical research, specifically in the domain of neuroimaging, the biggest problem is gathering an adequate quantity of images because of privacy concerns. Moreover, an insufficient and unbalanced dataset can cause overfitting problems, affecting the model's performance. In order to address these problems, data augmentation techniques are implemented on the original dataset. We experimented with all data augmentation techniques, such as adjusting brightness, rotation, zooming, etc., but we found that they did not improve the performance of our model. Therefore, we concluded to apply only the horizontal flipping technique [27] to enhance the size of the dataset as shown in Fig. 3. Post-processing, we ended with a balanced dataset consisting of two folders, "demented" and "non-demented," with 3200 images in each shown in Table II. We executed all preprocessing procedures using Python and built ML models by utilizing Keras libraries and Scikit-learn.

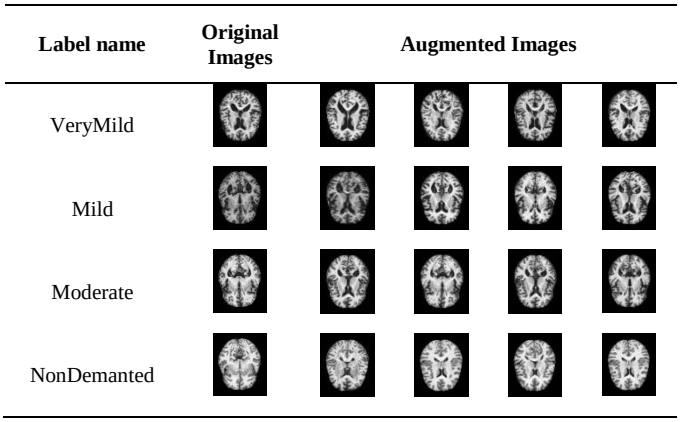


Fig. 3. Shows mirror images generated with horizontal flipping.

TABLE II. MRI DATASET AFTER DATA AUGMENTATION

Label Name	Number of Images
Non Demanted	3200
Demanted	3200
Total	6400

D. Feature Extraction

In the healthcare domain, where diagnosis involves image processing, CNN is highly effective in identifying patterns for classification [28] due to its ability to learn and extract relevant features directly from raw input data [29]. The proposed CNN model is structured with four convolutional blocks, each comprising two consecutive convolution operations followed by a pooling layer as shown in Fig. 4. The input layer accepts images of size 128x128 pixels with three color channels (RGB). The first convolutional block utilizes 64 filters and employs ReLU activation, Batch Normalization, MaxPooling, and Dropout to reduce spatial dimensions and prevent overfitting. This block outputs data in the shape of (64, 64, 64). The second block doubles the number of filters to 128 and follows a similar structure, outputting (32, 32, 128).

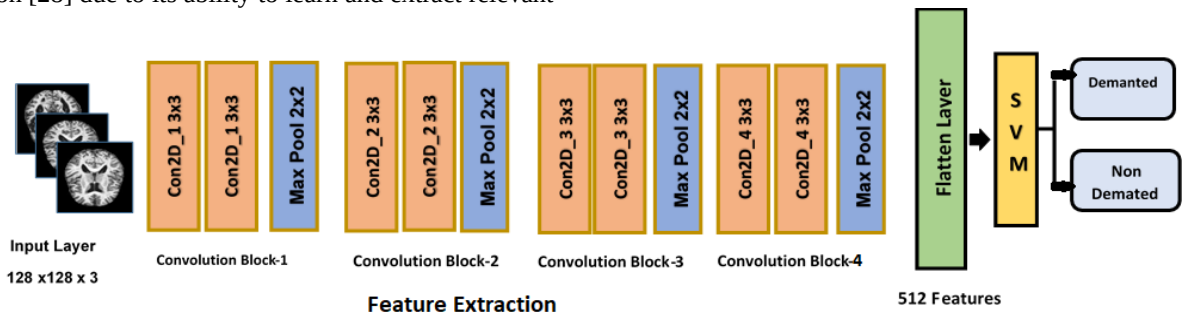


Fig. 4. Proposed model architecture.

The third convolutional block increases the filters to 256, maintaining the structure of convolutional layers, Batch Normalization, Max Pooling, and Dropout, resulting in an output shape of (16, 16, 256). The fourth block further increases the filters to 512 and follows the same pattern, producing an output shape of (8, 8, 512). Each block's design, with its combination of convolutional, normalization, pooling, and dropout layers, facilitates hierarchical feature extraction. This structure progressively enhances the model's ability to classify medical images by capturing increasingly complex patterns and features from the input data. The configuration of proposed CNN shown in Table III. The extracted features are given as input to classification algorithm for prediction.

TABLE III. PROPOSED CNN CONFIGURATION

Layer	Activation Function	Output Shape	Pool Size
2D Convolution Layer	ReLu	(128,128, 64)	
2D Max Pooling Layer	-	(64,64,64)	(2,2)
2D Convolution layer	ReLu	(64,64,128)	
2D Max Pooling Layer	-	(32, 32, 128)	(2,2)
2D Convolution Layer	ReLu	(32, 32, 256)	
2D Max Pooling Layer	-	(16, 16, 256)	(2,2)
2D Convolution Layer	ReLu	(16, 16, 512)	
2D Max Pooling Layer	-	(8, 8, 512)	(2,2)
Flatten Layer		(6400, 32768)	

Here, “Precn” – Precision, “Recl”- Recall,”Spcty”- Specificity, “Accry”- Accuracy,“F1S”- F1 score,”RF”- Random Forest,”LR”- Logistic Regression,DT- Decision Tree,NB- Naive Bayes Classifier,XGB- XGBoost,PM- Proposed Model.

E. Classification Algorithm

Support Vector Machines (SVM) [30] are a strong choice for binary classification tasks. SVM is well-suited for tasks where the objective is to distinguish between two classes, such as distinguishing between diseased and non-diseased patients in medical imaging. As shown in Fig. 5, the main goal is to create a hyperplane that effectively separates the input data points into two output classes. Eq. (1) illustrates the decision function is used to classify new data points.

$$f(x) = w \cdot x + b \quad (1)$$

Where,

The vector w corresponds to the weights, b to the bias, and x to the input features. We choose the hyperplane that gives the maximum distance between the two closest data points in each group. Eq. (2) represents the training dataset.

$$(x_1, y_1) \ldots\ldots, (x_{n,} y_n), x_i \in R_d \; and \; y_i \in (-1, +1) \quad (2)$$

Where

xi-feature vector and yi-output, Rd-set of feature vectors

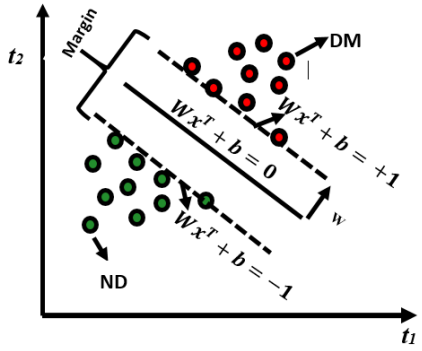


Fig. 5. Illustrates a linear SVM model with two classes.

For each element in the training dataset, the parameters 'b' and 'w' must satisfy the inequality specified in Eq. (3) and (4).

$$if \ \ f(x) > 0 \quad then \quad y_i = +1 \quad (Demanted) \tag{3}$$

$$if \ f(x) < 0 \ \ then \ \ y_i = -1 \quad (Non\ Demented) \tag{4}$$

Eq. (5) defines a function to classify a new data object (x).

$$f(x) = sign(w \cdot x + b) \tag{5}$$

Where sign (.) returns +1 or -1, indicating the class of the data point.

In this study, we evaluated various classification algorithms using extracted features. The Proposed model outperformed all other models across all metrics, showcasing its excellent ability in binary classification tasks. Proposed Model achieved the highest precision (99.35), recall (99.83), specificity (99.40), accuracy (99.60), and F1-score (99.58). Logistic Regression (LR) also performed well, though slightly lower than Proposed Model in specificity and recall. Random Forest (RF) and XGBoost (XGB) showed moderate results, while Naive Bayes (NB) and Decision Tree (DT) significantly lagged in most metrics, especially recall and F1-score. Overall, the Proposed Model demonstrates outstanding performance in binary classification shown in Table IV.

TABLE IV. COMPARING PROPOSED SVM CLASSIFIER WITH OTHER CLASSIFICATION ALGORITHMS

Model	Precn	Recl	Spcty	Accry	F1S
RF	87.64	89.34	86.29	87.89	88.48
LR	99.10	99.10	99.02	99.06	99.09
DT	76.47	76.01	74.55	75.31	76.23
NB	74.38	49.62	81.40	64.84	59.52
XGB	81.73	79.16	80.75	79.92	80.42
PM	99.35	99.83	99.40	99.60	99.58

F. SVM - Kernel Functions

A kernel function can boost the performance of a Support Vector Machine as shown in (6), it improves the performance of SVM by generating non-linear models in higher dimensions. It transforms complex problems into simpler ones by converting non-linear problems into linear ones. In a multi-dimensional space, this would necessitate sophisticated computations, but in this case, it speeds up the calculations.

$$K(t_1, t_2) = < f(t_1), f(t_2) > \tag{6}$$

K-kernel function, t1 and t2 are inputs that have M dimensions. Function f maps the input features from a space with M dimensions to a space with N dimensions. (t1, t2) is the dot product of the inputs.

We experimented with SVM with different kernels to achieve the best performance. The linear kernel performed exceptionally well, attaining the greatest accuracy. It surpassed other kernel functions. Conversely, The Polynomial and RBF kernels [31] are more appropriate for nonlinear data, although they exhibit somewhat poorer accuracy in comparison to the linear kernel. The sigmoid kernel performed poorly, indicating it is not suitable for this dataset. Fig. 6 shows accuracy comparison of various kernels.

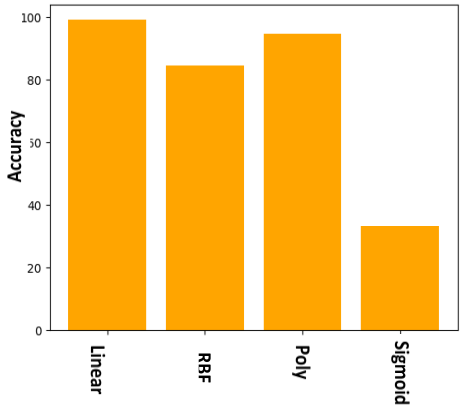


Fig. 6. Comparing the accuracy of different kernels using SVM.

G. Proposed Methodology Architecture

The key steps involved in the proposed methodology are depicted in Fig. 7. First, the data undergoes pre-processing to resize the images from 170x208 to 128x128 pixels. Data augmentation is then applied to increase the dataset size from 5,121 to 6,400 images, ensuring a balanced dataset. Key feature maps are extracted from each image using a CNN model, resulting in 512 features per image, yielding a resultant shape of (6400, 8, 8, 512). The dataset is then split into two sets: 80% for model training and 20% for testing. The SVM algorithm undergoes rigorous training to discover patterns in the data. Next, we perform model validation using a separate testing dataset. We conduct performance evaluation using widely recognized performance indicators like accuracy and F1 score.

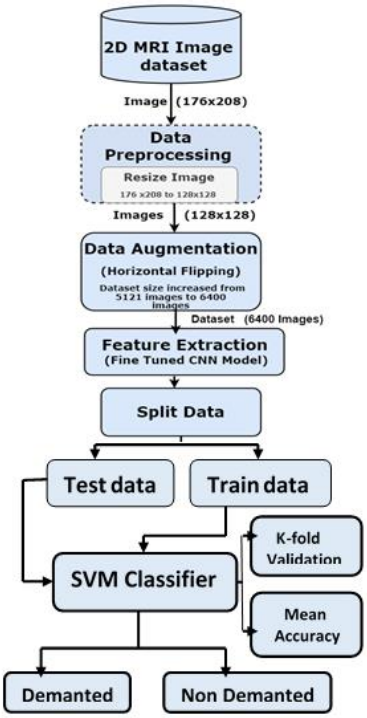


Fig. 7. The proposed methodology flowchart.

H. K-Fold Cross Validation

In order to mitigate the issue of overfitting, the model is tested with the cross-validation technique. This method is essential for evaluating the ability of a model by dividing the dataset into ‘t’ folds of equal size. In every iteration, the model trained on 't-1' folds, leaving one fold for the purpose of testing. This iterative process exposes the model to diverse data subsets, thereby mitigating overfitting. The StratifiedKFold cross-validation ensures consistent class distribution across folds. In this study, as illustrated in Table V, we used 15-fold cross-validation to demonstrate the model's consistent and robust performance across all folds, consistently observing high accuracy. This consistency indicates that the model can generalize to new data.

TABLE V. SHOWS 15-FOLD CROSS-VALIDATION RESULTS OF PROPOSED MODEL

Fold	AC(%)	AVG
1	99.63	
2	99.56	
3	99.73	
4	99.48	
5	99.63	
6	99.58	
7	99.58	
8	99.68	99.58
9	99.44	
10	99.68	
11	99.55	
12	99.30	
13	99.80	
14	99.66	
15	99.50	

Here, AC(%)-Accuracy, AVG-Average

IV. RESULTS

This section discusses model performance metrics, particularly those related to the health care domain, such as sensitivity and specificity. Sensitivity and specificity play crucial roles in assessing the model's accuracy in predicting positive and negative cases. Sensitivity (recall) measures the model's ability to correctly identify positive cases (Demanted) among all actual positive cases, while specificity gauges its accuracy in identifying negative cases (Non-Demanted).

The proposed model attains sensitivity of 99.83%, which implies its rare omission of AD cases and its accurate detection of positive cases, thereby potentially extending lifespan through early detection and appropriate treatment. Meanwhile, the model demonstrates a specificity of 99.40%, signifying high accuracy in identifying individuals without AD. Sensitivity helps us with early detection and treatment, while specificity provides a ratio of false alarms, reducing unnecessary interventions and costs.

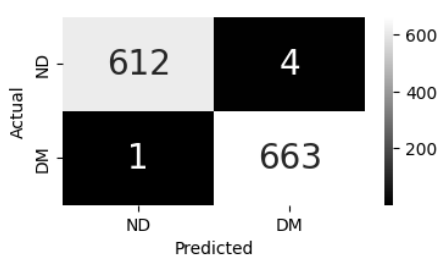
Precision is a quantitative measure that determines the ratio of correctly predicted positive cases to the total number of predicted positive cases. The model achieved a precision of 99.35%, indicating a minimal occurrence of incorrect positive predictions. The F1-score offers a comprehensive assessment of the model's performance by incorporating both precision and recall in its calculation. The model's F1-score of 99.58 indicates an excellent a balance between precision and recall, which is essential for accurate healthcare diagnosis.

V. DISCUSSION

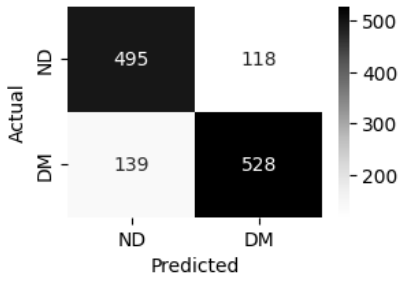
The proposed hybrid model, which uses fine-tuned CNN to pull out features from MRI images and SVM with a linear kernel for classification, achieved impressive metrics: 99.60% accuracy, 99.83% sensitivity, 99.40% specificity, and a 99.67 F1-score in detecting dementia patients. The above metrics collectively show how well the model is able to recognise AD cases while limiting false alarms, indicating its potential usefulness in clinical practice. Fig. 8 shows a comparison between our model's confusion matrix and the confusion matrix of other competitive classifiers. In the medical field, incorrect predictions can have serious consequences. The model we propose demonstrates a minimal number of such errors.

In this study, we applied only horizontal flipping for data augmentation to maintain the image's orientation. Other studies in the literature have employed various data augmentation techniques, which can cause their models to deviate in extracting key features. After several trial-and-error attempts, we determined that 128x128 was the optimal image size, whereas other studies used different sizes such as 64x64 or 256x256, which may result in the loss of important features or introduce noise into the image.

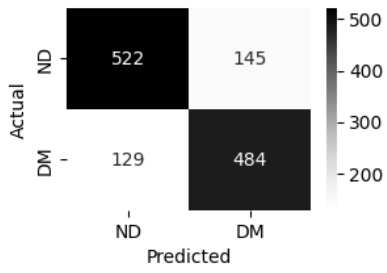
The proposed model successfully extracted crucial key features from well-preprocessed MRI images using a fine-tuned CNN model, outperforming state-of-the-art techniques mentioned in the literature. The proposed model produced superior accuracy compared to recent state-of-the-art techniques as displayed in Fig. 9. This model can serve as an automated computer-based decision support system, using structural MRI images to provide pattern-based decisions that assist manual systems in accurately identifying AD quickly, enabling timely medication to increase the lifespan of affected individuals.



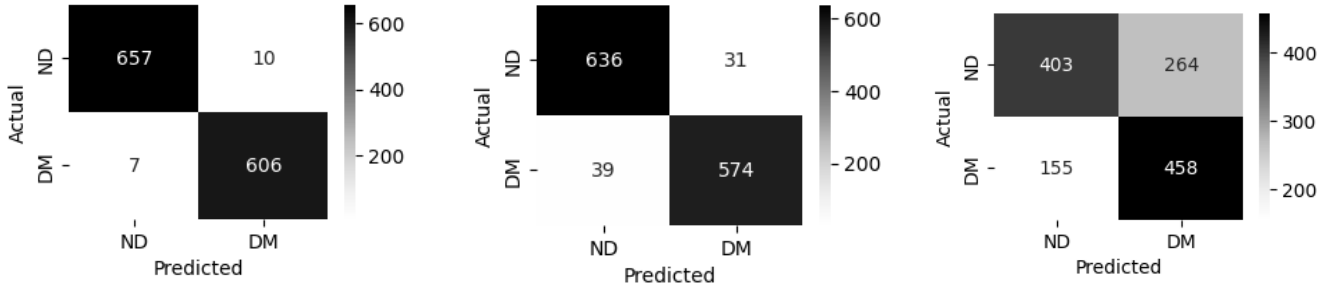
(i) Proposed



(ii) XG boost



(iii) Decision tree



(iv) Logistic regression (v) Random gorest (vi) Naïve bayes

Fig. 8. From (i) to (vi) demonstrates the CM for the proposed method and other classification algorithms. Here DM-"Demanted", ND-"Non_Demanted",CM-confusion matrix.

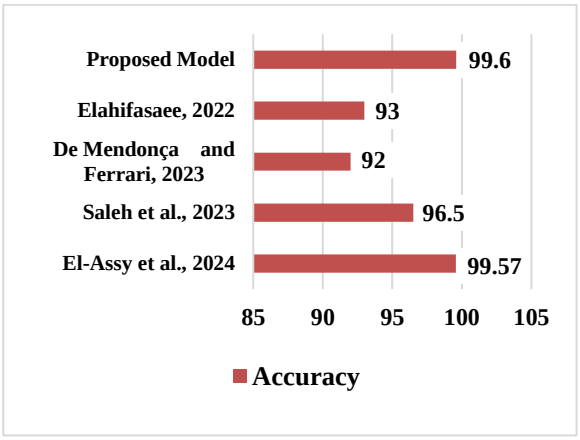


Fig. 9. Shows comparison of the proposed model's accuracy with state-of-the-art techniques.

VI. Conclusion

AD is a serious health concern, and early detection and treatment are crucial. The hybrid model is proposed to tackle this challenge. The model has demonstrated superior performance compared to contemporary techniques in various classification performance measures, indicating its potential as a robust tool for early diagnosis. Although the results are promising, there remains potential for additional enhancement and refinement. In this research, we used a dataset with sMRI images, integrating additional imaging modalities, such as functional MRI (fMRI), Fluid-Attenuated Inversion Recovery (FLAIR), Arterial Spin Labeling (ASL), Susceptibility Weighted Imaging (SWI), and Positron Emission Tomography (PET), could provide more comprehensive and precise insights into the disease's progression and characteristics. These modalities offer different perspectives on brain structure and function, which could enhance the accuracy and reliability of the diagnostic model. Furthermore, this study focused only on binary classification. Advanced deep learning techniques can extend this work to multistage classification using various modalities. This can lead to accurate, timely diagnosis, ultimately improving patient care and outcomes in the battle against Alzheimer's disease.

References

[1] Arafa, D.A., Moustafa, H.E.-D., Ali, H.A., Ali-Eldin, A.M.T., Saraya, S.F., 2023. A deep learning framework for early diagnosis of Alzheimer's disease on MRI images. Multimedia Tools and Applications 83, 3767–3799.https://doi.org/10.1007/s11042-023-15738-7.

[2] Kong, Z., Zhang, M., Zhu, W., Yi, Y., Wang, T., Zhang, B., 2022. Multi-modal data Alzheimer's disease detection based on 3D convolution. Biomedical Signal Processing and Control 75, 103565.https://doi.org/10.1016/j.bspc.2022.103565.

[3] Mehmood, A., Maqsood, M., Bashir, M., Shuyuan, Y., 2020. A Deep Siamese Convolution Neural Network for Multi-Class Classification of Alzheimer Disease. Brain Sciences 10, 84. https://doi.org/10.3390/brainsci10020084.

[4] 2023 Alzheimer's disease facts and figures, 2023. Alzheimer's & Dementia 19, 1598–1695. https://doi.org/10.1002/alz.13016.

[5] Hebert, L.E., Weuve, J., Scherr, P.A., Evans, D.A., 2013. Alzheimer disease in the United States (2010–2050) estimated using the 2010 census. Neurology 80, 1778–1783. https://doi.org/10.1212/wnl.0b013e31828726f5.

[6] Arevalo-Rodriguez, I., Smailagic, N., Roqué-Figuls, M., Ciapponi, A., Sanchez-Perez, E., Giannakou, A., Pedraza, O.L., Bonfill Cosp, X., Cullum, S., 2021. Mini-Mental State Examination (MMSE) for the early detection of dementia in people with mild cognitive impairment (MCI). Cochrane Database of Systematic Reviews 2021. https://doi.org/10.1002/14651858.cd010783.pub3.

[7] Harrell, L.E., Marson, D., Chatterjee, A., Parrish, J.A., 2000. The Severe Mini-Mental State Examination: A New Neuropsychologic Instrument for the Bedside Assessment of Severely Impaired Patients With Alzheimer Disease. Alzheimer Disease and Associated Disorders 14, 168–175. https://doi.org/10.1097/00002093-200007000-00008.

[8] Pangman, V.C., Sloan, J., Guse, L., 2000. An examination of psychometric properties of the Mini-Mental State Examination and the Standardized Mini-Mental State Examination: Implications for clinical practice. Applied Nursing Research 13, 209–213. https://doi.org/10.1053/apnr.2000.9231.

[9] Leifer, B.P., 2003. Early Diagnosis of Alzheimer's Disease: Clinical and Economic Benefits. Journal of the American Geriatrics Society 51. https://doi.org/10.1046/j.1532-5415.5153.x.

[10] Kavitha, C., Mani, V., Srividhya, S.R., Khalaf, O.I., Tavera Romero, C.A., 2022. Early-Stage Alzheimer's Disease Prediction Using Machine Learning Models. Frontiers in Public Health 10. https://doi.org/10.3389/fpubh.2022.853294.

[11] Tufail, A.B., Ma, Y.-K., Zhang, Q.-N., 2020. Binary Classification of Alzheimer's Disease Using sMRI Imaging Modality and Deep Learning. Journal of Digital Imaging 33, 1073–1090. https://doi.org/10.1007/s10278-019-00265-5.

[12] EL-Geneedy, M., Moustafa, H.E.-D., Khalifa, F., Khater, H., AbdElhalim, E., 2023. An MRI-based deep learning approach for accurate detection of Alzheimer's disease. Alexandria Engineering Journal 63, 211–221. https://doi.org/10.1016/j.aej.2022.07.062.

[13] Herrera, L.J., Rojas, I., Pomares, H., Guillen, A., Valenzuela, O., Banos, O., 2013. Classification of MRI Images for Alzheimer's disease Detection. 2013 International Conference on Social Computing. https://doi.org/10.1109/socialcom.2013.127.

[14] El-Assy, A.M., Amer, H.M., Ibrahim, H.M., Mohamed, M.A., 2024. A novel CNN architecture for accurate early detection and classification of Alzheimer's disease using MRI data. Scientific Reports 14. https://doi.org/10.1038/s41598-024-53733-6.

[15] Saleh, A.W., Gupta, G., Khan, S.B., Alkhaldi, N.A., Verma, A., 2023. An Alzheimer's disease classification model using transfer learning Densenet with embedded healthcare decision support system. Decision Analytics Journal 9, 100348. https://doi.org/10.1016/j.dajour.2023.100348.

[16] Prajapati, R., Kwon, G.-R., 2022. A Binary Classifier Using Fully Connected Neural Network for Alzheimer's disease Classification. Journal of Multimedia Information System 9, 21–32. https://doi.org/10.33851/jmis.2022.9.1.21.

[17] Rao, K.N., Gandhi, B.R., Rao, M.V., Javvadi, S., Vellela, S.S., Khader Basha, S., 2023. Prediction and Classification of Alzheimer's disease using Machine Learning Techniques in 3D MR Images. 2023 International Conference on Sustainable Computing and Smart Systems (ICSCSS). https://doi.org/10.1109/icscss57650.2023.10169550.

[18] de Mendonça, L.J.C., Ferrari, R.J., 2023. Alzheimer's disease classification based on graph kernel SVMs constructed with 3D texture features extracted from MR images. Expert Systems with Applications 211, 118633. https://doi.org/10.1016/j.eswa.2022.118633.

[19] Nanni, L., Interlenghi, M., Brahnam, S., Salvatore, C., Papa, S., Nemni, R., Castiglioni, I., 2020. Comparison of Transfer Learning and Conventional Machine Learning Applied to Structural Brain MRI for the Early Diagnosis and Prognosis of Alzheimer's Disease. Frontiers in Neurology 11. https://doi.org/10.3389/fneur.2020.576194.

[20] Lodha, P., Talele, A., Degaonkar, K., 2018. Diagnosis of Alzheimer's Disease Using Machine Learning. 2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA). https://doi.org/10.1109/iccubea.2018.8697386.

[21] K, U.R., S, Sharvari.S., G, Umesh.M., C, Vinay.B., 2021. Binary Classification of Alzheimer's disease using MRI images and Support Vector Machine. 2021 IEEE Mysore Sub Section International Conference (MysuruCon). https://doi.org/10.1109/mysurucon52639.2021.9641661.

[22] Shen, T., Jiang, J., Li, Y., Wu, P., Zuo, C., Yan, Z., 2018. Decision Supporting Model for One-year Conversion Probability from MCI to AD using CNN and SVM. 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). https://doi.org/10.1109/embc.2018.8512398.

[23] Dwivedi, S., Goel, T., Sharma, R., Murugan, R., 2021. Structural MRI based Alzheimer's Disease prognosis using 3D Convolutional Neural Network and Support Vector Machine. 2021 Advanced Communication Technologies and Signal Processing (ACTS). https://doi.org/10.1109/acts53447.2021.9708107.

[24] Thulasi N.P., K., Varghese, D., 2018. A Novel Approach for Diagnosing Alzheimer's Disease Using SVM. 2018 2nd International Conference on Trends in Electronics and Informatics (ICOEI). https://doi.org/10.1109/icoei.2018.8553789.

[25] Elahifasaee, F., 2022. Optimized SVM using AdaBoost and PSO to Classify Brain Images of MR. 2022 International Conference on Machine Vision and Image Processing (MVIP). https://doi.org/10.1109/mvip53647.2022.9738549.

[26] Alzheimer's Dataset (4 class of Images) [WWW Document], 2019. . Kaggle. URL https://www.kaggle.com/datasets/tourist55/alzheimers-dataset-4-class-of-images, Accessed on

[27] Mehmood, A., Maqsood, M., Bashir, M., Shuyuan, Y., 2020. A Deep Siamese Convolution Neural Network for Multi-Class Classification of Alzheimer Disease. Brain Sciences 10, 84. https://doi.org/10.3390/brainsci10020084.

[28] Mehmood A, Abugabah A, AlZubi AA, Sanzogni L (2022) Early Diagnosis of Alzheimer's Disease Based on Convolutional Neural Networks. Comput Syst Sci Eng 43(1):305–315. https://doi.org/10.32604/csse.2022.018520.

[29] "What is Feature Extraction? Feature Extraction in Image Processing | Great Learning." (n.d.) https:// www.mygreatlearning.com/blog/feature-extraction-in-image-processing/. Accessed 27 Feb 2024.

[30] Applications of Support Vector Machine (SVM) Learning in Cancer Genomics, 2018. Cancer Genomics & Proteomics 15. https://doi.org/10.21873/cgp.20063.

[31] Neffati, S., Taouali, O., 2017. An MR brain images classification technique via the Gaussian radial basis kernel and SVM. 2017 18th International Conference on Sciences and Techniques of Automatic Control and Computer Engineering (STA).
https://doi.org/10.1109/sta.2017.8314948.

Innovative Melanoma Diagnosis: Harnessing VI Transformer Architecture

Sreelakshmi Jayasankar[1]*, T. Brindha[2]
Research Scholar, Department of Computer Science and Engineering[1]
Associate Professor, Department of Information Technology[2]
Noorul Islam Centre for Higher Education, Tamil Nadu, India[1, 2]

***Abstract*—Melanoma, the most severe type of skin cancer, ranks ninth among the most prevalent cancer types. Prolonged exposure to ultraviolet radiation triggers mutations in melanocytes, the pigment -producing cells responsible for melanin production. This excessive melanin secretion leads to the formation of dark-colored moles, which can evolve into cancerous tumors over time and metastasize rapidly. This research introduces a Vision Transformer, revolutionizes computer vision architecture by diverging from traditional convolutional neural networks, employing transformer models to handle images as sequences of flattened, spatially-structured patches. The dermoscopy images sourced from the Kaggle repository, an extensive online database known for its diverse collection of high-quality medical imagery is utilized in this study. This novel deep learning model for melanoma classification, aiming to enhance diagnostic accuracy and reduce reliance on expert interpretation. The model achieves an accuracy of 96.23%, indicating strong overall correctness in classifying both Benign and Malignant cases. Comparative simulation of the proposed method against other methods in skin cancer diagnosis reveal that the suggested approach attains superior accuracy. These findings underscore the efficacy of the system in advancing the field of skin cancer diagnosis, offering promising prospects for enhanced accuracy and efficacy in clinical settings.**

Keywords*—Vision transformer; melanoma; convolutional neural networks; deep learning model; transformer encoder; dermoscopy image

I. Introduction

Prevalence of skin-related diseases has surged in recent years surpassing common conditions like hypertension and obesity [1]. Skin disorders account for approximately 12.4% of cases, affecting roughly one in every three individuals [2]. A concerning trend, a yearly increase of 1-2% in recorded skin diseases. Among these, melanoma stands out as the most aggressive form of skin cancer, capable of metastasizing through the lymphatic system and bloodstream to distant parts of the body. Melanoma arises from melanocytes, pigment-producing cells situated at the junction of the epidermis and dermis [3]. These cells are responsible for melanin production, when melanocytes undergo abnormal mutation, melanoma develops. This malignant condition poses a significant health risk due to its potential for rapid spread and invasive behaviour, underscoring the importance of early detection and effective treatment strategies.

Melanoma, a relatively uncommon form of skin cancer, poses a significant threat to mortality rates [4]. Although imaging studies can detect metastatic spread, the disease frequently goes undiagnosed until it progresses to an advanced stage or spreads to the bloodstream or lymph nodes [5]. It is essential to develop efficient computational techniques for early melanoma diagnosis. The five main forms of melanoma are nodular, lentigo maligna, Acral lentiginous, Subungual, and superficial spreading [6]. Each has a unique set of symptoms, interestingly amelanotic melanoma is a distinct subtype that occurs in people of different skin tones.

Conventional methods of melanoma diagnosis have limitations in terms of accuracy, accessibility, and scalability. Moreover, the increasing prevalence of melanoma underscores the urgent need for efficient and reliable diagnostic tools to address this public health challenge. In recent years, artificial intelligence (AI) and machine learning (ML) techniques have spurred the development of automated melanoma detection systems. These systems leverage computer vision algorithms, deep learning (DL) architectures, and large-scale datasets to analyze dermoscopy images and distinguish between benign and malignant lesions [7].

This paper aims to explore the current landscape of melanoma detection methodologies, highlighting the challenges and opportunities in this evolving field. Additionally, we present a comprehensive review of recent advancements in AI-based melanoma detection techniques, focusing on their strengths, limitations, and potential for clinical integration. The contributions of this work can be outlined as follows:

- Development of classification model based on DL aimed at effectively detecting and classifying melanoma, with a focus on enhancing detection performance.
- Assessment of the algorithm's performance using established benchmark metrics for evaluation.
- To assess its efficacy and explore its methodological strengths, compare the proposed model with existing models.

The rest of the paper is organized as follows: In Section II, a summary of literature is provided, highlighting areas that indicate a need for more investigation. In Section III, the methodology is explained in depth. Section IV goes into great detail about the results that the suggested strategy produced. A

discussion is provided in Section V and finally, a summary of the findings is included in Section VI, which gives a conclusion to the paper.

II. Literature Review

Melanoma detection has emerged as a significant global concern, drawing the interest of researchers worldwide who seek optimal methods for early identification of skin abnormalities to mitigate their progression. Numerous research endeavors have been initiated and continue to evolve in this field, aiming to improve patient outcomes and enhance the efficacy of medical interventions. This section provides a comprehensive overview of various research initiatives focused on melanoma detection.

Adla et al. [8] proposed a DL model for skin lesion detection. Tsallis entropy was utilized to identify the affected lesion areas in the dermoscopy images. Capsule Network in conjunction with class attention layer and Adagrad optimizer was utilized to extract features from the segmented lesions. The Convolutional Sparse Autoencoder, which was based on the Swallow Swarm Optimization algorithm, did the classification. The detection method exhibited limitations, particularly in its performance when presented with noisy images.

A CNN-based framework was presented by Shorfuzzaman et al. [9] to identify melanoma skin cancer. The final predictions were produced by a meta-learner, which incorporated all of the predictions from the sub models. The evaluation results demonstrated 95.76% accuracy in the ensemble model. A notable limitation of the paper lay in the extended duration necessitated for training, indicating a potential challenge in terms of resource allocation and efficiency within the framework.

To categorize the image samples of skin lesions, a framework was proposed by Khan et al. [10], consisting of two modules—the categorization and the localization of skin lesions. Transfer learning was used in the classification module to retrain a pre-trained DenseNet201 model on the segmented lesion images. The distribution stochastic neighbor embedding technique was used to downsample the features that were retrieved from the two fully connected layers. Using a fused vector, the highest accuracy on the ISBI2017 was 95.26%. One significant limitation of the work was that the model's training on localized regions entailed longer time in comparison to training on raw dermoscopy images, potentially impeding the scalability and practicality of the proposed approach.

Jiang et al. [11] introduced DRANet, a lightweight deep learning framework, for the classification of 11 types of skin diseases using real histopathological images. DRANet was the incorporation of a Squeeze Excitation Attention, which directed the framework's focus towards key areas crucial for identifying specific skin diseases. By employing stacked modules, the framework enhanced its capacity to learn from high-level features. Despite achieving an accuracy of 86.8%, the proposed approach was constrained by its inability to effectively diagnose images of poor quality.

Yacin Sikkandar et al. [12] introduced a model for skin lesion diagnosis, with an Adaptive NeuroFuzzy classifier merging a GrabCut algorithm. The model underwent simulation utilizing a benchmark ISIC dataset. Two significant limitations of the paper were the prolonged training time and the demand for substantial computational resources. The method relied on a large volume of data, posing a challenge in terms of data acquisition and processing.

Zghal et al. [13] sought to devise a straightforward model for detecting skin lesions from dermoscopy images, leveraging ABCD rules. Their approach consisted of five sequential stages: acquisition, pre-processing involving noise elimination and contrast enhancement techniques, and ultimately, classification via Total Dermoscopy Value computation. However, a notable constraint of their algorithm was its reliance on a substantial dataset for learning, which may not always be accessible.

Alwakid et al. [14] proposed a DL method for extracting a lesion zone in skin cancer diagnosis. ESRGAN was utilized to enhance image quality by generating high-resolution versions of low-resolution images. Melanoma and non-cancerous lesions could be distinguished using a ViT-based architecture suggested by Cirrincione et al. [15]. Based on the ISIC dataset, the suggested predictive model was evaluated with an accuracy of 94.8%.

An automated image-based method using ML classification techniques was presented by Inthiyaz et al. [16] for the diagnosis and classification of skin diseases. Their approach used the softmax classifier algorithm to identify images by leveraging Convolutional Neural Networks (CNNs). Six prevalent skin disorders were represented by images in the dataset, which showed different facial skin ailments. Their method showed significant effectiveness in skin condition detection and diagnosis with an obtained accuracy of 87%.

A. Research Gap

The research encounters challenges in performance attributed to the intricate visual attributes inherent in skin lesion images, characterized by diverse features and ambiguous boundaries. Detection accuracy notably diminishes for lesions smaller than 6mm, presenting a formidable hurdle in melanoma identification. Early melanoma symptoms often resemble benign skin conditions such as age spots and moles, underscoring deficiencies in early detection approaches. Furthermore, the subtle presentation of melanoma symptoms complicates early-stage detection for individuals, exacerbating gaps in effective detection strategies. Scarce access to medical data hampers algorithm development and training, highlighting deficiencies in data availability for research purposes. Moreover, the labor-intensive process of algorithmic development, validation, and deployment poses significant challenges, intensifying gaps in algorithmic implementation. The considerable variability in melanoma cases, including differences in size, shape, and color, poses a formidable obstacle for algorithms to achieve generalization, accentuating gaps in algorithmic robustness and adaptability.

Many models, such as those relying on capsule networks or CNN-based frameworks, exhibit performance degradation when faced with noisy or low-quality images, which is a significant drawback given the intricate visual characteristics

of melanoma lesions. Furthermore, models that require extensive training times or high computational resources, such as those employing ensemble learning or transfer learning, are impractical for real-time clinical applications where efficiency and scalability are crucial. Additionally, approaches like those utilizing ABCD rules or other manual feature extraction methods are limited by their dependence on large datasets, which are often not readily available, especially in diverse clinical environments. These limitations hinder the ability to achieve accurate and robust melanoma detection, particularly in early stages where symptoms may resemble benign conditions, thereby exacerbating gaps in effective diagnostic strategies. Consequently, a model that can address these challenges by providing reliable performance across varied image qualities, minimizing training time, and reducing dependence on extensive datasets is essential for improving melanoma detection and diagnosis. These limitations highlight the need for a more robust, adaptable, and efficient model capable of overcoming these challenges to provide accurate and timely melanoma detection.

III. Materials and Methods

The proposed methodology leverages the Vision Transformer (ViT) architecture as shown in Fig. 1, a cutting-edge approach in computer vision [17]. The Vision Transformer model was chosen for this research due to its innovative approach to image processing, which marks a significant departure from conventional convolutional neural networks (CNNs). ViT processes images by segmenting them into sequences of spatially-structured patches, which undergo linear embedding and positional encoding to preserve spatial information. This ability to handle complex, high-dimensional data makes ViT particularly well-suited for dermoscopy images, where subtle variations in color, texture, and structure are critical for accurate melanoma classification. These patches are then fed into a stack of transformer encoder blocks, allowing the model to capture global contextual dependencies and hierarchical features. Following this, the output undergoes global pooling to reduce spatial dimensions before being passed through a classification head comprising fully connected layers.

The classification phase translates spatial information into class predictions using softmax activation. Key components of the ViT architecture include patch extraction, patch embedding, and transformer encoder blocks. The self-attention mechanism within transformer encoder blocks allows the model to focus on relevant features while suppressing irrelevant ones. Residual connections and layer normalization facilitate gradient flow and stabilize input to subsequent layers. The Multi-Layer Perceptron (MLP head), the final component of the model, transforms aggregated representations from transformer encoder blocks into class predictions through fully connected layers and activation functions. Regularization techniques such as dropout layers employed to prevent over fitting. Ultimately, the methodology aims to learn representations for diverse visual patterns, fostering scalability and interpretability in image classification tasks.

A. Dataset

The data is collected from the Kaggle repository [18]. The dataset comprises a balanced collection of images representing two distinct categories: benign skin moles and malignant skin moles. Each category is represented by a folder containing 1800 images, with each image standardized to a size of 224x244 pixels. The standardized image size simplifies pre-processing tasks and ensures uniformity across the dataset, enabling straightforward integration into various ML pipelines. The balanced nature of the dataset ensures an equal representation of both benign and malignant cases, facilitating unbiased model training and evaluation. The inclusion of sample images, as depicted in the Fig. 2, provides a visual representation of the dataset contents, offering insight into the appearance and characteristics of both benign and malignant skin moles.

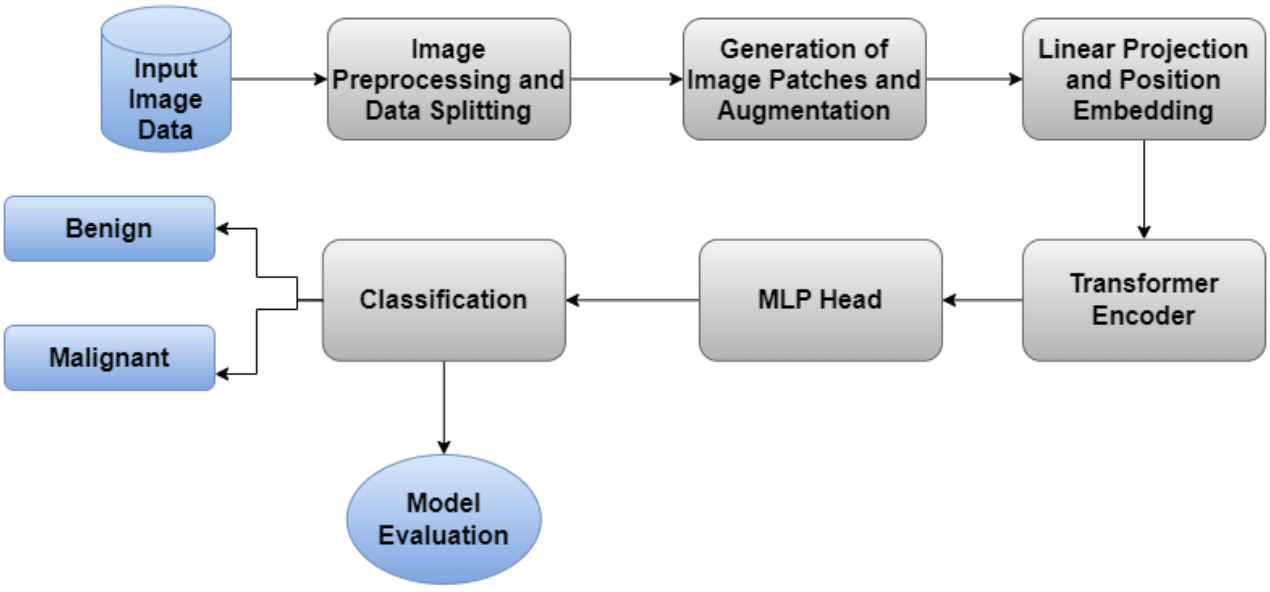


Fig. 1. Block diagram of the proposed system.

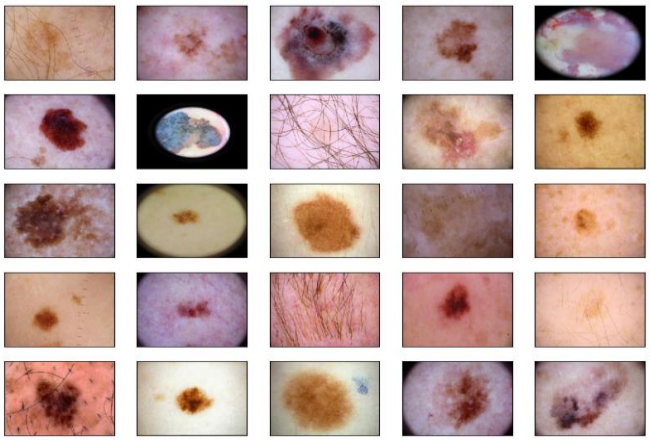


Fig. 2. Sample images from the dataset.

B. Pre-processing and Augmentation

Pre-processing involves preparing the input images for the model by standardizing various aspects to ensure consistency. One of the first steps in pre-processing is normalization, which calculates the mean and variance of the pixel values in the training images. By scaling the pixel values to a standard range, normalization minimizes the impact of intensity variations between different images. This is particularly important in medical imaging, where lighting conditions and image acquisition settings can vary significantly. Ensuring a uniform pixel intensity range helps the model focus on the relevant features of the melanoma lesions rather than being influenced by extraneous variations. Another key pre-processing step is resizing, where images are adjusted to a consistent size. This standardization ensures that all input images fit the model architecture requirements, making it easier for the model to process and analyze the images efficiently. Consistency in image dimensions also simplifies subsequent processing steps and improves the model's ability to learn from the data.

Data augmentation complements pre-processing by artificially expanding the dataset and introducing controlled variations to enhance model generalization. For melanoma detection, this often includes techniques such as horizontal flipping, rotation, and zooming. Horizontal flipping creates mirrored versions of the original images, introducing variability in image orientation. This technique helps the model learn to recognize melanoma lesions from different angles, enhancing its ability to generalize to real-world scenarios where lesions might not always be perfectly oriented. Random rotations further enrich the dataset by changing the angle at which the images are presented, making the model more adept at identifying melanoma regardless of its orientation in the image. Random zooming, on the other hand, introduces variations in the scale of the images. By simulating different distances from the lesion, zooming helps the model become proficient in detecting melanoma at various sizes and levels of detail.

C. Patch Generation

Patch generation phase extract patches from input images, a crucial process for uncovering localized features vital for a spectrum of computer vision tasks, including image segmentation, object detection, and image classification. Patch parameterization offers adaptability to various task requirements, allowing for tailored patch extraction, dictating the dimensions of the patches to be extracted. Leveraging the information including batch size, height, width, and channels alongside the specified `patch_size`, computes the number of patches extractable in both height and width dimensions, ensuring exhaustive coverage of the image data. This meticulous extraction process ensures uniform and controlled patch extraction across diverse datasets.

The extracted patches undergo reshaping process, transforming them into a 3D tensor format optimized for subsequent processing and analysis within the DL model. This transformation ensures seamless integration of patches into the larger computational framework, facilitating efficient feature extraction and model training. This serialization of patches is carried out to ensure consistency in patch generation across different model instances, facilitating seamless integration and reproducibility. Fig. 3 provides a visual comparison between a sample image and the corresponding image patches generated from it.



Fig. 3. Sample image and generated image patches.

D. Linear Projection and Positional Embedding

Patch encoding phase encode patches extracted from input images, a critical process that enriches the representation of localized features. Since transformers operate on fixed-size sequences and lack inherent understanding of spatial relationships, positional encoding is performed. During the initialization step, two important parameters are determined: the number of patches retrieved from the input images and the dimensionality of the projected feature space to which the patches are mapped. The initialization phase establishes the foundation for effective feature extraction and spatial representation within the encoded patches. Within the initialization, two sub layers are instantiated to facilitate the encoding process. It plays a role in protecting the input patches into a higher-dimensional feature space, enabling robust feature extraction.

The primary function of this layer is to provide position embeddings for every patch, thereby capturing essential spatial information within the encoded representation. By creating position indices for the patches, this technique makes sure that every patch is associated with a unique position. The input patches are then projected into the feature space that has more dimensions. The position embeddings generated are then added to the projected patches, effectively incorporating spatial

information into the encoded representation. The resulting encoded patches, enriched with both feature and positional information. The transformer processes these embeddings through multiple layers of attention mechanisms and feed forward networks.

The transformer encoder is a crucial component of the ViT architecture, responsible for processing and extracting features from input patches. It consists of several layers, each containing a set of modules such as Multi-Head Self-Attention Mechanism, Residual Connections and Layer Normalization, Feed forward Neural Network (MLP). The Fig. 4 illustrates the architectural framework of the ViT.

The mechanism of Self-attention enables the model to assign varying weights to different input patch embeddings, prioritizing relevant information while disregarding irrelevant parts. The mechanism of self-attention in the model enables differential weighting of input patch embeddings, prioritizing relevant information while downplaying irrelevant aspects. Each patch embedding undergoes linear transformation into key, query, and value vectors, which are then utilized to compute attention score. These scores are derived from the product of query and key vector, and processed with softmax function to yield attention weights. These weights are subsequently applied to the values to generate the attention output. To aid in gradient flow during training, residual connections are introduced, merging the attention output with the input and subjecting it to layer normalization. This normalization process stabilizes the output of each attention block, maintaining a consistent input range for subsequent layers. Following the attention mechanism, the output traverses through MLP consisting of two fully connected layers.

The MLP head within a ViT serves as the final stage of the model, responsible for converting the aggregated representations obtained from the transformer encoder blocks into class predictions. Typically, the MLP head comprises one or more fully connected layers, followed by an activation function. The input to this MLP head is the output derived from the last transformer encoder block, embodying the combined information extracted from the input image patches. Before being fed into the MLP head, these representations undergo flattening or global averaging. The resulting flattened or pooled representation is then passed through one or more dense layer, constituting the core of the MLP head. These layers enable the model to discern intricate non-linear relationships within the data. An activation function is applied after each fully connected layer to introduce non-linearity into the network.

E. Hardware and Software Setup

The research employed a computational setup that utilized a machine with powerful characteristics, including an Intel Core i7 processor. A powerful combination of 32GB RAM and the impressive NVIDIA GeForce GTX 1080Ti GPU. The model was implemented smoothly using the Keras library, which served as a prototype based on the Tensorflow architecture and executed using the flexible Python language. Keras, renowned for its intuitive interface and robust capabilities, played a crucial role in designing complex Neural Network structures. This framework guarantees optimal utilization of computational resources, effortlessly adapting to CPU, GPU, and TPU contexts. In order to take use of the powerful computational powers and optimise the process of training the model, the deployment was coordinated on Google Colab. Model training is made easier with this cloud-based Python notebook environment, which offers free access to powerful computing resources and supports interaction in development.

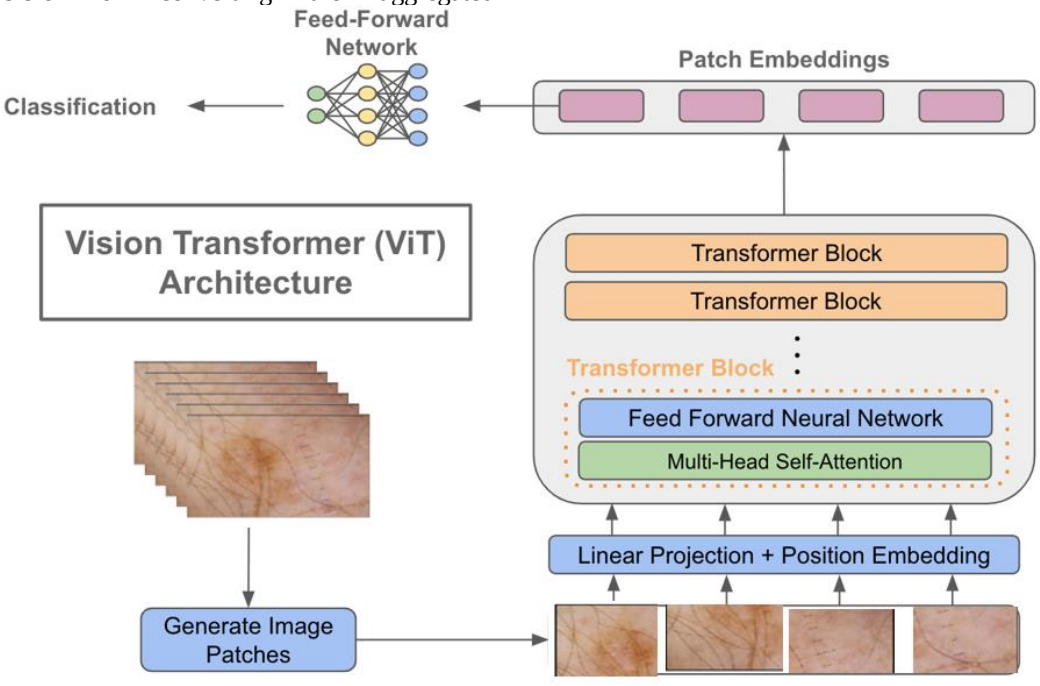


Fig. 4. Architecture of the VI transformer.

Hyper parameters are crucial configuration parameters that determine the behavior and attributes of a machine learning framework during the training phase. In contrast to the model's parameters, which are determined by the data itself, the user sets the hyper parameters prior to training. The neural network model utilizes the Adam optimizer. The training process is directed by the binary cross-entropy loss function. During the training process, the model handles input data in batches consisting of 32 samples per iteration. The training is conducted for 25 epochs, which represents the number of times the model processes the complete training dataset. The hyper parameter selections, including the optimizer, loss function, batch size, and number of epochs, determine the setup for training the neural network model. The goal is to optimise its performance in detecting melanoma. The proposed method's model configuration is presented in Table I.

TABLE I. MODEL CONFIGURATIONS

Parameters	Value
Learning rate	0.0001
Weight decay	0.00001
Image size	224
Patch size	6
Projection dimension	64
num_heads	6
transformer layers	6
mlp_head_units	[1024, 512]
Batch Size	32
Epochs	25

IV. EXPERIMENTAL RESULTS

The accuracy and loss plots are crucial for comprehending the performance and learning patterns of the proposed model. The accuracy plot visually depicts the model's ability to reliably predict data labels during training iterations on both the training and validation datasets. The alignment between the model's predictions and the actual labels is monitored to assess the model's performance throughout training.

The accuracy plot demonstrates the model's efficacy in differentiating between images containing signs of melanoma and those without, throughout the training process. Ideally, throughout the early stages, both the training and validation accuracies ought to increase simultaneously, demonstrating the model's ability to apply its knowledge beyond the training data. The trend illustrated in Fig. 5 indicates that the model is acquiring knowledge of fundamental patterns rather than only memorizing the instances presented in the training dataset.

The accuracy steadily increases from an initial value of approximately 0.8788 in the first epoch to around 0.9306 in the final epoch. This upward trajectory indicates that the model's performance improves over successive epochs as it learns from the training data. Notably, there are fluctuations in accuracy values throughout the training process, reflecting the dynamic nature of the optimization process and the model's adaptation to different patterns in the data.

A loss plot illustrates the trend of the model's loss function over different iterations or epochs during training. Fig. 6 illustrates the loss plot of the proposed model. The loss steadily decreases from an initial value of approximately 0.3062 in the first epoch to around 0.1707 in the final epoch. This downward trend signifies that the model's ability to minimize prediction errors improves over time. Lower loss values indicate better alignment between the model's predictions and the actual labels in the training data. Similar to accuracy, fluctuations in loss values are observed across epochs, reflecting the model's response to variations in the training data and optimization process.

An excellent way to assess the accuracy of the suggested model in detecting melanoma is by employing a confusion matrix. Fig. 7 presents the confusion matrix generated by the proposed model. The matrix offers a systematic summary of the model's performance by contrasting its predictions with the real labels across several classes. Essentially, it arranges the results in a tabular structure, with the rows representing the actual labels and the columns representing the predicted labels. Every individual cell in the matrix represents the number of occurrences where the model's predictions match or differ from the actual labels.

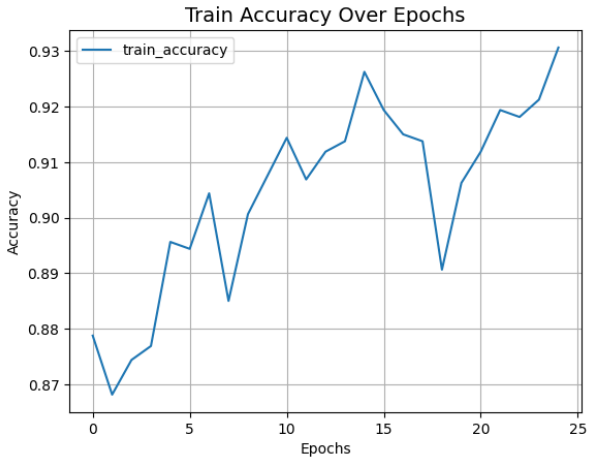


Fig. 5. Accuracy plot.

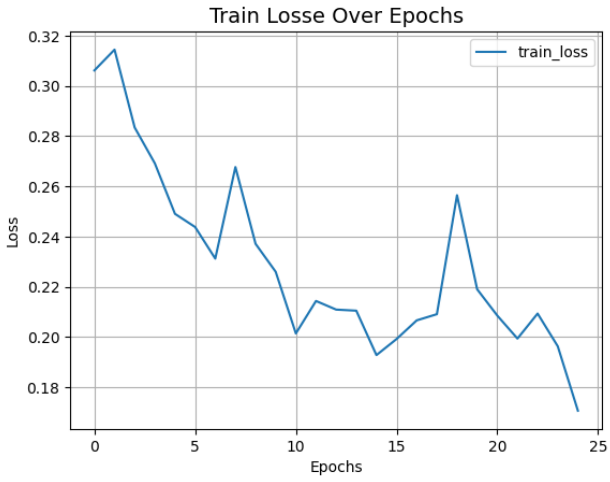


Fig. 6. Loss plot.

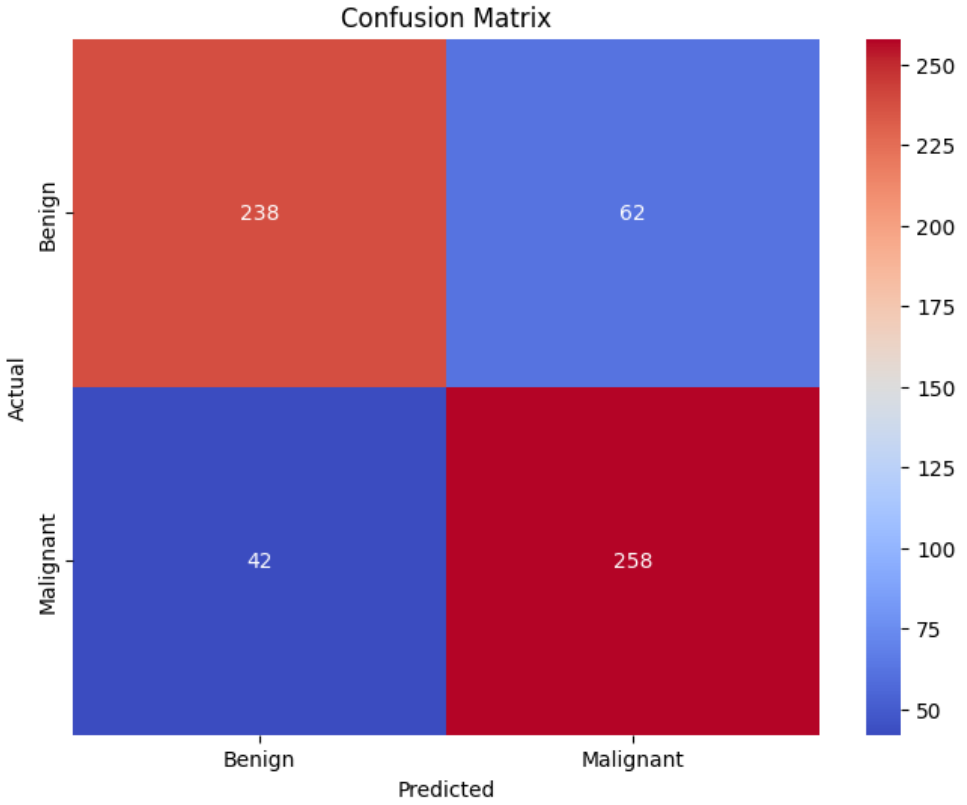


Fig. 7. Confusion matrix.

The confusion matrix is partitioned into four quadrants, where the items on the diagonal represent correct predictions and the elements off the diagonal represent cases of misclassification. This visual depiction allows for a comprehensive evaluation of the proposed model's efficacy in accurately detecting melanoma individuals. It reveals that the model accurately identifies 238 benign images as benign, but misclassifies 62 benign images as malignant. Similarly, it correctly identifies 258 malignant images as malignant, but erroneously classifies 42 malignant images as benign.

Performance metrics derived from the confusion matrix offer a thorough evaluation of the proposed model's efficacy in detecting melanoma. In order to thoroughly evaluate the efficacy and operational efficiency of the proposed model, the F1-score, accuracy, precision, and recall are the four primary metrics utilized. These measures, which are based on the concepts of False Positive (FP), False Negative (FN), True Negative (TN), and True Positive (TP), are essential for assessing the model's performance. These performance parameters have mathematical formulations that are shown in Eq. (1), (2), (3), and (4).

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

$$Precision = \frac{TP}{TP+FP} \quad (2)$$

$$Recall = \frac{TP}{TP+FN} \quad (3)$$

$$F1-score = 2 \times \frac{precision \times Recall}{Precision+Recall} \quad (4)$$

The obtained performance metrics, as shown in Fig. 8, highlight the exceptional efficacy of the developed model. An accuracy of 96.23% indicates that the model correctly identifies melanoma cases and non-melanoma instances with a high degree of reliability. The precision of 96.63% demonstrates the model's capability to accurately predict true positive cases of melanoma, minimizing the rate of false positives. The recall, or sensitivity, at 96.98% reflects the model's effectiveness in detecting almost all actual melanoma cases, ensuring a low rate of false negatives. The F1-Score, a harmonic mean of precision and recall, is 96.80%, underscoring the model's balanced performance in terms of both identifying true cases and excluding false alarms. These metrics collectively suggest that the model is robust and highly accurate, making it a reliable tool for the detection and classification of melanoma in clinical settings.

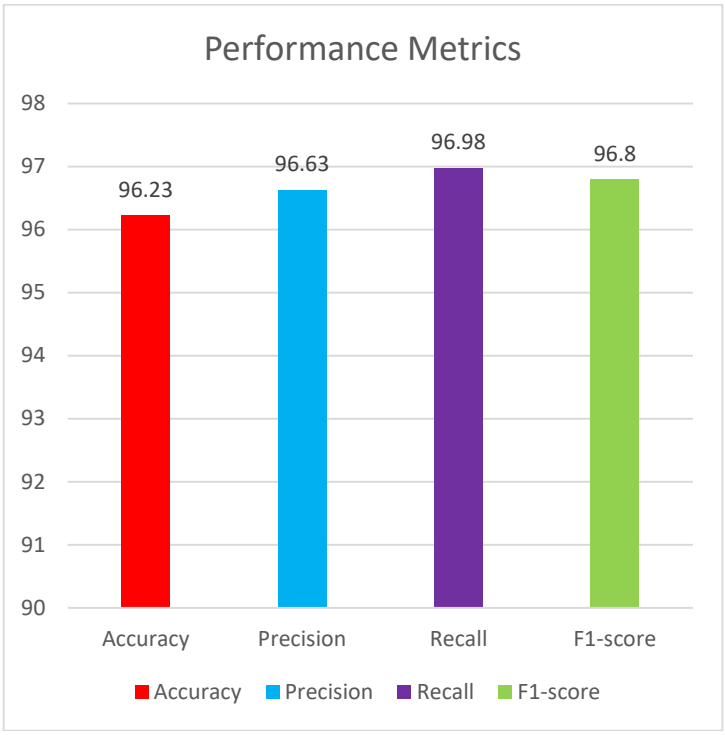


Fig. 8. Performance metrics.

Fig. 9 shows the Classification Output of the proposed system.

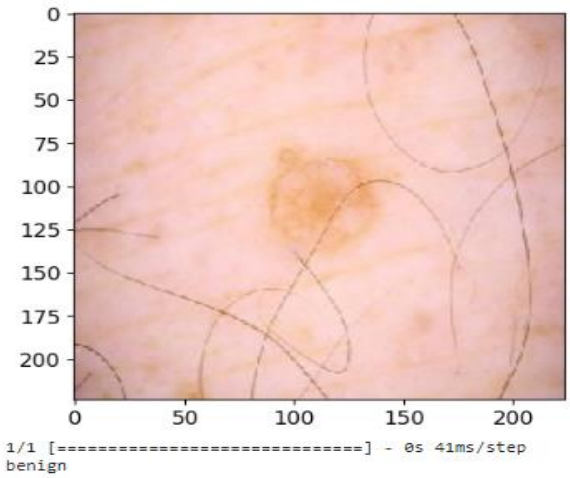


Fig. 9. Classification output of the system.

V. Discussion

The proposed Vision Transformer (ViT)-based deep learning model marks a significant advancement in the field of skin cancer diagnosis, particularly in melanoma detection. By leveraging the innovative architecture of the Vision Transformer, which processes images as sequences of spatially-structured patches rather than relying solely on convolutional layers, the model exhibits superior accuracy and robustness. As demonstrated in the Table II and Fig. 10, the ViT-based model achieves an accuracy of 96.23%, notably surpassing the performance of other state-of-the-art methodologies. Jiang et al.'s DRANet achieves an accuracy of 86.8%, while Shorfuzzaman et al.'s CNN-based stacked ensemble framework achieves a higher accuracy of 95.76%. Khan et al. employ DenseNet201 with transfer learning, achieving an accuracy of 95.26%. Inthiyaz et al.'s CNN method achieves an accuracy of 87%. This substantial improvement underscores the potential of the ViT-based model to enhance diagnostic accuracy, which is crucial for early detection and treatment of melanoma, ultimately contributing to better clinical outcomes.

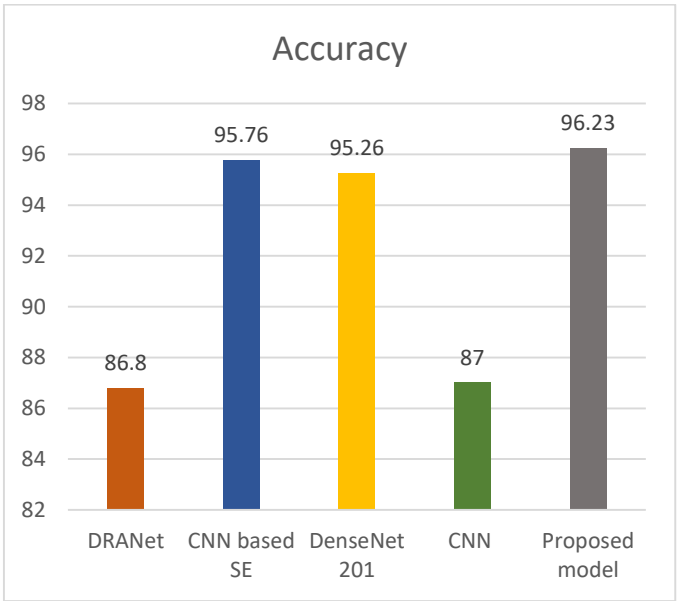


Fig. 10. Performance comparison.

The higher accuracy achieved by the ViT-based model is particularly noteworthy when considering the challenges associated with melanoma detection, including the complex and varied visual characteristics of skin lesions and the need for accurate differentiation between malignant and benign conditions. The model's ability to effectively handle these challenges, as evidenced by its outperformance of existing methods, highlights its robustness and adaptability. This advancement not only provides a powerful tool for clinicians but also addresses some of the key limitations of previous approaches, such as the extended training times, resource-intensive computations, and reduced efficacy in noisy or poor-quality images. The ViT-based model's success in overcoming these challenges and delivering high accuracy in skin cancer diagnosis positions it as a promising candidate for integration into clinical practice, where it could significantly improve the speed and accuracy of melanoma detection, ultimately saving lives through earlier and more precise intervention.

TABLE II. COMPARISON WITH EXISTING SYSTEM

Author	Methodology	Accuracy
Jiang et al	DRANet	86.8%
Shorfuzzaman et al	CNN based stacked ensemble framework	95.76%
Khan et al	DenseNet201 with transfer learning	95.26%.
Inthiyaz et al	CNN	87%
Proposed model	**Vi transformer based deep learning**	**96.23%**

VI. CONCLUSION

Melanoma, a form of skin cancer originating in melanocytes, presents a significant global health concern due to its aggressive nature and potential for metastasis. With its incidence steadily rising worldwide, melanoma detection and classification have become pivotal areas of research in the medical field. Early diagnosis plays a crucial role in improving patient outcomes, as timely intervention can significantly enhance treatment efficacy and prognosis. This research endeavours to develop a model capable of effectively detecting and classifying melanoma using publicly available datasets, with a focus on performance enhancement. The proposed approach introduces a Vision Transformer-based melanoma classification model adept at distinguishing between Benign and Malignant cases. With an accuracy of 96.23%, the model demonstrates a strong overall correctness in classification. These findings underscore the efficacy and potential of the proposed method in advancing the field of skin cancer diagnosis, offering promising prospects for enhanced accuracy and efficacy in clinical settings. These outcomes emphasize the effectiveness of the method in advancing skin cancer diagnosis, indicating promising prospects for heightened accuracy and efficacy in practical clinical settings. However, several avenues for future work could further advance the field. Future research could focus on expanding the dataset to include a more diverse range of skin types, lesion characteristics, and image quality conditions to improve the model's generalizability and robustness. Additionally, integrating the Vision Transformer with other advanced techniques, such as multi-modal data fusion or ensemble approaches, might enhance its performance further.

REFERENCES

[1] Parker, E. R., Mo, J., & Goodman, R. S. (2022). The dermatological manifestations of extreme weather events: a comprehensive review of skin disease and vulnerability. The Journal of Climate Change and Health, 8, 100162.

[2] Arnold, J. D., Yoon, S., & Kirkorian, A. Y. (2019). The national burden of inpatient dermatology in adults. Journal of the American Academy of Dermatology, 80(2), 425-432.

[3] Champsas, G., & Papadopoulos, O. (2020). The Role of the Sentinel Lymph Node Biopsy in the Treatment of Nonmelanoma Skin Cancer and Cutaneous Melanoma (pp. 647-704). Springer International Publishing.

[4] Saginala, K., Barsouk, A., Aluru, J. S., Rawla, P., & Barsouk, A. (2021). Epidemiology of melanoma. Medical sciences, 9(4), 63.

[5] Leong, S. P., Naxerova, K., Keller, L., Pantel, K., & Witte, M. (2022). Molecular mechanisms of cancer metastasis via the lymphatic versus the blood vessels. Clinical & Experimental Metastasis, 39(1), 159-179.

[6] Basurto-Lozada, P., Molina-Aguilar, C., Castaneda-Garcia, C., Vázquez-Cruz, M. E., Garcia-Salinas, O. I., Álvarez-Cano, A., ... & Robles-Espinoza, C. D. (2021). Acral lentiginous melanoma: Basic facts, biological characteristics and research perspectives of an understudied disease. Pigment cell & melanoma research, 34(1), 59-71.

[7] Zafar, K., Gilani, S. O., Waris, A., Ahmed, A., Jamil, M., Khan, M. N., & Sohail Kashif, A. (2020). Skin lesion segmentation from dermoscopic images using convolutional neural network. Sensors, 20(6), 1601.

[8] Adla, D., Reddy, G. V. R., Nayak, P., & Karuna, G. (2022). Deep learning-based computer aided diagnosis model for skin cancer detection and classification. Distributed and Parallel Databases, 40(4), 717-736.

[9] Shorfuzzaman, M. (2022). An explainable stacked ensemble of deep learning models for improved melanoma skin cancer detection. Multimedia Systems, 28(4), 1309-1323.

[10] Khan, M. A., Muhammad, K., Sharif, M., Akram, T., & de Albuquerque, V. H. C. (2021). Multi-class skin lesion detection and classification via teledermatology. IEEE Journal of Biomedical and Health Informatics, 25(12), 4267-4275.

[11] Jiang, S., Li, H., & Jin, Z. (2021). A visually interpretable deep learning framework for histopathological image-based skin cancer diagnosis. IEEE Journal of Biomedical and Health Informatics, 25(5), 1483-1494.

[12] Yacin Sikkandar, M., Alrasheadi, B. A., Prakash, N. B., Hemalakshmi, G. R., Mohanarathinam, A., & Shankar, K. (2021). Deep learning based an automated skin lesion segmentation and intelligent classification model. Journal of ambient intelligence and humanized computing, 12(3), 3245-3255.

[13] Zghal, N. S., & Derbel, N. (2020). Melanoma skin cancer detection based on image processing. Current Medical Imaging, 16(1), 50-58.

[14] Alwakid, G., Gouda, W., Humayun, M., & Sama, N. U. (2022, December). Melanoma detection using deep learning-based classifications. In Healthcare (Vol. 10, No. 12, p. 2481). MDPI.

[15] Cirrincione, G., Cannata, S., Cicceri, G., Prinzi, F., Currieri, T., Lovino, M., ... & Vitabile, S. (2023). Transformer-based approach to melanoma detection. Sensors, 23(12), 5677.

[16] Inthiyaz, S., Altahan, B. R., Ahammad, S. H., Rajesh, V., Kalangi, R. R., Smirani, L. K., ... & Rashed, A. N. Z. (2023). Skin disease detection using deep learning. Advances in Engineering Software, 175, 103361.

[17] Mao, X., Qi, G., Chen, Y., Li, X., Duan, R., Ye, S., ... & Xue, H. (2022). Towards robust vision transformer. In Proceedings of the IEEE/CVF conference on Computer Vision and Pattern Recognition (pp. 12042-12051).

[18] Fanconi, C. (2019). Skin cancer: malignant vs. benign. *Distributed by ISIC Archieve*.

Enhancing Safety for High Ceiling Emergency Light Monitoring

G. X. Jun[1], M. Batumalay[2], C. Batumalai[3], Prabadevi B[4]
Faculty of Data Science and Information Technology, INTI International University, Malaysia, Nilai, Malaysia[1, 2, 3]
School of computer Science Engineering and Information Systems, Vellore Institute of Technology, Vellore, Tamil Nadu, India[4]

***Abstract*—The performance of information technology has gradually improved and advanced during this period for safety management. Nonetheless, there is no disputing that during power outages, emergency lights continue to be crucial to people's safety and comfort. Regrettably, businesses don't give emergency lighting in buildings enough thought. The primary causes of the problem are expensive spending and long-term management for high ceiling safety. Additionally, one aspect that must be considered and ensured is the safety of maintenance personnel when the light is installed in high-rise locations. Thus, by creating wireless global control and monitoring via Android mobile phones, our effort intends to increase the availability and reliability of the emergency light. The suggested light monitoring system collects information from Internet of Things devices and transmits it to users' mobile phones over the Internet. Moreover, the risk of employees keeping the emergency light will be significantly reduced because it is monitored via the Internet on mobile devices. Additionally, by using the information the sensor inside the emergency light collects, it is possible to estimate its current condition, including its battery life. This repair will also improve everyone's safety within the building by increasing the emergency light's dependability with good process innovation.**

***Keywords*—Safety management; Internet of Things; high ceiling safety; high building safety; safety; process innovation**

I. INTRODUCTION

Buildings with several levels have been constructed more quickly and easily because of advances in current technology in the construction industry, which also included utilizing the power of materials and other clever technologies. Emergency lights are necessary to handle power outages in these facilities. The Department of Economic and Social Affairs' Goal 17 (2022) requires the installation of emergency lighting during building projects when a power outage could negatively affect the surrounding environment or even endanger human life. However, organizations may find it challenging to keep a sizable number of emergency lights maintained due to considerations including time consumption and safety. Furthermore, there is no denying that Internet of Things (IoT) technology has become indispensable in the current trend due to its automation and low participation control. Companies can get around these challenges by integrating Internet of Things technologies into emergency lighting systems. With Internet of Things capabilities, emergency lights may be remotely monitored and controlled, reducing the need for manual maintenance and ensuring quick problem-solving. Moreover, businesses can optimize power usage, save costs, and manage energy more effectively thanks to the automated features of IoT technology [1].

Building construction on a huge scale has become commonplace in the modern era, which is notable for its technological and economic advancements. However, it is imperative that emergency lights be installed in these structures because they are essential for surviving power outages. To prevent catastrophic catastrophes, emergency lighting must be provided in public or commercial buildings per government rules, with the preservation of human life and the environment receiving top priority [2, 3]. When there is a decrease in light or an emergency brought on by a power outage, emergency lighting turns on. Its primary job is to create enough light automatically so that everyone within the building can leave safely. Due to varying safety rules in different buildings and areas, there are numerous types of emergency lights available on the market. Dominik Pfaff has proposed four different types of emergency lights [2]. They are emergency escape lighting, which is designed to prevent dangerous accidents during emergencies; standby lighting, which continuously supports normal activities in specific locations and buildings; high-risk task area lighting, which is essential to the escape lighting system and provides enough illumination for people to reach specific areas; and escape route lighting, which guarantees clear visibility of evacuation routes.

It is obvious that emergency lights are necessary to guarantee people's safety in unexpected situations. Nonetheless, managing the availability and dependability of these lights is a big problem for businesses. The lack of practical means to ensure the continuous operation and reliability of emergency lights is one of the major problems [3]. Moreover, the financial aspect is a significant challenge since companies are often reluctant to bear the associated costs, such as wages, software development, and maintenance fees. Especially on construction sites where there are a lot of emergency lights, workers face several challenges throughout the maintenance process. Handling this big array is going to take a lot of work. The placement of emergency lights in elevated areas significantly increases the risk of injury or even death to a maintenance worker while they are performing their task. The last problem is that relying just on visual observation to assess emergency lights' functioning by personnel jeopardizes the accuracy of the test results and could lead to incorrect evaluations of the lights' operational state. Emergency lights play an important role in protecting and ensuring human safety during emergency events. However, companies often neglect the management of reliability and availability of emergency lights due to various challenges. Firstly, there is a lack of efficient ways to maintain the performance and reliability of emergency lights. Moreover, cost spending is also a factor that companies are unwilling to face, including employee fees,

software development, and maintenance fees. In addition, there are various challenges that employees face during maintenance of emergency lights. For example, due to the large number of emergency lights on the construction site, it requires a lot of time to manage the lights. Some emergency lights are installed at higher locations, which may cause injury or even loss of life for maintainers if they fall while maintaining the emergency lights. Lastly, the accuracy of examination results is low because employees can only tell if the emergency lights are working by observation.

Consider a scenario where a maintainer is inspecting emergency lights that are located at a height to provide another example that will help clarify this idea. Because of how challenging it is to reach these areas; maintenance personnel may need to employ ladders or other risky climbing equipment. Such actions constantly increase the risk of injury to maintenance personnel. Should the ladder be unstable, or the process not be carried out correctly, maintenance personnel could fall from a height and suffer severe injuries or even lose their lives. Working at these high altitudes therefore necessitates extra safety procedures as well as personnel training to minimize potential hazards. In addition to highlighting the urgent need to address these problems, this example illustrates the dangers that maintainers may face during inspections. For companies that value the safety of their maintenance staff, it is imperative to establish strict protocols and norms for operating at elevated heights. protecting the emergency lighting. Routine maintenance and inspections should be performed to ensure the stability and dependability of ladders and other machinery. Companies can lower the risk of accidents and protect the health of their maintenance workers by implementing these safety precautions. Furthermore, ladders and other equipment need to have routine maintenance and inspections performed to guarantee their stability and dependability. Organizations can mitigate the likelihood of accidents and safeguard the welfare of their maintenance staff by implementing these preventive measures.

The IoT technology is a network of physical things that have related to sophisticated software and high-tech sensors to facilitate data interchange with other devices and computers over the internet [2]. IoT devices link to a cloud, like Google Firebase or an IoT gateway, to exchange the gathered data for local data analysis. The Internet of Things (IoT) system, which includes low-power sensor technology, cloud computing platforms, and sophisticated connectivity, can benefit from the integration of artificial intelligence and machine learning to improve the data collection process. As a concrete illustration of an IoT system, modern applications include supply chains, robotics, smartwatches, and health monitoring devices [1].

Several businesses could be revolutionized by IoT technology. IoT devices, for instance, may monitor machinery in real-time in the manufacturing industry, identifying any problems before they arise and enabling preventative maintenance. This can boost output and drastically cut down on downtime. IoT devices in the healthcare industry are able to monitor patient vitals and notify medical personnel in the event of an emergency. The Internet of Things (IoT), a revolutionary concept that connects diverse products and systems through the internet, has evolved with the rapid growth of technology. This network's interconnection makes data sharing and communication easy, which has several advantages for a range of sectors. IoT technology has the power to completely change the way we live and work, from increasing productivity and efficiency to boosting safety and convenience. The capacity of IoT technology to facilitate predictive maintenance is one of its main benefits. IoT sensors, for instance, can be mounted on machines in the manufacturing sector to track performance and identify possible problems before they cause a breakdown [4]. Real-time data on variables like temperature, vibration, and power usage may be gathered by these sensors, and artificial intelligence algorithms can be used to assess the results. This minimizes downtime and increases overall equipment effectiveness by enabling maintenance professionals to proactively arrange repairs or replacements. IoT technology may also notify technicians about certain maintenance requirements, guaranteeing that the appropriate resources are available when needed [5]. IoT technology also has the potential to gather and evaluate data in order to spot patterns and trends, which helps maintenance specialists maximize maintenance plans and raise equipment dependability. An illustration of an IoT system can be found in Fig. 1.

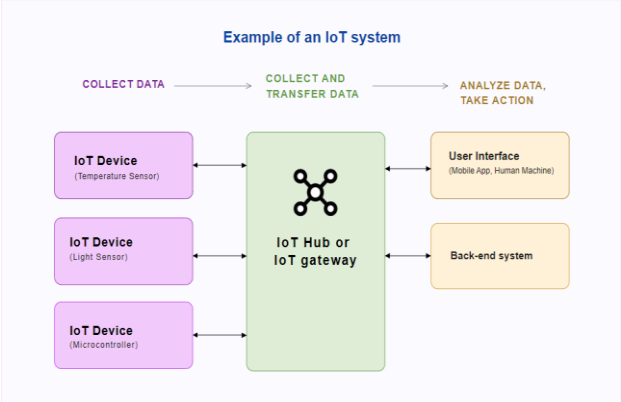


Fig. 1. Example of an IoT system.

The extensive attraction of IoT can be attributed to its pivotal role in intelligent management, which leverages automation to reduce human involvement and preserve time. This trend is further supported by the advantages it provides. Firstly, IoT increases productivity through task automation and simple connections between devices, systems, and processes. This reduces the requirement for manual intervention while optimising the utilisation of resources. Second, intelligent connectivity between IoT devices leads to enhanced decision-making capabilities and optimised operations. Real-time data monitoring and analysis enables people and businesses to respond swiftly to changing circumstances. Finally, IoT contributes to better resource management by enabling efficient asset monitoring and utilisation, which lowers costs and increases production. When everything is considered, the Internet of Things is not simply a fad but rather a revolutionary force that is facilitating state-of-the-art research and creating intelligent, effective, and adaptive products that satisfy the constantly shifting needs of our globalised society.

Large-scale construction projects are becoming more prevalent in the rapidly evolving 21st-century environment,

which is characterized by economic growth and technological advancements. Installing emergency lighting is also necessary in these kinds of locations to ensure residents' safety during power outages. Moreover, given the critical role emergency lights play in averting serious repercussions for human life and the environment, it is important to resolve maintenance-related concerns surrounding them (Goal 17 | Department of Economic and Social Affairs, 2022). Firstly, the proposed system addresses these problems head-on by providing a unique solution: a smartphone app designed to monitor emergency light conditions without forcing users to approach close. This promises to address concerns related to safety and enhance the efficiency of maintenance protocols.

The core of this innovative system is the Internet of Things (IoT) technology, which generates critical data through sensors [6]. Long-distance monitoring and control are also made possible by the data that is safely kept on the Firebase server and sent. Furthermore, because of its low cost and the availability of open-source tools and software, Arduino is selected as the Internet of Things device for data collection. This solution guarantees affordability and provides a wealth of resources through open-source, comprehensive software development modules.

The proposed approach extends its UI to an Android app, going beyond only integrating hardware. Users of this app can also easily regulate operations and monitor the status of emergency lights. While speed of performance is not the primary factor, the Android operating system was specifically selected for the application, emphasizing its ability to display data in real-time [6]. Because it makes use of Google's Firebase cloud server and online storage platform, which are both efficient and readily available, the system is also the best choice. The platform's easy setup and installation processes, in addition to its free plan for students and learners, contribute even more to the overall effectiveness of the recommended system. In essence, this innovative strategy uses contemporary technology to support accessible, safe, and effective building management while concurrently solving issues with emergency light maintenance. To improve routine check efficiency and human safety, this project aims to develop a system that integrates Internet of Things (IoT) sensors with a smartphone application to remotely monitor emergency light status.

II. METHODOLOGY

The waterfall model is one of the straightforward techniques in the Software Development Life Cycle (SDLC), an organised methodology used to produce high-quality products. The requirement analysis, design, development, testing, deployment, and maintenance phases are the six stages that this paradigm usually entails. But just the first four stages of the project—analysis, design, development, and testing—have been completed thus far. The waterfall model is a step-by-step process where the output from one phase becomes the input for the subsequent phase. It's important to remember that a variety of factors, including stakeholder preferences and project complexity, influence the decision between the waterfall model and alternative techniques. Phase 1, data collection, which included requirements analysis and design; Phase 2, system development, which covered the development activities; and Phase 3, testing, which concentrated on the testing phase, comprised the three phases of the research approach. This methodology follows a step-by-step process, with each phase building on the output of the one before it. It is similar to the waterfall approach as mentioned by another team [7].

A. Phase 1: Data Collection

Initially, in-depth perspectives and opinions on the idea from Malaysian citizens were obtained through interviews, a qualitative data gathering technique. To ascertain whether the public appendices and papers in the report were acceptable to them, three individuals were interviewed. The purpose of this interview was to ascertain the significance of emergency lighting and the suggested system, as well as to assess people's comprehension of them. In addition, it aids in determining the qualities that people find appealing. Furthermore, the interviews also helped in identifying any potential concerns or issues that may arise with the proposed emergency lighting system. This feedback will be crucial in refining and improving the system before implementation. Overall, the interviews provided valuable insights that will inform the development and implementation of the emergency lighting system. By addressing any concerns raised and incorporating desirable features, the final product will be better tailored to meet the needs and expectations of the user.

Knowledge of the suggested high-ceiling emergency light monitoring system was obtained from talks with professionals in the field. An expert emergency light vendor stressed the significance of emergency lights in guaranteeing security during blackouts. The expert also voiced excitement about the potential advantages of an IoT system, emphasizing how it might improve system reliability by lowering the need for physical involvement during maintenance, thereby increasing safety. Furthermore, it highlighted how crucial practical aspects like battery life monitoring and long-distance control are in huge facilities. First, in a similar vein, an emergency mechanical engineer emphasized the need for emergency lighting for safe building evacuation in times of crisis. The engineer also said that the system would be more reliable and that inspectors would not have to be there in person as much. He suggested features like automated testing that could generate reports and long-range control by supporting the use of an IoT monitoring system. Finally, the suggestions made by industry professionals will have a significant impact on the design and development of the High Ceiling Emergency Light Monitoring System.

The use case diagram in Fig. 2, which is a representation of the functionality and extent of the system using the Unified Modelling Language (UML) language, then illustrates the interaction between the actors and various use cases [8]. The use case diagram provided below illustrates how actors and use cases relate to one another in the proposed system. The functioning of the high-ceiling emergency light monitoring system is also depicted in the diagram. It describes how communication between the three primary components—Firebase, IoT devices, and the Android app—works. The diagram emphasizes the seamless flow of commands and data between different components, highlighting their interconnectedness. It explains the intricate communication routes and shows how Firebase serves as a central hub for data storage and retrieval, enabling real-time control and monitoring.

As the system's senses, the Internet of Things devices communicate back and forth with Firebase and the Android app to maintain a continuous flow of critical data. [9]. Furthermore, the Android app acts as the user interface, facilitating a variety of operations for consumers. By illustrating the transparent communication channels and the relationship between user actions and system reactions, the graphic highlights the system's efficiency and eventually enhances the monitoring and control of emergency lights in situations with high ceilings.

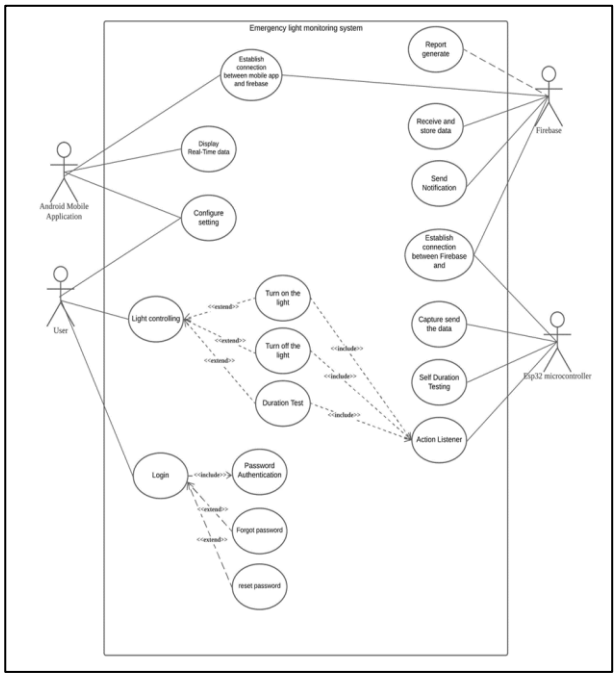


Fig. 2. Use case diagram for solid waste classification system.

B. Phase 2: System Development

Fig. 3 depicts the block diagram of the prototype system, which utilizes the ESP32 module as the core controlling and monitoring component. The system is composed of three distinct sections: the controlling board (navy blue), the controlling medium (green), and the original target device (orange).

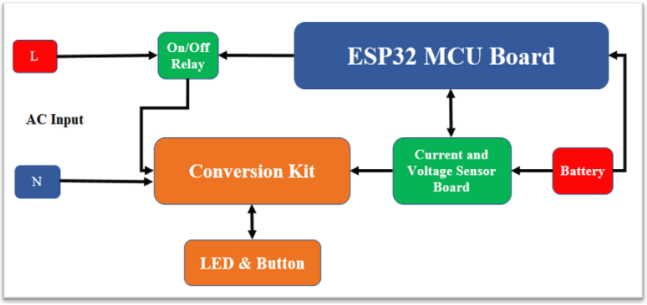


Fig. 3. Block diagram of prototype system.

As illustrated by Fig. 4, the suggested emergency light monitoring system is organized around three main elements: data display via Android mobile phones, data storage in Firebase, and data sensing through IoT devices [10]. Furthermore, the system uses the Arduino ESP32 module as its microprocessor, which makes data management and wireless network connections like Wi-Fi and Bluetooth possible. Furthermore, in order to improve its functionality, the ESP32 module supports a wide range of sensors, including voltage, light, and temperature. First off, the ESP32 may be connected to the internet and Firebase server networks thanks to Arduino's open-source software development tools, which include Wi-Fi, Bluetooth, and Firebase modules. Additionally, users can permanently store configuration data, such as the positions of emergency lights and Wi-Fi connections, using the ESP32's ROM storage. The ESP32's LED lights make effective performance testing possible [10]. After sensor data is gathered, it is sent over the internet to the Firebase server platform for real-time monitoring and archiving. In addition, the Firebase server has cloud messaging and cloud function messaging features that allow users to communicate with one another directly from the server [10]. The system's third component is an Android mobile application that lets users keep an eye on emergency light data. Additionally, the application gives users control over the emergency light switching during system inspections by displaying all collected data parameters saved in Firebase. The application also highlights the emergency light's condition with colour changes. Last but not least, since Google developed Android, connecting the Android application to Firebase is a simple process, ensuring seamless integration and user-friendly access to real-time emergency light monitoring [10, 11]. This comprehensive system design not only leverages IoT technology for efficient data sensing but also employs Firebase and Android integration to provide a user-friendly and effective emergency light monitoring solution [12].

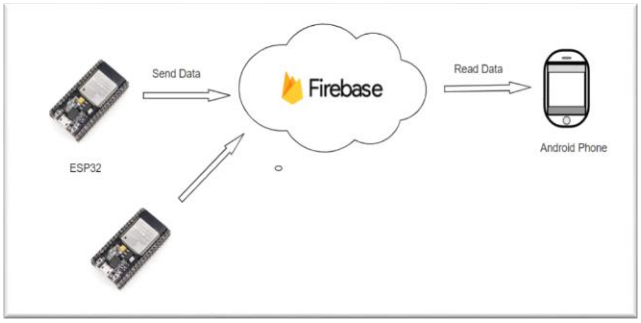


Fig. 4. Structure of proposed system development .

C. Phase 3: Testing

The High Ceiling Emergency Light Monitoring System went through a thorough testing process, including unit, integration, and acceptance testing to check its usability, reliability, and performance. In the unit testing phase, specific criteria for each component, from the ESP32 microcontroller [13] to various sensors and user interface activities, were carefully examined. Additionally, the successful completion of tasks, such as connecting to the internet, capturing sensor data accurately, and performing user actions, demonstrated the individual components' functionality. Moreover, integration testing further strengthened the system's robustness, ensuring smooth communication between modules. The system demonstrated cohesive behaviour and met the predetermined pass criteria, regardless of whether it was internet connectivity, real-time database interactions, or user interface functionalities. Furthermore, acceptance testing provided a complete

evaluation, confirming that the system effectively met user requirements and expectations [14, 15]. Documented procedures and results from each testing phase attest to the system's overall success, confirming its high performance, reliability, and user satisfaction and highlighting its readiness for deployment in monitoring emergency lights in high-ceiling environments. The system's ability to meet the pass criteria in various areas demonstrates its robustness and adaptability. Additionally, the positive feedback received from users during acceptance testing further validates its effectiveness and usability in real-world scenarios.

III. Results and Discussion

The High Ceiling Emergency Lighting Monitoring System's useful features and characteristics make it a valuable tool for emergency lighting system maintenance is captured in Fig. 5.

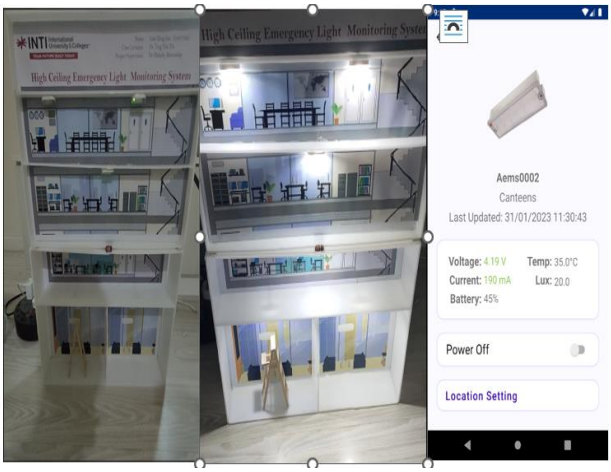


Fig. 5. High ceiling emergency light monitoring system prototype.

One notable feature that guarantees emergency light dependability during power outages is the automatic testing feature. This feature allows for regular assessments of emergency lights without requiring operator intervention. Additionally, maintenance staff can effectively operate emergency lights from a distance thanks to the system's support for remote control. Additionally, a reliable testing function tracks battery life in compliance with safety standards. As a result, real-time monitoring features give precise information about the state of the system and enable prompt problem identification. The report-generating capability also gathers diagnostic data to facilitate prompt remediation and ongoing enhancement. All these carefully thought-out features together increase the system's ability to dependable sustain emergency lighting systems in rooms with high ceilings. Furthermore, the sophisticated technology of the system guarantees effective coordination in emergency scenarios by enabling a smooth interface with other building management systems. Furthermore, the interface is easy to use, making it simpler to adjust settings to suit individual requirements and tastes. IoT based monitoring supports easily scalable and can handle expanding or complex systems without significant additional infrastructure. It also help to reduces maintenance efforts with automated alerts, predictive analytics, and remote management capabilities. Finally, IoT also adapts easily to changes in system requirements or facility needs with minimal adjustments.

Still, there are a few things that could be done better with the High Ceiling Emergency Light Monitoring System. The first requirement for the system's correct operation is a reliable internet connection. The Internet of Things devices accountable for gathering data, storing it in Firebase, and receiving commands require a dependable and continuous internet connection. Furthermore, because the system only supports Android devices at this time, users of other mobile operating systems may find their accessibility limited [13]. Because of this, the overall efficacy and viability of putting the High Ceiling Emergency Light Monitoring System into place should be carefully considered while weighing these limitations.

IV. Conclusion and Future Enhancement

As the High Ceiling Emergency Light Monitoring System is developed further, a few possible enhancements become apparent. Initially and foremost, it is necessary to tackle the significant issue of enhancing the status reporting procedure. The current system can identify defective circuits, for example, but it may be enhanced by providing more precise information, even though it already provides a rudimentary diagnosis report. In addition, this comprehensive information would enable maintainers to carry out repairs with greater accuracy and efficiency. In addition, it would be very important to have an urgent notification system. Ignoring the lack of a notification system for users in the event of an emergency, like a fire tragedy or ongoing issues, would be one method to ensure prompt remediation and enhance overall safety. This means that looking into more advanced IoT device connectivity perhaps through a mesh network like Bluetooth mesh for the ESP32 which might end up being a game-changer. In the end, this would increase system responsiveness by enhancing group control, saving energy, and reducing network traffic. The installation process also needs to be optimized for the user's convenience. Adding features like auto-configuration and self-diagnosis could improve the installation process' efficiency and usability in the future. Keeping the system at the forefront of emergency light monitoring technology is the goal of all these significant enhancements. All things considered, the suggested improvements present a system that could significantly improve emergency light performance and reliability, setting a higher standard for general human safety.

References

[1] Gokhale, P., Bhat, O., & Bhat, S. (2018). Introduction to IOT. International Advanced Research Journal in Science, Engineering and Technology, 5(1), 41-44.

[2] Madakam, S., Lake, V., Lake, V., & Lake, V. (2019). Internet of Things (IoT): A literature review. Journal of Computer and Communications, 3(05), 164.

[3] Cameron, R. (2021, July). Emergency Warning Light Technology. In Transportation Research Circular 475: 11th Equipment Management Workshop (pp. 52-58).

[4] Ullo, S. L., & Sinha, G. R. (2020). Advances in smart environment monitoring systems using IoT and sensors. Sensors, 20(11), 3113.

[5] Methul, S., & Kaswa, S. (2023). Green Lights Ahead: An IoT Solution for Prioritizing Emergency Vehicles. Journal of Ubiquitous Computing and Communication Technologies, 5(3), 250-266.

[6] Pravalika, V., & Prasad, C. R. (2019). Internet of things based home monitoring and device control using Esp32. International Journal of Recent Technology and Engineering, 8(1S4), 58-62.

[7] Chun, Y. K., Thinakaran, R., & Nagalingham, S. (2022, May). Real-Time Face Mask Detector Using Deep Learning. In 2022 Applied Informatics International Conference (AiIC) (pp. 159-164). IEEE.

[8] Fauzan, R., Siahaan, D., Rochimah, S., & Triandini, E. (2019, July). Use case diagram similarity measurement: A new approach. In 2019 12th International Conference on Information & Communication Technology and System (ICTS) (pp. 3-7). IEEE.

[9] Goadrich, M. H., & Rogers, M. P. (2021, March). Smart smartphone development: iOS versus Android. In Proceedings of the 42nd ACM technical symposium on Computer science education (pp. 607-612).

[10] Nishadha (2022). A comprehensive survey on real-time applications of WSN. Future internet, 9(4), 77.

[11] Manvi, M. M., & Maakar, M. S. K. (2020). Implementing Wireless Mesh Network Topology between Multiple Wi-Fi Powered Nodes for IoT Systems. International Research Journal of Engineering and Technology, 7(10), 1242-1244.

[12] Li, W. J., Yen, C., Lin, Y. S., Tung, S. C., & Huang, S. (2018, February). JustIoT Internet of Things based on the Firebase real-time database. In 2018 IEEE International Conference on Smart Manufacturing, Industrial & Logistics Engineering (SMILE) (pp. 43-47). IEEE.

[13] Babiuch, M., Foltýnek, P., & Smutný, P. (2019, May). Using the ESP32 microcontroller for data processing. In 2019 20th International Carpathian Control Conference (ICCC) (pp. 1-6). IEEE.

[14] Rogers, R., Lombardo, J., Mednieks, Z., & Meike, B. (2009). Android application development.

[15] Mohammad, M., Pagkale, P. J., Abd Rahman, N. F., & Shariff, M. S. M. (2022). Hydrological Safety of Vaturu Dam by Evaluating Spillway Adequacy. The Eurasia Proceedings of Science Technology Engineering and Mathematics, 21, 349-355.

Data-Driven Approaches to Energy Utilization Efficiency Enhancement in Intelligent Logistics

Xuan Long
School of International Trade Hainan College of Economics and Business, Hai'kou 571127, Haikou, China

***Abstract*—With the rapid development of intelligent logistics, new challenges and opportunities are presented for energy utilization efficiency improvement. This study explores the feasibility and effectiveness of using data-driven methods to improve energy utilization efficiency in an intelligent logistics environment and provides theoretical support and practical guidance for achieving the sustainable development of optimized logistics management procedures. First, a dataset was established by collecting relevant data in the optimized logistics management procedure, including transportation information and energy consumption data. Then, data analysis and mining techniques are used to conduct an in-depth dataset analysis to reveal the influencing factors of energy utilization efficiency and potential optimization directions. Then, strategies and methods for energy utilization efficiency improvement are designed by combining intelligent optimization algorithms. Finally, simulation experiments and case studies are utilized to verify the effectiveness and feasibility of the proposed methods. The results show that using data-driven methods can significantly improve the energy utilization efficiency of optimized logistics management procedures, reduce logistics costs, and enhance the sustainability and competitiveness of the system. Through in-depth analysis and empirical research, a series of actionable optimization strategies are proposed, providing new ideas and methods for optimizing energy and logistics management procedures. These results significantly promote the sustainable development of optimized logistics management procedures and enhance competitiveness.**

***Keywords*—Intelligent logistics; energy; utilization efficiency; data-driven**

I. Introduction

As of 2021, China's digital economy is expected to reach $45.5 billion, according to a report by China's Ministry of Industry and Information Technology. With China's economy expected to grow by 4.5% in 2024, the digital economy is already playing an integral role in production and life [1]. Since the new Crown Pneumonia epidemic in 2019, the digital economy has contributed significantly to China's economic recovery and pushed various industries toward digital transformation. China's economy is developing coordinated and systematically, with the digital economy playing an increasingly prominent role in this process. China is in a critical period of industrial reform, and the digital economy will play a vital role in this process, driving the development of new industries and upgrading and transforming traditional ones. However, China faces many challenges to its long-term economic development, including colossal energy consumption and environmental concerns [2]. The international energy situation has recently been unstable, and China's dependence on imported resources is under pressure. It has therefore become particularly urgent to ensure national energy security, avoid new energy bottlenecks, utilize digital technologies to improve energy efficiency, develop new clean energy technologies, improve the environment, and promote the development of a sustainable green society [3]. China needs to deepen the reform of its energy system, strengthen the clean and efficient use of coal, and accelerate the design and construction of new energy systems while promoting the two main objectives of coal. The Global Digital Energy Conference, Carbon Summit, and Carbon Neutral Forum 2022 emphasized accelerating the convergence of digital energy and innovative energy technologies [4]. The Chinese government has emphasized promoting the development of digital transformation in the green energy industry, accelerating the development of the clean energy industry, and building a safe and efficient electrical system. Applying the digital economy in the energy sector is critical to China's digitalization and achieving its "dual-carbon" goals, improving new clean energy technologies and enhancing green energy efficiency [5]. These initiatives are crucial to promoting green transformation and realizing high-quality economic development in China. Therefore, studying the factors affecting green energy efficiency in the digital economy is essential. In-depth research in this area can help address China's challenges in sustainable development and green transformation while providing essential references and guidance for future policymaking.

The rapid rise of the digital economy has attracted widespread attention globally, as it changes production methods and daily life and profoundly impacts overall energy utilization. With the widespread use of digital technologies, concerns about energy efficiency are increasing [6]. Improving energy utilization efficiency is urgently needed to ensure high-quality economic development and a sustainable social environment. Research on the impact of the digital economy on green energy efficiency is of dual significance, not only to deepen theoretical understanding but also to provide valuable insights for practice [7]. Although research on the impact of the Internet or ICT on global energy efficiency has been relatively adequate, research on the specific impact of the digital economy on urban energy efficiency still needs to be developed. Therefore, it is necessary to combine theoretical and empirical analyses to explore specific ways to improve the energy efficiency of various elements of the digital economy [8]. This research results will not only enrich the theoretical foundation of the digital economy field in terms of improving the eco-efficiency of various energy elements but also provide essential references for the formulation of specific policy recommendations in various regions.

II. BACKGROUND

In recent years, China has made some progress in restructuring its energy supply side. It has optimized its energy mix by strengthening its green policy and financial system [9]. Despite the gradual replacement of some of the use of coal by cleaner energy sources, total carbon emissions have continued to rise, resulting in China still needing to reach peak carbon emissions. This suggests that improving the efficiency of energy use remains an integral part of China's energy strategy. China faces an unbalanced energy mix, with shortages of traditional energy sources such as coal, oil, and natural gas, and coal dominates the energy consumption mix. Although China's energy mix has improved from 2011 to 2021, with coal's share of the energy mix decreasing, it has remained the dominant carbon in the overall energy mix [10]. With the restructuring of energy policies, CO2 emissions are expected to increase further by 2030, posing challenges for China's environment and sustainable development. China's energy infrastructure is unlikely to be effectively upgraded in the short term, resulting in lower than global average energy utilization efficiency. This situation may not only hinder business development but also exacerbate environmental concerns. While most modern countries globally rely on clean and renewable energy sources, China's energy mix relies heavily on traditional carbon-intensive sources [11]. This means China still has a long way to go in its energy transition. Therefore, China needs to intensify its efforts further to improve energy utilization efficiency and actively promote energy structure transformation to achieve sustainable economic development and environmental improvement. Through innovative science and technology, policy support, and international cooperation, China can achieve a more significant transformation of its energy mix and contribute to the goal of carbon neutrality.

Improving energy efficiency involves reducing carbon dioxide emissions and addressing environmental and climate challenges. As China has gradually become one of the world's major CO_2 emitters, the international community's attention to reducing emissions has increased. In some developed countries, such as the EU member states, initiatives such as promoting CO_2 tariff programs and establishing green barriers have become essential strategies to combat climate change. These measures are expected to be further strengthened, and China, one of the world's largest greenhouse gas emitters, will be under even greater international pressure [12]. In addition, national economic issues are increasingly taking on an international dimension, such as trade wars and technological embargoes, and economic development is no longer confined to the domestic arena but is closely linked to the global economic landscape. Therefore, adapting to China's new economic development standards and realizing sustainable development has become a top priority [13]. Improving energy utilization efficiency reduces carbon emissions, solves the problems of unbalanced energy consumption and insufficient energy reserves in China, and promotes economic restructuring and upgrading. By adopting clean and efficient energy technologies, such as renewable and energy storage technologies, energy utilization can be maximized while reducing reliance on traditional fossil energy sources, providing solid support for China's sustainable economic development.

China's solemn commitment to achieving the "double coal" goal reflects its active role in global environmental governance and demonstrates its determination to build an ecological civilization. Achieving the "dual coal" goal is necessary for China's sustainable development and an essential contribution to the global ecological environment and climate stability [14]. The Chinese government has put forward a series of plans, including promoting the establishment of a zero-carbon energy sector, which will promote the optimization and transformation of the energy structure by encouraging enterprises to adopt advanced technologies, shift to cleaner energy sources, and significantly increase the use of renewable and cleaner energy sources [15]. However, despite a series of measures already taken, it will take time to realize the shift in energy consumption patterns, and there are several challenges to achieving the "double coal" goal. The Chinese government has implemented environmental protection documents on the agenda, but there still needs to be a gap in addressing the problem [16]. The inconsistent pace of market reforms is also a challenge, so the Chinese government is gradually exploring market-based environmental regulatory tools, including the introduction of a carbon planning system, an energy exchange system, and a carbon swap system, in order to adapt to market trends and facilitate the development and improvement of the carbon market. Despite the many challenges, improving energy efficiency remains one of the keys to realizing China's development goals in the new era. This will address China's energy consumption imbalance and reserve shortage, effectively respond to environmental and climate threats, and enhance the country's coping capacity [17]. Therefore, the Chinese government will continue to promote energy efficiency, realize the "double coal" goal, make more remarkable contributions to global sustainable development, and demonstrate the responsibility and commitment of a great country (Table I).

TABLE I. FACTORS AFFECTING ENERGY EFFICIENCY

Variant	Brochure	Average value	S.D	Minimum value	Maximum values
T	251	3.62	0	0.25	60.36
SO_2	251	555.36	0	43.21	68423
CO_2	251	0.516	0	0.0021	0.654
economic structure	251	47.62	0.36	12.36	651.52
GPD	251	983.36	0.54	44.636	66517
GNP	251	44.63	77.88	11.25	11158.32

Ensuring the proper functioning of the carbon dioxide trading system requires establishing a fair carbon dioxide allocation mechanism and effective carbon market regulation to ensure that enterprises comply with the relevant regulations. The fairness of such a mechanism directly affects the stability and sustainable development of the carbon trading market. In carbon markets set up abroad, the impact of the CO_2 quota method on company emissions and CO_2 intensity could be more precise, so further research and practice are needed to determine the most effective allocation method [18]. In addition, paying for CO_2 does not automatically change the transfer of costs to the energy sector, so regulatory mechanisms are needed to ensure that these costs are used correctly to promote the development and use of

clean energy. Different carbon allocation methods impact innovation, operational viability, and product prices. For example, historical intensity increases carbon prices, while reference rules incentivize firms to promote innovation. Therefore, the most suitable carbon allocation method needs to be selected by considering the actual situation of enterprises and the development trend of the market. Recently, China's carbon allocation mechanism has been widely discussed. Some scholars have developed carbon trading decision-making models and studied the impact of different carbon allocation methods on the economic efficiency of the carbon market. These studies provide essential references for constructing China's carbon market and help promote the regular operation and development of the carbon trading system.

III. Methodology

A. Study Design

The main reasons for choosing data from 2011 to 2019 to assess the digital economy and environmental efficiency can be summarized in two ways. First, 2011 marked the launch of China's 12th Five-Year Plan, which promulgated energy-saving and emission-reduction policies and new environmental and energy requirements. Implementing these policies had a profound impact on the economic structure, industrial layout, and energy consumption patterns of cities, so the data from this period can more accurately reflect the effects of these policies. Second, 2019, as the last year of data collection, helps to avoid the interference of epidemics in the data and ensure the accuracy and comparability of the results.

The level of the digital economy is an essential criterion for assessing the degree of digitization of a city or region. It plays a crucial role in evaluating the degree of economic development of a city. China's digital economy development over the past timeframe has shown an average annual growth rate of 0.122. Although the overall level of development in the digital economy is relatively low, its potential is still enormous. Specifically, from 2011 to 2019, China's digital economy index grew from 0.07 to 0.162, indicating that the level of the digital economy has shown a steady improvement and that relevant policies have achieved significant results. Among the top 20 cities in Guangdong Province, six possess rich political, workforce, and resource advantages and excel in the digital economy. This phenomenon reflects the differences in the level of digital economy among cities and suggests the importance of factors such as politics, workforce, and resources in urban development. In order to promote sustainable economic development, innovative cooperation with high-level cities should be strengthened to enhance the digitalization level of cities, targeting low-level cities.

Intelligent logistics transportation process, as shown in Fig. 1.

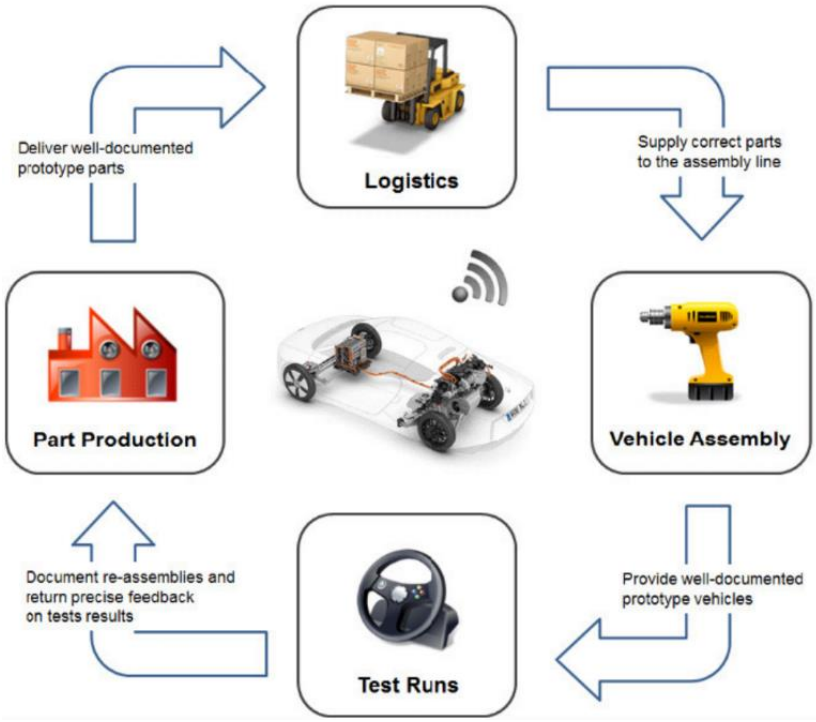


Fig. 1. Intelligent logistics transportation.

B. Indicators and Input-Output Ratios

In conducting the energy use efficiency assessment, the results of previous research were taken into account, and a system of assessment indicators was developed to analyze energy use in Chinese cities. These assessment indicators cover three main input variables: labor, capital, and energy. When selecting the type of energy, the current energy structure of the cities was fully considered, with particular attention paid to the total amount of natural gas, gas, and liquefied natural gas (LNG) delivered, as well as the electricity consumption of the whole society. In order to accurately estimate the total energy consumption, the standard carbon dioxide conversion rates of various energy sources were used as reference standards. In

order to eliminate the influence of price distortions on the results, the real GDP of each city was adjusted to match its expected economic output, thus ensuring a more objective and accurate assessment.

Industrial waste and greenhouse gas emissions have also been introduced as indicators of adverse impacts to enhance the reliability and accuracy of the assessment results. These indicators include industrial waste emissions such as sulfur dioxide, water, and soot and greenhouse gas emissions such as carbon dioxide. These data were mainly obtained from the Annual Statistical Reports of Chinese Cities, and the vacuum coefficients were estimated using linear interpolation. In calculating carbon dioxide emissions, reference was made to the calculation methods of other countries, specifically, the carbon dioxide emission coefficients published by the Natural Gas Commission in 2006 and the emission coefficients of the power grids of six regions in China, which were used. These comprehensive data and methods enable a more comprehensive assessment of the energy utilization efficiency of Chinese cities and provide valuable references for policymaking and urban development. The flow of the open and closed logistics system is shown in Fig. 2.

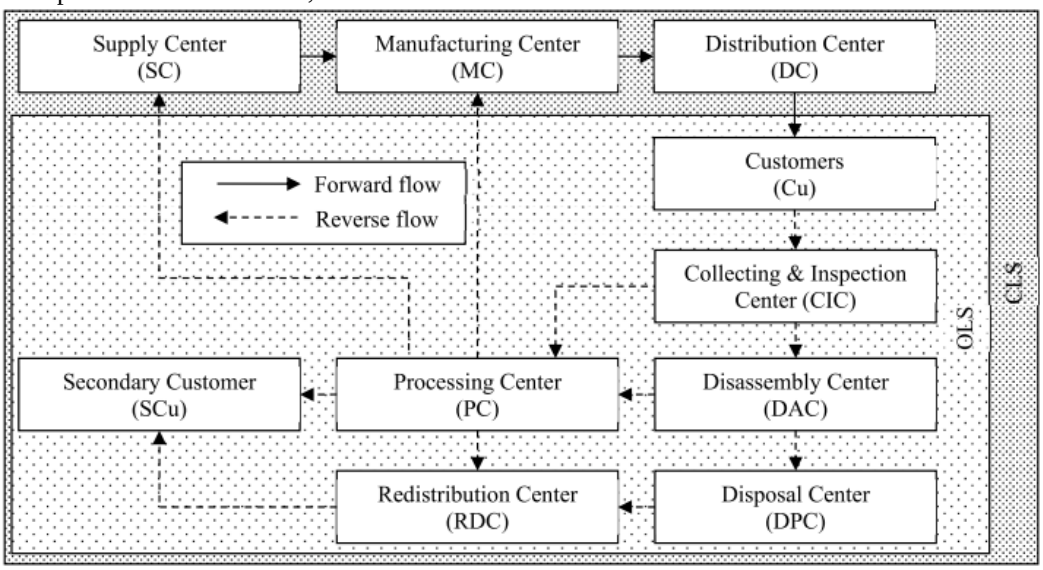


Fig. 2. Flow of open and closed logistics system.

A data-driven methodology model for energy use efficiency:

$$\pi = p(q_0 + \eta_1 e_1 + \eta_2 e_2) - \alpha e_1^2 - \frac{\beta}{e_1} e_2^2 \quad (1)$$

π in Eq. (1) is the unknown variable for efficiency improvement;

$$e_1^{1^*} = \frac{1}{2\alpha} p\eta_1 + K(1-D) \quad (2)$$

(1-D) in Eq. (2) is to get the actual distance of TOPSIS.

$$e_2^{1^*} = e_1^{1^*} \frac{p\eta_2 + KD + \theta(1+g)}{2\beta} \quad (3)$$

$(1+g)$ Eq. (3) is the coefficient of θ.

Analyzing the diversity of urban agglomerations is critical to understanding urban energy use. In this study, we consider different city sizes and delve into the impact of the digital economy on green energy utilization. The results show that the level of digital economy in medium and large cities contributes significantly to green energy utilization, being able to increase green energy utilization by 1%. However, differences in energy efficiency between different groups of cities were also identified. This triggered a further examination of energy efficiency between groups to assess its impact on environmental performance more fully. Although the model's p-value of 0.0681 fails at all levels of statistical significance, it passes the 10% group rate test, suggesting some unevenness in the impact of direct investment on environmental protection across different city sizes. We conducted a heterogeneity test to compare the regressors and obtained a high clustering value (0.594). This indicates significant differences between city groups, especially regarding battery eco-efficiency. Small city clusters have a higher impact on battery eco-efficiency in large cities (0.274), while small and medium-sized cities have a lower impact. This phenomenon can be explained by the fact that large cities have more developmental advantages, including geographic location, transportation infrastructure, technological innovation, and human resources, and are therefore more likely to achieve high efficiency in energy use.

Multiple dimensions characterize the indirect impact of the digital economy on urban green energy efficiency. First, the dynamic development of the digital economy promotes the rapid advancement and broad application of technological innovation, improving the efficiency of green energy utilization. Second, the rise of the digital economy has changed the industrial structure, promoted the rise and development of green industries, and indirectly improved the green energy efficiency of cities. The theoretical analysis identifies technological innovation, industrial structure, and rational industrial structure as the primary mechanisms by which the digital economy affects urban

green energy efficiency. Recent studies have shown that technological innovation is critical in improving energy efficiency and creating more favorable conditions for the innovation and application of low-carbon energy technologies. Therefore, the impact of the digital economy on urban green energy efficiency involves technological innovation, adjustment and optimization of industrial structure, and inherent economic mechanisms. Empirical evidence of energy utilization efficiency improvement in intelligent logistics is shown in Table II.

TABLE II. EMPIRICAL EVIDENCE OF ENERGY UTILIZATION EFFICIENCY IMPROVEMENT IN THE CONTEXT OF INTELLIGENT LOGISTICS

	(1)*Ee*	(2)*Tfee*	(3)*Ee*	(4)*Tfee*
Treat×post	0.5817***	0.0514***	0.517***	0.0514**
Cons	0.871***	1.214***	11.214***	1.251***
Year	Y	Y	Y	Y
Province	Y	Y	Y	Y
Control	Y	Y	Y	Y
N	251	251	251	251
R^2	0.627	0.817	0.817	0.985

IV. RESULTS AND DISCUSSION

A. Discussion of Empirical Results on Green Total Factor Energy Efficiency (GTFE)

China faces energy and environmental challenges, including lower energy efficiency and environmental quality. In the context of green development, accelerating the development of the digital economy is critical to China's modernization and improvement of green energy efficiency. The rise of the digital economy provides new opportunities and impetus for China to realize its green transformation. On the energy front, China is developing traditional green energy production technologies, such as solar and wind power, and promoting low-carbon oil extraction technologies to reduce emissions from traditional energy use. The development of intelligent transportation is also vital, with digital technologies enabling the intelligent management and optimization of transportation systems, reducing traffic congestion and pollution, and improving the efficiency of urban energy use [19].

Regarding the environment, China uses digital technology to build a digital environmental management platform to monitor environmental conditions in real-time and provide accurate early warning and rapid response to environmental problems, thereby controlling the behavior of high-energy-consuming and high-emission enterprises and promoting sustainable environmental development. Digital means can realize comprehensive monitoring and analysis of environmental data, providing a scientific basis and technical support for environmental governance. China also needs to actively promote the development of the digital economy, accelerate the integration of digital technology with traditional industries, and especially promote the digital transformation of traditional industries. This process requires focusing on developing advanced industries, such as the modern service industry, and accelerating the intelligent development of traditional industries. Integrating digital technologies with traditional industries can help promote the digital transformation of industrial structures, improve the green energy efficiency of cities, realize the green and low-carbon transformation of cities, and inject new vitality into China's sustainable development.

Cities rich in natural resources, although endowed with abundant natural resources, have a relatively low level of economic development, a unique industrial structure, and a high degree of dependence on natural resources. Over-exploitation of resources may not only lead to resource depletion but also cause damage to the environment, thus affecting the sustainable development of society. Therefore, there is an urgent need to promote the digital transformation of these resource-based cities into resource-saving cities and actively explore helpful regional resources. These resource-centric cities are improving resource efficiency and driving the digital revolution. In general, the industrial structure of resource-intensive cities is dominated by resource-saving industries. However, these resource-centric cities should seize the significant opportunity to develop a digital economy and promote the upgrading and transformation of traditional industries into new digital industries to unleash the potential of the digital economy while reducing energy consumption. Realizing the digital transformation of cities requires an overall improvement in green energy efficiency. Some cities have steel as their primary industry, while others have wood as their primary industry. These resources could be used to establish innovative partnerships with other cities, which would not only help to promote economic development but also change the integrated industrial structure and facilitate the digital transformation of resource-efficient cities. Strengthening collaboration with resource-efficient cities to support their digital transformation is crucial. Resourceful cities often depend solely on natural resources and need more geographic and economic advantages, leading to a widespread brain drain. Technical cooperation is critical. Cities without natural resources are rich in technical resources and can provide technical support and guidance to accelerate the digital transformation of resource-saving cities. At the same time, resource-saving cities can also learn from the experience of non-resource-saving cities to find the most suitable path of digital transformation. The second is talent cooperation. Resource-poor cities have sufficient human resources to support the recruitment and training of resource-saving cities. At the same time, resource-saving cities can attract and train high-quality digital change experts to improve the efficiency and quality of digital change. Energy use efficiency data map.

Modernizing the energy mix and strengthening technological innovation are critical. Although China currently relies on traditional energy sources, such as coal, its reserves must be increased to meet demand. At the same time, more oil and natural gas inventories are required to meet market demand, and it relies heavily on imports, leading to an imbalance between supply and demand. The use of traditional energy sources generates large amounts of harmful gases, causing severe damage to the environment. Under the guidance of the green development concept, strengthening technological innovation and promoting energy conversion and modernization are vital ways to achieve the goals of dual-carbon and high-quality economic development. To this end, it is necessary to strengthen energy technology innovation, optimize energy structure, reduce

traditional energy consumption, and improve energy efficiency. In particular, it should focus on developing traditional innovative energy exploration technologies, new energy technologies, clean energy technologies, and digital energy system technologies, as well as the construction of digital energy systems to gradually promote energy structure transformation. In addition to technological innovation, policy support and financial investment are equally crucial. The government should increase investment in R&D in the energy sector and provide sufficient funds for technicians to enhance the vigor of technological innovation in enterprises and promote the transition to clean and low-carbon energy, as shown in Fig. 3.

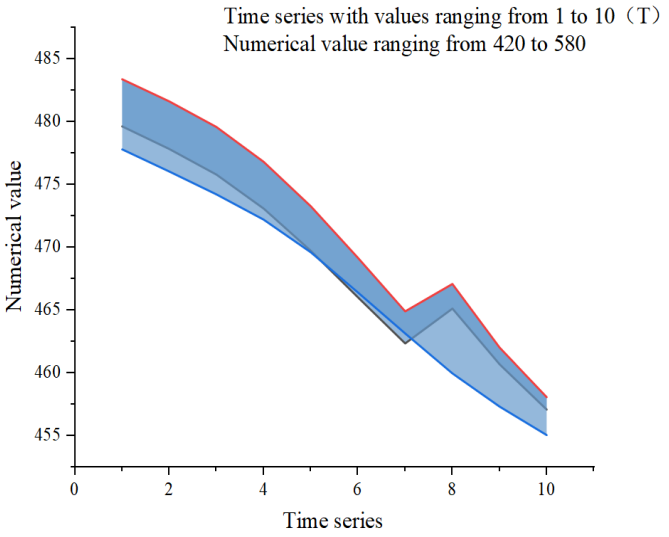


Fig. 3. Energy use efficiency data.

On the other hand, the government should pay attention to regions with high carbon consumption and provide appropriate policy incentives to promote the promotion and application of clean energy and new energy technologies to promote the green energy transition. Accelerating the modernization of the industrial structure and optimizing energy consumption is also crucial. Currently, the secondary industry occupies a larger share of China's economy, and its energy consumption per unit of GDP is much higher than that of the service sector, which means that industrial energy consumption is relatively high. However, the industry is the foundation of the real economy and the cornerstone of modernization, and reducing the proportion of the industry is not realistic. Improving industrial infrastructure and optimizing industrial energy use is the only way to improve environmental energy efficiency and promote sustainable urban development. In addition, according to relevant national laws and regulations, it is necessary to eliminate high-energy-consuming and high-polluting projects, eliminate excess capacity, rationally allocate social resources, and promote the green transformation of steel and other industries. Promoting the modernization of the energy structure reduces dependence on traditional energy sources, reduces environmental pollution, improves energy use efficiency, and lays a solid foundation for the sustainable development of China's economy and the realization of the dual-carbon goal [20].

B. Time-Space Evolution of Energy Use Efficiency

Based on the SBM super-efficiency model and MaxDE7 super software, the green energy efficiency of cities was calculated from 2011 to 2019. The study aims to explore the regional differences in the overall eco-energy efficiency of cities by analyzing them from a temporal and spatial perspective. During the study period, China's average energy efficiency was 0.508, indicating that China still has the potential to improve energy efficiency. Much work remains to be done to realize the goal of "dual-carbon" and sustainable green development of society. The temporal characteristics of the development of the national green energy efficiency coefficient show an increasing volatility trend. 2011-2016 is considered a period of energy efficiency improvement, 2016-2017 a period of energy efficiency decline, and 2016-2019 a period of energy efficiency improvement. 2016 marks the beginning of the 13th Five-Year Plan, which aims to reduce individual energy consumption and manage overall energy consumption. Although there is a time lag between policy liberalization and implementation, it will take time for regional industrial reforms and modernization to become established, which will reduce overall environmental impacts in the short term. As policies are implemented and local development adapts, the Green TFE will gradually improve its energy efficiency after 2017. To further improve China's energy efficiency, attention needs to be paid to the research, development, and application of green energy technologies, increasing the utilization rate of renewable energy and reducing dependence on traditional energy sources; optimizing the energy structure, encouraging clean energy to replace traditional energy sources, and improving the efficiency of energy use; and strengthening energy management, setting up a comprehensive energy monitoring and evaluation system, promoting changes in energy consumption patterns, and promoting green and low-carbon development. The government should introduce more robust policy measures to guide and support enterprises and residents in saving energy and reducing emissions and jointly promote improving China's energy efficiency level. Energy improvement efficiency data map, as shown in Fig. 4.

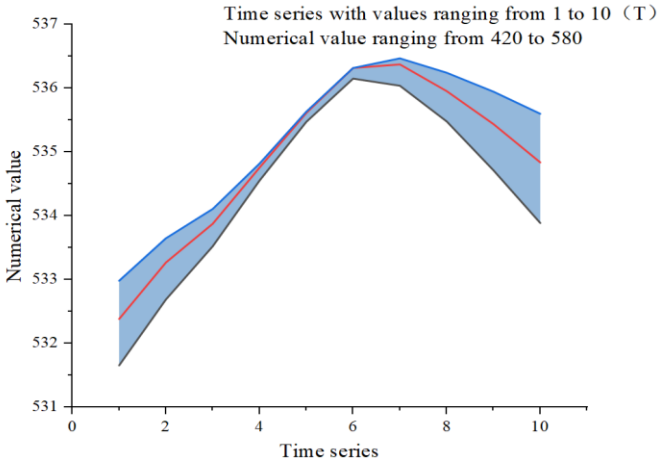


Fig. 4. Energy improvement efficiency data.

Analyzed from the perspective of control variables, the level of economic growth has a significant positive impact on all aspects of environmental effects at a significance level of 1%. Further, economic growth also attracts more investment, which

increases enterprises' production capacity and competitiveness and promotes technological innovation and the application of environmental management technologies. The capital mobilization of enterprises can be used not only to promote technological innovation but also to improve the production process and enhance resource utilization efficiency. On the other hand, population density (POP) significantly negatively impacted the overall ecological efficiency (significance level of 1%), indicating that high population density may lead to higher energy consumption and more severe pollution problems, thus hindering the improvement of ecological efficiency. Therefore, densely populated areas must adopt more effective environmental management and resource utilization measures to reduce environmental pressure and improve ecological performance. This finding underscores the positive impact of economic growth on environmental and energy efficiency, as economic progress helps to provide more resources for technological upgrading and industrial modernization, which in turn leads to the adoption of cleaner sources of energy, reduced use of traditional energy sources and pollution, and effective improvement of infrastructure. This further highlights the positive correlation between economic development and environmental protection and the mutually reinforcing relationship between economic prosperity and environmental sustainability. Infrastructure quality has a significant positive impact on the overall energy efficiency of green energy, with better quality infrastructure reducing transportation costs, improving energy use efficiency, and promoting technological advances that increase energy efficiency. Therefore, building high-quality infrastructure is one of the most important ways to improve environmental energy efficiency, which is of positive significance for promoting economic growth, improving the quality of life, and environmental protection.

Intelligent logistics construction principles and processes:

$$\pi_m = p_1(q_0 + \eta_1 e_1 + \eta_2 e_2) + K - \alpha e_1^2 \quad (4)$$

Thein Eq. (4) is a free variable introduced to guarantee the error value.

$$\pi_m^{2*} = p_2 q_0 + \frac{(p_2\eta_2 + \theta(1+g))^2}{4\beta} e_1^{2*} \quad (5)$$

The $p_2 q_0$ in Eq. (5) is the residual term of the least squares method.

Excellent infrastructure plays a significant role in reducing the transport coefficient, which helps optimize resource allocation and improve overall productivity and environmental sustainability. The entropy method was used to explain the potential regressors of economic levels, enabling the most influential variables to be extracted from many economic indicators, thus validating the stability of the model. The results show that PCADI significantly impacts green energy efficiency, with 0.00335 units per increase, which is significant for low-level regression, indicating the model's reliability. In addition, the study introduced foreign direct investment (FDI) as a new control variable to avoid the possible impact of omitted variables on the empirical results. The analysis results show that the level of the digital economy significantly negatively impacts overall eco-efficiency, implying that technological development is only sometimes beneficial to eco-efficiency.

Meanwhile, FDI negatively impacts overall energy efficiency, especially when developed countries transfer highly polluting and energy-intensive industries to developing countries. To validate these findings, static panel data were used to test the regression model's robustness and investigate possible endogeneity issues. Local government policies and interventions significantly impact environmental energy efficiency in a given developing urban setting, suggesting that small changes in government policies may significantly impact overall energy efficiency. The p-value of 0.0092 for the two direct investments suggests that they may significantly impact energy efficiency. The analysis of the differences between the different factors shows that direct investment's impact on the urban environment varies between developed and developing countries.

Regarding recovery factors, developed cities have a significant positive impact on energy efficiency because they have advantageous resources such as high levels of digitalization, economic development, and technological innovation. Therefore, implementing environmental management strategies and the increased use of clean energy will help improve these cities' energy efficiency and promote their sustainable development. The weight of each type of energy is shown in Fig. 5.

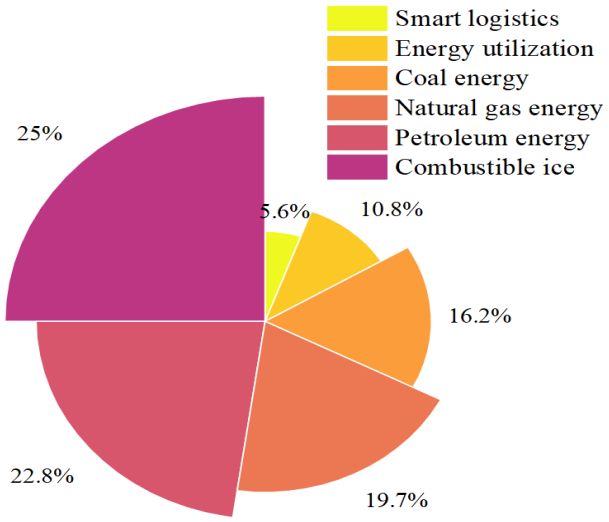


Fig. 5. Share of various energy sources.

In this study, 272 heterogeneous samples were grouped. Although the effect of the inhaler on the energy efficiency of the green factor was insignificant, there was a significant difference between the two groups. In order to ensure the reliability of the results of the different studies, the researchers delved into the factors that differed between the two groups. This showed that the model had passed the 5% group gap test, again validating the inconsistency of the impact of non-resource direct investment on green energy efficiency. The diversity recession factor reveals that direct investment significantly impacts the overall environmental energy efficiency of cities that do not consume large amounts of resources. This is because resource-efficient

cities, despite having significant resources, are mainly made up of less developed cities with less digitization—global energy consumption shares, as shown in Fig. 6.

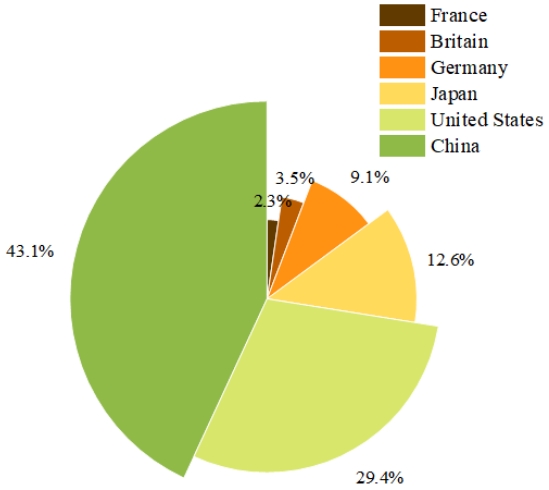


Fig. 6. Global energy consumption shares.

Over time, the industry has become the engine of economic development in resource-intensive cities; in the process of industrialization, although their economies are booming, these cities consume large amounts of national resources and bear environmental costs that do not contribute to sustainable development and more efficient use of society's resources. Resource-efficient cities are inherently less affluent than non-resource-efficient cities that rely on energy supplies or imports for economic development and daily life. However, in resource-scarce cities, industries change relatively quickly. While other cities rely on primary and secondary industries to drive GDP growth, some, such as Beijing, Shanghai, and Shenzhen, have shifted from primary and secondary industries to more environmentally friendly ones. These high-industry-consuming cities consume less energy, produce less waste and CO2 emissions, and have reduced energy waste and emissions in their cities. At the same time, the high degree of digitization, digitalization, and rapid digitization of natural resource-based cities has improved the eco-energy efficiency of various factors.

A comparison of domestic energy efficiency is shown in Fig. 7.

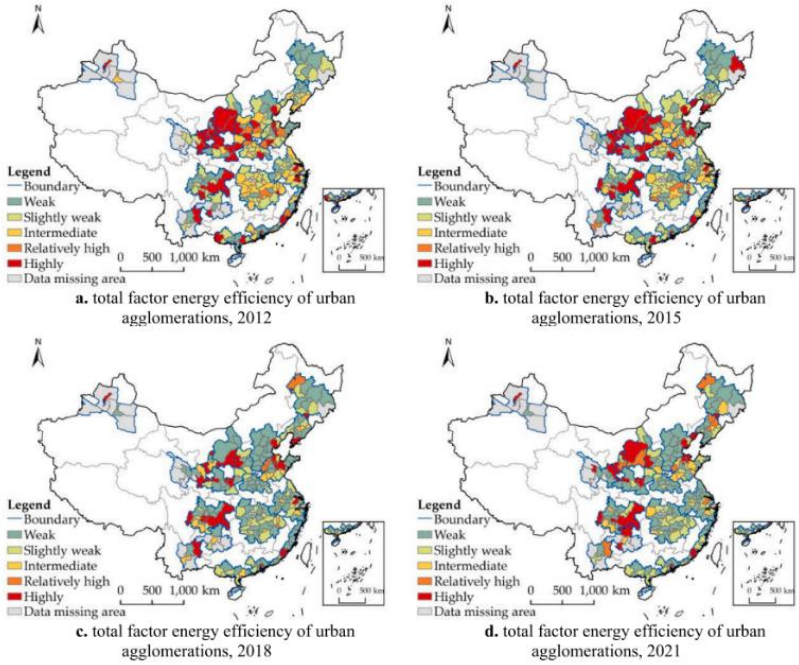


Fig. 7. Comparison of domestic energy efficiency.

V. Conclusion

Intelligent logistics is a current trend in the logistics industry, and one of its core objectives is to enhance energy utilization efficiency to reduce costs, minimize resource consumption, and improve the environment. This study systematically explores the feasibility and effectiveness of improving energy utilization efficiency through a data-driven approach in the context of global intelligence in various industries. The following conclusions are drawn from the introduction, purpose, methodology, and results: First, the large amount of accumulated data in the intelligent logistics environment becomes a valuable resource for improving energy utilization efficiency. Through the collection and analysis of logistics transportation information, energy consumption data, and other data, the operational status and energy utilization of the optimized logistics management procedures are deeply understood, providing a solid foundation for subsequent optimization. Secondly, key factors affecting energy utilization efficiency and potential optimization directions are revealed through data analysis and mining techniques. For example, optimizing transportation routes and improving vehicle load factors are essential in improving energy utilization efficiency. These findings provide theoretical guidance for the development of targeted optimization strategies. A series of data-driven

optimization strategies and methods were designed at the methodological level by incorporating intelligent optimization algorithms. These methods not only consider the actual situation of optimizing logistics management procedures but also make full use of the advantages of big data and artificial intelligence technology to improve the precision and operability of the optimization effect. The effectiveness and feasibility of the proposed methods are verified through simulation experiments and case studies. The experimental results show that adopting the data-driven approach can significantly improve the energy utilization efficiency of the optimized logistics management procedure, reduce logistics costs, and improve the sustainability and competitiveness of the system.

In summary, through in-depth analysis and empirical research, this study demonstrates the feasibility and effectiveness of using data-driven methods to improve energy utilization efficiency in the context of global intelligence in various industries. The proposed series of optimization strategies provides new ideas and methods for energy management and optimization of optimized logistics management processes. These results significantly promote the sustainable development of optimized logistics management procedures and enhance overall competitiveness. In future research, optimizing other aspects of optimized logistics management procedures, such as environmental impact assessment and supply chain visualization, can be further explored to achieve the overall optimization and sustainable development of optimized logistics management procedures.

REFERENCES

[1] Sarker, I. H. (2022). Smart City Data Science: Towards data-driven smart cities with open research issues. Internet of Things, 19, 100528. https://doi.org/10.1016/j.iot.2022.100528

[2] Ahmad, T., Zhang, D., Huang, C., Zhang, H., Dai, N., Song, Y., & Chen, H. (2021). Artificial intelligence in the sustainable energy industry: Status Quo, challenges, and opportunities. Journal of Cleaner Production, p. 289, 125834. https://doi.org/10.1016/j.jclepro.2021.125834

[3] Tambare, P., Meshram, C., Lee, C.-C., Ramteke, R. J., & Imoize, A. L. (2021). Performance measurement system and quality management in data-driven Industry 4.0: A review. Sensors, 22(1), 224. https://doi.org/10.3390/s22010224

[4] Sarmas, E., Marinakis, V., & Doukas, H. (2022). A data-driven multicriteria decision-making tool for assessing investments in energy efficiency. Operational Research, 22(5), 5597–5616. https://doi.org/10.1007/s12351-022-00727-9

[5] Ohalete, N. C., Aderibigbe, A. O., Ani, E. C., Ohenhen, P. E., Daraojimba, D. O., & Odulaja, B. A. (2023). AI-driven solutions in renewable energy: A review of data science applications in solar and wind energy optimization. World Journal of Advanced Research and Reviews, 20(3), 401–417. https://doi.org/10.30574/wjarr.2023.20.3.2433

[6] Tan, J., Xie, S., Wu, W., Qin, P., & Ouyang, T. (2022). Based on a data-driven approach, evaluating and optimizing the cold energy efficiency of power generation and wastewater treatment in LNG-fired power plants. Journal of Cleaner Production, 334, 130149. https://doi.org/10.1016/j.jclepro.2021.130149

[7] Mohapatra, S. K., Mishra, S., Tripathy, H. K., Bhoi, A. K., & Barsocchi, P. (2021). Using predictive intelligence approaches, a pragmatic investigation of energy consumption and utilization models in the urban sector. *Energies*, *14*(13), 3900. https://doi.org/10.3390/en14133900

[8] Akhtar, S., Sujod, M. Z. B., & Rizvi, S. S. H. (2022). An intelligent data-driven approach for electrical energy load management using machine learning algorithms. *Energies*, *15*(15), 5742. https://doi.org/10.3390/en15155742

[9] Majeed, A., & Hwang, S. O. (2021). Data-driven analytics leveraging artificial intelligence in the era of COVID-19: An insightful review of recent developments. *Symmetry*, *14*(1), 16. https://doi.org/10.3390/sym14010016

[10] Ala'raj, M., Radi, M., Abbod, M. F., Majdalawieh, M., & Parodi, M. (2022). Data-driven based HVAC optimization approaches A systematic literature review. *Journal of Building Engineering*, *46*, 103678. https://doi.org/10.1016/j.jobe.2021.103678

[11] Ma, S., Zhang, Y., Lv, J., Ren, S., Yang, H., & Wang, C. (2022). Data-driven cleaner production strategy for energy-intensive manufacturing industries: Southern and Northern China case studies. *Advanced Engineering Informatics*, p. *53*, 101684. https://doi.org/10.1016/j.aei.2022.101684

[12] Alrashidi, M., Alrashidi, M., & Rahman, S. (2021). Global solar radiation prediction: Application of novel hybrid data-driven model. *Applied Soft Computing*, *p. 112*, 107768. https://doi.org/10.1016/j.asoc.2021.107768

[13] Lazaroiu, G., Androniceanu, A., Grecu, I., Grecu, G., & Neguriță, O. (2022). Artificial intelligence-based decision-making algorithms, IoT sensing networks, and sustainable cyber-physical management systems in big data-driven cognitive manufacturing. *Oeconomia Copernicana*, *13*(4), 1047–1080. https://doi.org/10.24136/oc.2022.030

[14] Kahraman, A., Kantardzic, M., Kahraman, M. M., & Kotan, M. (2021). A data-driven multi-regime approach for predicting energy consumption. *Energies*, *14*(20), 6763. https://doi.org/10.3390/en14206763

[15] Zheng, Z., Wang, F., Gong, G., Yang, H., & Han, D. (2023). Intelligent technologies for construction machinery using data-driven methods. *Automation in Construction*, *147*, 104711. https://doi.org/10.1016/j.autcon.2022.104711

[16] Bachmann, N., Tripathi, S., Brunner, M., & Jodlbauer, H. (2022). The contribution of data-driven technologies in achieving sustainable development goals. *Sustainability*, *14*(5), 2497. https://doi.org/10.3390/su14052497

[17] Strielkowski, W., Vlasov, A., Selivanov, K., Muraviev, K., & Shakhnov, V. (2023). Prospects and challenges of the machine learning and data-driven methods for the predictive analysis of power systems: A review. *Energies*, *16*(10), 4025. https://doi.org/10.3390/en16104025

[18] Bousdekis, A., Lepenioti, K., Apostolou, D., & Mentzas, G. (2021). A review of data-driven decision-making methods for industry 4.0 maintenance applications. *Electronicsweek*, *10*(7), 828.

[19] Bahramian, M., Dereli, R. K., Zhao, W., Giberti, M., & Casey, E. (2023). Data to intelligence: The role of data-driven models in wastewater treatment. *Expert Systems with Applications*, *p. 217*, 119453. https://doi.org/10.1016/j.eswa.2022.119453

[20] Bibri, S. E. (2021). Data-driven smart, sustainable cities of the future: An evidence synthesis approach to a comprehensive state-of-the-art literature review. *Sustainable Futures*, p. *3*, 100047. https://doi.org/10.1016/j.sftr.2021.100047

Design and Application of the DPC-K-Means Clustering Algorithm for Evaluation of English Teaching Proficiency

Mei Niu
Basic Courses Department, Jiyuan Vocational and Technical College, Jiyuan 459000, Henan, China

***Abstract*—Effective and precise methodologies for evaluating the proficiency in English language instruction are instrumental in enhancing educators' competencies and the effectiveness of educational administrative processes. The objective of this paper is to refine the neutrality and precision of such assessments by introducing a novel approach that leverages an advanced K-means algorithm in conjunction with convolutional neural networks (CNNs). Initially, a thorough examination of the issue at hand leads to the formulation of an assessment framework that integrates both a clustering algorithm and a CNN, with a comprehensive elucidation of the pivotal technical aspects. Subsequently, the paper introduces a data clustering and categorization technique grounded in the DPC-K-means methodology, specifically tailored for indices that measure English teaching proficiency, and employs CNNs to devise a model for evaluating these competencies. The integration of these two components—data clustering and the assessment model—gives rise to an innovative technique. Ultimately, the proposed method is implemented and its practicality is substantiated through an analysis of empirical data from educators' teaching proficiency indices. A comparative analysis with existing algorithms reveals that the proposed method achieves superior clustering performance and the lowest margin of error in predictive assessments.**

***Keywords*—K-Means; density-peak clustering algorithm; ELT competency assessment; convolutional neural network**

I. Introduction

As educational reforms progress and evolve, the caliber of education has emerged as a pivotal topic of societal interest [1]. Proficiency in teaching represents a critical facet of educational quality, reflecting not just the professional acumen of educators but also the trajectory of educational management's advancement [2]. English, being the lingua franca of the modern world, occupies a significant place among the mandatory courses within the academic curricula of higher education institutions [3]. The evaluation of English teaching capabilities is a crucial component in the pedagogical framework of universities and colleges, serving as a vital mechanism for educators to gather feedback, refine their instructional strategies, and uphold educational standards, as well as for students to refine their study approaches, enhance learning techniques, and boost academic performance [4]. In recent years, research into the assessment of English language teaching proficiency has predominantly focused on two key areas: the development of assessment indices and the formulation of algorithms for evaluating teaching proficiency [5]. Existing systems for evaluating English teaching skills are often examined from dual viewpoints—that of the educator and the student—yet they often fall short in terms of being comprehensive, standardized, and providing timely feedback [6]. Current research into algorithms for assessing English teaching skills typically employs the assessment index system as input and the level of proficiency as output, with common methodologies including fuzzy logic [7], hierarchical analysis [8], clustering algorithms [9], and neural networks [10]. However, methods relying on fuzzy logic and hierarchical analysis are limited in their capacity to address intricate and varied mapping relationships and are generally straightforward to compute and implement [11]. On the other hand, clustering-based algorithms for assessing English teaching skills offer a temporal perspective on the assessment challenge, addressing the dynamic nature of teaching data and establishing a big data model for English teaching proficiency through quantitative recursive analysis [12]. Neural network-driven algorithms, meanwhile, are adept at managing the intricate and shifting relationships between the assessment indices and the proficiency levels, offering a degree of generalization [13]. With the advancement of intelligent algorithms, the integration of clustering methodologies with neural networks to cluster and train the indices of English teaching proficiency has become a burgeoning research avenue, aimed at achieving a quantifiable measure of teaching skills [14]. Despite these developments, there remain issues within the algorithms for assessing English teaching skills, including the need for improved generalization and the absence of well-defined quantitative criteria [14].

This paper addresses the aforementioned challenges in the construction of an English teaching proficiency assessment system and the design of its algorithms by integrating a clustering algorithm with a convolutional neural network. It proposes a method for developing an English teaching proficiency system grounded in an enhanced clustering approach and an assessment algorithm based on CNNs. Concentrating on the evaluation of English teaching skills, the paper delves into the conceptualization and resolution of the issue at hand, scrutinizing the critical technical aspects. It utilizes an improved clustering algorithm for the aggregation and synthesis of English teaching proficiency indices and employs a CNN to assimilate and refine the clustered data, ultimately verifying and analyzing the proposed methodology using a test dataset.

II. Programme Research Design

A. Line of Research

To address the problem of assessing English teaching ability, this paper follows the basic research idea of "identifying

problems - analysing problems - proposing solutions - verifying solutions" [15] to study the English teaching ability index system (see Fig. 1). The paper follows the basic research idea of "finding the problem - analysing the problem - proposing the solution - verifying the solution" [15] to study the indicator system of English teaching competence, and at the same time, it adopts the data learning model to construct and analyse the model for assessing English teaching competence.

By integrating a clustering methodology with the advanced computational framework of convolutional neural networks (CNNs), a comprehensive scheme for the evaluation of teaching capabilities has been meticulously crafted, as delineated in Fig. 1. As depicted in Fig. 1, this scheme for assessing English teaching proficiency, grounded in both clustering techniques and CNNs, adheres to the foundational principles of assessment system development. It initiates with a thorough examination of the challenges inherent in evaluating English teaching skills, followed by the establishment of pertinent assessment criteria. The data undergoes a rigorous pre-processing phase that includes the mitigation of outliers, the imputation of missing values, normalization to ensure consistency, and a thorough analysis of inter-variable correlations and the reduction of data dimensionality to enhance computational efficiency. Subsequently, the refined indices are aggregated and scrutinized through the application of a clustering algorithm, which facilitates the identification of distinct levels of teaching proficiency. The subsequent phase involves the utilization of a CNN to train a labeled dataset, thereby fostering the development of a robust model capable of accurately assessing English teaching competencies.

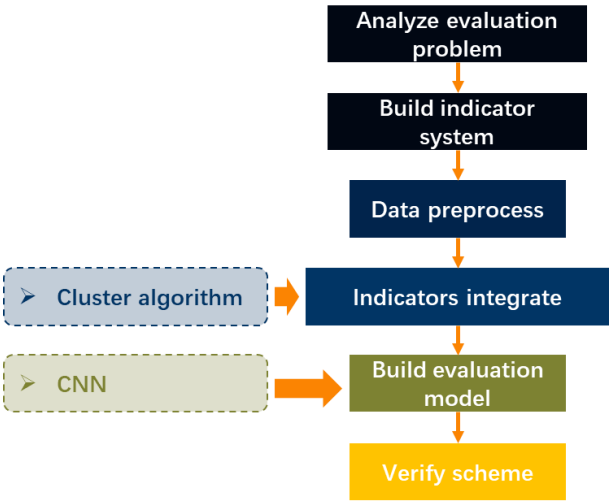


Fig. 1. ELT competency assessment research programme.

B. Analysis of Key Technologies

The system for evaluating English teaching ability, which integrates a clustering algorithm with a convolutional neural network, encompasses several pivotal technological components. These include the development of an assessment framework for English teaching proficiency, the initial handling of gathered data, the organization of data indices through clustering, the assembly of a model for gauging teaching capabilities, and the verification of the system's design, as illustrated in Fig. 2.

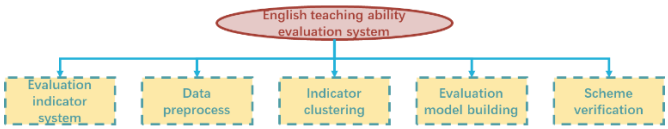


Fig. 2. Key technologies of the EFL assessment system.

1) Establishment of an assessment indicator system: According to the principles of objectivity, orientation, wholeness, operability and English characteristics [16] of the construction of assessment indicators (as shown in Fig. 3), through consultation, investigation and modification, we construct an assessment indicator system of English teaching competence that contains 20 indicators in five aspects, such as the purpose of teaching, the content of teaching, the language of teaching, the method of teaching, and the effect of teaching, as shown in Fig. 4.

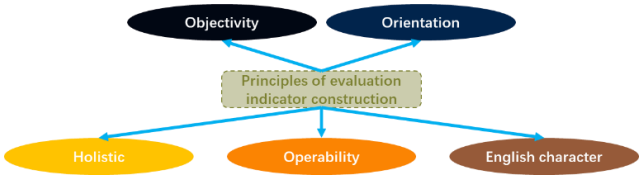


Fig. 3. Principles for selecting assessment indicators.

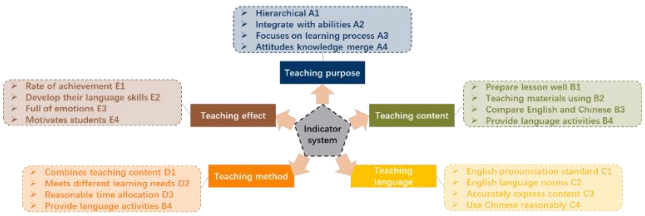


Fig. 4. System of assessment indicators.

2) Data pre-processing: The precision and reliability of the English Language Teaching (ELT) competency evaluation model are often compromised by the absence of comprehensive raw data, the presence of anomalies, discrepancies in scaling, and significant data redundancy, which result in the model's performance falling short of the expected benchmark [17]. To address these issues and enhance the fidelity and exactitude of the assessment model, it is imperative to implement data preprocessing techniques on the collected data. In dealing with anomalies, the paper employs the 3σ rule [18], which designates any data point that lies beyond the mean by more than three standard deviations as an outlier, subsequently eliminating such points. For addressing missing data, the paper utilizes a proximity filling technique [19], which estimates the missing entries based on adjacent data points. To standardize the quantitative index values, the *Z*-score normalization method [20] is applied. Furthermore, to refine the accuracy of the assessment algorithm and mitigate computational overhead, the paper utilizes Pearson's correlation coefficient to assess the interdependencies among the assessment indicators, subsequently employing principal component analysis [21] to streamline the dimensions of the indicators and distill the core variables that capture the most variance in the data.

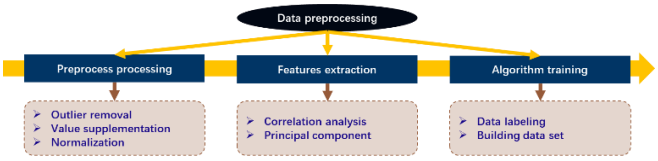


Fig. 5. Flow of data pre-processing technology.

3) Indicator clustering integration: In order to carefully classify the assessment values and grades of English teaching ability, this paper adopts the unsupervised learning method based on clustering algorithm to cluster and analyse the indicator data of English teaching ability assessment. Indicator data clustering integration mainly includes two steps, i.e., indicator data clustering and assessment score labelling division, as demonstrated in Fig. 5.

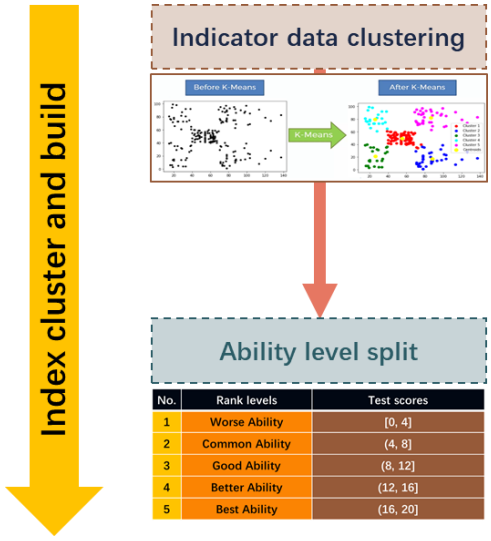


Fig. 6. Flow of indicator clustering integration technique.

For the indicator data clustering problem, this paper adopts the improved K-means clustering algorithm based on the assessment of English teaching ability indicator data clustering analysis, through the input of assessment indicator data, to represent the minimum relative distance between the sample point and the data sample point as the optimization goal, after many iterations of search, the output of clustering centre and the assessment of the indicator data division, the specific principle is displayed in Fig. 6 and Fig. 7. The data clustering model is calculated as follows:

$$X_{all-cluster} = f_{DPC-K-means}\left(X\right) \tag{1}$$

where, $X_{all-cluster}$ denotes the data segmentation results, $f_{DPC-K-means}$ denotes the improved K-means clustering algorithm, and X denotes the input indicator data.

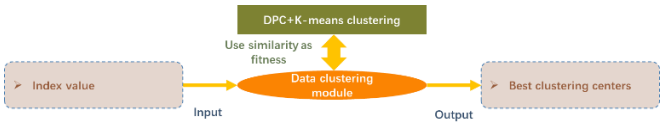


Fig. 7. Data clustering model for assessment indicators.

To address the issues of assessment score labelling and grade division, according to the clustering results, this paper divides the assessment value of English teaching ability into five grades, and the corresponding scores are shown in Table I.

TABLE I. CORRESPONDENCE BETWEEN ENGLISH PROFICIENCY ASSESSMENT SCORES AND CLASSIFICATION LEVELS

No.	Rank Levels	Test Scores
1	Worse	[0,4]
2	Common	(4,8]
3	Good	(8,12]
4	Better	(12,16]
5	Best	(16,20]

$$Y_{rank} = \begin{cases} 5 & Y_{score} \geq 16 \\ 4 & 12 < Y_{score} \leq 16 \\ 3 & 8 < Y_{score} \leq 12 \\ 2 & 4 < Y_{score} \leq 8 \\ 1 & Y_{score} \leq 4 \end{cases} \tag{2}$$

where, Y_{score} denotes the ELT assessment score and Y_{rank} denotes the ELT level.

4) Evaluation model construction: In order to construct the mapping relationship between the index value of English teaching ability and the assessment score, this paper adopts the convolutional neural network to construct the English teaching ability assessment model, as shown in Fig. 8.



Fig. 8. Schematic diagram of the construction of the EFL assessment model.

$$Y_{score} = f_{CNN}\left(X\right) \tag{3}$$

where, Y_{score} denotes the ELA score, f_{CNN} denotes the convolutional neural network, and X denotes the input indicator data.

5) Programme validation techniques: In order to fairly analyse the performance of each capability assessment model, this paper analyses and compares the clustering delineation capability and assessment prediction capability, and the specific technical ideas are shown in Fig. 9. The clustering delineation ability uses the evaluation indexes such as contour coefficient SI, variance ratio criterion CHI, and Davis-Boulding index DBI to analyse the results [22]. The assessment and prediction ability is used to analyse and compare the results using evaluation indexes such as mean absolute error MAE, root mean square error RMSE, mean absolute percentage error MAPE [23].

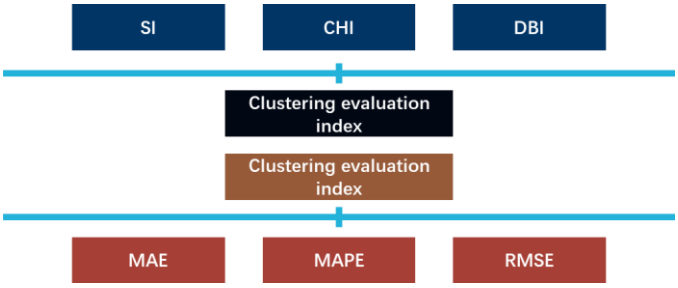


Fig. 9. Programme validation technology roadmap.

III. Data Clustering and Competency Assessment Algorithms

A. Clustering of English Teaching Competence Division

The clustering algorithm's function is pivotal in categorizing the vast and intricate dataset of English Language Teaching (ELT) proficiency indicators into coherent groups based on their inherent characteristics [23]. This process is fundamental to establishing the groundwork for the subsequent development of the ELT proficiency assessment model. The efficacy of the clustering process is paramount, as the accurate segregation of data is directly proportional to the quality of the sample set formation. Consequently, identifying a clustering algorithm with superior performance is of utmost importance to ensure the robustness of the model.

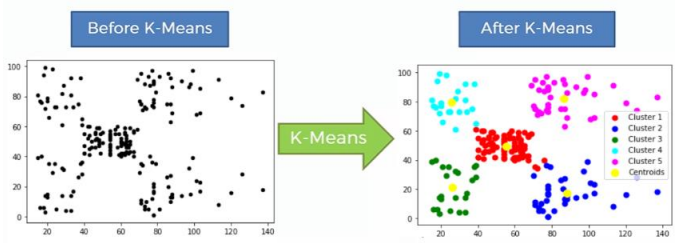


Fig. 10. Principle of K-means clustering algorithm.

1) K-means clustering algorithm: The K-means methodology stands as a prevalent technique for partitioning datasets into distinct clusters, particularly adept at classifying extensive volumes of data [24]. This algorithm operates by identifying the optimal positions for cluster centroids and assigning data points to the nearest of these, thereby minimizing an objective function that is predicated on the sum of squared differences. The objective is to maximize the distance between centroids while ensuring that each data point is linked to the closest centroid. Within the K-means framework, the Euclidean distance serves as the standard metric for gauging the similarity between data points, where a shorter distance signifies a higher degree of resemblance and a longer one suggests dissimilarity, as depicted in Fig. 10.

The objective function of the K-means algorithm is defined as follows:

$$J=\sum_{i=1}^{K}\left(\sum_{k}\left\|x_k-c_i\right\|^2\right) \tag{4}$$

where, K is the number of clusters, c_i is the centre of the cluster, and x_k is the k th data point in the i th cluster.

The algorithm proceeds as follows (see Fig. 11):

- Step 1: Determine the total number of categories K and randomly select K cluster category centres $C=(c_1,c_1,\cdots,c_K)$.
- Step 2: Compute the partition matrix. A data point belongs to the cluster whose centre is closest to that data point. Therefore, the clusters are represented by the binary division matrix U . U The elements in it are determined as follows:

$$u_{ij}=\begin{cases}1 & if\ \left\|x_j-c_i\right\|^2\le\left\|x_j-c_t\right\|^2,\forall t\ne i\\ 0 & otherwise\end{cases} \tag{5}$$

where u_{ij} indicates whether the j th data point belongs to the i th cluster class.

- Step 3: Update the cluster centres. Define each cluster class centre c_i that minimizes the objective function as follows:

$$c_i=\frac{\sum_{j=1}^{N}u_{ij}x_j}{\sum_{j=1}^{N}u_{ij}} \tag{6}$$

where, N denotes the number of samples.

- Step 4: Compute the objective function using equation (4). Verify that the function converges or the difference between two neighbouring values of the objective function is less than a given threshold and stop; otherwise repeat step 2.

2) DPC-K-means clustering algorithm: Given that the initial selection of cluster centroids in the K-means algorithm is arbitrary, there is a propensity for the algorithm to converge on a local rather than global optimum, potentially resulting in erroneous classification outcomes [25]. To counteract this, the present study introduces an enhanced version of the K-means algorithm, aimed at elevating the precision of the clustering results. This enhancement is achieved by integrating the Density Peaks Clustering (DPC) approach [26], which facilitates the identification of more accurate initial centroids for the K-means clustering process.

a) Peak density clustering algorithm: The Density Peak Clustering (DPC) algorithm is based on two assumptions: 1) The local density of the cluster centre is higher than that of its neighbourhood samples. 2) The cluster centre is farther away from other high local density samples. According to the above assumptions, given a dataset $X=\left[x_1,x_2,\cdots,x_N\right]$ with a

sample size of N ,the attribute (dimension) of each sample x_i is D , the DPC process is shown in Fig. 12. The clustering process of the standard DPC algorithm is divided into three main steps:

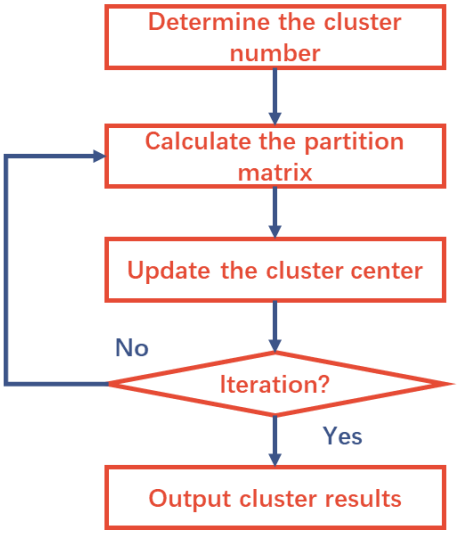


Fig. 11. K-means clustering algorithm.

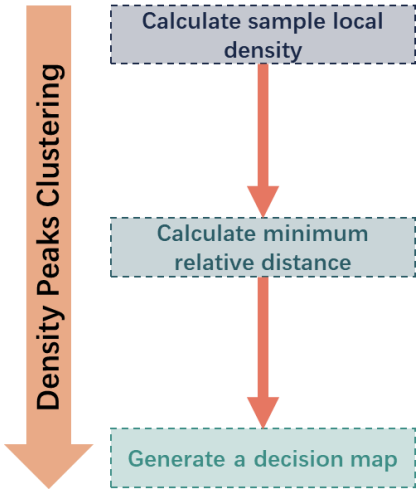


Fig. 12. DPC flowchart.

- Step 1: Calculate the local density of the sample point. For the sample point x_i , the local density ρ_i is described as follows:

$$\rho_i = \sum_j \chi(d_{ij} - d_c) \tag{7}$$

where, $\chi(x) = \begin{cases} 1, & x < 0 \\ 0, & x \geq 0 \end{cases}$. d_{ij} denotes the Euclidean distance between x_i and x_j . d_c denotes the stage distance, the exact value of which needs to be set manually.

- Step 2: Calculate the minimum relative distance. The minimum relative distance between the sample point x_i and any other sample point δ_i that has a higher local density than it is calculated as follows:

$$\delta_i = \begin{cases} \min(d_{ij}), & \rho_j > \rho_i \\ \max(d_{ij}), & \rho_j < \rho_i \end{cases} \tag{8}$$

where, if x_i is a sample point with a local or global maximum in the density, its relative distance δ_i is much larger than the relative distance of the neighbouring sample points.

- Step 3: Generate decision diagram. Select the sample points where local density ρ_i is large while the minimum distance δ_i is large as the cluster centre and generate a decision diagram with ρ_i as the x coordinates and δ_i as the y coordinates. After selecting the cluster centres in the decision diagram, the remaining sample points are assigned to different cluster classes based on the minimum distance.

b) Improvement strategies: To address the issue of clustering inaccuracies in the K-means algorithm that arise from the suboptimal identification of cluster centroids, the paper presents a novel adaptation of the K-means algorithm, underpinned by the Density Peaks Clustering (DPC) strategy, termed DPC-K-means. This refined algorithm is structured into two distinct phases: initially, the DPC algorithm is deployed to pinpoint the precise locations of cluster centroids within the raw dataset; subsequently, these centroids are integrated into the K-means algorithm, which then proceeds through iterative refinements to achieve a more nuanced clustering resolution.

Stage 1 is the prerequisite basis for ensuring the clustering accuracy. In order to confirm the initial clustering centre more accurately, the DPC algorithm is improved in this paper. The standard DPC algorithm is very subjective about the selection of d_c . And different d_c often leads to a large variability in the final results of clustering. Since the standard deviation can reflect the degree of dispersion of the data set, d_c is redefined as:

$$d_c = \omega \frac{\sqrt{\sum_{j=1}^{m} (\sigma_j / \mu_j)} \sum_{j=1}^{m} \mu_j (m-1)}{2m^2} \tag{9}$$

where, σ_j and μ_j are the standard deviation and mean of attribute j, respectively. $\omega \in (0,1]$ is the weight parameter that controls the size of the cutoff distance. d_c Approaching the 2-parameter of the standard deviation vector, which represents the standard deviation of all attributes, the cut-off distance is computed under the same mean criterion.

Facing multi-dimensional data, the metric error of standard Euclidean distance is large. Aiming at this problem, this paper designs a dynamic weight to correct the Euclidean distance in order to improve the final clustering accuracy. The basic design idea is: according to the magnitude of the difference between the corresponding attributes between two different data, matching different size weights. When the values of two attributes are more similar, the greater the proportion of in the overall similarity measure should be, and the greater the weights of should be assigned accordingly. The more distant the values of the two attributes are, the less weight should be assigned to the overall similarity measure, and therefore the less weight should be assigned accordingly. Dynamic Weight Correction Euclidean The exact form of the distance is as follows:

$$d_{ij}=\sqrt{\sum_{m=0}^{D}\frac{w_m}{W}\left(x_{im}-x_{jm}\right)^2} \tag{10}$$

where, $W=\sum_m w_m$ is the normalisation factor. w_m is the dynamic weights and its expression is

$$w_j=e^{-\left(\frac{R_m-R_{MIN}}{R_{MAX}-R_{MIN}}\right)} \tag{11}$$

In the Eq. (11), $R_m=\left|x_{im}-x_{jm}\right|$, $R_{MAX}=\max\sum_{m=0}^{D}\left(x_{im}-x_{jm}\right)$, $R_{MIN}=\min\sum_{m=0}^{D}\left(x_{im}-x_{jm}\right)$.

c) Algorithm flow: In summary, the flowchart of the DPC-K-means algorithm is shown in Fig. 13 and the overall process is:

- Step 1: Calculate the dynamic weight-corrected Euclidean distance d_{ij} ;
- Step 2: Calculate the stage distance d_c ;
- Step 3: Calculate the local density ρ_i and the relative minimum distance δ_i ;
- Step 4: Generate a decision diagram to identify the clustering centres K ;
- Step 5: Calculate the division matrix U ;
- Step 6: Update the clustering centre c_i ;
- Step 7: Compute the objective function. Verify that the function converges and no longer changes, then stop; otherwise repeat step 2.

3) Algorithmic applications: To enhance the precision of the model assessing English teaching capabilities, this study employs a clustering approach that segments the indicators of English teaching proficiency using the DPC-K-means methodology. The algorithm processes the input values representing the indicators of English teaching ability and yields both the categorized indicator data and the determined centroids of the clusters.

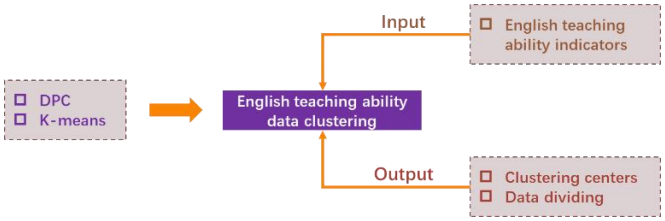


Fig. 13. Application of DPC-K-means clustering algorithm.

B. Algorithm for Assessing Competence in Teaching English

The assessment of English teaching ability is inherently a factual representation of the intrinsic laws governing teaching proficiency. To achieve autonomous capability evaluation, this study integrates convolutional neural networks to formulate a model for assessing English teaching ability.

1) Convolutional neural network: A CNN is an artificial neural network (structure as in Fig. 14) that consists of one or more convolution layers (convolution layer) [27], including two most important operations: convolution and pooling. The convolution layer uses the convolution operation instead of the matrix multiplication operation, which serves to detect the local connectivity of the features in the previous layer; the pooling layer serves to merge similar features [27].The features of a CNN include 1) sparse connectivity, 2) weight sharing, and 3) pooling. The convolution operation of the convolution layer is defined as follows:

$$z_j^{(l)}=\sum_{i=1}^{I}w_i^{(l)}a_{i+j-1}^{(l-1)}+b^{(l)} \tag{12}$$

In this context, *i* signifies the identifier for the convolutional kernel, while l serves as a marker for the active convolutional layer, with $l-1$ representing its preceding layer. The dimension of the convolutional kernel is denoted by I. Post-convolution operations, the resultant feature map is referred to as $z_j^{(l)}$, which is derived from the shared weights $w_i^{(l)}$, the activation outputs of the preceding layer $a_{i+j-1}^{(l-1)}$, and adjusted by the bias term $b^{(l)}$ associated with the convolutional layer

2) Algorithm application: In order to portray the mapping relationship between the constructed ELT competence assessment indicators and the assessment scores, this paper adopts CNN to construct the ELT competence assessment model, as shown in Fig. 18. The algorithm applies the input as the value of English teaching ability index and the output as the assessment score and ability level.

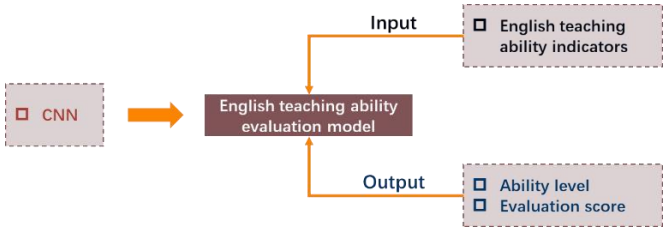


Fig. 14. CNN network application.

IV. Algorithmic Application Process Methodology

This manuscript introduces an innovative approach to evaluating the proficiency in English language instruction by fusing the DPC-K-means clustering technique with convolutional neural networks (CNNs), as depicted in Fig. 15. The methodology for appraising English teaching skills is delineated through the following sequential stages:

- Step 1: A thorough analysis of the English teaching proficiency assessment is conducted, identifying key features across five dimensions—objectives, content, language, methodology, and outcomes—to forge a comprehensive indicator system for evaluation purposes;
- Step 2: Data pertaining to English Language Teaching competencies are sourced through expert deliberation, literature synthesis, and surveys, followed by a meticulous preprocessing regimen involving the 3σ criterion, imputation of missing data via proximity methods, standardization through Z-score normalization, and dimensionality reduction via principal component analysis;
- Step 3: The DPC-K-means algorithm is harnessed to perform clustering on the meticulously preprocessed indicator data, thereby categorizing the data into distinct groups;
- Step 4: The CNN framework is then engaged to cultivate the correlation between the quantified indicators of English teaching proficiency and their corresponding assessment scores, effectively mapping the data to defined scoring echelons;
- Step 5: Validation analysis of the proposed ELT competency assessment method using the test set.

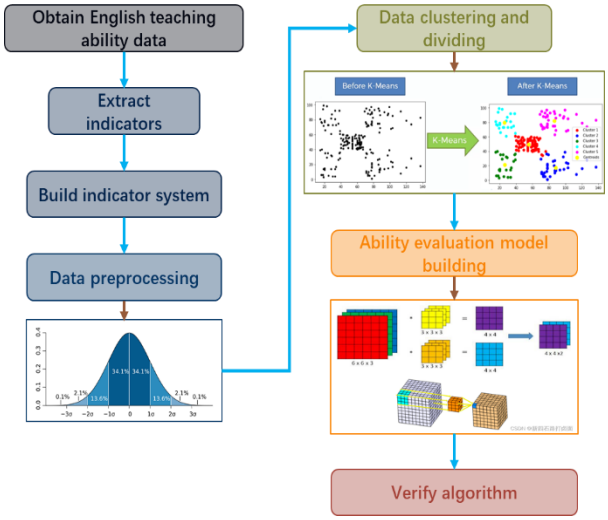


Fig. 15. Flowchart of algorithm application.

V. Experimental Analysis

A. Experimental Environment Design

To substantiate the practicality of the English teaching ability assessment method that integrates DPC-K-means and CNN algorithms, this study scrutinizes the data collected from English teaching index measurements. In this evaluation, a quintet of algorithms has been selected for comparative analysis. Table II enumerates the configuration parameters for these comparative algorithms, with the DPC-K-means-CNN being the algorithm posited by this paper, alongside others for comparative purposes.

TABLE II. Parameter Settings of Clustering Algorithm

Name	Composition method	Parameterisation
DPC-K-means-SVR	DPC-K-means SVR	SVR regularisation factor 1, kernel function is radial basis kernel function The number of K-means clusters is 5
DPC-K-means-RBM	DPC-K-means RBM	RBM with a learning rate of $10e^{-3}$ K-means clusters set to 5
DPC-K-means-ELM	DPC-K-means ELM	ELM featuring 50 nodes in the hidden layer K-means clusters set to 5
DPC-K-means-BP	DPC-K-means BP	Backpropagation (BP) training with 60 hidden layer nodes K-means clusters set to 5
K-means-CNN	K-means CNN	CNN with 100 nodes in the hidden layer network is 100, utilizing Adam's optimization K-means clusters set to 5
DPC-K-means-CNN	DPC-K-means CNN	CNN with 100 nodes in the hidden layer network is 100, employing Adam's method K-means clusters set to 5

The experimental simulation environment uses Matlab programming language and the system is Win10.

B. Parametric Analysis

The determination of clustering precision is significantly influenced by the selection of the number of cluster centers. With the aim of enhancing the accuracy of the model and accelerating the efficiency of the assessment process, this study delves into the impact of varying the number of clusters from two to ten on the model's accuracy, with the outcomes graphically represented in Fig. 16. A visual analysis of the figure reveals that at the threshold of five clusters, the DPC-K-means-CNN algorithm achieves the optimal precision in evaluating the capabilities of English language instruction Fig. 17 gives the results of the principal component analysis of the indicators of English proficiency assessment based on PCA technique. From

Fig. 17, when the indicator reaches 15, the cumulative contribution reaches 90 per cent.

C. Contribution of Indicators

An evaluation was conducted on the models DPC-K-means-SVR, DPC-K-means-RBM, DPC-K-means-ELM, DPC-K-means-BP, K-means-CNN, and DPC-K-means-CNN using a dataset comprising the profiles of 30 educators. The findings are detailed in Fig. 18 (a)-(f) and Tables III and IV. A review of Fig. 18 indicates that the method underpinned by the DPC-K-means-CNN algorithm outperforms others in terms of precision, with its assessment of English teaching proficiency closely mirroring the actual levels of competence.

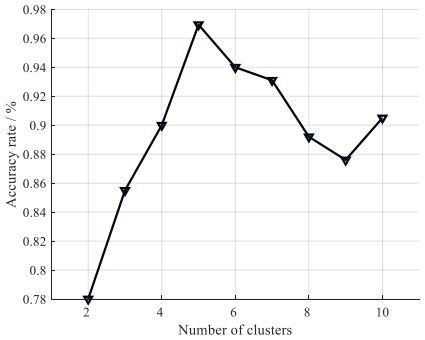


Fig. 16. Accuracy with different number of clustering centres.

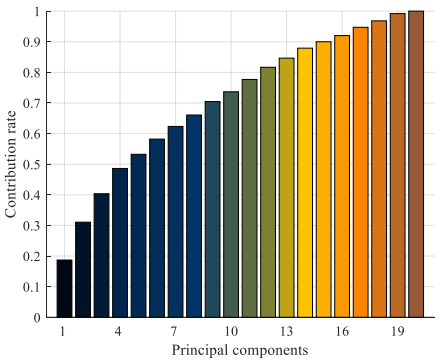
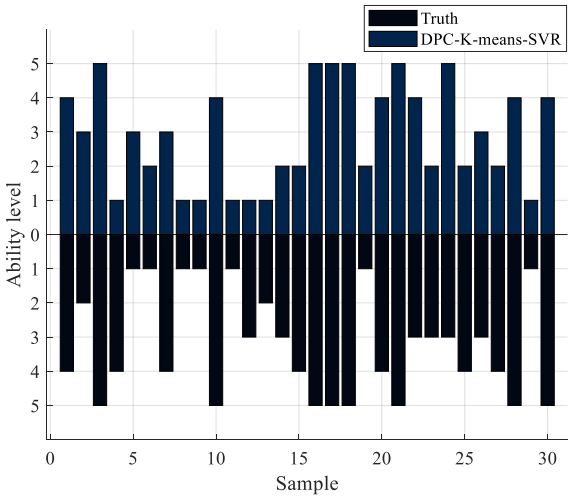
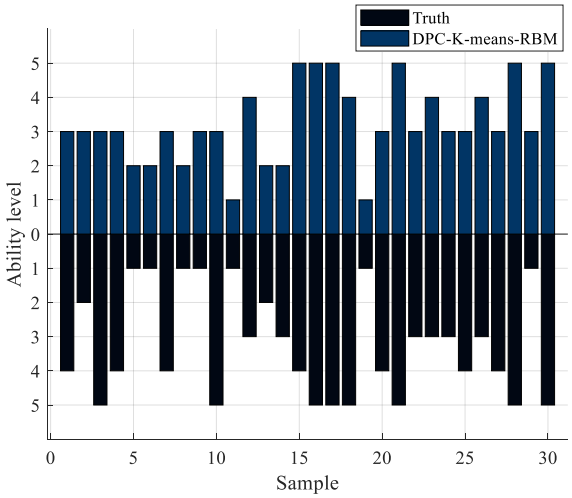


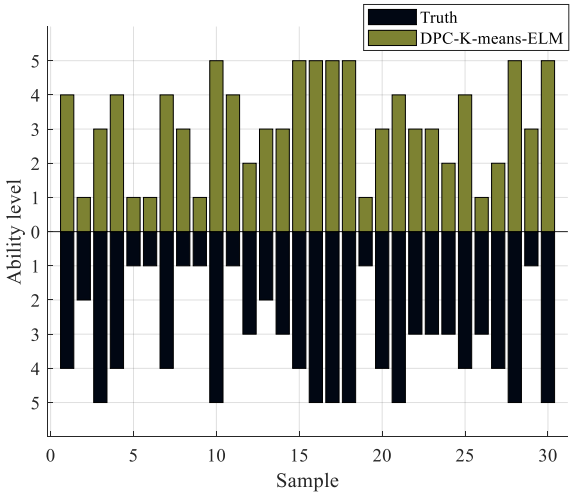
Fig. 17. Contribution of indicators.



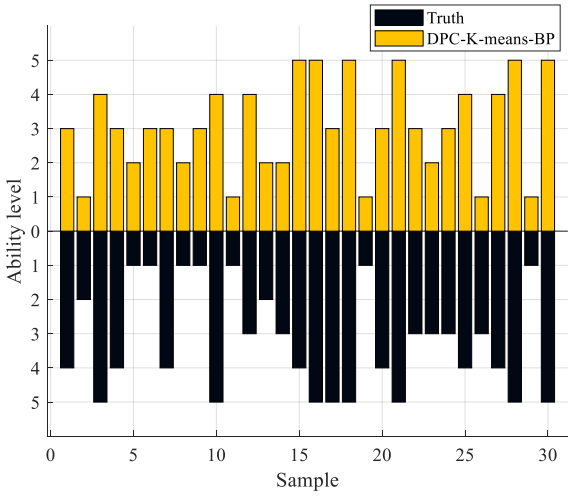
(a) DPC-K-means-SVR.



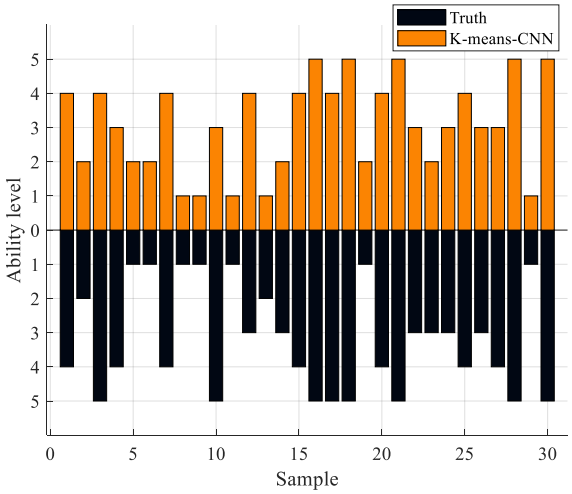
(b) DPC-K-means-RBM.



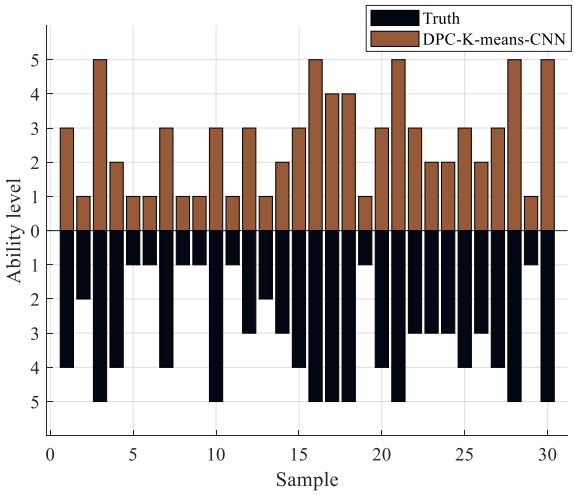
(c) DPC-K-means-ELM.



(d) DPC-K-means-BP.



(e) K-means-CNN.



(f) DPC-K-means-CNN.

Fig. 18. Comparison of assessment level results for different fusion assessment algorithms.

Upon examination of Table III, it is evident that the clustering of teaching ability data utilizing the DPC-K-means-CNN model yields the highest Silhouette Index (SI) value, with K-means-CNN, DPC-K-means-BP, DPC-K-means-RBM, DPC-K-means-ELM, and DPC-K-means-SVR trailing in sequence; concerning the CHI value, the DPC-K-means-CNN model demonstrates the most substantial ratio of inter-cluster to intra-cluster distance, peaking at 81.4, ahead of the DPC-K-means-BP, K-means-CNN, DPC-K-means-RBM, DPC-K-means-ELM, and DPC-K-means-SVR algorithms; regarding the Davies-Bouldin Index (DBI), the DPC-K-means-CNN algorithm exhibits the lowest average similarity score of 0.79 when comparing each cluster to its nearest cluster, outperforming alternative methodologies. A holistic assessment confirms that the DPC-K-means-CNN algorithm achieves superior clustering performance.

From Table IV, it can be seen that the DPC-K-means-CNN evaluates the best score prediction performance when analysed in terms of MAE, MAPE, and RMSE, with values of 0.0108, 2.7053e^{-04}, and 0.0122, respectively.

TABLE III. DATA CLUSTERING RESULTS

Arithmetic	SI	CHI	DBI
DPC-K-means-SVR	0.12	34.2	1.18
DPC-K-means-RBM	0.29	56.1	1.41
DPC-K-means-ELM	0.28	37.3	1.23
DPC-K-means-BP	0.31	78.3	1.21
K-means-CNN	0.37	76.1	0.83
DPC-K-means-CNN	0.40	81.4	0.79

TABLE IV. RESULTS OF THE CAPACITY ASSESSMENT

Arithmetic	MAE	MAPE	RMSE
DPC-K-means-SVR	0.0506	1.2650e-03	0.0589
DPC-K-means-RBM	0.0323	8.0814e-04	0.0388
DPC-K-means-ELM	0.0492	1.2300e-03	0.0590
DPC-K-means-BP	0.0390	9.7548e-04	0.0447
K-means-CNN	0.0248	6.2057e-04	0.0291
DPC-K-means-CNN	0.0108	2.7053e-04	0.0122

VI. CONCLUDING REMARKS

The pursuit of an intelligent and precise system for evaluating English teaching skills stands as a pivotal aspect of the reform in English education, with the dual benefit of elevating the instructional proficiency of teachers and solidifying the neutrality of the criteria used in educational administrative decisions. This paper advances an assessment methodology that integrates clustering algorithms with advanced recognition techniques, thereby enhancing both the precision and the objectivity of the evaluation process. By dissecting the complexities inherent in assessing English teaching capabilities, the paper formulates a strategy that leverages the strengths of data clustering and optimized training routines. It presents a scheme for evaluating English teaching ability that synergizes the DPC-K-means clustering approach with the analytical prowess of convolutional neural networks (CNNs), scrutinizing the critical technologies underpinning this method. The paper introduces a novel clustering and categorization technique grounded in the DPC-K-means algorithm, complemented by an assessment methodology that harnesses the predictive capabilities of CNNs for English teaching proficiency indicators. Utilizing a dataset comprising the teaching ability indicators of 30 educators, the paper validates the proposed method's feasibility and scrutinizes its performance through established clustering and predictive metrics. The analysis affirms the proposed method's preeminence in clustering efficacy, reflected in the superior values of SI, CHI, and DBI—0.40, 81.4, and 0.79 respectively. Furthermore, the method demonstrates the minutest error in predictive assessment, with MAE, MAPE, and RMSE values recorded at 0.0108, 2.7053e-04, and 0.0122 respectively.

REFERENCES

[1] Gu Z .The Evaluation Method of English Teaching Ability in View of Internet of Things[J].Wireless Communications and Mobile Computing, 2022.

[2] Wu X .Establishing an Operational Model of Rating Scale Construction for English Writing Assessment[J].English Language Teaching, 2022, 15.

[3] Wang G .Formative Assessment System of VR Teaching in English Translation Class[J].OALib, 2022.

[4] Zhang L , Li Q .Improved Collaborative Filtering Automatic Assessment System for Teaching English Writing in College[J].Advances in multimedia,. 2022(5):1.1-1.9.

[5] Yan Z , Zhao S .The Relationship Between School-Based Research and Preschool Teachers' Teaching Ability: The Mediating Role of Constructivist Beliefs in Teaching[J].Frontiers in psychology, 2022, 13:814521.

[6] Honma Y , Nagao T , Aiga K .Evaluation of a Moral Education Programme Teaching Role-Taking Ability to Youth at a Juvenile Training School:[J]. Japanese Journal of Educational Psychology, 2022.

[7] Lu J , Gao H .Online Teaching Wireless Video Stream Resource Dynamic Allocation Method considering Node Ability[J].Scientific Programming, 2022,. 2022:1-8.

[8] Xie Q .Using business negotiation simulation with China's English-major undergraduates for practice ability development[J].Heliyon, 2023, 9(6) .

[9] Zhang Y .The Research on Critical Thinking Teaching Strategies in College English Classroom [J].Creative Education, 2022.

[10] Syafitri W .Learning Experiences in Small Group Discussion in the Third Semester of English Education Students[J]. Language Teaching, 2023.

[11] Li H .A Multirate Cognitive-Based Approach for Optimal Dynamic Allocation of English Online Teaching Resources and IoT Applications[J].Wireless Communications and Mobile Computing, 2022.

[12] Ma Y .The Use of Cooperative Learning in English Writing Teaching[J].Journal of Higher Education Research, 2022, 3(2):163-165.

[13] Liao Q .Study on Students' Learning Ability from the Design and Analysis of Primary English Test Paper[J]. 2023.

[14] Bagheridoust E , Khairullah Y K .A Comparative-Correlative Study of Test Rubrics Used as Benchmarks in Assessing IELTS and TOEFL Speaking Skills[J]. .English Language Teaching, 2023.

[15] Vidhiasi D M .The Implementation of Reading Assessment Method[J].International Journal of English and Applied Linguistics (IJEAL), 2022.

[16] Jufang M A .Research on Integrating Moral Education Into College English Course on the Basis of Outcomes-Based Teaching Mode[J].Psychology Research, 2022, 12(10):865-873.

[17] Zhu M , Sun G .Factors Influencing Analysis for Level of Engineering English Education Based on Artificial Intelligence Technology[J]. Mathematical Problems in Engineering, 2022, 2022.

[18] SHAO-MING XU,YU LI,QING-LONG YUAN. A combinatorial pruning method based on reinforcement learning and 3σ criterion[J]. Journal of Zhejiang University (Engineering Edition),2023,57(03):486-494.

[19] Yang Tonglai, Geng Yueqi. Research on data auto-filling method[J]. Communication World,2024,31(01):148-150.

[20] Li S.P.,Meng Q.Z.. Measurement of operational risk of unlisted banks in China - based on Z-score method[J]. Journal of Shandong University of Finance and Economics,2018,30(03):61-71.

[21] S.J. Hsing, W. Ma, J.K. Xue. Payload health parameter extraction based on nonlinear PCA[J]. Space Electronics Technology,2024,21(03):99-104.

[22] CHEN Wen-Jin, YANG Xiao-Feng, QI Wei-Wen, WANG Jian-Jun, ZHAO Feng, CHEN Jianguo, WANG Jian. A fast cluster division method for photovoltaic power plants based on prototype extraction and clustering[J]. Zhejiang Electric Power,2024,43(04):74-84.

[23] Peng Cheng. Real-time monitoring method of lift energy consumption based on K-means clustering and BP neural network[J]. Journal of Tonghua Normal College,2024,45(04):50-56.

[24] Yin Li-Feng,Li Q.J.. Improvement study of heuristic k-means clustering algorithm[J]. Journal of Dalian Jiaotong University,2024,45(02):115-119.

[25] Wang Huili,Qiao Yongyi. An emergency material deployment model considering priority and time window constraints[J]. Science, Technology and Engineering,2024,24(12):5069-5075.

[26] W.J. Liu, H.J. Yang, S.S. Yang, W.D. Qin. Multi-criteria ABC inventory classification model based on density peak clustering method[J]. Journal of Heihe College,2024,15(06):67-71.

[27] Lin W, Chen Y. Sentiment analysis of Chinese microblogs by fusing BERT-BiGRU and multi-scale CNN[J]. Journal of Chinese Academy of Electronic Science, 2023, 18(10):939-945.

Enhancing Indonesian Text Summarization with Latent Dirichlet Allocation and Maximum Marginal Relevance

Muhammad Faisal[1*], Bima Hamdani Mawaridi[2], Ashri Shabrina Afrah[3], Supriyono[4],
Yunifa Miftachul Arif[5], Abdul Aziz[6], Linda Wijayanti[7], Melisa Mulyadi[8]
Department of Informatics Engineering, Universitas Islam Negeri Maulana Malik Ibrahim, Malang, Indonesia[1, 2, 3, 4, 5]
Faculty of Humanities, Universitas Islam Negeri Maulana Malik Ibrahim, Malang, Indonesia[6]
Profesional Engineer Program, Universitas Katolik Indonesia Atma Jaya, Jakarta, Indonesia[7, 8]

***Abstract*—Maximum Marginal Relevance (MMR) Summarization of text is very important in grasping quickly long articles particularly for people who are very busy. In this paper, we use LDA to give topic queries for news articles, which then become inputs to the MMR method. According to this paper's summarization system, the ROUGE metric is employed to evaluate the summaries of news articles with 30 percent compression and 50 percent compression. Experimental findings show that the LDA-MMR combination outperforms MMR on its own in all our tests across all query lengths or number of sentences used and gives highest average ROUGE value of 0.570 for a 50% compression rate; 0.547 at 30% This implies that our system efficiently produces meaningful summaries using content-based keywords rather than click bait titles, which should not lead to complaints about misleading advertisements. This summarizer can convey the main points of a piece of news coverage in a concise form, thus offering people useful new tools for quickly digesting information.**

Keywords*—Indonesian summarization; LDA; MMR; ROUGE evaluation

I. Introduction

Natural Language Processing (NLP) is a field that combines computer methods and cognitive, seeking to make computers understand, process, or produce human language. This field entails such tasks as sentence analysis, terminology analysis, and decision-making. The usefulness of NLP is many including machine translation for instant translation between languages, electronic mail spam recognition and rejection of unwanted messages, information mining to recover relevant information from large text repositories, and chatbots as a kind of automatic customer service. An important application of NLP is the generation of automatic text summaries to produce shorter, more understandable summaries of long texts, while retaining all their essential meaning. This is particularly important with the explosion of textual data on the Internet and in digital archives [1].

By exploring and applying different methods or algorithms, automatic text summary aims to produce shorter versions of texts. These methods can be divided according to the input type (single document and multiple documents) and output type (extractive or abstractive summaries) [2]. Extractive summarization means picking sentences, phrases or sections out of the original text, while abstractive summarization involves constructing new sentences in one's own voice which interpret or compress the essence of the original text. Extractive methods are often preferred for their simplicity and lower computational requirements vs. the more sophisticated natural language understanding and generation capabilities that are needed in abstractive techniques.

An important technique in extractive summarization is called Maximum Marginal Relevance (MMR). MMR checks sentence for their relevance to a given query and removes redundancy in a dataset containing similar content [3] while it uses the cosine similarity matrix plus Vector Space Model (VSM) to assess sentence significance. It is well suited to making summaries from both single documents and multiple. However, text queries must be made by hand, taking up a lot of time. And given the arrival of large-scale editorial systems that are now reaching their limits on efficiency through human interaction alone, automated query generation methods will therefore be needed to raise productivity and meet higher quality levels. Latent Dirichlet Allocation (LDA), a popular topic modeling technique, can uncover topics from a text corpus without any human intervention. Efficiently raising queries, the second use for LDA is to find topics hidden in a data set and model them [4]. The utilization of LDA can facilitate the streamlining of manual query generation, thereby enhancing the efficacy of the summarization process. Empirical evidence has demonstrated that the integration of LDA with other summarization techniques can markedly enhance the quality of the resulting summaries. For instance, the conjunction of LDA with MMR has been observed to yield outcomes that are superior to those obtained by either method in isolation [5].

The method proposed in this study serves an inventive approach to use LDA in conjunction with MMR constructed along the lines of an algorithm for efficient summary-making Indonesian news articles. It begins by using LDA to reveal the most important themes present in each article and then builds queries for MMR onto these constituent word distributions This no-nonsense approach is designed to make the summaries both brief and germane to the essence of the articles themselves, so that even if they give little by way of clue,

* Corresponding author

within five minutes readers will already gain some understanding about what contents this news offers [6].

In summarizing Indonesian text, LDA and MMR approaches have never been comprehensively challenged. Although past studies have demonstrated that both methods are effective enough in themselves, the uniqueness of news topics and characteristics requires their combined use to completely handle [7]. Our aim is to bridge this gap, utilizing the strengths of both LDA and MMR with the same final goal of getting better quality and relevance on resulting summaries.

The present research aims at creating an automatic summarization system that employs LDA for topic modeling and then MMR for extractive summarization, to generate accurate summaries of Indonesian news articles. By offering concise language summaries focused on the topic, this method is intended to improve the efficiency of retrieving information and also promote a better reading experience for people who need it. The novelty and contribution of this paper lies in combining LDA and MMR. This is expected to push forward the development of text summarization models as an efficient approach for managing large amounts of data.

The remainder of this paper is organized as follows: Section II provides an in-depth literature review, analyzing various current methods and their limitations. In Section III, we describe the method we propose by combining Latent Dirichlet Allocation (LDA) with Maximum Marginality Relevant (MMR) for text summarization. Section IV outlines the experimental settings, while in Section V we report results along with a discussion of them. Comparison is given in Section VI. Finally, Section VII and VIII concludes the paper and points out future research directions.

II. Literature Review

Most of the contemporary means developed for automatic summarization are made to supply summaries at least on par with these extracted by people. Most of this research has focused on high-resource languages, although there are some studies for low-resource language such as Indonesian. This summarization has shown state of the art results in LDA, MMR: two example techniques that have successfully been applied to automatic text summarization across various languages.

Saikumar and Subathra (2020) introduced a set of summarization method using LDA, MMR and Text Rank (TR), proving that the generated summaries are much precise comparing to standalone use of MMR or TR techniques. Finally, the performance of this two-level document summarization (DS) method with LDA was compared to that based on MMR and TR [3]. TextRank and MMR were integrated to be used for summarization of Indonesian news articles by Gunawan, Harahap & Rahmat (2019) [8].

Tuhpatussania, Utami and Hartanto in 2022 [9]: In their work on summarization of online Indonesian news text they have compared the performance between LexRank dan MMR Algorithm: proof that mmmr is better than lexrank for precision, recall and f-measure. Musyaffanto, Herwanto and Riasetiawan (2019), on the other hand integrated MMR with Nonnegative Matrix Factorization (NMF) to ensure precision of online news articles [10].

LDA has been applied to text analysis problems in other research areas as well. To summarize Malayalam news documents closer to what human makes, Kondath, Suseelan and Idicula (2022) used LDA [5]. Rahman et al. (2021) used LDA to create summaries from Malay news documents that showed how system-generated news can save readers' time [6].

The other one is hybrid models for text summarization. Hybrid Approach Gurusamy, Rengarajan and Srinivasan [7] proposed a hybrid approach to this problem that combines semantic LDA and sentence concept mapping with transformer models for generation of coherent text summaries. LISJANA et al. [11], 2020 Classifiers used: They have also applied LDA classifiers with Latent Semantic Indexing (LSI), Similarity-Based Histogram Clustering (SHC) for multi-document text summarization.

Studies for enhancing MMR have started gaining momentum as well. Zheng, Liu, and Qin proposed an improved MMR algorithm that uses Word2Vec, TextRank with semantic information to make text summarization more efficient. [12]. Ramezani et al. (2023) [13] compared LSA vs. MMR in summarizing Persian broadcast news transcriptions, demonstrating LSA's superiority in generic summarization [13].

The above studies point to several research gaps which this study intends to address. Several studies have shown that LDA or MMR, both individually and combined with the other methods, are effective at detecting significant information in a document, but there is no unified approach combining these two aspects: prior to this study not yet proposed any research for Indonesian text summarization task. Besides, current studies usually do not consider the practical difficulty to generate concise summaries over articles of different subjects. This research tries to solve this problem by creating a new summarization method that combines the use of LDA (TextSum) and MMR, for Indonesian article summary generation can be done better with more relevant topics.

III. Proposed Methodology

Fig. 1 shows our proposed hybrid approach using Latent Dirichlet Allocation (LDA) and Maximum Marginal Relevance (MMR) in an Indonesian article summarisation system. It starts with the pre-processing of text documents, which consists of performing all the important steps needed to clean the data and make it ready for analysis. First, the text is broken down into small units such as sentences or words. Next, lowercase (): applies case folding to normalise the text by converting all characters to lowercase. Cleanup: removes all characters irrelevant to the analysis, such as unimportant symbols, punctuation or special characters. This is followed by the removal of stop words, which are common words that do not add any useful information to the summarisation process. The last step in the pre-processing is the stemming, where words are reduced to their root forms, so that different variants of a word can be treated equally.

In this step, the processed text is analysed using a technique called Latent Dirichlet Allocation and then Maximum Marginal

Relevance. We start by using LDA to find the natural topics in the text, which gives us an experienced explorer's view of what kind of information we want. A generative probability model helps to identify the content of the document, and it uses only key topics or words used to describe a single topic from the rest that was scattered. On the other hand, Maximum Marginal Relevance will select the sentences that are most relevant to the given query, i.e. the summary will be both comprehensive and contextually satisfying for the user. The combination of LDA and MMR helps to summarise the given text by capturing what is important in the text data. This dual approach not only increases summarisation accuracy, but also ensures that the output is contextually relevant and insightful optimizing real-world performance.

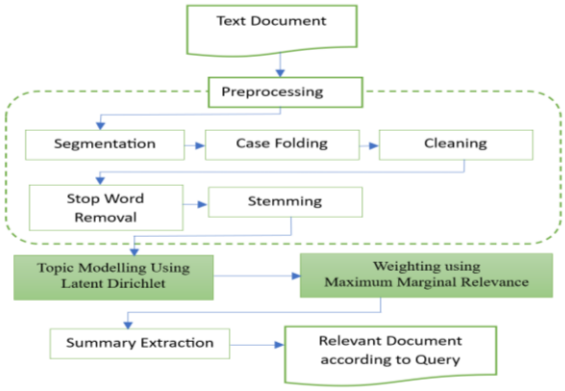


Fig. 1. Proposed system.

A. Dataset

Experimentally, we conducted the research on IndoSum dataset [14]. It was used in the current research. This corpus consists of around 14,290 news articles along with titles in which all are categorized into six classes (see Fig. 2) from ten different sources of the Indonesian press released to public [14]. The article URL and summary abstracts separately written by two native Indonesian speakers [14]. Rini Wijanti et al. used this dataset in their research [15]. Fig. 3 shows source of news.

The study carried out two testing experiments. This shown in experiment 2, reaching the best ROUGE measure by applying stemming without stopwords removal on test data.

B. Pre-processing

The preprocessing step is responsible for improving the structure of the input data. Preprocessing in NLP often involves tokenization. However, in this paper, we used data from the IndoSum dataset where tokenization has already been performed. Therefore, we can avoid repeating the process of tokenization in this study. Each paragraph consists of a list of clauses, with each clause containing a list of words (token). Segmentation, tokenization, case folding, stop word removal, stemming.

C. Segmentation

In the segmentation process, any paragraph separator is removed so that articles are divided directly on per sentence basis for further processing. During this stage, all paragraphs within each article are combined and then divided into individual sentences by the segmentation process. As a result, each article has sentences, and in turn the sentence will contain words or Tokens.

D. Case Folding

Stage 2 - Case folding at this stage of the preprocessing, we have to convert all uppercase words to lowercase. This process will help to avoid any confusion in the meaning of a word based on whether it's capitalized or not. Step 6 - Lower case processing of the segmentation output (sentence list with tokenized sentence) at the end of this process, we have a list of lowercase words.

E. Cleaning

After tokenization, the third step is data cleaning: This is because only characters are needed for input, not punctuation or numerals, so characters other than letters are removed from the record. And then they have a data cleansing purpose to remove other academic input needed for the program by keeping only clean text as input.

F. Stop Removal

The fourth stage is the elimination of stop words, which is the identification and elimination of words that are common and occur frequently in the text, but often without providing important information. Removing stop words is primarily aimed at cleaning up text and improving text analysis/modelling quality. It is useful to remove keywords to focus the analysis on more relevant or significant words. An example of stopword is a conjunction. Each token is checked whether it is on the stopword list or not, if it is, the token is deleted and not included in the next process. If it is not on the stopword list, it is passed on to the next process.

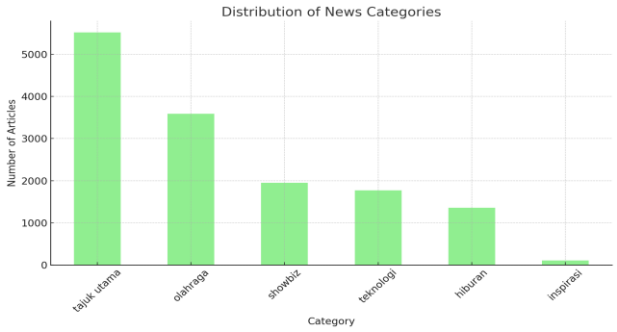


Fig. 2. Category of news.

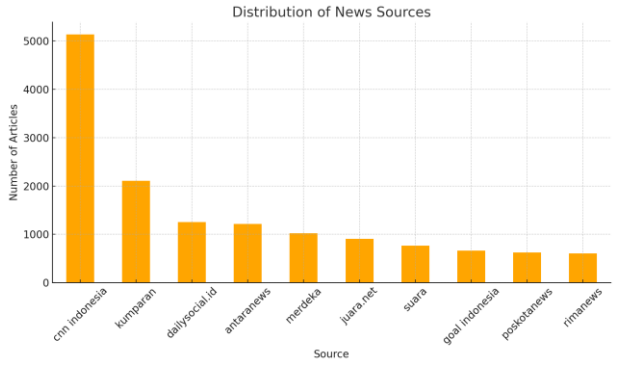


Fig. 3. Source of news.

G. Stemming

The next stage is stemming, which is a process in which words are transformed into their most basic forms. Removing inflections or affixes to ensure a consistent representation of words with a common root is the main goal of stemming. Each token is cross-referenced with the base word lexicon. If a token is missing, it is identified as an affixed word and stemming is initiated by deleting suffixes (-lah, -kah, -ku, -mu, -nya, -tah or -pun). Next, derivative affixes (-i, -kan, -an) are removed, followed by the removal of prefix affixes (be-, di-, ke-, me, pe-, se- and te-).

H. Final Preprocessing

The stemming process is carried out in the final preprocessing stage. The list of tokens will be transformed in this way. All empty strings and lists left over from the previous steps are removed in this final preprocessing step.

I. Latent Dirichlet Allocation

LDA is commonly used when the topics of the papers tend to cluster around a single focus. They are also used to produce topic model results in information technology papers, including domain-specific documents like research papers, news stories, and patents [16] [17] [18]. By using LDA on a set of documents, we can determine the distribution of hidden topics both across the set and within each individual paper. Each topic has its own probability distribution of words attached to it. It is a type of statistical modelling designed to represent the probability distribution of a set of data. This model is useful for generating new data [19], as it can generate data similar to the training data.

LDA models are shown in Fig. 4.

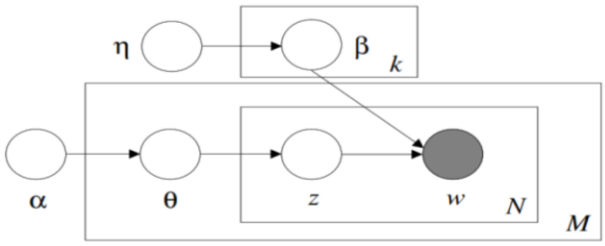


Fig. 4. Graphical representation of the smoothed LDA model [20].

The image shows the Latent Dirichlet Allocation (LDA) model, a generative statistical approach to uncover hidden thematic structure in text documents. The main elements of the model are α, θ, η, βk, z and w. The hyperparameters α and η affect the distributions of topics and words respectively. θ denotes the topic proportions for a document derived from a Dirichlet distribution with parameter α. βk represents the word distribution for a topic derived from a Dirichlet distribution with parameter η. The generative process involves selecting a topic z from θ and then a word w from βk, thereby linking words to topics and topics to documents [21].

Recent advances in LDA have addressed several challenges and expanded its applications. The empirical prior LDA (epLDA) model, which uses latent semantic indexing to obtain priors from the data, has shown notable improvements over the traditional LDA model [22].

The equation illustrates the joint probability distribution in Latent Dirichlet Allocation (LDA), a generative probabilistic approach to topic modelling. This model reveals thematic structures within a collection of documents by decomposing the joint probability into several elements: β_K (word distributions for each topic), θ_D (topic distributions for each document), Z_D (topic assignments for each word), and W_D (observed words). The hyperparameters α and η determine the Dirichlet priors for the topic and word distributions. This relationship is expressed in Eq. (1).

$$p(\beta_K, \theta_D, Z_D, W_D \mid \alpha, \theta) = \prod_{k=1}^{K} p(\beta_k|\eta) \prod_{d=1}^{D} p(\theta_d|\alpha) \prod_{d=1}^{N} p(Z_{d,n}|\theta_d) p(W_{d,n}|Z_{d,n}, B_{d,k}) \quad (1)$$

This product-based formula incorporates several essential elements: the distributions of words within topics (β_k), the distributions of topics within documents (θ_d the topic assignments for each word in each document ($z_{d,n}$), and the observed words in the documents ($w_{d,n}$).

In this model, $\beta_{1:K}$ represent the topic-word distributions, which are influenced by the parameter η and which determine the likelihood of words given topics. The term $\theta_{1:D}$ denotes the document-topic distributions, which are influenced by the parameter α and which determine the likelihood of topics given documents. The term $p(\beta_{k/\eta})$ denotes the probability of the word distribution for topic k given the Dirichlet prior η, while $p(\theta_d/\alpha)$ is the probability of the topic distribution for document d given the Dirichlet prior α. The term $p(z_{d,n}/\theta_d)$ represents the probability of assigning the n-th word in document d to a topic, based on the topic distribution for that document.

Finally, $p(w_{d,n} / z_{d,n}, \beta_{d:K})$ represents the probability of the n-th word in document d being assigned to topic $Z_{d,n}$, given the aforementioned topic assignment and the word distribution for that topic, β_k. Recent advances in the latent Dirichlet allocation (LDA) approach have led to the introduction of enhanced models and methodologies, including the empirical prior latent Dirichlet allocation (epLDA) and StreamFed-LDA. The epLDA model improves the computation of topics by employing latent semantic indexing to derive priors from data, thereby enhancing prediction accuracy [22].

- Hyperparameter Selection and Impact

The selection of hyperparameters in Latent Dirichlet Allocation (LDA) has a significant impact on the quality of the results, particularly in terms of topic coherence and relevance. In this study, the value of α (the Dirichlet prior for the distribution of topics per document) was set to 1/K, where K is the number of topics, and η (the Dirichlet prior for the distribution of words per topic) was set to 1/V, where V is the vocabulary size. These values were selected to ensure a balanced distribution of topics across documents and words across topics. The parameter α regulates the diversity of topics within a document; higher values of α result in more uniform topic distributions, thereby enabling documents to encompass a broader range of topics. Conversely, the influence of the parameter η on the distribution of words within each topic is inverse. Lower values of η result in sparser distributions, which

in turn produce more focused and distinctive topics. The preliminary experiments demonstrated that varying α and η has a significant impact on the coherence of the topics and the relevance of the summaries produced. Further research could involve a more comprehensive investigation of these hyperparameters to enhance LDA performance in diverse contexts.

J. Maximum Marginal Relevance

The Maximum Marginal Relevance (MMR) algorithm is a well-established method in the field of information retrieval. The algorithm calculates a linear combination that includes both the relevance of the documents to the query and their similarity to previously chosen documents for summarization. This measure, known as 'edge correlation', is optimized during the retrieval and summarization processes to refine the final summary iteratively [23][24]. MMR summarizes text by evaluating the similarity between different parts of the text, showing efficiency in retrieving relevant information and uncluttering content. It identifies documents that match a specific query by combining two criteria: relevance and heterogeneity.

MMR summarizes text by evaluating the similarity between different parts of the text, thereby demonstrating efficiency in the retrieval of relevant information and the uncluttering of content. The method identifies documents that match a specific query by combining two criteria: relevance and heterogeneity.

In this framework, a linear combination is calculated in order to integrate a document's relevance to the query and its similarity to pre-selected documents for summarization. This metric, designated as 'edge correlation', is calibrated during the retrieval and summarization phases to incrementally refine the final summary. MMR employs a methodology whereby content is described by assessing the degree of similarity between text segments. This demonstrates the ability of the method to retrieve related data and avoid redundancy.

MMR employs a ranking system based on a combination of cosine similarity matrices in response to a given query. The calculation entails a comparison of the results pertaining to the relevance of the query with those concerning the similarity of sentences. A document is deemed to possess high marginal relevance if it exhibits a strong alignment with the document content and a high degree of similarity with the query. The MMR score can be calculated using the following Eq. (2) [25].

$$MMR = argmax\ [\lambda * Sim_1(S_i, Q) - (1-\lambda) * maxSim_2(S_i, Summ)] \quad (2)$$

In this context, Si represents the candidate sentence, and Q is the query or main topic. The parameter λ (which ranges from 0 to 1) serves to regulate the equilibrium between relevance and diversity. The term Sim1(Si,Q) is employed to ascertain the degree of similarity between the candidate sentence Si and the query Q, thereby ensuring that the selected sentences are highly relevant. The term $(1 - \lambda) \cdot$ maxSim2(Si, Summ) is employed to ascertain the maximum similarity between the candidate sentence Si and the sentences that have already been included in the summary. This serves to minimize redundancy.

In this context, the term "*Si*" represents a sentence within the document, whereas "Summ" refers to the selected or extracted sentences. The relevance of a sentence is determined, and redundancy is minimized through the utilization of the coefficient λ.

The parameter λ is defined in the range of 0 to 1. When λ equals one, the MMR value is more pertinent to the original document. Conversely, when λ equals zero, the MMR value is more aligned with the previously extracted sentences. It is therefore recommended that λ be adjusted within this range to achieve optimal summarization. In the case of shorter texts, such as articles, the optimal value for λ is generally considered to be 0.7, which yields effective summary results [25].

IV. Experimental Results

The experiment was conducted on the initial 50 article data entries within the train.03 JSON file, which forms part of the IndoSum dataset. The test scenario was conducted on articles 1 to 50 to ascertain the ROUGE-1 value for each system-generated summary. Prior to the generation of summaries, each article was subjected to topic modelling using the Latent Dirichlet Allocation (LDA) technique, which yielded one topic and ten keywords. In the LDA topic modelling process, the alpha value was set to 1/K and the eta value was set to 1/V. The ten keywords were then employed as queries to generate summaries utilizing the Maximum Marginal Relevance (MMR) method. The MMR summarization process was conducted with three different lambda values, as follows: The values of 0.5, 0.7, and 0.9 were employed. The resulting summaries comprised either 50% or 30% of the total sentences in the original text.

A series of tests were conducted to compare the quality of human-generated summaries with those produced by the system. To obtain recall, precision, and F1-score values, this research employs the ROUGE-1 evaluation method. The greater the degree of alignment between the system summary and the human summary, the higher the recall value. Should the recall value attain its maximum value or a value of 1, it signifies that the entirety of the human summaries will be incorporated into the system summary. Conversely, if the precision value reaches the maximum value or 1, then the entire system summary will be included in the human summary. The combination of recall and precision, namely the F1-score, provides an overall picture of the system's ability to capture and present appropriate and relevant information in its summary.

The experiment was conducted using three values of λ: 0.5, 0.7, and 0.9. The results of Experiment 1 are presented in Table I.

TABLE I. ROUGE-1 Evaluation Result Experiment 1LDRMMR

λ	Compression rate		
	50%		
	Average Recall	Average Precision	Average F1-Score
λ = 0.5	0.863	0.402	0.534
λ = 0.7	0.865	0.404	0.536
λ = 0.9	0.864	0.402	0.535

Table I illustrates the Rouge-1 assessment outcomes for Experiment 1LDRMMR, which evaluates a document summarisation system at a 50% compression rate utilising diverse values of the smoothing parameter, lambda (λ). In this comparison, the following metrics are considered: recall average, precision average and F1 score average for λ values of 0.5, 0.7 and 0.9. As λ increases, there is a modest enhancement in recall average, from 0.863 to 0.865, before a slight decline to 0.864. The value of the Precision Average is observed to increase from 0.402 to 0.404 at λ = 0.7 and then to remain at this level of 0.402 when λ = 0.9. The F1-score average, which weighs precision and recall equally, demonstrates the most optimal performance at λ=0.7. It increases from 0.534 to 0.536 and then experiences a slight decrease to 0.535.

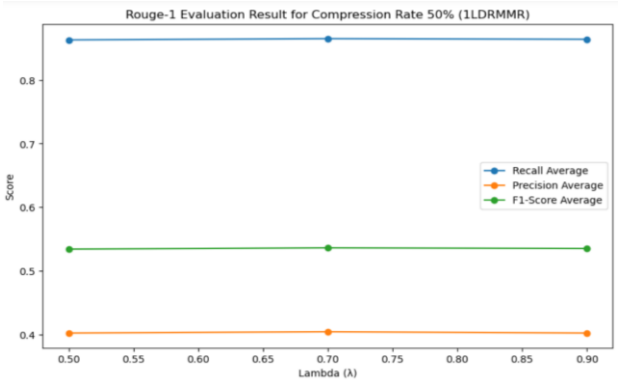


Fig. 5. Rouge-1 evaluation result experiment 1LDRMMR.

The graph in Fig. 5 demonstrates these trends, indicating that an increase in λ generally enhances the system's performance, with the optimal λ value being 0.7. At this value, the system attains the optimum balance between recall and precision, resulting in the highest F1 score. These findings suggest that while higher λ values produce marginal improvements in performance, the gains plateau beyond λ=0.7, indicating an optimal range for λ to optimise summarisation quality.

Additionally, in the second experiment, an evaluation was conducted with the results presented in Table II.

TABLE II. ROUGE-1 EVALUATION RESULT EXPERIMENT 2LDAMMR

λ	**Compression rate**		
	30%		
	Recall Average	**Precision Average**	**Average F1-Score**
λ = 0.5	0.778	0.485	0.580
λ = 0.7	0.775	0.484	0.578
λ = 0.9	0.781	0.486	0.581

The results of the Rouge-1 evaluation for the 2LDAMMR experiment, with a compression rate of 30%, are presented in Table II. The table presents a comparison of three distinct λ values: These values were 0.5, 0.7, and 0.9. For λ = 0.5, the recall average is 0.778, the precision average is 0.485, and the average F1-measure is 0.580. When λ=0.7, a slight decrease is observed in the Recall Average (to 0.775), while the Precision Average remains almost unchanged (at 0.484). The Average F1-Measure also decreases, albeit to a lesser extent (to 0.578). At λ=0.9, the recall average increases to 0.781, the precision average rises slightly to 0.486, and the average F1-measure also increases slightly to 0.581.

It can be concluded from these results that the λ value influences the evaluation metrics. While the changes observed are minor, higher λ values tend to result in slight improvements in both the Recall Average and Precision Average. However, these changes are not significant and there is consistency in the Average F1-Measure, indicating that the model demonstrates stable performance across different λ values. For a more comprehensive visual representation, please see the graph below, which illustrates the comparison of the metrics for each λ value.

Here is the graph that illustrates the comparison of metrics for each λ value in greater detail, as shown in Fig. 6.

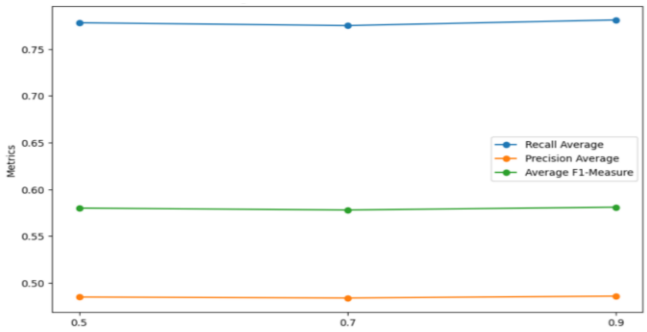


Fig. 6. Graph rouge-1 evaluation result experiment 2 LDAMMR.

Fig. 6 depicts the performance of Recall Average, Precision Average, and Average F1-Measure across λ values of 0.5, 0.7, and 0.9 for the 2LDAMMR experiment with a 30% compression rate. The Recall Average demonstrates a slight increase with elevated λ values, indicating enhanced recall capability, whereas the Precision Average remains stable, suggesting consistent precision. The Average F1-Measure, representing the harmonic mean of precision and recall, also shows minimal variation, indicating balanced performance. Overall, increasing λ results in minor improvements in recall and F1-Measure, demonstrating the model's robustness and consistent performance across the tested λ range.

Experiments were also conducted using MMR with title queries with the same dataset and λ value of 1, the results of which can be seen in Table III and Table IV.

TABLE III. ROUGE-1 EVALUATION RESULT EXPERIMENT 1MMR

Experiment 1MMR	**Compression rate**		
	50%		
	Average Recall	**Average Precision**	**Everage F1-score**
1 (λ = 0.5)	0.843	0.407	0.536
2 (λ = 0.7)	0.843	0.407	0.536
3 (λ = 0.9)	0.843	0.407	0.536

Table III presents the findings of the evaluation conducted on the 1MMR experiment, utilizing a compression rate of 50%. Three distinct λ values (0.5, 0.7, and 0.9) were subjected to

evaluation. The results for Recall Average, Precision Average, and Average F1-Measure are consistent across all λ values. The Recall Average remains at 0.843, the Precision Average at 0.407, and the Average F1-Measure at 0.536 for each λ value. This consistency suggests that the λ parameter does not significantly affect the model's performance under these conditions.

TABLE IV. ROUGE-1 EVALUATION RESULT EXPERIMENT 2MMR

Experiment 2MMR	Compression rate		
	30%		
	Average Recall	Average Precision	Average F1-score
1 (λ = 0.5)	0.680	0.460	0.536
2 (λ = 0.7)	0.680	0.460	0.536
3 (λ = 0.9)	0.680	0.460	0.536

Table IV presents the results of the evaluation of the 2MMR experiment with a 30% compression rate. Once more, three distinct λ values (0.5, 0.7, and 0.9) are subjected to examination. As with the 1MMR experiment, the results for Recall Average, Precision Average, and Average F1-Measure are consistent across all λ values. The recall average is 0.680, the precision average is 0.460, and the average F1-measure is 0.536 for each λ value. The consistency across different λ values indicates that the λ parameter does not significantly impact the model's performance in the 2MMR experiment.

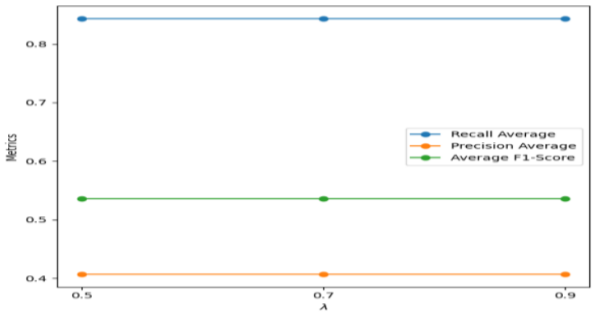


Fig. 7. Rouge-1 evaluation result experiment 1MMR.

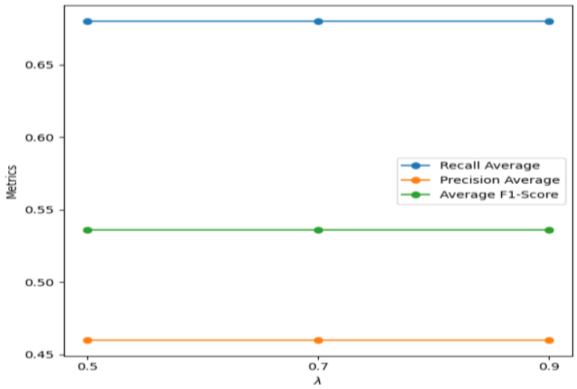


Fig. 8. Rouge-1 evaluation result experiment 2MMR.

Fig. 7 and Fig. 8 present a graphical representation of the performance of Recall Average, Precision Average, and Average F1-score across λ values of 0.5, 0.7, and 0.9 for both the 1MMR experiment with a 50% compression rate and the 2MMR experiment with a 30% compression rate. In the case of the 1MMR experiment, all metrics remain constant regardless of λ. The values for these metrics are as follows: Recall Average at 0.843, Precision Average at 0.407, and Average F1-Measure at 0.536. Similarly, in the 2MMR experiment, the metrics demonstrate no variation with different λ values, maintaining a Recall Average of 0.680, a Precision Average of 0.460, and an Average F1-Measure of 0.536. This consistency across both experiments indicates that λ has a negligible impact on model performance in these scenarios. At a 50% compression rate, 1MMR achieves a higher recall than 2MMR at a 30% compression rate.

TABLE V. COMPARISON BETWEEN MMR AND LDAMMR

Methods	F1-score Average	
	Compression Rate 50%	Compression Rate 30%
MMR	0.536	0.536
LDA & MMR	0.536	0.581

Table V presents a comparison of the average F1-score between two methods. The study compares the effectiveness of two approaches to text compression: Maximal Marginal Relevance (MMR) and a combined approach of MMR and Latent Dirichlet Allocation (LDA), applied at two different compression rates, 50% and 30%.

The average F1-score for the MMR method remains constant at 0.536 for both compression rates of 50% and 30%. This consistency indicates that the performance of MMR alone is not affected by different levels of compression, suggesting that MMR is robust in maintaining its effectiveness regardless of the compression rate applied.

In contrast, the combination of MMR and LDA yielded a notable enhancement in the average F1-score at the 30% compression rate, which increased to 0.581. However, at the 50% compression rate, the combination yields the same average F1-score of 0.536 as MMR alone. This suggests that the incorporation of LDA with MMR improves performance, particularly at the lower compression rate of 30%. This implies that LDA provides supplementary contextual information, enhancing the model's efficacy when the data is more condensed.

The experiments conducted utilising MMR with LDA queries yielded superior ROUGE-1 evaluation scores in comparison to those employing MMR with title queries. However, both systems exhibit a commendable ROUGE-1 score. According to Deutsch, the discrepancy in ROUGE-1 scores below 0.5 between systems is less indicative of the human perception of the same two systems [26].

The results presented in Table V illustrates that while the MMR method demonstrates consistent performance across varying compression rates, the integration of MMR with LDA markedly enhances the Average F1-Score at the 30% compression rate. This suggests that LDA improves MMR's capacity to capture pertinent information in a more condensed dataset. The consistency of results at the 50% compression rate indicates that the advantages of LDA are more evident when dealing with higher levels of data compression.

Fig. 9 illustrates the comparison of the Average F1-Score for each method under both compression rates.

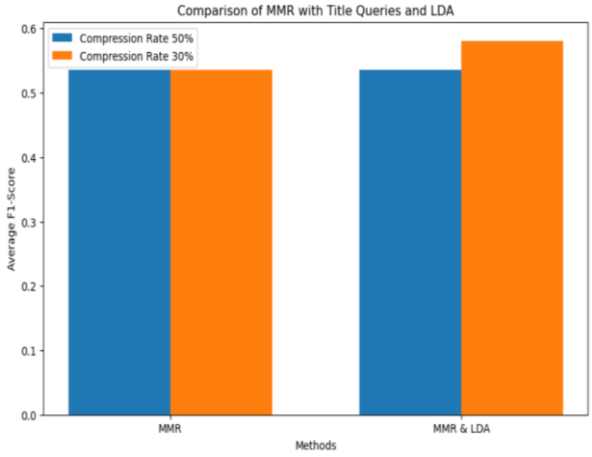


Fig. 9. Comparison of MMR with title query and LDA.

V. Discussion

The combination of Latent Dirichlet Allocation (LDA) and Maximal Marginal Relevance (MMR) offers notable advantages that make it a more effective approach than traditional text summarisation techniques. Although MMR has been extensively employed for the selection of pertinent sentences based on title queries, its principal limitation resides in its reliance on manually crafted or title-based queries, which frequently impede the precision of the resulting summaries, particularly when the article title does not fully encapsulate the content.

The incorporation of LDA enables the generation of contextually rich topic queries, which results in summaries that more accurately represent the contents and themes of the articles. This is especially advantageous in cases where article titles may be deceptive or inadequate in representing the actual content, such as in the context of clickbait headlines.

The combination of LDA and MMR has been demonstrated to achieve a higher F1-score at a 30% compression rate, indicating an improvement in both the quality and relevance of the summaries. These findings indicate that the proposed method is not only effective for summarising Indonesian news articles but could also be adapted to other languages and document types, thereby enhancing its overall applicability and versatility.

This approach represents a significant step forward in the creation of more accurate and contextually relevant summaries, particularly in cases where traditional methods may be inadequate.

VI. Comparison

The objective of this research is to develop a document summarization system that can extract the essential information from documents. The study presents a distinctive profile when compared to the various methods and tests identified in other studies. Table VI provides a summary of these differences. In their study [3] employs the Two-Level LDA method with customer opinion data on products and hotels, comparing LDA with MMR and TR. In contrast, our study employs the LDAMMR method and compares it solely with MMR. In contrast to the study [9] employs solely the MMR method, without comparison to other techniques. In contrast to the studies [6] does not undertake a comparative analysis; instead, it focuses on summarizing Malay language news articles using LDA. In their study [25] employed MMR and VSM to summarize students' final project abstracts. While existing research has demonstrated the effectiveness of LDA and MMR individually or in combination with other methods, there remains a lack of comprehensive approaches that specifically integrate LDA with MMR for summarizing Indonesian news articles. Additionally, existing studies often fail to thoroughly address the challenges of summarizing articles that cover a wide range of topics.

TABLE VI. Text Summarization Comparative Study

References	Object	Methods	Comparison methods	ROUGE-1 Results (F1-Score)
[3]	Two categories: products, hotels	Two Level LDA	Compare with MMR and TR summarization techniques	Not provided
[9]	Indonesian News Article	MMR	Not compare	Not provided
[6]	Malay News Article	LDA	Not compare	Not provided
[25]	One category: Students Final Project Abstracts	MMR & VSM	Not compare	Not provided
Ours	Four Category Indonesian News Article	LDAMMR	Compare LDA vs LDAMMR	0.536 (50% compression) 0.581 (30% compression)

VII. Conclusion

The integration of Latent Dirichlet Allocation (LDA) with Maximum Marginal Relevance (MMR) has been demonstrated to enhance text summarization. This is achieved by generating more accurate and relevant queries, reducing redundancy, and providing a contextual understanding of the document's themes. This combination of techniques improves efficiency through the automatic generation of queries, while maintaining a balance between precision and recall. The results of the research demonstrate that while MMR exhibited a constant average F1-score of 0.536, the integration of LDA resulted in an increase to an average F1-score of 0.581 at a 30% compression rate. This illustrates that LDA augments MMR's capacity to capture pertinent information in a more efficacious manner, thereby rendering summaries more succinct and contextually pertinent, particularly in the case of diverse and evolving subject matter such as Indonesian news articles.

The potential for application in multiple languages is a further advantage of this approach.

This passage presents the findings of a study on the summarization of Indonesian news articles, employing a methodology that is not language specific. The key techniques employed, namely Latent Dirichlet Allocation (LDA) and Maximal Marginal Relevance (MMR), are language-agnostic, meaning that they can be applied to different languages with some adjustments. These modifications include the adaptation of preprocessing procedures, such as tokenization, stopword removal and stemming, to align with the linguistic characteristics of the target language. The study posits that this approach could prove beneficial for languages with limited resources, where sophisticated text summarization tools are not as readily accessible. Furthermore, it urges future research to apply this methodology to multilingual datasets, which could facilitate the advancement of more versatile and globally applicable summarization techniques.

VIII. Future Work

The encouraging outcomes of this study suggest several avenues for future research. One avenue for further research would be to explore alternative topic modelling techniques, such as non-negative matrix factorisation (NMF) or latent semantic analysis (LSA), to ascertain whether they can enhance the quality of summaries even further. Furthermore, applying this method to a broader range of document types, including legal texts, scientific articles, or social media content, could serve to test its versatility and robustness across different contexts. Another promising avenue for future research is the integration of this method with transformer-based models, such as BERT or GPT, to develop a hybrid approach that combines the strengths of both extractive and abstractive summarisation. This could result in the generation of more coherent and contextually rich summaries, thereby advancing the state of the art in automatic text summarisation. Furthermore, adapting this model for real-time or streaming data could make it a valuable tool for dynamic content summarisation, providing immediate insights in fast-paced environments such as newsrooms or social media monitoring.

References

[1] V. Agate, S. Mirajkar, and G. Toradmal, “Book Summarization using NLP,” International Journal of Innovative Research in Engineering, pp. 476–480, Apr. 2023, doi: 10.59256/ijire.2023040218.

[2] U. Rani and K. Bidhan, “Comparative Assessment of Extractive Summarization: TextRank, TF-IDF and LDA,” Journal of scientific research, vol. 65, no. 01, pp. 304–311, 2021, doi: 10.37398/JSR.2021.650140.

[3] D. Saikumar and P. Subathra, “Two-Level Text Summarization Using Topic Modeling,” 2021, pp. 153–167. doi: 10.1007/978-981-15-5400-1_16.

[4] C. P. George and H. Doss, “Principled Selection of Hyperparameters in the Latent Dirichlet Allocation Model,” J. Mach. Learn. Res., vol. 18, pp. 162:1-162:38, 2017.

[5] M. Kondath, D. P. Suseelan, and S. M. Idicula, “Extractive summarization of Malayalam documents using latent Dirichlet allocation: An experience,” Journal of Intelligent Systems, vol. 31, no. 1, pp. 393–406, Mar. 2022, doi: 10.1515/jisys-2022-0027.

[6] N. A. Rahman, S. N. A. Ramlam, N. A. Azhar, H. M. Hanum, N. I. Ramli, and N. Lateh, “Automatic Text Summarization for Malay News Documents Using Latent Dirichlet Allocation and Sentence Selection Algorithm,” in 2021 Fifth International Conference on Information Retrieval and Knowledge Management (CAMP), IEEE, Jun. 2021, pp. 36–40. doi: 10.1109/CAMP51653.2021.9498029.

[7] B. M. Gurusamy, P. K. Rengarajan, and P. Srinivasan, “A hybrid approach for text summarization using semantic latent Dirichlet allocation and sentence concept mapping with transformer,” International Journal of Electrical and Computer Engineering (IJECE), vol. 13, no. 6, p. 6663, Dec. 2023, doi: 10.11591/ijece.v13i6.pp6663-6672.

[8] D. Gunawan, S. H. Harahap, and R. F. Rahmat, “Multi-document summarization by using textrank and maximal marginal relevance for text in Bahasa Indonesia,” in 2019 International conference on ICT for smart society (ICISS), 2019, pp. 1–5.

[9] S. Tuhpatussania, E. Utami, and A. D. Hartanto, “Comparison Of Lexrank Algorithm And Maximum Marginal Relevance In Summary Of Indonesian News Text In Online News Portals,” Jurnal Pilar Nusa Mandiri, vol. 18, no. 2, pp. 187–192, 2022.

[10] I. R. Musyaffanto, G. Budi Herwanto, and M. Riasetiawan, “Automatic Extractive Text Summarization for Indonesian News Articles Using Maximal Marginal Relevance and Non-Negative Matrix Factorization,” in 2019 5th International Conference on Science and Technology (ICST), IEEE, Jul. 2019, pp. 1–6. doi: 10.1109/ICST47872.2019.9166376.

[11] O. A. LISJANA, D. P. RINI, and N. YUSLIANI, “Multi-Document Text Summarization Based on Semantic Clustering and Selection of Representative Sentences Using Latent Dirichlet Allocation,” in Proceedings of the Sriwijaya International Conference on Information Technology and Its Applications (SICONIAN 2019), Paris, France: Atlantis Press, 2020. doi: 10.2991/aisr.k.200424.029.

[12] Y. Zheng, Y. Liu, and H. Qin, “Chinese News Text Abstract Extraction Using Improved MMR,” in 2021 International Conference on Electronic Information Engineering and Computer Science (EIECS), IEEE, Sep. 2021, pp. 601–607. doi: 10.1109/EIECS53707.2021.9587964.

[13] M. Ramezani, M.-S. Shahryari, A.-R. Feizi-Derakhshi, and M.-R. Feizi-Derakhshi, “Unsupervised Broadcast News Summarization; a Comparative Study on Maximal Marginal Relevance (MMR) and Latent Semantic Analysis (LSA),” in 2023 28th International Computer Conference, Computer Society of Iran (CSICC), IEEE, Jan. 2023, pp. 1–7. doi: 10.1109/CSICC58665.2023.10105403.

[14] K. Kurniawan and S. Louvan, “Indosum: A new benchmark dataset for Indonesian text summarization,” in 2018 International Conference on Asian Language Processing (IALP), 2018, pp. 215–220.

[15] R. Wijayanti, M. L. Khodra, and D. H. Widyantoro, “Single document summarization using bertsum and pointer generator network,” International Journal on Electrical Engineering and Informatics, vol. 13, no. 4, pp. 916–930, 2021.

[16] H. Jelodar et al., “Latent Dirichlet allocation (LDA) and topic modeling: models, applications, a survey,” Multimed Tools Appl, vol. 78, pp. 15169–15211, 2019.

[17] C. Tian, J. Zhang, D. Liu, Q. Wang, and S. Lin, “Technological topic analysis of standard-essential patents based on the improved Latent Dirichlet Allocation (LDA) model,” Technol Anal Strateg Manag, pp. 1–16, 2022.

[18] J. Kim et al., “Trend Research on Maritime Autonomous Surface Ships (MASSs) Based on Shipboard Electronics: Focusing on Text Mining and Network Analysis,” Electronics (Basel), vol. 13, no. 10, 2024, doi: 10.3390/electronics13101902.

[19] H. Liu, T. Zhang, F. Li, M. Yu, and G. Yu, “A probabilistic generative model for tracking multi-knowledge concept mastery probability,” Front Comput Sci, vol. 18, no. 3, p. 183602, 2024.

[20] D. M. Blei, A. Y. Ng, and M. I. Jordan, “Latent dirichlet allocation,” Journal of machine Learning research, vol. 3, no. Jan, pp. 993–1022, 2003.

[21] D. M. Blei, A. Y. Ng, and M. I. Jordan, “Latent dirichlet allocation,” Journal of machine Learning research, vol. 3, no. Jan, pp. 993–1022, 2003.

[22] M. A. Adegoke, J. O. A. Ayeni, and P. A. Adewole, “Empirical prior latent Dirichlet allocation model,” Nigerian Journal of Technology, vol. 38, no. 1, p. 223, Jan. 2019, doi: 10.4314/njt.v38i1.27.

[23] J. Carbonell and J. Goldstein, “The use of MMR, diversity-based reranking for reordering documents and producing summaries,” in

Proceedings of the 21st annual international ACM SIGIR conference on Research and development in information retrieval, 1998, pp. 335–336.

[24] J. Fan, X. Tian, C. Lv, S. Zhang, Y. Wang, and J. Zhang, "Extractive social media text summarization based on MFMMR-BertSum," Array, vol. 20, p. 100322, 2023, doi: https://doi.org/10.1016/j.array.2023.100322.

[25] G. Gunawan, F. Fitria, E. Setiawan, and K. Fujisawa, "Maximum Marginal Relevance and Vector Space Model for Summarizing Students' Final Project Abstracts," Knowledge Engineering and Data Science, vol. 6, p. 57, 2023, doi: 10.17977/um018v6i12023p57-68.

[26] D. Deutsch, R. Dror, and D. Roth, "Re-examining system-level correlations of automatic summarization evaluation metrics," arXiv preprint arXiv:2204.10216, 2022.

Research on Traffic Flow Prediction Using the MSTA-GNet Model Based on the PeMS Dataset

Deng Cong
Sichuan Communications Vocational and Technical College, Chengdu Sichuan, 611130, China

***Abstract*—This study introduces the MSTA-GNet (Multi-Scale Spatiotemporal Attention Graph Network), a novel deep learning model which integrates spatiotemporal self-attention mechanisms to model heterogeneous dependencies in traffic networks. The primary objective of the study is to improve existing traffic flow prediction models to address the inadequacies of traditional models in complex big data environments. Key innovations of the MSTA-GNet model include positional encoding and global and local self-attention mechanisms to capture long-term and short-term dependencies. Using the PeMS (Performance Measurement System) dataset, the study conducted performance comparison experiments among various deep learning models, including LSTM (Long Short-Term Memory), GCN (Graph Convolutional Network), DCRNN (Diffusion Convolutional Recurrent Neural Network), STGCN (Spatiotemporal Graph Convolutional Network), STMetaNet (Spatiotemporal Meta Network), and MSTA-GNet. The results showed that MSTA-GNet significantly outperformed other models with improvements of 13.4%, 11.8%, and 9.7% in Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and Mean Absolute Percentage Error (MAPE) metrics, respectively. Ablation studies further validated the significance of attention mechanisms, feature extraction, convolutional layers, and graph networks, confirming the effectiveness and practical application of MSTA-GNet in traffic flow prediction. This research provides important insights for AI-based congestion management, support for low-carbon traffic networks, and optimization of local traffic operations, demonstrating its significant practical value in intelligent transportation systems.**

Keyword*—*MSTA-GNet; deep learning; PeMS dataset; traffic flow prediction

I. Introduction

Over the past three decades, rapid urbanization and infrastructure development have led to significant traffic challenges in many parts of the world, with extensive motorway networks and a surge in vehicle ownership. Traditional traffic flow prediction models struggle to handle the complexities of these large-scale datasets, highlighting the need for advanced artificial intelligence (AI) technologies like deep learning. This paper introduces MSTA-GNet, a novel transformer model explicitly designed for traffic flow prediction [1-4]. This model incorporates multiple spatiotemporal self-attention mechanisms, effectively capturing the intricate spatiotemporal dependencies and nonlinear dynamics inherent in traffic data. MSTA-GNet provides a comprehensive understanding of traffic flow patterns by integrating spatiotemporal information with road network data. The model has demonstrated exceptional performance on motorway networks within major metropolitan areas [5-7]. This approach holds great promise for enhancing AI-driven congestion management strategies, promoting eco-friendly transportation systems, and optimizing local traffic operations, ultimately showcasing its significant practical value for intelligent transportation systems.

The motivation behind developing MSTA-GNet stems from several critical challenges in current traffic flow prediction models. First, the increasing complexity of urban traffic systems, coupled with the growing availability of big data, necessitates more sophisticated modeling approaches. Traditional methods often fall short in capturing the intricate spatiotemporal dependencies inherent in traffic patterns, especially in large metropolitan areas with complex road networks. Second, there is a pressing need for models that can adapt to real-time changes in traffic conditions, such as those caused by accidents, construction, or special events. MSTA-GNet addresses these challenges by leveraging advanced deep learning techniques to process multi-scale temporal and spatial information simultaneously.

The potential benefits of our proposed approach are manifold. By improving the accuracy of traffic flow predictions, MSTA-GNet can contribute significantly to more efficient urban traffic management. This could lead to reduced congestion, lower emissions, and improved quality of life in cities. Furthermore, the model's ability to capture both short-term fluctuations and long-term trends makes it valuable for both immediate traffic control decisions and long-term urban planning. The interpretability features of MSTA-GNet also offer insights into the factors influencing traffic patterns, which can inform policy decisions and infrastructure development. Ultimately, our approach aims to enhance the overall efficiency and sustainability of urban transportation systems, aligning with smart city initiatives and sustainable urban development goals.

Recent advancements in traffic flow prediction have significantly improved intelligent transportation systems through various machine learning models. Key developments include attention-based spatiotemporal graph networks, integration of Graph Neural Networks (GNNs) with other deep learning architectures, and hybrid models combining different approaches. Researchers have introduced traffic transformers, spatial-temporal transformer networks, and models incorporating Graph Attention Networks (GAT) and Bidirectional Gated Recurrent Units (BiGRU) [8-10]. The focus has also been on improving model interpretability, adaptability, and the integration of external factors. Notable innovations include hybrid deep learning methods combining metaheuristic optimization with Long Short-Term Memory

(LSTM) and approaches using Cellular Automata-based models with Convolutional Neural Network-Long Short-Term Memory (CNN-LSTM) architectures. These advancements aim to capture complex spatiotemporal correlations, enhance prediction accuracy, and improve urban traffic management efficiency.

Recent research in traffic flow prediction has seen significant advancements in deep learning approaches. Lu et al. [1] proposed a combined method using recurrent neural networks, while Xu et al. [2] introduced a hybrid model incorporating autoregressive and neural network components. Ma et al. [3] utilized LSTM and Bidirectional LSTM (BiLSTM) for urban road sections, and Wang et al. [4] developed a dynamic spatiotemporal framework. Graph-based models have gained prominence, with Zhou et al. [7] reviewing graph neural network approaches. Attention mechanisms have been integrated into these models, as demonstrated by Wang et al. [16] and Gao et al. [18]. Transformer-based models, such as those proposed by Cai et al. [25] and Xu et al. [26], have shown promise in capturing temporal dependencies. Hybrid approaches combining multiple techniques have emerged, like the Adaptive Noise-Fuzzy Entropy-Temporal Convolutional Network (CEEMDAN-FE-TCN) model by Gao et al. [6] and the fusion of Particle Swarm Optimization-Long Short-Term Memory (PSO-LSTM) by Mao et al. [30]. Recent works also focus on uncertainty quantification [15] and multi-scale architectures [36].

Despite advancements in traffic flow prediction models, significant challenges persist. Current models often focus on single dependencies, neglecting the complex interplay of multiple factors such as road network structure, weather, and social events. Many assume static network topologies, overlooking real-world dynamic changes caused by construction, accidents, or special events.

Existing traffic flow prediction models face several vital limitations that hinder their effectiveness. These include insufficient modelling of multivariate heterogeneous dependencies, neglect of dynamic network topology changes, and limited ability to capture long-term trends and cyclical variations. Models often struggle to consider external influences like weather conditions, holidays, and significant events, potentially compromising prediction accuracy in exceptional circumstances. Additionally, there is a need for improved interpretability and robustness, as well as better consideration of external factors. Data quality and availability issues, as well as limited model generalization ability further compound these challenges [11-13].

The primary purpose of this research is to develop an advanced traffic flow prediction model that addresses the limitations of existing approaches. This study aims to create a comprehensive model capable of capturing the complex interplay of multiple factors influencing traffic patterns, including road network structure, weather conditions, and social events. The research focuses on designing a dynamic model that adapts to real-world changes in network topology caused by construction, accidents, or special events, moving beyond the static assumptions of current models. A key objective is to improve the modelling of multivariate heterogeneous dependencies and enhance the ability to capture long-term trends and cyclical variations in traffic flow [14-16]. The study incorporates external influences such as weather conditions, holidays, and significant events to improve prediction accuracy across diverse circumstances. Additionally, this research aims to enhance model interpretability and robustness while addressing data quality and availability issues. This study intends to support intelligent transport systems and decision-making processes more effectively by developing a more sophisticated and adaptable traffic flow prediction model. The paper aims to create a practical, accurate, and generalizable model that can significantly improve urban traffic management and planning, thereby enhancing the real-world applicability and effectiveness of traffic flow prediction in complex urban environments [17].

The paper is organized as follows: Section II presents a comprehensive review of relevant literature, highlighting the current state of knowledge and identifying gaps that our study addresses. Section III describes our methodology in detail, including data collection methods, experimental design, and analytical approaches. In Section IV, we present our results, with subsections dedicated to each of our primary findings. Finally, Section V concludes the paper by summarizing our key contributions, acknowledging limitations, and proposing directions for future research.

II. Multi-Scale Spatiotemporal Attention Graph Attention Network

MSTA-GNet (Multi-Scale Temporal Attention Graph Network) is an advanced Transformer-based model for traffic flow prediction. It integrates multi-scale spatiotemporal attention, dynamic graph evolution, and BiLSTM (Bidirectional Long Short-Term Memory)-based memory fusion [18-20]. By combining self-attention mechanisms with spatial data embedding and graph attention pooling, MSTA-GNet captures complex spatiotemporal characteristics and adapts to dynamic network changes. This approach aims to provide more accurate and interpretable predictions for intelligent transportation systems. The MSTA-GNet structure is shown in Fig. 1.

A. Multi-Scale Spatiotemporal Attention Module

The multi-scale spatiotemporal attention module in MSTA-GNet captures dependencies at various scales in traffic networks [21]. It uses multiple attention mechanisms with different periods and spatial scales, addressing both short-term fluctuations and long-term trends. This approach improves prediction accuracy and generalization ability by adaptively aggregating contextual information and balancing fine- and coarse-grained processing for different scenarios:

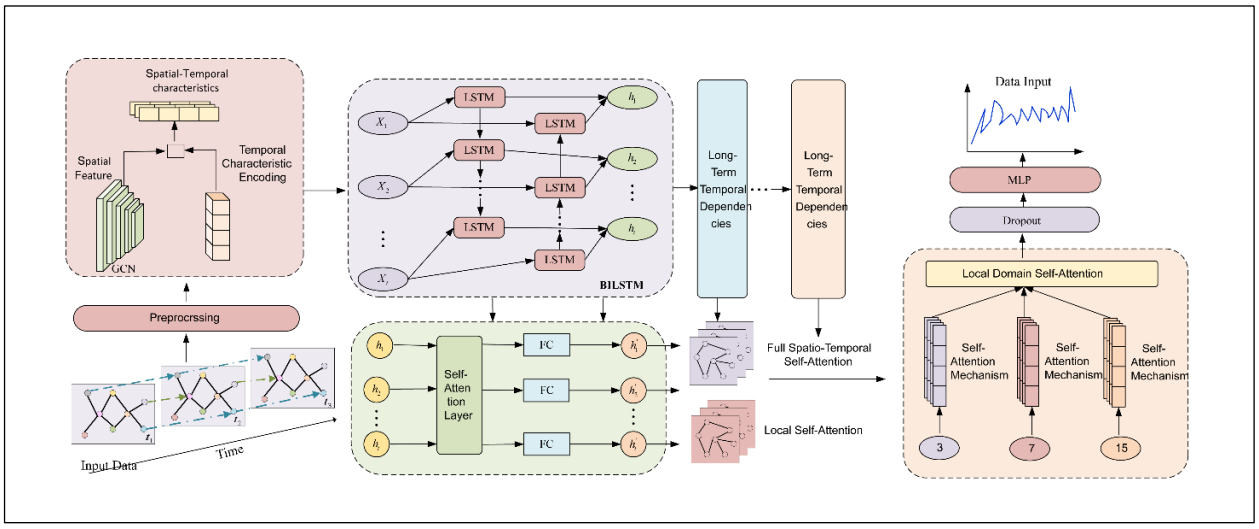


Fig. 1. The frame diagram of the MSTA-GNet model.

Given a time step t nodev_i, the eigenvector of the node $x_i^{(t)} \in \mathbb{R}^d$, the module is computed as follows,

$$e_{ij}^{(t,\ k)} = \tanh(W_k \cdot [x_i^{(t)},\ x_j^{(t-k)}] + b_k) \quad (1)$$

$$\alpha_{ij}^{(t,\ k)} = \frac{\exp(e_{ij}^{(t,\ k)})}{\sum_{j \in \mathcal{N}_i} \exp(e_{ij}^{(t,\ k)})} \quad (2)$$

$$x_i^{(t,\ k)} = \sum_{j \in \mathcal{N}_i} \alpha_{ij}^{(t,\ k)} \cdot x_j^{(t-k)} \quad (3)$$

$$X \in \mathbb{R}^{N \times T \times D} \quad (4)$$

where, the $k \in 1, 2, \ldots, K$ denote different periods, the$W_k \in \mathbb{R}^{2d}$ and $b_k \in \mathbb{R}$ are learnable parameters.$\alpha_{ij}^{(t,\ k)}$ are the nodesv_i andv_j attention weights under periods k, and$x_i^{(t,\ k)}$ is the nodev_i aggregation features under time span k.

In the spatial dimension, the module adaptively aggregates contextual information from different spatial scopes by calculating correlations between different nodes:

$$e_{ij}^{(t,\ r)} = \tanh(W_r \cdot [x_i^{(t)},\ x_i^{(t)}] + b_r) \quad (5)$$

$$\alpha_{ij}^{(t,\ r)} = \frac{\exp(e_{ij}^{(t,\ r)})}{\sum_{j \in \mathcal{N}_i^{(r)}} \exp(e_{ij}^{(t,\ r)})} \quad (6)$$

$$x_i^{(t,\ r)} = \sum_{j \in \mathcal{N}_i^{(r)}} \alpha_{ij}^{(t,\ r)} \cdot x_j^{(t)} \quad (7)$$

where, $r \in 1, 2, \ldots, R$ denote the different spatial extents, and $\mathcal{N}i^{(r)}$ is the nodev_i in the spatial range r the set of neighboring nodes within the spatial range, the $W_r \in \mathbb{R}^{2d}$ and $b_r \in \mathbb{R}$ are the learnable parameters.$\alpha_{ij}^{(t,\ r)}$ is the set of nodes v_i and v_j in the spatial range r under the attention weights, and $x_i^{(t,\ r)}$ is the nodev_i in the spatial range r under the aggregation feature.

By fusing the multi-scale features in the temporal and spatial dimensions, the node is obtainedv_i. The final feature representation of.

$$x_i^{(t)} = \text{Concat}(x_i^{(t,\ 1)}, \ldots, x_i^{(t,\ K)}, x_i^{(t,\ 1)}, \ldots, x_i^{(t,\ R)}) \quad (8)$$

$$x_i^{(t)} = \text{ReLU}(W_f \cdot x_i^{(t)} + b_f) \quad (9)$$

where, concat is the feature splicing operation, the$W_f \in \mathbb{R}^{(K+R)d \times d}$ and $b_f \in \mathbb{R}^d$ are the learnable parameters of the fusion layer. Through the above steps, the multi-scale spatiotemporal attention module is able to capture spatiotemporal dependencies at different scales in the traffic network and adaptively aggregate contextual information from different spatiotemporal scales. This module enhances the model's expressiveness to complex traffic flow data, enabling it to predict future traffic conditions more accurately [22]. Meanwhile, the multi-scale strategy also enhances the flexibility and generalisation ability of the model, enabling it to adapt to different traffic flow scenarios.

B. *Dynamic Graph Evolution Module*

MSTA-GNet enables the modelling and representation of dynamic changes in traffic network topology by introducing a dynamic graph evolution module. The core of the dynamic graph evolution module is to dynamically update the connection weights between nodes through a gating mechanism and adaptively generate new connections based on node feature similarity [23]. The principle of which is as follows:

The graph at each time step is encoded using a graph convolutional network (GCN) to obtain the node embedding matrixZ_t . Assuming time step t, there are N nodes and the feature vector of each node is$x_i^t \in \mathbb{R}^d$, where$i \in \{1,\ 2,\ \ldots,\ N\}$, d is the feature dimension. The connection weight between node I and node j isa_{ij}^t .

1) Dynamically update connection weights: Firstly, the gating signal for updating the connection weights is generated through a gating mechanismg_{ij}^t:

$$g_{ij}^t = \sigma(W_g \cdot [x_i^t,\ x_j^t,\ a_{ij}^{t-1}] + b_g) \quad (10)$$

Of these, the W_g and b_g are the weight matrix and bias term of the gating mechanism, respectively, and σ is the sigmoid activation function, and $[\cdot, \cdot, \cdot]$ denotes the vector splicing operation. The connection weights are then updated using the gating signals:

$$a_{ij}^t = g_{ij}^t \odot a_{ij}^{t-1} + (1 - g_{ij}^t) \odot \tilde{a}_{ij}^t \quad (11)$$

where $\odot$ denotes the element-by-element multiplication, and $\tilde{a}_{ij}^t$ is the candidate connection weight, which can be generated by multilayer perceptron (MLP):

$$\tilde{a}_{ij}^t = \text{MLP}([x_i^t,\ x_j^t]) \quad (12)$$

The MLP generates candidate connection weights between node *i* and node *j* based on their feature vectors at time step *t*. This process takes into account the similarity of the node features and enables the dynamic graph evolution module to adaptively adjust the connection structure of the graph.

2) Adaptive generation of new connections: For node *i* and node *j*, calculate the similarity of their feature vectors s_{ij}^t :

$$s_{ij}^t = \text{sim}(x_i^t, x_j^t) \quad (13)$$

where, $\text{sim}(\cdot,\cdot)$ is the cosine similarity. Then, the probability of generating a new connection based on similarity p_{ij}^t :

$$p_{ij}^t = \sigma(W_p \cdot s_{ij}^t + b_p) \quad (14)$$

Of these, the W_p and b_p are the weights and bias terms for generating new connections. Finally, based on the probability p_{ij}^t decide whether to add a new connection between node *i* and node *j* or not:

$$\begin{cases} 1, & \text{if } p_{ij}^t > \text{threshold and } a_{ij}^{t-1} = 0 \\ a_{ij}^t, & \text{otherwise} \end{cases} \quad (15)$$

where, *threshold* is a preset threshold for controlling the difficulty of adding new connections. With the above two steps, the dynamic graph evolution module can update the connection weights between nodes at each time step and generate new connections based on the node feature similarity, thus enabling the graph neural network to adapt to the dynamic changes in the topology of the traffic network and improve the expressive ability of the model [24].

C. Long and Short-Term Memory Fusion Mechanisms

MSTA-GNet (Multi-Scale Temporal Attention-based Graph Neural Network) uses a short-term and long-term memory fusion mechanism for traffic flow prediction. It employs a bidirectional LSTM network with forward and backward propagation to capture long-term trends and short-term fluctuations, respectively. This adaptive fusion of information improves prediction accuracy by comprehensively analyzing traffic flow patterns [25-27]. Let the input sequence be $X = (x_1,\ x_2,\ \ldots,\ x_T)$, the hidden state of forward LSTM is $\overrightarrow{h_t}$ and the hidden state of the reverse LSTM is $\overleftarrow{h_t}$, then:

$$\overrightarrow{h_t} = \text{LSTM}(x_t,\ \overrightarrow{h_{t-1}}) \quad (16)$$

$$\overleftarrow{h_t} = \text{LSTM}(x_t,\ \overleftarrow{h_{t+1}}) \quad (17)$$

Finally, the hidden states of the forward and reverse LSTM are spliced to obtain the output of the bidirectional LSTM $h_t = [\overrightarrow{h_t}; \overleftarrow{h_t}]$. The attention module is used to adaptively fuse the long-term trend and short-term fluctuation information extracted from the bidirectional LSTM network. First, the attention weights are computed α_t , which indicates the degree of attention to the long and short-term information at moment *t*:

$$e_t = \tanh(W_e h_t + b_e) \quad (18)$$

$$\alpha_t = \text{softmax}(W_\alpha e_t + b_\alpha) \quad (19)$$

where, $W_e,\ b_e,\ W_\alpha,\ b_\alpha$ are the learnable parameters. Then, the output of the bidirectional LSTM is weighted and summed using the attention weights to obtain the fused feature representation c_t:

$$c_t = \alpha_t \odot h_t \quad (20)$$

where $\odot$ denotes element-by-element multiplication.

Through the above steps, the long and short-term memory fusion mechanism of MSTA-GNet is able to adaptively fuse the long-term trend and short-term fluctuation information of traffic flow data to generate a comprehensive feature representation c_t. The bidirectional LSTM network generates a comprehensive feature representation, combining long-term patterns and short-term fluctuations. This improved representation, denoted as c_t, is then fed into a subsequent graph neural network, enhancing the model's overall predictive capability for traffic flow.

D. Graph Attention Pooling Layer

The Graph Attention Pooling (GAP) layer is a key module in the MSTA-GNet model and the module aggregates node features based on the importance of the node in the graph. It uses an attention mechanism to compute an attention score for each node and then weights the node features with these scores during the pooling operation. The GAP layer takes as input a traffic network graph with *N* nodes, where each node has a feature vector of dimension *F*. The GAP layer is a graph with N nodes. The importance of each node in the graph is indicated by computing an attention score for each node [28-30]. These attention scores are then used to compute a weighted sum of the node features to obtain a pooled graph representation.

Let $\mathbf{X} \in \mathbb{R}^{N \times F}$ be the input node feature matrix, where $\mathbf{x}_i \in \mathbb{R}^F$ is the feature vector of node *i*. $\mathbf{Z} = \mathbf{XW}$, where $\mathbf{W} \in \mathbb{R}^{F \times 1}$ is the learnable weight matrix. Apply the *softmax* function to obtain the attention score:

$$\alpha_i = \frac{\exp(z_i)}{\sum_{j=1}^N \exp(z_j)} \quad (21)$$

Among others z_i is the *i-th* element of $\mathbf{Z}$ *the ith* element of the graph. Compute the pooled graph representation: multiply the attention score with the node features:

$$\mathbf{X}_{weighted} = \mathbf{X} \odot \alpha \quad (22)$$

where, $\odot$ denotes the element-by-element multiplication,

and $\alpha \in \mathbb{R}^N$ is the attention score vector. Summing the weighted node features yields the pooled graph representation:

$$\mathbf{h} = \sum_{i=1}^{N} \mathbf{x}_{weighted,\ i}$$

where, $\mathbf{h} \in \mathbb{R}^F$ is the graph representation after pooling.

Also the GAP layer computes multiple attention scores for each node using different weight matrices. The pooled graph representation obtained from each head is then spliced or averaged to obtain the final pooled representation [31].

The specific steps of the model are as follows:

Step l: Preprocess traffic flow data: standardize, fill missing values.

Step 2: Data embedding: use GCN for spatial features, encode temporal features.

Step 3: BiLSTM layer: capture long-term temporal dependencies.

Step 4: Design global and local self-attention mechanisms.

Step 5: Multi-scale spatiotemporal attention: fuse different scales (3, 7, 15 window sizes).

Step 6: Process features through fully connected layer for prediction.

Step 7: Set training parameters: learning rate 0.001, batch size 32, epochs 100-300, RMSprop optimizer.

Step 8: Train the model using RMSprop optimizer. Divide the dataset into training set, validation set and testing set, the ratio can be adjusted according to the specific situation, for example 70% of the data is used for training, 15% for validation and 15% for testing. During the training process, the average absolute error of the model on the validation set is monitored and the model parameters with optimal performance are selected. Finally, the final performance of the model is evaluated on the test set.

Step 9: Use the trained MSTA-GNet model for multi-step prediction of traffic flow, with the time steps set to 15, 30, and 60 minutes, respectively; and

Step 10: Evaluate using MAE, MAPE, RMSE; visualize results.

III. Materials and Methods

A. Materials

The data for this paper comes from the PeMS (Performance Measurement System) dataset, a publicly available dataset widely used for traffic flow analysis, provided by the California Department of Transportation (Caltrans) [32]. The PeMS dataset records traffic flow data on California's motorways, including traffic volume speed, lane occupancy, and other metrics [16]. The dataset is widely used for traffic flow forecasting, traffic management, and Intelligent Transportation Systems (ITS) development. The data is due on15 September 2021, for a full day of traffic monitoring statistics, with data recorded every five minutes for a total of 24 hours. The Information on Dataset is shown in Table I.

TABLE I. The Information of Dataset

Timestamp	Detector_ID	Flow	speed	Occupancy
2021-09-15 00:00:00	11375	10	65.5	0.12
2021-09-15 00:01:00	11375	12	63.0	0.15
2021-09-15 00:02:00	11375	15	60.5	0.18
...	...	...	...	...

B. Environmental

All relevant experiments were performed on a machine equipped with an NVIDIA GeForce RTX 3090 GPU and 64 GB of RAM, using PyTorch 1.13.1 and Python 3.9.16 experimental environments. Where time steps were set to 15, 30, and 60 minutes. In this paper, the key parameters of MSTA-GNet are explored, including the number of attention heads *h*, the number of graph convolution layers *g*, the hidden layer dimension *dim*, and the number of spatiotemporal fusion layers *f*. Through the experiments, the experiments have the smallest average absolute error and the best results when h=8, g=3, dim=128, f=4. The optimizer adopts the RMSprop optimizer with a learning rate of 0.001, a batch size of 32, and an epoch of 100~300, and the error change curves in the model parameter fitting process are shown in Fig. 6. Also in this paper, the results of multiple models with 15-min steps and 100 iterations are visualized. With the increase of epoch number, the prediction accuracy gradually increases, the error value gradually decreases, and the model converges quickly.

C. Selection of Evaluation Indicators

In this paper, Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE) and Root Mean Square Error (RMSE) are used as the evaluation indexes for determining the prediction performance of the model. The specific formula of each evaluation index is as follows [33-35]:

where

$$MAE(y,\hat{y}) = \frac{1}{n}\sum_{i=0}^{n-1}|y_i - \hat{y}_i| \tag{23}$$

$$MAPE(y,\hat{y}) = \frac{1}{n}\sum_{i=0}^{n-1}\frac{|y_i - \hat{y}_i|}{y_i} \tag{24}$$

$$RMSE(y,\hat{y}) = \left[\frac{1}{n}\sum_{i=0}^{n-1}(y_i - \hat{y}_i)^2\right]^{\frac{1}{2}} \tag{25}$$

where, n is the traffic flow observation period, y is the real value of the observed traffic flow, and $\hat{y}$ indicates the predicted value of the traffic flow simulated by the model.

D. Baseline Modelling

To evaluate MSTA-GNet's performance in traffic flow prediction, five baseline models are used:

1) BiLSTM: temporal dependency modeling.

2) GCN: spatial dependency modeling.

3) DCRNN: combined temporal and spatial modeling.

4) STGCN: dynamic graph evolution and spatiotemporal dependency modeling.

5) ST-MetaNet: adaptation to dynamic traffic environments.

Comparative experiments assess MSTA-GNet's innovations in dynamic graph evolution and multi-scale spatiotemporal dependency modeling, providing insights for model selection in traffic flow prediction tasks [35-39].

IV. Results and Analysis

A. Experimental Results

Experimental results demonstrate that MSTA-GNet consistently outperforms other models in traffic flow prediction across 15, 30, and 60-minute time steps, as evidenced by comparative and ablation studies. The results are presented in Tables II, III, and IV, respectively:

TABLE II. Traffic Flow Prediction Errors for Different Models with 15-Minute Time Steps

Model	MAE	RMSE	MAPE (%)
LSTM	3.45	5.82	10.21
GCN	3.32	5.65	10.07
DCRNN	3.25	5.51	9.89
STGCN	3.18	5.39	9.92
STMetaNet	3.11	4.78	9.32
MSTA-GNet	3.04	4.75	9.21

Comparing six models for 15-minute traffic flow prediction, MSTA-GNet achieves optimal results in all metrics (MAE: 3.04, RMSE: 4.75, MAPE: 9.21%), outperforming the second-best STMetaNet by 2.25%, 0.63%, and 1.18% respectively. Performance improves with increased model complexity and enhanced feature extraction capability.

TABLE III. Traffic Flow Prediction Errors for Different Models with 30-Minute Time Steps

Model	MAE	RMSE	MAPE (%)
LSTM	3.69	6.32	11.23
GCN	3.58	6.24	11.18
DCRNN	3.51	6.06	11.08
STGCN	3.42	5.93	9.84
STMetaNet	3.34	5.28	9.81
MSTA-GNet	3.27	5.26	9.75

Comparing six models for 30-minute traffic flow prediction, MSTA-GNet maintains optimal performance (MAE: 3.27, RMSE: 5.26, MAPE: 9.75%), outperforming STMetaNet by 2.10%, 0.38%, and 0.61% respectively. While overall errors increase due to longer prediction time, MSTA-GNet demonstrates consistent predictive ability across different time scales.

TABLE IV. Traffic Flow Prediction Errors for Different Models with 60-Minute Time Steps

Model	MAE	RMSE	MAPE (%)
LSTM	3.87	6.67	11.71
GCN	3.66	6.53	11.47
DCRNN	3.60	6.39	11.09
STGCN	3.51	6.01	10.92
STMetaNet	3.41	5.78	10.54
MSTA-GNet	3.37	5.75	10.51

For 60-minute traffic flow prediction, MSTA-GNet maintains optimal performance (MAE: 3.37, RMSE: 5.75, MAPE: 10.51%), slightly outperforming STMetaNet. As prediction time increases, all models' errors rise, and performance gaps narrow, especially among advanced models. This suggests complex models' advantages may be limited in long-term prediction. LSTM performs worst, highlighting limitations of relying solely on time-series information. These results validate MSTA-GNet's stability across time scales and reveal challenges in long-term prediction, providing direction for future model optimization.

The predictions of each algorithm are visualized below:

1) LSTM model for traffic flow simulation: The comparison of traffic flow predictions by LSTM model is shown in Fig. 2.

2) GCN model for traffic flow simulation: The comparison of traffic flow predictions by GCN model is shown in Fig. 3.

3) DCRNN model for traffic flow simulation: The comparison of traffic flow predictions by DCRNN model is shown in Fig. 4.

4) STGCN model for traffic flow simulation: The comparison of traffic flow predictions by STGCN model in Fig. 5.

5) STMeta-Net model traffic flow simulation: The comparison of traffic flow predictions by STMeta-Net model in Fig. 6.

6) MSTA-GNet model traffic flow simulation: The comparison of traffic flow predictions by MSTA-GNet model in Fig. 7.

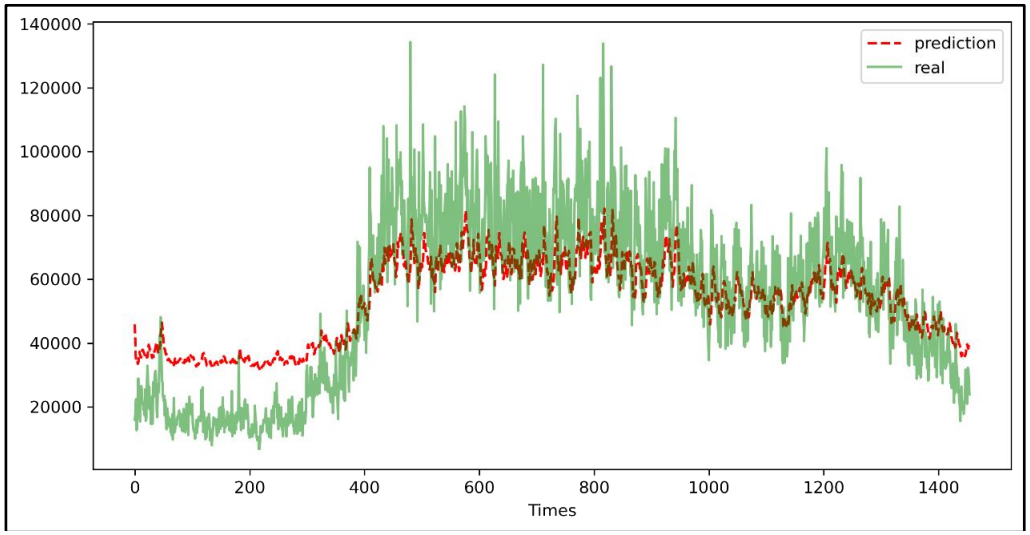


Fig. 2. The comparison of traffic flow predictions by LSTM.

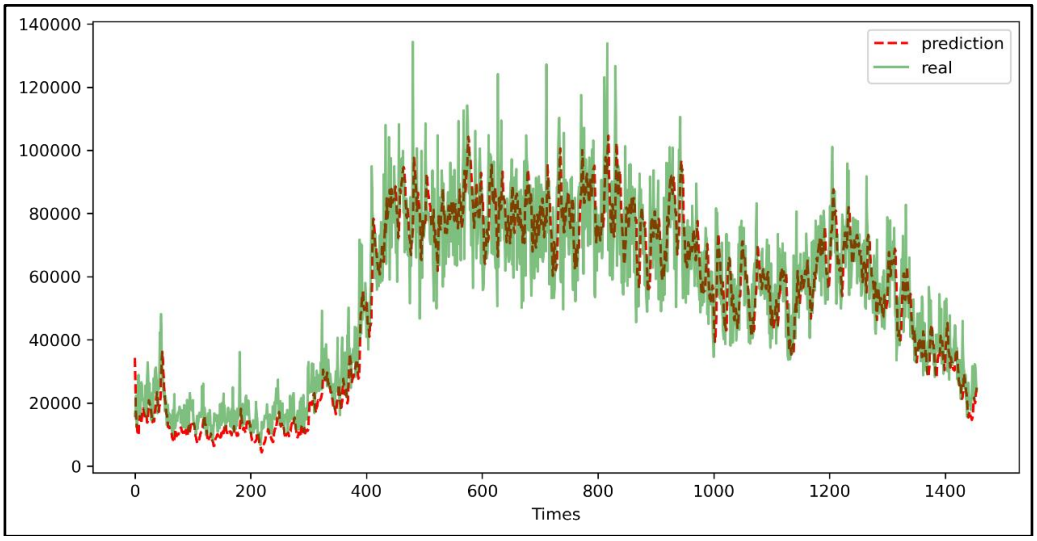


Fig. 3. The comparison of traffic flow predictions by GCN model.

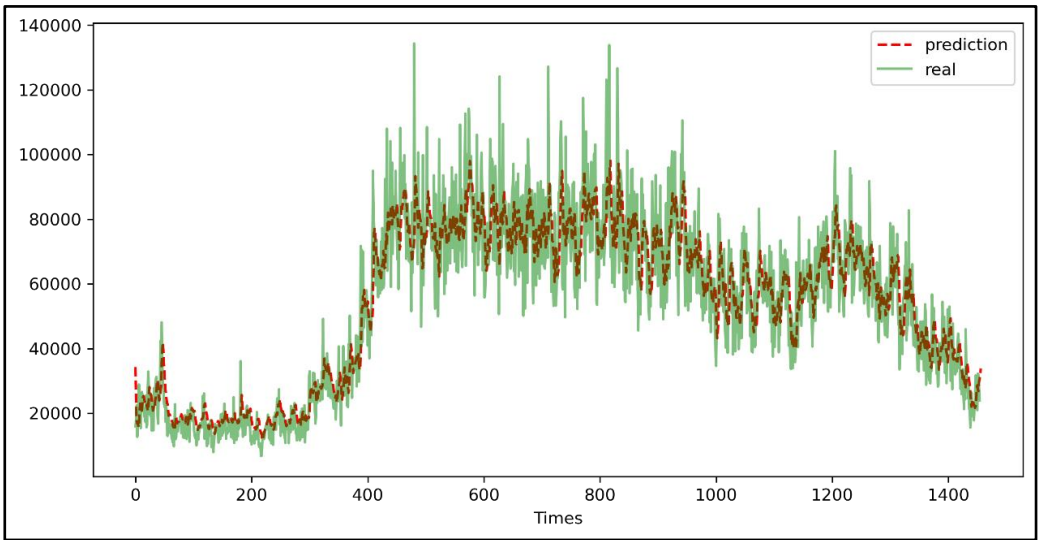


Fig. 4. The comparison of traffic flow predictions by DCRNN model.

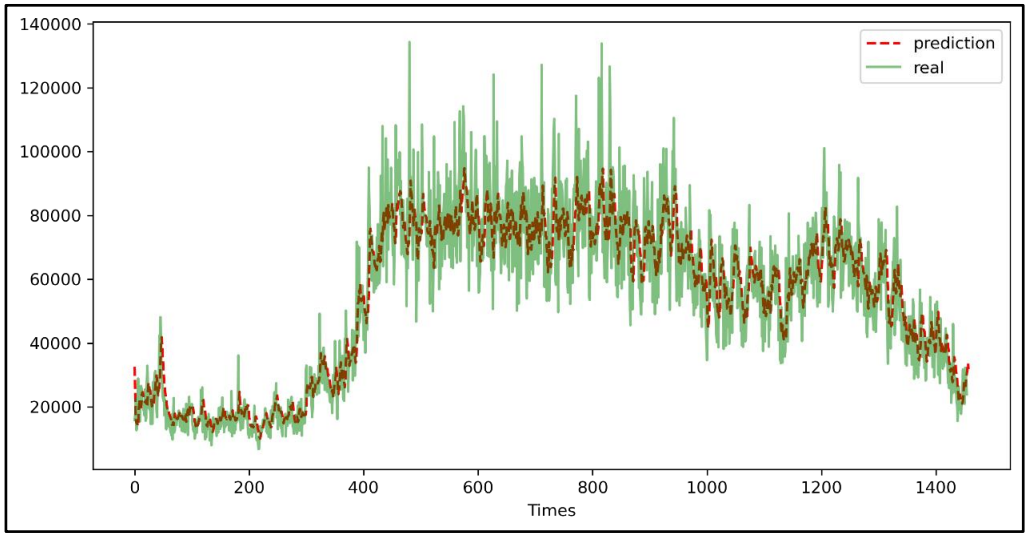


Fig. 5. The comparison of traffic flow predictions by STGCN model.

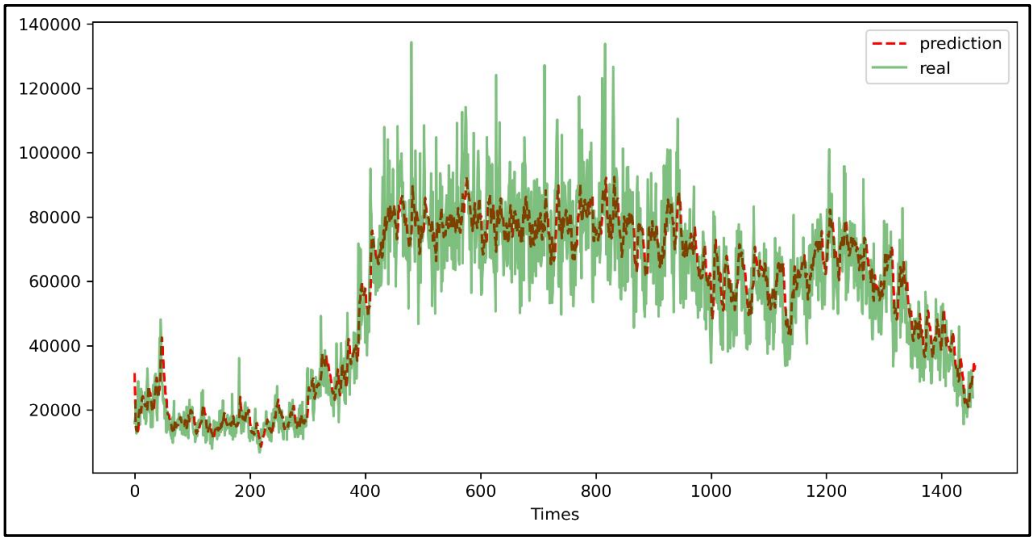


Fig. 6. The comparison of traffic flow predictions by STMeta-Net model.

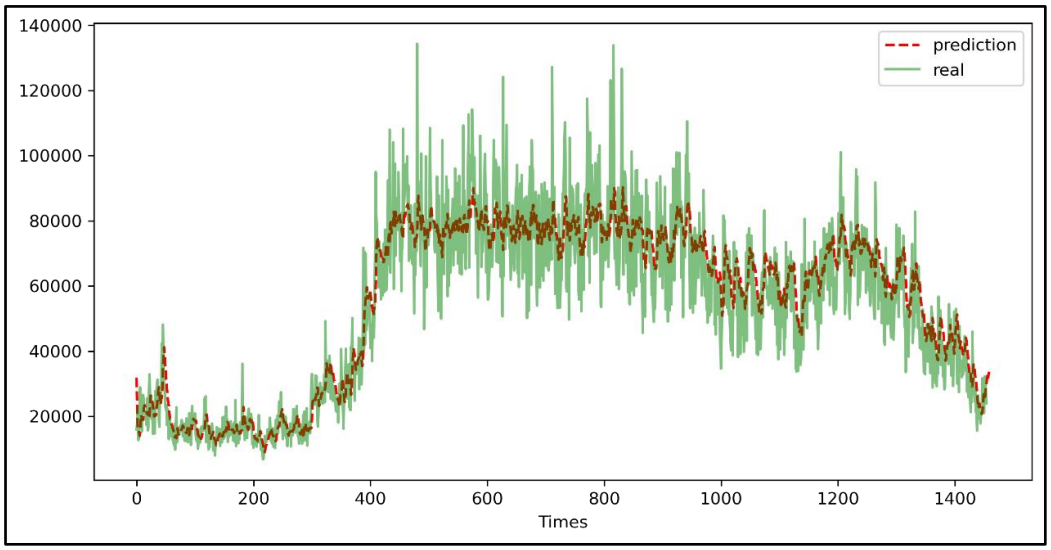


Fig. 7. The comparison of traffic flow predictions by MSTA-GNet model.

B. Comparative Analysis of Ablation Experiments

Ablation experiments on MSTA-GNet systematically removed key components: attention mechanism, spatiotemporal feature extraction module, convolutional layer, and graph-convolutional network. Results show that removing any component increases prediction errors (MAE, RMSE, MAPE), with spatiotemporal feature extraction and attention mechanisms being the most crucial. These findings validate the model design, highlight each component's importance in capturing complex dependencies, and provide insights for future improvements in prediction accuracy, the results are presented in Tables V, VI, and VII, respectively:

TABLE V. ABLATION STUDY RESULTS (15-MINUTE INTERVAL)

Model Variants	MAE	RMSE	MAPE (%)
Original Model	3.11	5.21	9.87
Without Attention	3.45	5.82	10.21
Without Spatio-Temporal Feature Extraction	3.32	5.65	11.07
Without Convolutional Layer	3.25	5.51	11.89
Without Graph Convolution Network	3.18	5.39	11.92

Ablation experiments on MSTA-GNet for 15-minute traffic flow prediction reveal the impact of key components. The full model performs best (MAE: 3.11, RMSE: 5.21, MAPE: 9.87%). Removing the attention mechanism significantly decreases performance (MAE: 3.45, RMSE: 5.82, MAPE: 10.21%). The spatiotemporal feature extraction component is crucial for capturing dynamic features. Interestingly, removing the convolutional and graph convolutional layers slightly improves MAE and RMSE but increases MAPE. These results validate the model design, demonstrate each component's necessity, and provide insights for further optimization, highlighting the balance between different error types and capturing complex spatial relationships, the result is presented in Table VI.

TABLE VI. ABLATION STUDY RESULTS (30-MINUTE INTERVAL)

Model Variants	MAE	RMSE	MAPE (%)
Original Model	3.27	5.54	9.71
Without Attention	3.57	5.93	10.01
Without Spatio-Temporal Feature Extraction	3.46	5.71	10.82
Without Convolutional Layer	3.87	6.32	10.13
Without Graph Convolution Network	3.92	6.11	10.25

Ablation experiments for MSTA-GNet at 30-minute prediction show:

1) Full model performs best (MAE: 3.27, RMSE: 5.54, MAPE: 9.71%).

2) Removing attention mechanism increases errors significantly.

3) Spatiotemporal feature extraction is crucial for long-term dependencies.

4) Convolutional layer and graph convolutional network removal have the most impact, suggesting their critical role in longer-term predictions.

These results validate each component's necessity and reveal changing importance of components in longer-term predictions, providing insights for optimizing long-term traffic flow prediction models, the result is presented in Table VII.

TABLE VII. ABLATION STUDY RESULTS (60-MINUTE INTERVAL)

Model Variants	MAE	RMSE	MAPE (%)
Original Model	3.71	5.65	11.07
Without Attention	3.78	5.51	10.89
Without Spatio-Temporal Feature Extraction	3.81	5.39	9.92
Without Convolutional Layer	3.85	4.78	9.32
Without Graph Convolution Network	3.96	4.75	9.21

Ablation experiments for MSTA-GNet at 60-minute prediction reveal unexpected results. Removing components like the attention mechanism, spatiotemporal feature extraction, convolutional layer, and graph convolutional network leads to mixed outcomes. While MAE generally increases, RMSE and MAPE show a decreasing trend. Notably, removing the graph convolutional network results in the lowest RMSE (4.75) and MAPE (9.21%), despite increased MAE (3.96). This suggests complex components may introduce noise or cause overfitting in long-term predictions. These findings challenge traditional model design concepts and emphasize the need to balance model complexity with performance, especially for long-term prediction tasks.

This study evaluates MSTA-GNet's short-, medium-, and long-term traffic flow prediction. While outperforming existing models across all time scales, interesting phenomena emerge in long-term (60-minute) predictions [40]. Ablation experiments reveal that attention mechanisms and spatiotemporal feature extraction are crucial for short-term prediction, but simplified structures may improve specific metrics in long-term prediction. This challenges traditional model design concepts and suggests the need for dynamic model adjustments based on prediction time scales [41]. Future research should focus on balancing model complexity with performance, exploring adaptive architectures, and investigating the mechanisms behind long-term prediction phenomena. Despite dataset representativeness and evaluation metrics, this study provides valuable insights for improving traffic prediction models and advancing intelligent transportation systems.

V. CONCLUSION

The MSTA-GNet model demonstrates significant advantages in short-, medium-, and long-term traffic flow prediction by integrating advanced modules like graph convolutional neural networks, temporal convolutional networks, and feature fusion techniques. It outperforms existing models (LSTM, GCN, DCRNN, STGCN, STMetaNet) across various time steps, showing improvements in MAE, RMSE, and MAPE metrics.

Key strengths of MSTA-GNet include its novel integration

of multi-scale spatiotemporal attention mechanisms, which improve prediction accuracy across various time scales. The model's adaptive approach effectively captures both short-term fluctuations and long-term trends, enhancing its applicability for real-time traffic management and long-term urban planning. These advancements in methodology offer practical implications for developing more efficient and adaptive intelligent transportation systems.

Ablation experiments provide critical insights into the model's components, revealing their importance for different prediction horizons and challenging traditional model design assumptions. The spatiotemporal feature extraction and attention mechanisms prove crucial for short-term predictions, while the results for long-term predictions suggest the need for dynamic model adjustments based on prediction time scales.

However, limitations exist, such as not considering external factors (weather, holidays) and relying on traditional evaluation metrics. Future research should address these limitations, explore dynamic model adjustments for long-term predictions, develop new evaluation metrics, and investigate counterintuitive phenomena in long-term forecasting.

MSTA-GNet provides valuable insights for improving traffic prediction models and advancing intelligent transportation systems.

REFERENCES

[1] S. Lu, Q. Zhang, G. Chen, D. Seng, "A combined method for short-term traffic flow prediction based on recurrent neural network," Alexandria Engineering Journal, vol. 60, no. 1, pp. 87-94, 2021. Doi: 10.1016/j.aej.2020.06.008

[2] X. Xu, X. Jin, D. Xiao, C. Ma, , &S. C. Wong, "A hybrid autoregressive fractionally integrated moving average and nonlinear autoregressive neural network model for short-term traffic flow prediction," Journal of Intelligent Transportation Systems, vol. 27, no. 1, pp. 1 -18, 2023. Doi: 10.1080/15472450.2021.1977639

[3] C. Ma, G. Dai, J. Zhou, "Short-Term Traffic Flow Prediction for Urban Road Sections Based on Time Series Analysis and LSTM_BILSTM Method," IEEE Transactions on Intelligent Transportation Systems, vol. 23, no. 6, pp. 5615-5624, 2022. Doi: 10.1109/TITS.2021.3055258

[4] J. Wang, W. Zhu, Y. Sun, C. Tian, "An effective dynamic spatiotemporal framework with external features information for traffic prediction," Applied Intelligence, vol. 51, pp. 3159-3173, 2021. Doi: 10.1007/s10489-020-02043-1

[5] I. Lana, J. Del Ser, M. Velez, E. I. Vlahogianni, "Road traffic forecasting: recent advances and new challenges," IEEE Intelligent Transportation Systems Magazine, vol. 10, no. 2, pp. 93-109, 2018. Doi: 10.1109/MITS.2018.2806634

[6] H. Gao, H. Jia, L. Yang, "An Improved CEEMDAN-FE-TCN Model for Highway Traffic Flow Prediction," Journal of Advanced Transportation, vol. 2022, no. 1, pp. 2265000, 2022. Doi: 10.1155/2022/2265000

[7] Y. Zhou, S.T. Hu, W. Li, N. Cheng, N. Lu, X.M. Shen, "Graph neural network driven traffic prediction technology: review and challenge," Chinese Journal on Internet of Things, vol. 5, no. 4, pp. 1-16, 2021.

[8] B. Medina-Salgado, E. Sánchez-DelaCruz, P. Pozos-Parra, J. E. Sierra, "Urban traffic flow prediction techniques: a review," Sustainable Computing: Informatics and Systems, vol. 35, pp. 100739, 2022. Doi: 10.1016/j.suscom.2022.100739

[9] A. A. Kashyap, S. Raviraj, A. Devarakonda, S. R. Nayak K, S.KV, S. J. Bhat, "Traffic flow prediction models-A review of deep learning techniques," Cogent Engineering, vol. 9, no. 1, pp. 2010510, 2022. Doi: 10.1080/23311916.2021.2010510

[10] W. Lu, Y. Rui, Z. Yi, B. Ran, Y. Gu, "A hybrid model for lane-level traffic flow forecasting based on complete ensemble empirical mode decomposition and extreme gradient boosting," IEEE Access, vol. 8, no. 1, pp. 42042-42054, 2020. Doi: 10.1109/ACCESS.2020.2977219

[11] Y. Lv, Y. Duan, W. Kang, Z. Li, F. Y. Wang, "Traffic flow prediction with big data: a deep learning approach," IEEE Transactions on Intelligent Transportation Systems, vol. 16, no. 2, pp. 865-873, 2014. Doi: 10.1109/TITS.2014.2345663

[12] D. Zhang, H. Lan, Z. Ma, Z. Yang, X. Wu, X. Huang, "Spatial-temporal gated graph convolutional network: A new deep learning framework for long-term traffic speed forecasting," Journal of Intelligent &Fuzzy Systems, vol. 44, no. 6, pp. 10437-10450, 2023. Doi: 10.3233/JIFS-224285

[13] A. Sharma, A. Sharma, P. Nikashina, V. Gavrilenko, A. Tselykh, A. Bozhenyuk, H. Meshref, "A graph neural network (GNN)-based approach for real-time estimation of traffic speed in sustainable smart cities," Sustainability, vol. 15, no. 15, pp. 11893, 2023. Doi: 10.3390/su151511893

[14] E. V. N. Jyothi, G. S. Rao, D. S. Mani, C. Anusha, M. Harshini, M. Bhavsingh, A. Lavanya, "A Graph Neural Network-based Traffic Flow Prediction System with Enhanced Accuracy and Urban Efficiency," Journal of Electrical Systems, vol. 19, no. 4, pp. 336, 2023.

[15] Y. Wang, S. Ke, C. An, Z. Lu, J. Xia, "A Hybrid Framework Combining LSTM NN and BNN for Short-term Traffic Flow Prediction and Uncertainty Quantification," KSCE Journal of Civil Engineering, vol. 28, no. 1, pp. 363-374, 2024. Doi: 10.1007/s12205-023-2457-y

[16] Y. Wang, C. Jing, S. Xu, T. Guo, "Attention based spatiotemporal graph attention networks for traffic flow forecasting," Information Sciences, vol. 607, pp. 869-883, 2022. Doi: 10.1016/j.ins.2022.05.127

[17] H. Wang, R. Zhang, X. Cheng, L. Yang, "Hierarchical traffic flow prediction based on spatial-temporal graph convolutional network," IEEE Transactions on Intelligent Transportation Systems, vol. 23, no. 9, pp. 16137-16147, 2022. Doi: 10.1109/TITS.2022.3148105

[18] Y. Gao, L. Zhao, J. Du, J. Wang, "Spatial-temporal Traffic Flow Prediction Model Based on the GAT and BiGRU," Journal of Physics: Conference Series, vol. 2589, no. 1, pp. 012024, 2023. Doi:10.1088/1742-6596/2589/1/012024

[19] H. Yi, H. Jung, S. Bae, "Deep neural networks for traffic flow prediction," 2017 IEEE International Conference on Big Data and Smart Computing (BigComp), Jeju, IEEE, pp. 328-331, 2017. Doi: 10.1109/BIGCOMP.2017.7881687

[20] A. Karbasi, E. Sherafat, "Short-term prediction of traffic flow based on gated recurrent unit neural networks," National Conference on New Studies and Findings in the Field of Civil Engineering, Architecture and Urban Planning in Iran, 2021.

[21] D. Yang, S. Li, Z. Peng, P. Wang, J. Wang, H. Yang, "MF-CNN: traffic flow prediction using convolutional neural network and multi-features fusion," IEICE TRANSACTIONS on Information and Systems, vol. E102-D, no. 8, pp. 1526-1536, 2019. Doi: 10.1587/transinf.2018EDP7330

[22] B. Yu, H. Yin, Z. Zhu, "Spatio-temporal graph convolutional networks: a deep learning framework for traffic forecasting," arXiv preprint arXiv:1709.04875, 2017. Doi: 10.48550/arXiv.1709.04875

[23] Y. Li, R. Yu, C. Shahabi, Y. Liu, "Diffusion convolutional recurrent neural network: data-driven traffic forecasting," arXiv preprint arXiv:1707.01926, 2017.Doi: 10.48550/arXiv.1707.01926

[24] P. Veličković, G. Cucurull, A. Casanova, A. Romero, P. Lio, Y. Bengio, "Graph attention networks," arXiv preprint arXiv:1710.10903, 2017. Doi: 10.48550/arXiv.1710.10903

[25] L. Cai, K. Janowicz, G. Mai, B. Yan, R. Zhu, "Traffic transformer: Capturing the continuity and periodicity of time series for traffic forecasting," Transactions in GIS, vol. 24, no. 3, pp. 736-755, 2020. Doi: 10.1111/tgis.12644

[26] M. Xu, W. Dai, C. Liu, X. Gao, W. Lin, G. J. Qi, H. Xiong, "Spatial-temporal transformer networks for traffic flow forecasting," arXiv preprint arXiv:2001.02908, 2020. Doi: 10.48550/arXiv.2001.02908

[27] B. Naheliya, P. Redhu, K. Kumar, "A hybrid deep learning method for short-term traffic flow forecasting: GSA-LSTM," Indian Journal of Science and Technology, vol. 16, no. 46, pp. 4358-4368, 2023.

[28] W. Zhuang, Y. Cao, "Short-term traffic flow prediction based on a

K-Nearest Neighbor and bidirectional Long Short-Term Memory model," Applied Sciences, vol. 13, no. 4, pp. 2681, 2023. Doi: 10.3390/app13042681

[29] Z. Yang, K. Jerath, "Deep Learning for Traffic Flow Prediction using Cellular Automata-based Model and CNN-LSTM architecture," arXiv preprint arXiv:2403.18710, 2024.Doi: 10.48550/arXiv.2403.18710

[30] Y. Mao, G. Qin, P. Ni, Q. Liu, "Analysis of road traffic speed in Kunming plateau mountains: a fusion PSO-LSTM algorithm," International Journal of Urban Sciences, vol. 26, no. 1, pp. 87-107, 2022.Doi: 10.1080/12265934.2021.1882331

[31] L.Y. Li, "Traffic flow prediction based on spatiotemporal dynamic graph convolutional network," Intelligent City, vol. 8, no. 1, pp. 1-7, 2022.Doi: 10.19301/j.cnki.zncs.2022.01.001

[32] W.Z. Zhao, G. Yuan, Y.M. Zhang, S.J. Qiao, S.Z. Wang, L. Zhang, "Multi-view Fused Spatial-temporal Dynamic GCN for Urban Traffic Flow Prediction," Journal of Software, vol. 35, no. 4, pp. 1751-1773, 2024.Doi: 10.13328/j.cnki.jos.007018

[33] W.D. Li, L. Yu, "Construction of urban traffic flow prediction model under neural network model," Cyberspace Security, no. Z3, pp. 56-59+77, 2021.

[34] X. T. Zhang, J. Y. Zheng, Q. Shen, D. F. Sun, Y. L. Jiang, "Multi-channel spatial-temporal traffic flow prediction based on hybrid static-dynamic graph convolution," Telecommunications Science, vol. 39, no. 9, pp. 97-110, 2023.

[35] P. Qiao, "Research on traffic flow detection based on deep learning and edge task offloading [D]," Xi'an Electronic Science and Technology University, 2020.Doi: 10.27389/d.cnki.gxadu.2019.001182

[36] R. Tian, C. Wang, J. Hu, Z. Ma, "MFSTGN: a multi-scale spatial-temporal fusion graph network for traffic prediction," Applied Intelligence, vol. 53, no. 19, pp. 22582-22601, 2023.Doi: 10.1007/s10489-023-04703-4

[37] Y. Qin, H. Luo, F. Zhao, Y. Fang, X. Tao, C. Wang, "Spatio-temporal hierarchical MLP network for traffic forecasting," Information Sciences, vol. 632, pp. 543-554, 2023.Doi: 10.1016/j.ins.2023.03.063

[38] Y.P. Liang, Z.Y. Mao, W.B. Zou, R. Xu, "Short-term Traffic Flow Prediction Based on Similar Data Aggregation and KNN with Varying K-value," Journal of Geo-Information Science, vol. 20, no. 10, pp. 1403-1411, 2018.

[39] X.L. Zhang, "A study on short-time traffic volume forecasting based on non-parametric regression [D]," Tianjin University, 2007.

[40] S. Liu, H. Chen, X.Y. Chen, J.J. He, "Double Branch Spatial-temporal Graph Convolutional Neural Network for Traffic Flow Prediction," Information and Control, no. 3, pp. 391-404+416, 2023. Doi: 10.13976/j.cnki.xk.2023.2092

[41] L.L. Wu, L.L. Yin, Q.L. Ren, "A Short-Term Traffic Flow Forecasting Method Based on EMD and DE-BPNN Combined Optimization," Journal of Chongqing University of Technology (Natural Science), vol. 35, no. 12, pp. 155-163, 2021.

Laboratory Abnormal Behavior Recognition Method Based on Skeletal Features

Dawei Zhang*
School of Information Engineering, Liaodong University, Dandong, China

***Abstract*—The identification of abnormal laboratory behavior is of great significance for the safety monitoring and management of laboratories. Traditional identification methods usually rely on cameras and other equipment, which are costly and prone to privacy leakage. In the process of human body recognition, they are easily affected by various factors such as complex backgrounds, human clothing, and light intensity, resulting in low recognition rates and poor recognition results. This article investigates a laboratory abnormal behavior recognition method based on skeletal features. One is to use Kinect sensors instead of traditional image sensors to obtain characteristic skeletal data of the human body, reducing external limitations such as lighting and increasing effective data collection. Then, the collected data is smoothed, aligned, and image enhanced using moving average filtering, Discrete Fourier Transform, and contrast, effectively improving data quality and helping to better identify abnormal behavior. Finally, the OpenPose algorithm is used to construct a laboratory anomaly behavior recognition model. OpenPose can be used to connect the entire skeleton through the relationships between points during the process of extracting human skeletal points, and combined with multi-scale pyramid networks to improve the network structure, effectively improving the accuracy and recognition speed of laboratory abnormal behavior recognition. The experiment shows that the accuracy, precision, and recall of the behavior recognition model constructed by the algorithm are 95.33%, 96.68%, and 93.77%, respectively. Compared with traditional anomaly detection methods, it has higher accuracy and robustness, lower parameter count, and higher operational efficiency.**

***Keywords*—Skeletal features; abnormal behavior recognition; OpenPose algorithm; Kinect sensor; Discrete Fourier Transform**

I. Introduction

In recent years, the rapid development of computer vision and artificial intelligence technology has brought major changes to many industries. Among them, motion recognition technology, as a key branch, is gradually being integrated into all aspects of people's lives. In the scientific research laboratory environment, efficient and accurate identification of abnormal behaviors is essential to ensure the safety of personnel, the normal operation of equipment, and the reliability of experimental results. However, the traditional recognition method based on video surveillance has shortcomings in dealing with lighting changes, line of sight blocking and viewing angle restrictions, which affects the accuracy and stability of recognition. To meet these challenges, this study proposes a strategy for identifying abnormal behaviors based on human skeletal Features. By using the OpenPose algorithm and the Kinect depth sensor, it is possible to accurately capture and extract the three-dimensional bone structure data of the experimenter. As the internal manifestation of human movement, bone characteristics can not only effectively reflect the relative position and trajectory of joints, but also have a higher tolerance and anti-interference ability to light, color changes and background complexity, and show excellent robustness in complex laboratory environments.

By constructing a laboratory abnormal behavior recognition model based on skeletal Features, it aims to break through the limitations of traditional video surveillance technology, realize in-depth analysis of the movements and attitudes of experimenters, and detect and warn of potential safety hazards or operational errors in a timely manner. This research is expected to significantly improve the level of safety management in the laboratory, and provide strong support for the optimization and efficiency improvement of the experimental process. Therefore, the introduction of skeletal features into the field of abnormal behavior recognition in laboratories is not only an inevitable trend of technological development, but also a practical need to ensure the smooth progress of scientific research activities and promote the sustainable prosperity of scientific undertakings.

II. Related Work

Abnormal behavior recognition is the process by which a monitoring system compares and identifies behaviors or events that do not conform to known conventional behavior patterns [1-2]. This method can analyze the patterns existing in the data, identify some behaviors that do not conform to the norm, and timely discover and respond to potential problems or dangers. Li Lin developed an abnormal behavior recognition method based on the spatiotemporal background of learning behavior. The specific recognition method adopts a top-down strategy and evaluates the effectiveness of local behavior themes and abnormal behavior recognition using spatiotemporal context learning [3]. Guan Yepeng proposed an abnormal behavior recognition method based on three-dimensional convolutional neural network (CNN) and long short-term memory (LSTM), which replaces the three primary color images with a feature image composed of optical flow and motion history images as input [4]. Hao Yixue combined Graph Convolutional Network (GCN) and 3D Convolutional Neural Network to propose an end-to-end anomaly behavior detection framework from a new perspective. Specifically, a class of classifiers is trained to extract features and estimate anomaly scores to improve the performance of anomaly behavior detection [5]. Shi Xiaonan proposed an underground abnormal behavior recognition method based on optimized Alphapose-GCN. Firstly, he defogged and enhanced the image set captured in the

underground monitoring video. Secondly, he optimized Alphapose object detection using the YOLOv3 model. Finally, he performed abnormal behavior recognition [6]. Bae Hyun-Jae proposed a method to identify abnormal behavior using only joint keypoints and joint motion information. He extracted joint keypoints of body parts through AlphaPose and sequentially inputted the extracted joint keypoints into the LSTM model for recognition [7]. Lee Jiyoo proposed an anomaly behavior detection model based on deep learning models, which reflects the accuracy of this research method for anomaly behavior detection by detecting violent and fainting behaviors in videos [8]. In summary, although the existing methods can cope with the challenges of lighting, occlusion, and viewing angle changes, the recognition accuracy rate may decrease in extremely complex scenes, and enhancing the robustness of the algorithm in complex environments is an urgent problem to be solved. At the same time, the definition of abnormal behavior in different scenes and contexts is different, and there are ambiguities and ambiguities. Building a common and highly explanatory abnormal behavior recognition framework to adapt to different application scenarios is another important issue. To this end, this article will use the bone feature research method to study it, hoping to improve the accuracy of recognition.

Behavior recognition based on bone features is a combination of machine vision and deep learning, which can analyze and recognize human motion and behavior [9-10]. This method monitors and analyzes human movement by detecting key parts in the human skeleton. TIAN Zhiqiang designed a temporal divergence model based on skeleton points to describe the motion state of skeleton points, amplifying the inter class differences in different human behaviors. At the same time, he also designed an attention mechanism with temporal divergence features to highlight key skeleton points, and this algorithm has a relatively high accuracy in authoritative human behavior datasets [11]. Liu Yuchao extracted key points of the human 3D skeleton in time series using YOLOv4, and applied the Meanshift object tracking algorithm to convert the key points into spatial tricolors. He put it into a multi-layer convolutional neural network for recognition, which can quickly identify various abnormal behaviors [12]. Li Maosen used two chart scales to clearly capture the relationship between body joints and body parts, and conducted experiments on skeleton-based motion recognition and prediction on four datasets, achieving good results [13]. Gao Guohong introduced an extraction technique that combines jawbone skeleton features, using skeleton heatmap descriptors and Kalman filtering algorithm to extract skeleton features of the upper and lower jaw bones. This method has better recognition accuracy than other models [14]. Shu Xiangbo proposed a new skeleton joint method to capture spatial consistency between joints and temporal evolution between skeletons. At the same time, he found through experiments on human motion prediction that the proposed method is superior to other methods on the feature maps that are of common concern to the skeleton joints [15]. Through the integration of machine vision and deep learning, the behavior recognition method based on bone characteristics has significantly improved the accuracy of human movement and behavior analysis. The current research mainly focuses on the precise detection of key bone points and the establishment of motion state models. Technical methods such as time divergence model, attention mechanism, and tracking of key points of the three-dimensional skeleton are used to further highlight the differences between different behaviors, thereby improving the accuracy of recognition. These studies have not only enhanced bone feature recognition technology, but also provided solid support for practical application scenarios such as laboratory abnormal behavior monitoring.

Today, with the increasing awareness of laboratory safety, it is crucial to effectively detect and prevent abnormal laboratory behavior. Although traditional video monitoring technology can monitor the situation inside the laboratory in real time, it often requires manual monitoring and analysis, which is not only inefficient but also prone to missed detections. The abnormal behavior recognition method based on bone features automatically identifies abnormal behaviors by capturing and analyzing bone movements, to improve the efficiency of laboratory safety management. This project aims to identify abnormal behaviors in the laboratory based on bone characteristics.

III. Bone Feature Extraction Technology

This article uses bone feature extraction technology to extract bone points from abnormal behaviors in the laboratory. Firstly, the video data in the laboratory is captured by a camera, and more features are extracted from human bones using Kinect depth sensors. By analyzing the collected bone structure data, important evidence can be provided for subsequent identification of abnormal behavior [16-17]. Skeleton points are a type of landmark data. The human skeleton is the internal framework of the human body, which generally includes two elements: joints and bones. Bones are the lines connecting points and edges, and joints and bone elements correspond exactly to edges and points in the shape [18]. Therefore, the skeleton structure is shown in Fig. 1.

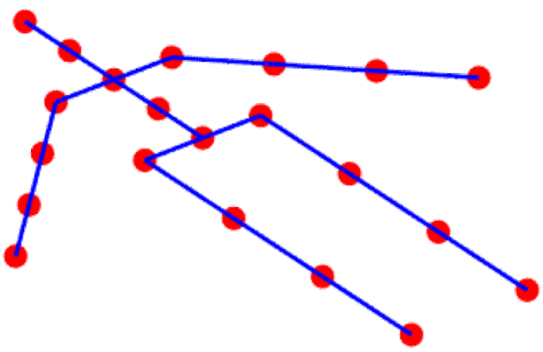


Fig. 1. Skeleton diagram.

There are two main methods for obtaining human skeletal information: one is posture estimation [19], and the other is motion capture devices [20-21]. The human skeletal sequence is based on the two-dimensional or three-dimensional coordinates of human joints. On this basis, the coordinate vectors of each node can be correlated and an independent feature vector can be generated between each frame. This article utilizes Kinect depth sensors for bone feature extraction.

Kinect depth sensors first obtain a continuous stream of human body sequence data based on the actions of the experimenters. Through this stream of sequence data, corresponding depth images can be obtained. By preprocessing the obtained images, the coordinates of the actions in three-dimensional space and time series can be obtained. By collecting joint point data through Kinect, the first step is to segment the collected human body contours. During this process, Kinect filters and distinguishes each pixel point to find the pixel that is most likely to be the human body boundary. Then, edge detection algorithms are used to identify the contour area of the target. On this basis, the extracted human body contours were used to segment the torso and limbs of the human body.

When performing behavior recognition, only representative skeletal parts such as hands and feet are usually needed to make corresponding recognition and judgment of actions. These skeletal points all have one thing in common, which is that they are far from the center of gravity of the body, have a larger range of motion and higher flexibility. Compared to the more fragile skeletal parts of the human body, the wrist and ankle joints play a more important role in human behavior recognition. Taking the center of gravity of the human body as a reference, the line h connecting the center of gravity to other bone points represents the relative distance between each bone point and the center of gravity of the human body. Therefore, at adjacent moments $[t + \Delta t]$, the average distance H from bone point j_i to the center of gravity can be expressed as:

$$\bar{h}_i^t = \left(|h_i^t| + |h_i^{t+\Delta t}|\right)/2 \quad (1)$$

Among them, the $\bar{h} \in [0,1]$ and $\bar{h}$ values are too small to serve as characterizations. Here, β is used to activate the function and normalize h¯to a new interval [x, y]. The specific expression is as follows:

$$\beta(a) = ln(a + q) + p \quad (2)$$

Let $a = H$ analyze the original interval [0,1] and the generated inter new area $[x, y]$ to solve.

$$q = 1/(e^{y-x} - 1) \text{ and } p = y - ln\left(\left(e^{y-x}/\left((e^{y-x} - 1)\right)\right)\right).$$

Human abnormal behavior is very random and difficult to predict. It is impossible to describe the entire abnormal behavior of the human body well if only one feature is extracted. So, this article uses two skeletal features, namely the aspect ratio of the human body and the structural vector, to study, and the expression for the aspect ratio of the human body is as follows:

$$x = \frac{a}{b} \quad (3)$$

Among them, x represents the aspect ratio of the human body, a represents the width of the smallest outer rectangle, and b represents the height of the smallest outer rectangle.

IV. Data Collection and Preprocessing

A. Data Collection

With the rapid development of the chemical industry, the number of chemical laboratories is gradually increasing, and their safety issues are also increasingly attracting social attention. To ensure safety during the experiment, it is necessary to detect and identify potential abnormal behaviors of the experimental personnel. A unique chemical laboratory dataset was established by collecting and cleaning chemical experimental data from public datasets. To improve the efficiency and accuracy of data classification and labeling, YOlOv5 is introduced in the process of establishing the behavior model to complete relevant behavior classification and labeling, thereby enhancing the overall recognition performance of the model. The tag files used in this article are in the format of the dataset and are marked using current open and convenient tagging tools. In behavior recognition experiments based on human bones, the NTU RGB+D dataset has the largest number of samples so far, including multiple video sequences. This article collects a total of 60 different action categories. This dataset was collected by Kinect sensors and synchronously observed from multiple perspectives. The collected data is shown in Fig. 2.



Fig. 2. Partial dataset collection.

B. Data Preprocessing

1) Data smoothing processing: Due to the performance of Kinect sensors themselves, there may be some additional noise in the collected data. In addition, when experimental participants engage in behavioral movements, noise may also be present in the data due to factors such as body tremors. To eliminate noise in the data, it is necessary to perform smoothing processing, i.e. filtering, on the data. For this purpose, this article uses the moving average filtering method [22-23] to process the collected data. This method averages a fixed length signal to achieve the goal of eliminating noise. Take the data segment $A_1, A_2, \cdots, A_H$ with a fixed length of H, and the specific calculation formula is as follows:

$$\bar{A} = \frac{1}{H}\sum_{i=0}^{H} A_{H-1} \quad (4)$$

Among them, $\bar{A}$ is the average value of H data points, which can be used to replace the current position of A_1 data points. Then, starting from A_2, H data points are taken, and the average value is calculated to replace A_2. After n iterations of the above process, n filtering smoothing results are obtained.

The data smoothing effect based on the moving average method is shown in Fig. 3. It can be found that using moving average to smooth the data has achieved very good results, and the data can retain more original features.

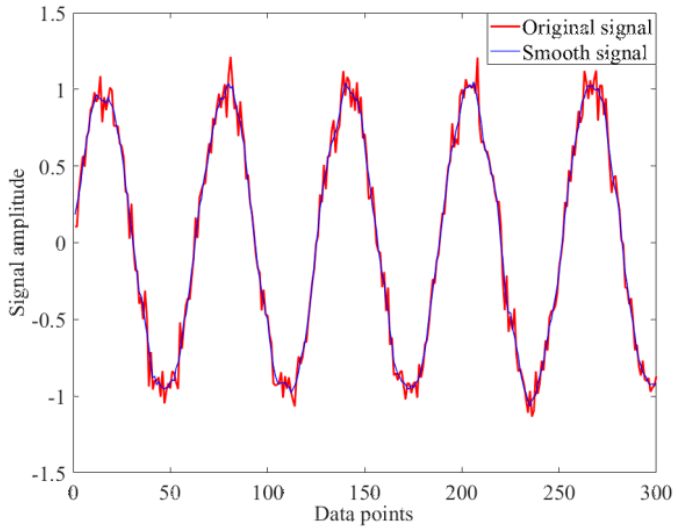


Fig. 3. Data smoothing effect based on moving average method.

2) Data alignment processing: The process of obtaining human bone data based on Kinect sensors is real-time, and different action intervals result in different lengths of each set of data even at the same sampling frequency [24-25]. In addition, the influence of factors such as the technical level and operating habits of the experimenters results in different completion times for the same action participants, leading to different data lengths. Therefore, it is necessary to align the collected bone data to achieve effective learning and prediction of human behavior. Discrete Fourier Transform is the process of converting n signals that were originally in the time domain to obtain an equal number of signals in the frequency domain [26-27]. Assuming that a non-periodic continuous time signal is denoted as y(t), the Fourier transform expression for this signal is as follows:

$$Y(\beta) = \int_{-\infty}^{+\infty} y(t) e^{-jet} dt \quad (5)$$

This article adopts the data alignment method of Fourier transform, and the specific effect diagram is shown in Fig. 4.

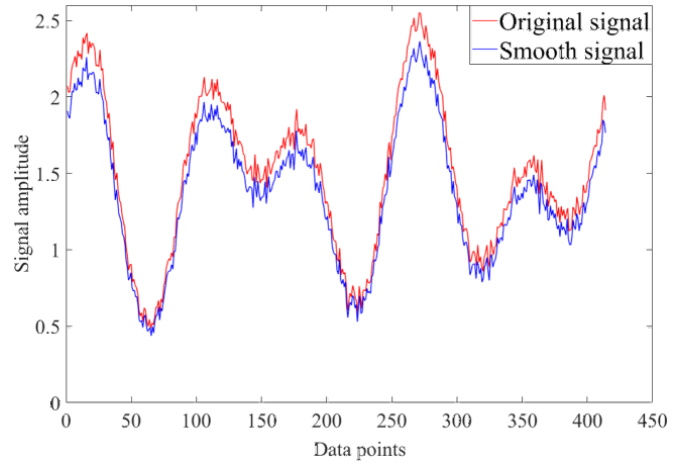


Fig. 4. Data alignment effect based on fourier transform.

As shown in Fig. 4, the length of the bone action data signal is 414, and the target length is 450. By interpolation, human behavior data of the same length can be obtained. Therefore, this also indicates that using Fourier transform can obtain human behavior feature data of the same length, achieving the purpose of data alignment.

3) Contrast enhancement: Image contrast refers to the difference in brightness of an image, while contrast enhancement refers to the difference in grayscale colors in an image, making the content of the image clearer [28-29]. The expression for contrast enhancement is as follows:

$$h|(a) = xg(a) + y \quad (6)$$

Among them, g(a) represents the input data; g(x) is the output data; x is the gain, which can set the contrast of the image; y is paranoia, which can set the brightness of the image.

If it is necessary to enhance the contrast of non-linear images, x and y methods are usually automatically selected. Assuming the height of the input matrix Q is H and the width is W, $Q(s,t)$ represents the grayscale value of the s row and t column. The minimum grayscale value in matrix Q is Q_{min}, and the maximum value is Q_{max}, which means the grayscale value of the matrix is $Q_{min} \leq Q(s,t) \leq Q_{max}$. The range of the output matrix R is $[R_{max}, R_{min}]$, and the expression is as follows:

$$R(s,t) = \frac{R_{max} - R_{min}}{Q_{max} - Q_{min}} (Q(s,t) - Q_{min}) + R_{min} \quad (7)$$

Among them, $0 \leq s \leq H$, $0 \leq t \leq W$, and $R(s,t)$ represent the grayscale values of the s row and t column. Generally, $R_{min} = 0$ and $R_{max} = 255$ are set.

The effect of using contrast enhancement to enhance the collected laboratory behavior images is shown in Fig. 5.

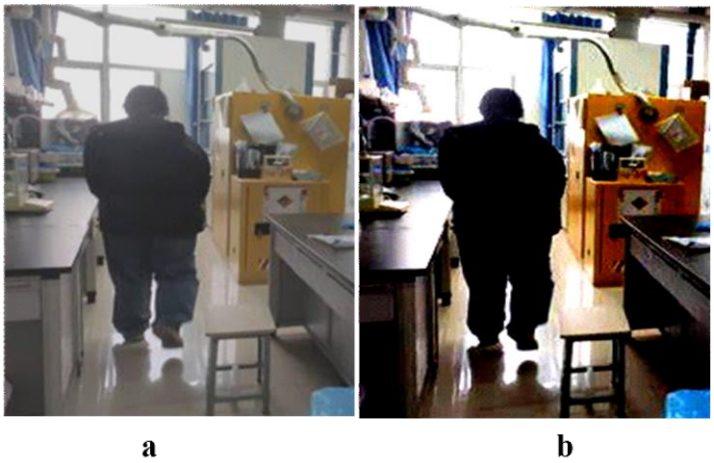


Fig. 5. Effect image based on contrast enhancement.

As shown in Fig. 5, 'a' is the original image, and 'b' is the image after contrast enhancement. The use of contrast enhancement methods can effectively enhance the clarity of images, making their brightness more pronounced and highlighting key information.

V. Construction of a Laboratory Abnormal Behavior Recognition Model Based on OpenPose

Compared with 2D images, bone sequences can more accurately reflect the spatial distribution positions of various joints in the human body, but cannot effectively characterize the connection status between joints in the human body. To better express the semantics of complex interaction behaviors, this article can use graph data structures to characterize the connection states between joints, and based on this, study a behavior feature description method based on time-space graphs. On the one hand, spatial graph connections can be designed based on the characteristics of individual bones in the frame and the corresponding joint interactions between two individuals; On the other hand, connectivity methods for temporal graphs can be studied based on the mapping relationships between nodes in each frame. Through these operations, the connectivity status of each bone node in the spatial region can be achieved, achieving spatiotemporal description of interaction behavior. The OpenPose algorithm is a bottom-up approach for behavior recognition. This algorithm can recognize human body movements, facial expressions, etc., and has good noise resistance [30-31]. The backbone extraction network of OpenPose adopts a variant based on ResNet, mainly composed of deep residual neural networks. This network can accept input from images of any size and convert the input images into feature maps of the same size. In the transformed feature map, multi-layer convolution and pooling layers are used to extract features from the image. At the final layer of the network, the obtained feature map is transformed into a local thermal spectrum for extracting bones, thereby achieving localization of various parts of the human body. This method maps the temperature distribution of various parts of the human body to obtain the temperature distribution of each part of the human body, thereby completing the pose estimation of the human body.

When using preprocessed data with OpenPose to identify abnormal behavior, although the novelty of this method may be affected by the wide application of OpenPose, it does not prevent innovation in data processing and model construction. As a key link to ensure the performance of the model, data preprocessing explores a more efficient bone feature extraction technology for the data generated by OpenPose, and tries to incorporate novel features to enhance the expressive force of the data. The construction of recognition model is the core of this study. Based on the skeleton information extracted by OpenPose, a variety of machine learning and deep learning models are skillfully designed and trained to achieve a breakthrough in the field of abnormal behavior recognition. These measures not only highlight the innovation of research, but also are expected to significantly improve the accuracy of abnormal behavior identification.

OpenPose has excellent ability to identify and locate key points of human body, and can capture multiple key points of face, hands, feet and main parts of the body at the same time, with a total of 25 points. This comprehensive detection enables the algorithm to grasp the human posture more accurately, and then accurately judge the abnormal behavior in the laboratory. At the same time, the algorithm uses the human body component decoder to determine the relationship between key points, which helps the algorithm to understand the human posture structure more deeply, thus improving the recognition accuracy. This design makes the algorithm perform better in dealing with complex human movements and posture changes, and effectively reduces misjudgment. In addition, OpenPose also uses multi-scale pyramid network to detect the human contour and key points, and then improves the positioning accuracy through gradual thinning. This structure is efficient and accurate. In practical application, lightweight OpenPose models can be selected to reduce the computational load and improve the recognition speed. These models significantly reduce the demand for computational resources while maintaining high accuracy. In addition, OpenPose also adopts efficient strategies such as parallel processing and cache optimization to further reduce the amount of computation and time consumption, and ensure that the algorithm can quickly process the input and output the recognition results.

In this article, human behavior and actions are recognized based on human bone data, mainly through the acquisition of human bone information through Kinect sensors. The obtained bone information is used as the three-dimensional coordinate values of various joint nodes in the human body. After completing a behavior, the experimental participants can obtain the corresponding bone data stream and save this bone data with a dimension of 48 * L, where L refers to the length taken during the data alignment process. OpenPose can be used to connect the entire skeleton through the relationship between points during the extraction process of human skeletal points [32-33]. The coordinate vectors of S joint points can be used to express the initial input bone data within one frame. For any t-the frame, the initial input data can be represented as $A_t \in Q^{Z_0 * S}$, Z_0 represents the coordinate dimension of the input original bone data, and t represents the t-th frame bone data. To better capture the relationship between the corresponding joint connections and behavioral semantics between adjacent frames, a frame-to-frame time graph connection design is performed for each frame of joint data, combined with the joint data of the previous and subsequent frames. The expression is:

$$m_{a,b} = \begin{cases} \alpha & (a,b) \in \aleph_5 \\ \beta & (a,b) \in \aleph_6 \\ 0 & a = b \end{cases} \tag{8}$$

Among them, $m_{a,b}$ describes the different weights α and β given to the edge response calculation when the joint point a and the joint point b belong to different connection categories ℵ_5 and ℵ_6 in the connection diagram.

Therefore, the adjacency matrix of the inter-frame time graph can be expressed as:

$$W_{t,t+1} = \begin{bmatrix} E_1 & E_2 \\ Q_1 & Q_2 \end{bmatrix} \tag{9}$$

Among them, E_1 and E_2 describe the joint relationship between the interaction parties between frames and themselves, as well as the joint connection relationship between Q_1 and Q_2 interaction parties and the interaction object between frames, jointly constructing a time graph description of interaction behavior between frames.

When human behavior is different, there can be a certain angle relationship between limbs due to deformation, and the

angle information has good scale invariance. In addition, in laboratory settings, abnormal behavior exhibits strong mobility in many aspects compared to normal behaviors such as standing or standing on the left. Therefore, using joint angles as separate data to analyze abnormal behavior is very useful. The formula for calculating the joint angle is as follows:

$$\theta_{n,m} = arccos\frac{\gamma\cdot\delta}{|\gamma||\delta|}, |\gamma| \neq 0, |\delta| \neq 0 \quad (10)$$

Among them, $\theta_{n,m}$ represents the angle value of the nth joint in frame m, m is the inner product of the vector, $\|$ is the modulus of the vector, γ and δ represent the limb vectors corresponding to the joint angle, respectively.

The appearance characteristics of the human body reflect the proportion of the body shape during movement, which is a highly recognizable information. In laboratory settings, abnormal human behavior exhibits more significant movement characteristics than normal behavior, that is, during the movement process, abnormal behavior can undergo corresponding changes in its appearance due to the significant movement of limbs. The expression for calculating the proportion of human body shape is as follows:

$$ratio_i^k = \frac{max(l_1^k,\cdots,l_m^k)-min(l_1^k,\cdots,l_m^k)}{max(d_1^k,\cdots,l_m^k)-min(d_1^k,\cdots,d_m^k)} \quad (11)$$

Among them, i represents the current frame rate, k represents the kth person in the current frame, $(d_1^k,\cdots,d_m^k)$ and $(l_1^k,\cdots,l_m^k)$ represent the d and l coordinates of the kth human joint point, respectively.

This article can input preprocessed data into the OpenPose model. Firstly, the image features can be extracted through the backbone extraction network. Then, the key points in the image can be associated with each other to extract the confidence interval of the bone points. The optimal algorithm can be used to locate the key points, resulting in a bone structure map. The identification results of key areas for abnormal behavior (sleep) of experimental personnel are shown in Fig. 6.

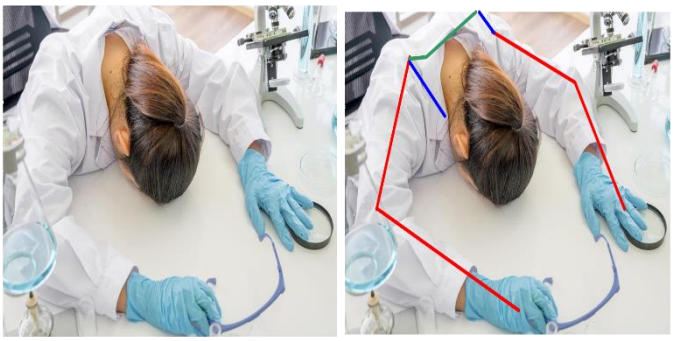


Fig. 6. Comparison of abnormal behavior before and after processing.

As shown in Fig. 6, the OpenPose model can effectively identify abnormal behaviors of experimenters. When abnormal behaviors occur during sleep, the OpenPose model can quickly capture and process corresponding bone information. On this basis, the OpenPose model utilizes convolutional neural networks to extract bone features and inputs them into a recurrent neural network to model the sequence. Therefore, it can effectively obtain time series information of human posture and improve the estimation accuracy of moving targets.

VI. Experiment of Laboratory Abnormal Behavior Recognition Model Based on OpenPose Algorithm

A. Experimental Preparation

The specific environment configuration of the server is shown in Table I.

TABLE I. Model Parameter Settings

Serial number	Experimental environment	Parameter
1	Central processing unit	Intel Xeon(R) E5-2620
2	Operating system	Ubuntul6.04 LTS
3	Deep learning framework	Pytorch 0.4.0
4	Programming language	Python 3.5
5	Initial learning rate	0.01
6	Weight attenuation coefficient	$1*10^{-4}$

Based on the characteristics and safety requirements of the laboratory itself, abnormal behavior can be classified according to specific regulations and laboratory safety needs. Meanwhile, corresponding discrimination criteria were established based on different types of abnormal behaviors. The specific abnormal targets that need to be detected for each type of experiment are shown in Table II.

TABLE II. Requirements for Detecting Abnormal Targets in Different Experiments

Serial number	Abnormal behavior	Test set	Experiment set	Verification set	Total
1	Sleep	1040	130	130	1300
2	Long hair	640	80	80	800
3	Without gloves	720	90	90	900
4	Illegal disposal of waste	400	50	50	500
5	Eat	960	120	120	1200
6	Smoking	1120	140	140	1400
7	Play with mobile phone	1760	220	220	2200
8	Playing and running	2400	300	300	3000
Total		9040	1130	1130	11300

As shown in Table II, it can be observed that abnormal behavior is divided into eight categories, each of which is divided in a certain proportion into test set, experimental set, and validation set. Among them, the abnormal behavior of playing and running has the highest amount of data, while the abnormal behavior of illegal disposal of waste has the lowest amount of data.

B. Experimental Analysis

To achieve more accurate and scientific experimental results, multiple iterative experiments can be conducted. The 11,300 image data extracted in Table II are studied. In order to make the experimental results more accurate and scientific, multiple iterative experiments will be conducted, and the accuracy, accuracy, recall rate, and F1-Measure (F1 value) obtained from the laboratory abnormal behavior recognition model constructed based on the proposed method will be experimentally compared with the convolutional neural network (CNN) [4], Spatial-Temporal Context visual tracking (Spatio-Temporal Context, STC) [3], AlphaPose algorithm [6] , Long Short-Term Memory network (Long Short-Term Memory, LSTM) [7] , Graph Convolutional Network (GCN) [5] and Meanshift [12] abnormal behavior recognition model constructed by the algorithm is compared, and the specific model performance comparison results are shown in Table III.

TABLE III. PERFORMANCE COMPARISON OF DIFFERENT ABNORMAL BEHAVIOR RECOGNITION MODELS

Algorithm	Accuracy (%)	Precision (%)	Recall (%)	F1 value (%)
Proposed	95.33	96.68	93.77	94.38
CNN	87.41	88.12	84.14	85.31
STC	86.25	84.91	80.47	81.59
AlphaPose	80.94	81.48	77.75	75.09
LSTM	86.39	89.12	82.66	80.67
GCN	82.63	84.39	83.97	82.07
Meanshift	84.61	82.77	81.73	80.33

As shown in Table III, the performance of the laboratory abnormal behavior recognition model constructed by the proposed algorithm in this paper is better than that of other algorithms in all aspects, which shows that the method studied in this paper can better detect and identify abnormal behaviors that occur in the laboratory. Behavior. The accuracy rate of the behavior recognition model constructed based on the proposed method is 95.33%, which is higher than that of the abnormal behavior recognition model constructed based on CNN, STC, AlphaPose, LSTM, GCN, and Meanshift algorithms. 7.92%、9.08%、14.39%、8.94%、12.7% And 10.72%.The high accuracy rate of the proposed method not only reflects its strong feature extraction and classification capabilities, but also its effectiveness in dealing with diverse abnormal behaviors in a laboratory environment. Compared with traditional visual methods such as CNN and STC, the proposed method can more accurately capture subtle changes in key points of human bones, so as to accurately distinguish between normal behavior and abnormal behavior. The accuracy rate of the method studied in this paper is 96.68%, which is 8% higher than the accuracy rate of the model constructed based on CNN, STC, AlphaPose, LSTM, GCN and Meanshift algorithms, respectively..56%、11.77%、15.2%、7.56%、12.29% and 13.91%, which means that the model has a lower false positive rate when identifying abnormal behavior, that is, it is less likely to misjudge normal behavior as abnormal.

At the same time, the recall rate of up to 93.77% ensures that the model can detect most real abnormal behaviors, significantly reducing the risk of missed inspections. This dual guarantee makes the proposed method of great practical value in laboratory safety control. The F1 value, as the reconciled average of the accuracy rate and recall rate, can more comprehensively show the overall performance of the model. The F1 value of the proposed method as high as 94.38% shows that it has done an excellent job in balancing accuracy and recall, once again confirming its superiority as a laboratory abnormal behavior recognition tool. The proposed algorithm can extract key point data of human bones from videos or images in real time and accurately, providing a valuable source of information for analyzing human posture and movement patterns. Although the comparison algorithms such as LSTM and GCN in Table III also incorporate time and graph structure information, the method relies on its unique attitude evaluation architecture to more naturally combine space-time background information, so as to more effectively capture the dynamic evolution of behavior evolution.

Research can be conducted on the extracted abnormal behavior data in Table II. Skeleton point extraction can be performed on the extracted abnormal behavior data images. The faster the extraction speed, the better for subsequent behavior recognition. This article can use the OpenPose algorithm to extract bone points from the collected images. The required extraction time can be compared with models constructed based on CNN, STC, AlphaPose, LSTM, GCN, and Meanshift algorithms. The specific comparison results are shown in Table IV.

TABLE IV. COMPARISON OF BONE POINT EXTRACTION SPEEDS AMONG DIFFERENT MODELS

Algorithm	1	2	3	4	5	6	7	8	Average value
Proposed	1.95	1.84	2.16	2.49	1.74	1.91	2.31	3.08	2.19
CNN	6.87	6.17	5.86	4.69	7.08	5.47	6.22	7.96	6.29
STC	3.34	2.96	2.88	4.02	4.81	4.39	5.93	5.22	4.19
AlphaPose	4.91	3.34	3.17	4.49	4.02	5.17	4.61	5.06	4.35
LSTM	5.64	4.06	5.49	6.07	4.86	6.02	4.77	5.83	5.34
GCN	5.13	4.46	4.97	6.39	4.08	3.84	5.06	6.28	5.03
Meanshift	3.58	3.27	4.51	3.28	5.18	4.63	5.88	6.31	4.58

As shown in Table IV, the comparative experimental results of different algorithms on the extraction speed of bone points are shown. First of all, the proposed method showed a significant speed advantage when extracting bone points of abnormal behavior images in Table II. The average extraction time is only 2.19 seconds, which is 4.1 seconds, 2 seconds, 2.16 seconds, 3.15 seconds, 2.84 seconds and 2.39 seconds lower than the time required for models built based on CNN, STC, AlphaPose, LSTM, GCN, and Meanshift algorithms, respectively. The time required for the model is 4.1 seconds, 2 seconds, 2.16 seconds, 3.15 seconds, 2.84 seconds, and 2.39 seconds. This advantage is not only reflected in the average time, but also in the stability and efficiency of the proposed method when dealing with various types of abnormal behaviors. Further analysis found that although the number of slapsticks running images numbered 8 is large, the proposed method can still complete bone point extraction within 3.08 seconds, compared to other algorithms such as CNN's 7.96 seconds, LSTM's 5.83 seconds, etc., its time advantage is particularly obvious. This proves the efficiency and robustness of the method in handling complex dynamic scenes. On the contrary, for the eating image numbered 5, although the number of images is large, the amplitude of its movement is relatively clear. The method can quickly capture these subtle changes and complete the bone point extraction at an extremely fast speed of 1.74 seconds, which further verifies the advantages of the algorithm in accurately capturing the posture of the human body. In addition, the proposed method can maintain a relatively stable time consumption when extracting bone points of different abnormal behavior images, without significant fluctuations, which shows that the algorithm has good generalization ability and stability, which is conducive to dealing with diverse scenarios and needs in practical applications. In summary, the proposed method with its excellent extraction speed and stability, provides strong support for the rapid and accurate identification of bone points in abnormal behavior images, and lays a solid foundation for subsequent behavior recognition work.

Eight possible abnormal behaviors that may occur in the laboratory were selected in Table II, and the eight selected abnormal behaviors were numbered 1-8. A laboratory abnormal behavior recognition model based on OpenPose algorithm can be used to detect these eight abnormal behaviors, and the accuracy of detecting each abnormal behavior can be obtained. The detected experimental results can be compared with the accuracy of models constructed based on CNN, STC, AlphaPose, LSTM, GCN, and Meanshift algorithms. The specific comparison results are shown in Fig. 7.

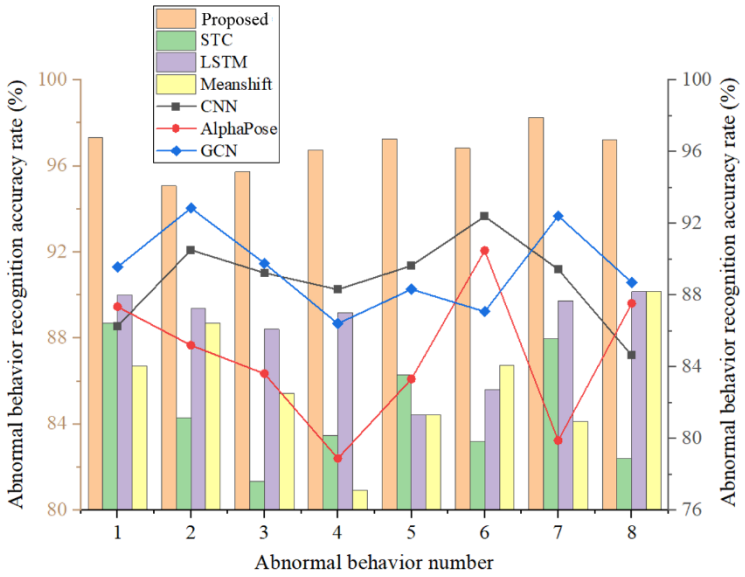


Fig. 7. Comparison of accuracy of different abnormal behavior recognition models for abnormal behavior recognition.

In Fig. 7, the x-axis represents the abnormal behavior number, and the left and right y-axis represent the abnormal behavior recognition accuracy. Among them, the recognition accuracy of the proposed method, STC, LSTM, and Meanshift algorithms refers to the left y-axis, while the CNN, AlphaPose, and GCN algorithms refer to the right y-axis. As shown in Fig. 7, the method has excellent performance in identifying these abnormal behaviors, and its average recognition accuracy rate is as high as 96.81%.Compared with other algorithms, the accuracy rate of the method is 8% higher than that of CNN, STC, AlphaPose, LSTM, GCN, and Meanshift, respectively..01%、12.1%、12.27%、8.44%、7.41% and 10.91%. Among them, the method's recognition rate exceeded 95.07% when dealing with all eight behaviors. Especially when it comes to identifying complex and subtle abnormal behaviors such as playing with mobile phones, its accuracy rate is as high as 98.27%, far surpassing other algorithms. Compared with other algorithms, the proposed method not only leads in the average accuracy rate, but also performs well in the recognition of each type of abnormal behavior. Compared with CNN and STC, the average accuracy rate of the method has improved significantly, showing high reliability and accuracy in dealing with diverse and complex abnormal behaviors in the laboratory. Compared with AlphaPose, which is also an attitude estimation

algorithm, the method has more obvious advantages in identifying subtle movements and behaviors in complex scenes. Although the recognition rate of GCN in some cases is close to that of the method, overall, the proposed method still has an advantage in recognition accuracy and stability. In summary, this study experimentally verifies the excellent performance of the abnormal behavior recognition model based on the method in the laboratory environment. The model is not only efficient in bone point extraction, but also has a significant breakthrough in recognition accuracy, which provides strong support for behavioral monitoring fields such as laboratory safety management.

When key points cannot be detected, in order to accurately identify character behavior, we need to combine other characteristics, not just rely on posture estimation results. However, the introduction of non-skeletal posture features may have an impact on the real-time performance of behavior recognition algorithms. Through the observation of the NTU RGB+D data set, it is found that the camera is installed in a position with a large field of view and can avoid looking down or looking up from a large angle. This good camera position can effectively reduce occlusion, thereby capturing clearer images. Therefore, the requirements for the installation angle of the camera are put forward, which can not only meet the needs of real-time detection, but also improve the accuracy of multi-target behavior recognition to a certain extent. After obtaining the model-related data and the test data of the simple scene video, some materials in the laboratory environment were selected to test the algorithm. Using the laboratory abnormal behavior recognition model constructed by the method, the renderings of the abnormal behavior bone extraction of laboratory personnel are shown in Fig. 8.

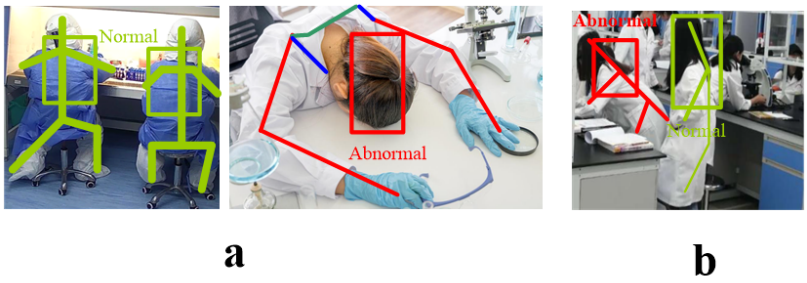


Fig. 8. Recognition effect of the proposed laboratory abnormal behavior recognition model.

As shown in Fig. 8, normal is marked in Fig. 8(a), which indicates the normal laboratory behavior of the experimenter, while abnormal is marked next to Fig. 8(a), which represents the abnormal behavior of the experimenter. In this paper, the proposed method is used to effectively identify the abnormal behavior of laboratory personnel. In Fig. 8(b), the standing position of the experimenter in the laboratory is identified. The algorithm studied in this paper can be well identified for the laboratory behavior of more complex personnel. Normal is marked for normal experimental behavior, while abnormal experimental behavior is marked as abnormal. The proposed method has good motion and posture estimation capabilities, and can accurately detect and track human joints in real time. This allows the model to accurately identify the posture and movements of personnel in the laboratory. At the same time, this method can also achieve multi-person posture estimation, without knowing the specific location of the human body in advance, multiple human bodies can be positioned.

C. Discussion

The obtained results were discussed in the context of previous research findings and methods. Comparative analysis highlights that the proposed method has better performance advantages in detecting abnormal laboratory behavior compared to conventional detection methods. The proposed method focuses on the research of laboratory abnormal behavior detection methods for human skeletal features, which better reduces the interference of complex background environments and other factors in the behavior recognition process. The method uses Kinect sensors to obtain human skeletal feature information, which has a smaller data volume and higher security compared to conventional image data acquisition. The use of moving average filtering, Discrete Fourier Transform, and contrast to preprocess data has improved data quality and recognition performance. Using OpenPose algorithm to achieve fast and stable extraction and classification of behavioral skeletal features. The laboratory abnormal behavior detection model based on skeletal features has higher accuracy and robustness, lower parameter count, and higher operational efficiency compared to traditional laboratory abnormal behavior detection methods.

VII. Conclusions

In recent years, with the continuous deepening of laboratory safety management, effective identification of abnormal behaviors during the experimental process has become one of the current research hotspots. This article takes human bones as the research object, analyzes the characteristics of human bones, and constructs a laboratory abnormal behavior recognition model based on bone features. The model mainly uses OpenPose algorithm and sensor devices to complete behavior recognition. Meanwhile, the model studied in this article mainly analyzes the changes in the spatial trajectory of human skeletal joints, which can effectively identify abnormal behavior in the laboratory. In this study, personnel behavior data was collected in the laboratory, and the YOLOv5 algorithm was used to extract bone features. Then, a laboratory abnormal behavior recognition model was established based on the OpenPose algorithm. The experimental results show that this method can effectively identify abnormal behavior in the laboratory, and performs excellently in accuracy and recall. The main research results obtained are as follows: 1) A model for identifying abnormal

behavior can be established, which has a very high accuracy in identifying abnormal behavior; 2) The recognition performance of the constructed model is superior to other models; 3) The recognition model based on OpenPose algorithm can extract human bone feature points more quickly. In summary, the method studied in this article not only has high recognition accuracy, but also has good real-time and stability, which can provide better assistance for laboratory safety management work.

When discussing the superiority and effectiveness of the laboratory abnormal behavior recognition model based on skeletal features studied in this article, it is also necessary to consider its limitations and future improvement directions. The following are several main limitations: Limitations of the dataset: The dataset used in the study may be limited to specific types of laboratory environments and behavioral patterns, thus the model's generalization ability may be limited. Algorithm complexity and computational resources: Although OpenPose algorithm performs well in bone feature extraction and behavior recognition, they have higher computational complexity, which in turn puts higher demands on the hardware resource conditions of the application environment. In the future, further research will be conducted to optimize model performance, enhance model adaptability, and expand the applicability of the model.

ACKNOWLEDGMENT

This work was supported by the Liaoning Provincial Department of Education's Basic Research Project for Universities [Grant No. JYTMS20230711] and Liaoning Province Science and Technology Plan Joint Program (Fund) Project [Grant No. 2023JH2/101700009].

REFERENCES

[1] K. Rezaee, M. R. Khosravi, and M. S. Anari, "Deep-Transfer-learning-based abnormal behavior recognition using internet of drones for crowded scenes," IEEE Internet of Things Magazine, vol. 5(2), pp. 41-44, 2022.

[2] X. P. Zhang, J. H. Ji, L. Wang, Z. H. He, and S. D. Liu, "Review of video-based identification and detection methods of abnormal human behavior," Control and decision-making, vol. 37(1), pp. 14-27, 2022.

[3] L. Li, Z. Liu, and G. Liu, "Abnormal behavior recognition based on spatio-temporal context,". Journal of Information Processing Systems, vol. 16(3), pp. 612-628, 2020.

[4] Y. Guan, W. Hu, and X. Hu, "Abnormal behavior recognition using 3D-CNN combined with LSTM," Multimedia Tools and Applications, vol. 80(12), pp. 18787-18801, 2021.

[5] Y. Hao, Z. Tang, B. Alzahrani, R. Alotaibi, R. Alharthi, M. Zhao, and A. Mahmood, "An end-to-end human abnormal behavior recognition framework for crowds with mentally disordered individuals," IEEE Journal of Biomedical and Health Informatics, vol. 26(8), pp. 3618-3625, 2021.

[6] X. Shi, J. Huang, and B. Huang, "An underground abnormal behavior recognition method based on an optimized alphapose-st-gcn," Journal of Circuits, Systems and Computers, vol. 31(12), pp. 2250214, 2022.

[7] H. J. Bae, G. J. Jang, Y. H. Kim, and J. P. Kim, "LSTM (long short-term memory)-based abnormal behavior recognition using AlphaPose," KIPS Transactions on Software and Data Engineering, vol. 10(5), pp. 187-194, 2021.

[8] J. Lee, and S. J. Shin, "A study of video-based abnormal behavior recognition model using deep learning," International journal of advanced smart convergence, vol. 9(4), pp. 115-119, 2020.

[9] C. Li, C. Xie, B. Zhang, J. Han, X. Zhen, and J. Chen, "Memory attention networks for skeleton-based action recognition," IEEE Transactions on Neural Networks and Learning Systems, vol. 33(9), pp. 4800-4814, 2021.

[10] P. T. Sheeba, and S. Murugan, "Fuzzy dragon deep belief neural network for activity recognition using hierarchical skeleton features," Evolutionary Intelligence, vol. 15(2), pp. 907-924, 2022.

[11] Z. TIAN, C. DENG, and J. ZHANG, "Human behavior recognition algorithm based on skeletal temporal divergence feature," Journal of Computer Applications, vol. 41(5), pp. 1450, 2021.

[12] Y. Liu, S. Zhang, Z. Li, and Y. Zhang, "Abnormal behavior recognition based on key points of human skeleton," IFAC-PapersOnLine, vol. 53(5), pp. 441-445, 2020.

[13] M. Li, S. Chen, X. Chen, Y. Zhang, Y. Wang, and Q. Tian, "Symbiotic graph neural networks for 3d skeleton-based human action recognition and motion prediction," IEEE transactions on pattern analysis and machine intelligence, vol. 44(6), pp. 3316-3333, 2021.

[14] G. Gao, C. Wang, J. Wang, Y. Lv, Q. Li, X. Zhang, and G. Chen, "UD-YOLOv5s: Recognition of cattle regurgitation behavior based on upper and lower jaw skeleton feature extraction," Journal of Electronic Imaging, vol. 32(4), pp. 043036-043036, 2023.

[15] X. Shu, L. Zhang, G. J. Qi, W. Liu, and J. Tang, "Spatiotemporal co-attention recurrent neural networks for human-skeleton motion prediction," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 44(6), pp. 3300-3315, 2021.

[16] M. Li, F. Wei, Y. Li, S. Zhang, and G. Xu, "Three-dimensional pose estimation of infants lying supine using data from a Kinect sensor with low training cost," IEEE Sensors Journal, vol. 21(5), pp. 6904-6913, 2020.

[17] Y. Tian, G. Wang, L. Li, T. Jin, F. Xi, and G. Yuan, "A universal self-adaption workspace mapping method for human–robot interaction using kinect sensor data," IEEE Sensors Journal, vol. 20(14), pp. 7918-7928, 2020.

[18] H. Wang, and H. Yuan, "A motion recognition method based on the fusion of bones and apparent features," Journal of Communications, vol. 43(1), pp. 138-148, 2022.

[19] C. Zheng, W. Wu, C. Chen, T. Yang, S. Zhu, J. Shen, and M. Shah, "Deep learning-based human pose estimation: A survey," ACM Computing Surveys, vol. 56(1), pp. 1-37, 2023.

[20] T. Petrosyan, A. Dunoyan, and H. Mkrtchyan, "Application of motion capture systems in ergonomic analysis," Armenian journal of special education, vol. 4(2), pp. 107-117, 2020.

[21] G. R. Dalla Bernardina, T. Monnet, H. T. Pinto, R. M. L. de Barros, P. Cerveri, and A. P. Silvatti, "Are action sport cameras accurate enough for 3D motion analysis? A comparison with a commercial motion capture system," Journal of applied biomechanics, vol. 35(1), pp. 80-86, 2019.

[22] M. Li, X. Liu, and F. Ding, "The filtering-based maximum likelihood iterative estimation algorithms for a special class of nonlinear systems with autoregressive moving average noise using the hierarchical identification principle," International Journal of Adaptive Control and Signal Processing, vol. 33(7), pp. 1189-1211, 2019.

[23] X. Liu, and Y. Fan, "Maximum likelihood extended gradient-based estimation algorithms for the input nonlinear controlled autoregressive moving average system with variable-gain nonlinearity," International Journal of Robust and Nonlinear Control, vol. 31(9), pp. 4017-4036, 2021.

[24] A. Bilesan, S. Komizunai, T. Tsujita, & A. Konno, "Improved 3D human motion capture using Kinect skeleton and depth sensor," Journal of robotics and mechatronics, vol. 33(6), pp. 1408-1422, 2021.

[25] C. Novo, R. Boss, P. Kyberd, E. Biden, J. Taboada, & M. Ricardo, "Testing the Microsoft kinect skeletal tracking accuracy under varying external factors," MOJ App Bio Biomech, vol. 6(1), pp. 7-11, 2022.

[26] K. Wang, H. Wen, & G. Li, "Accurate frequency estimation by using three-point interpolated discrete fourier transform based on rectangular window," IEEE Transactions on Industrial Informatics, vol. 17(1), pp. 73-81, 2020.

[27] L. Kämmerer, "Constructing spatial discretizations for sparse multivariate trigonometric polynomials that allow for a fast discrete

Fourier transform," Applied and Computational Harmonic Analysis, vol. 47(3), pp. 702-729, 2019.

[28] S. Agrawal, R. Panda, P. K. Mishro, and A. Abraham, "A novel joint histogram equalization based image contrast enhancement," Journal of King Saud University-Computer and Information Sciences, vol. 34(4), pp. 1172-1182, 2022.

[29] Y. Zhang, P. Wu, S. Chen, H. Gong, and X. Yang, "FCE-Net: a fast image contrast enhancement method based on deep learning for biomedical optical images," Biomedical optics express, vol. 13(6), pp. 3521-3534, 2022.

[30] Wai, C. Y., & Ngali, M. Z. B. "The Biomechanics Analysis: Development of Biomechanics Analysis Algorithm with OpenPose Motion Capture System," Research Progress in Mechanical and Manufacturing Engineering, vol. 2(2), pp. 658-668, 2021.

[31] M. Yamamoto, K. Shimatani, M. Hasegawa, Y. Kurita, Y. Ishige, and H. Takemura, "Accuracy of temporo-spatial and lower limb joint kinematics parameters using OpenPose for various gait patterns with orthosis," IEEE Transactions on Neural Systems and Rehabilitation Engineering, vol. 29, pp. 2666-2675, 2021.

[32] F. Angelini, Z. Fu, Y. Long, L. Shao, S. M. Naqvi, "2D pose-based real-time human action recognition with occlusion-handling," IEEE Transactions on Multimedia, vol. 22(6), pp. 1433-1446, 2019.

[33] F. C. Lin, H. H. Ngo, C. R. Dow, K. H. Lam, and H. L. Le, "Student behavior recognition system for the classroom environment based on skeleton pose estimation and person detection," Sensors, vol. 21(16), pp. 5314, 2021.

Twitter Truth: Advanced Multi-Model Embedding for Fake News Detection

Yasmine LAHLOU, Sanaa El FKIHI, Rdouan FAIZI
IRDA Group-ADMIR Laboratory-Rabat IT Center-ENSIAS, Mohammed V University in Rabat, Morocco

***Abstract*—The identification of fake news represents a substantial challenge within the context of the accelerated dissemination of digital information, most notably on social media and online platforms. This study introduces a novel approach, entitled " MT-FND: Multi-Model Embedding Approach to Fake News Detection," which is designed to enhance the detection of fake news. The methodology presented here integrates the strengths of multiple transformer-based models, namely BERT, ELECTRA, and XLNet, with the objective of encoding and extracting contextual information from news articles. In addition to transformer embeddings, a variety of other features are incorporated, including sentiment analysis, tweet length, word count, and graph-based features, to enrich the representation of textual content. The fusion of signals from diverse models and features provides a more comprehensive and nuanced comprehension of news articles, thereby improving the accuracy of discerning misinformation. To evaluate the efficacy of the approach, a benchmark dataset comprising both authentic and fabricated news articles was employed. The proposed framework was tested using three different machine-learning models: Random Forest (RF), Support Vector Machine (SVM), and XGBoost (XGB). The experimental results demonstrate the effectiveness of the multi-model embedding fusion approach in detecting fake news, with XGB achieving the highest performance with an accuracy of 87.28%, a precision of 85.56%, a recall of 89.53%, and an F1-score of 87.50%. These findings signify a notable improvement over traditional machine learning classifiers, underscoring the potential of this fusion approach in advancing methodologies for combating misinformation, promoting information integrity, and enhancing decision-making processes in digital media landscapes.**

Keywords*—*Fake news detection; transformer-based models; text classification; sentiment analysis

I. Introduction

In recent years, the rapid spread of fake news on social media has emerged as a significant challenge to public opinion and even to democratic processes. The widespread dissemination of fake news can cause considerable societal damage, including political polarisation, the erosion of trust in legitimate sources of information and the manipulation of public behavior. In order to meet this challenge effectively, it is essential to implement robust and accurate systems capable of detecting fake news in real time, with the aim of preventing its spread and mitigating its effects.

The traditional methods for the detection of fake news, which often rely on manual verification and heuristic approaches, are no longer sufficient in the context of the current volume and speed with which information is propagated on social media.

In contrast, advanced machine learning techniques offer a promising solution, automating the detection process and improving the accuracy of the results. These techniques facilitate the rapid analysis of large data sets, enabling the identification of patterns and anomalies that may indicate the presence of misinformation. As Lahlou et al. [1] have observed the deployment of machine learning and natural language processing techniques is vital for the identification and categorisation of fake news, given their capacity to handle extensive datasets and extract meaningful attributes. The objective of this study is to utilize advanced methods for optimizing the detection of fake news, particularly on Twitter. Similarly, Shu et al. [2] emphasize the potential of data mining and machine learning in combating the dissemination of fake news on social media platforms.

A number of studies demonstrate the effectiveness of NLP transformation models in identifying fake news, with promising results. In the field of natural language processing (NLP), BERT (Bidirectional Encoder Representations from Transformers) has emerged as a seminal model, distinguished by its capacity to discern intricate contextual nuances in textual data. Developed by Devlin and colleagues [3], BERT has achieved notable levels of accuracy in various Natural Language Processing (NLP) benchmarks through the utilisation of a bidirectional training process and a transformer architectural approach. The utilisation of BERT in the detection of fake news involves the analysis of textual content in an effort to identify any subtle linguistic cues that may indicate misinformation. Another advance in the field of transformer models is represented by the introduction of ELECTRA (Efficiently Learning an Encoder that Classifies Token Replacements Accurately), as proposed by Clark et al. [4]. In comparison to the conventional masked language modelling approach used by BERT, ELECTRA offers a more streamlined training mechanism. This methodology concentrates on the identification of substituted tokens within the textual data, thereby enhancing the model's capacity to discern intricate textual subtleties. In the domain of fake news detection, ELECTRA's discriminative training has demonstrated efficacy in discriminating between authentic and disingenuous content with enhanced precision.

The XLNet approach, as described by Yang et al. [5], integrates autoregressive and auto-encoding methodologies to capture bidirectional contextual dependencies, representing a significant advancement over previous models. By overcoming the limitations of traditional pre-training tasks, the XLNet model

achieves a superior performance in understanding complex linguistic structures and capturing subtle semantic relationships. In the domain of fake news detection, XLNet's capacity to model bidirectional dependencies allows it to identify intricate patterns of misinformation, enhancing the precision and dependability of detection. Recent research has further explored the integration of these transformation models into sophisticated frameworks for the detection of fake news. For instance, Liu et al. [6] illustrated the efficacy of integrating BERT and ELECTRA into a hybrid model. This model combines the respective strengths of both models, thereby achieving enhanced robustness in the detection of various types of fake news.

Despite the advances made to detect fake news, there are still some limitations to existing approaches. Many studies focus only on textual content, neglecting the rich contextual information available from user behavior and social context features as described by Lahlou et al. [7]. Furthermore, a single-model approach may not capture all language features needed for accurate detection [8] [9].

To address these limitations, we propose an advanced, representative approach to detecting fake news by harnessing the collective power of cutting-edge NLP transformative models and additional contextual features. These models, including BERT, ELECTRA, and XLNet, are at the forefront of natural language understanding, each offering unique capabilities to capture complex linguistic nuances [3] [4] [8]. BERT features context-sensitive language modelling using bidirectional transformers, while ELECTRA improves efficiency through a discriminative training approach. XLNet comprehensively captures bidirectional contextual dependencies by incorporating autoregressive and auto-encoding mechanisms.

The objective of this study is to integrate advanced transformer models derived from deep learning with traditional machine learning classifiers for classifying news articles into fake or real categories. A novel approach is introduced that harnesses the capabilities of BERT, ELECTRA, and XLNet transformers to extract contextual embeddings from tweets. These embeddings, in conjunction with sentiment analysis-derived features, serve as comprehensive input features for the classifiers.

The value of our work lies in its innovative approach to integrating state-of-the-art transformer models with traditional machine learning classifiers to tackle the ubiquitous problem of detecting fake news. The study exploits models such as BERT, ELECTRA and XLNet, which are capable of understanding and extracting complex contextual information from news headlines. This in-depth understanding of context is essential for accurately discerning the nuances between true and fake news, which often require a nuanced interpretation of language and sentiment. Furthermore, the incorporation of sentiment analysis-derived features enhances the model's capacity to discern subtle emotional nuances embedded in text, thereby providing a more comprehensive foundation for decision-making in classification tasks.

This study is structured as follows: Section II, titled "Related Works," provides an overview of existing AI techniques and methodologies employed to detect fake news on twitter. Section III delves into the problem formulation, detailing the architectural design and the proposed approach. Section IV outlines the methodology used in this research, describing the experimental setup, data preprocessing techniques, sentiment analysis, and the feature extraction process. It also covers the evaluation metrics used. Section V presents the experimental results, including a detailed comparison of classifier performance on both fused and individual model embeddings, assessing the effectiveness of the fusion approach. Finally, Section VI concludes the research by summarizing key findings, discussing the limitations of the current approach, and proposing potential directions for future work.

II. Related Works

In the past ten years, there has been a great deal of scientific research conducted on the detection of fake news on social media platforms. This study will present an overview of the key scientific research models developed to detect fake news on Twitter. The aim is to provide a comprehensive summary of these models. A threefold classification of the models in question is proposed: traditional machine learning models, deep learning models and transformation models.

With regard to machine learning, varieties of techniques are available for the differentiation between true and fake news on Twitter. These techniques include, but are not limited to, logistic regression, long-term memory, K-mean, support vector machine (SVM), random forest (RF), and Naive Bayes (NB). These techniques employ data pre-processing, feature extraction and sentiment analysis in order to enhance the accuracy of classification models. The linear SVM classification algorithm, which employs TF-IDF feature extraction, demonstrated the highest accuracy of 99.36% [10] [11]. Similarly, Raja [12]achieved a high level of accuracy (93%) using TF-IDF and an SVM classifier. Srinivas [13] employed TF-IDF in conjunction with MNB, RF, SVC and LR classifiers, achieving an accuracy of 79.05%.

Recent advancements in automated fake news detection have increasingly emphasized the integration of contextual features, including temporal patterns, social context, and cross-platform data. The analysis of temporal patterns, particularly in terms of how information spreads over time, can provide invaluable insights into the veracity of the content in question. For example, the rate and timing of tweet propagation frequently display anomalous spikes when fake news is disseminated.

The incorporation of contextual data into the process of feature extraction represents a pivotal stage in aligning the core intent of tweets with their content. This approach facilitates the enhanced detection of misinformation propagation on Twitter [14]. Zhou and Zafarani [15] demonstrated that temporal dynamics, when analysed with machine learning models, enhance the detection of fake news by capturing the evolving nature of misinformation. Furthermore, the utilisation of social context is an efficacious methodology for the identification of misinformation. The analysis of social signals, including user interactions, follower networks and retweet patterns, can provide valuable additional information that can be used to distinguish between genuine and fabricated content. As demonstrated by Shu et al. [16], machine-learning models trained on social network data are capable of effectively assessing the credibility of users disseminating information,

which facilitates improved detection of fake news. The study demonstrated that incorporating social context into feature extraction enables models to not only detect fake news but also trace its propagation paths, thereby providing valuable insights into the dynamics of misinformation spread.

Shetty et al. [17] put a comprehensive framework that incorporates linguistic features, user engagement models, and network analysis, thereby demonstrating effective detection of fake news. Bhogi et al. [18], employed a range of machine learning (ML) techniques, including natural language processing (NLP), classification algorithms, and anomaly detection, achieving an accuracy of 93% with the passive-aggressive classifier.

Another promising approach to the automatic detection of fake news is cross-platform analysis. Misinformation frequently disseminates across a multitude of social media platforms; thus, cross-platform analysis can discern patterns that may evade detection on a single platform. In a related study, Nguyen et al. [19], developed a framework that integrates data from various social media platforms with the objective of detecting inconsistencies in the manner in which news is presented and shared. The results indicate that machine-learning models incorporating cross-platform data are more resilient and effective in identifying fake news, as they are capable of detecting anomalies characteristic of coordinated misinformation campaigns.

Collectively, these studies demonstrate the potential of ML in detecting fake news on social networks.

The transition from machine learning to more sophisticated models, including those based on deep learning, has been instrumental in advancing the field. This has been achieved through the exploitation of the abilities of convolutional neural networks (CNNs), recurrent neural networks (RNNs) and long-term memory networks (LSTMs). Such models can successfully identify and analyse complex patterns within tweet content, user behavior and propagation dynamics. For example, Manaskasemsak and Rungsawang [20].proposed a deep neural network model that leverages tweet content published time and social graph features for the detection of fake news sources on Twitter with high accuracy.

Furthermore, Alghamdi, Lin, and Luo explored the integration of users' posting behavior clues with deep learning techniques, namely Convolutional Neural Networks (CNNs) and Bidirectional Gated Recurrent Units (BiGRUs), for the improvement of fake news detection on social media platforms [21]. Naik and Kumar emphasised the effectiveness of Long Short-Term Memory (LSTM) models in the automatic detection of fake news. They demonstrated the ability of these models to address the challenges presented by the rapid dissemination of misinformation on social media platforms [22]. Additionally, Monti and Sahoo [23] [24] emphasize the significance of considering social network structure and user behavior, in addition to content-based analysis. Monti's model, based on geometric deep learning, achieved an accuracy of 92.7% in detecting fake news on Twitter. In a further development, Kaliyar proposed a deep convolutional neural network (FND Net), [25] that attained a remarkable 98.36% accuracy in detecting fake news. Sedik [26] subsequently improved upon these results by proposing a deep learning-based system that combines concatenated and recurrent modalities, achieving an impressive 99.6% accuracy.

The recent advancements in deep learning have significantly enhanced the ability to detect fake news by integrating textual and contextual features. Mouratidis et al. [27] proposed a schema for textual inputs that are pairwise in nature, incorporating both the content of the news items in question and their context. The combination of content and context enables the model to more effectively capture the semantics and detect inconsistencies that are often indicative of fake news.

In a recent study, Bhatia et al. [28] introduced a deep neural network model that integrates multiple contextual features, including tweet content, publishing time, and social graph information. This multi-modal approach enables the model to consider a range of elements pertaining to the news, including the timing of publication and the social network of users sharing the content, in addition to the textual content itself. The combination of these diverse features enables the model to achieve an accuracy rate of 98.7% on the FakeNewsNet dataset. The incorporation of temporal and social graph features facilitates an enhanced comprehension of the broader context of news dissemination, thereby augmenting the model's capacity to identify fake news with greater reliability. This comprehensive approach underscores the significance of integrating diverse contextual data to enhance the efficacy of detection.

In the same context, Rajakumaran et al. [29] concentrated their attention on the possibility of early detection of fake news by examining propagation patterns and user interactions. The deep learning model employs contextual features, including user behavior and the propagation dynamics of news articles, to identify instances of fake news at the earliest possible stage of their dissemination. By monitoring early signals and interactions, their approach enables detection within hours of initial propagation, thereby providing a timely response to emerging misinformation. This capability is crucial for mitigating the impact of fake news and ensuring that accurate information prevails. The integration of contextual features related to propagation patterns enhances the model's sensitivity to early indicators of fake news, thereby demonstrating the importance of timely detection in combating misinformation.

In general, these deep learning techniques provide effective solutions for the automatic detection of fake news on Twitter, outperforming traditional classifiers and potentially complementing content-based approaches.

Transformer models, a subset of deep learning architectures, differ significantly from traditional deep learning models like Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs). While CNNs excel at handling spatial data and capturing local dependencies, and RNNs are adept at processing sequential data by maintaining temporal dependencies, they both have limitations in capturing long-range dependencies efficiently [30]. Transformers, on the other hand, utilize a self-attention mechanism that allows them to consider the entire context of a sentence simultaneously, making them highly effective at capturing long-range dependencies and contextual relationships across a text c.

The combination of the self-attention mechanism with the parallel processing capabilities of transformers results in more efficient training and a greater suitability for large datasets than that observed with RNNs, which are sequential in nature [31].

Furthermore, transformers benefit greatly from transfer learning through pre-training on extensive corpora followed by fine-tuning on specific tasks, leading to significant performance improvements in natural language processing tasks [32] [33]. Notable examples of transformer models include BERT, GPT, and T5, which have set new benchmarks in the field by leveraging these advanced techniques [32] [34].

Varieties of approaches are currently under consideration, including the TSRI model, which combines the characteristics of the text, source, and user response in order to obtain accurate predictions [35].

Recent research has focused on transformer-based approaches for fake news detection, leveraging both news content and social contexts. These models aim to address challenges such as early detection and limited labeled data [36].The proposed frameworks utilize transformer architectures to learn representations from news articles and social media data [37].

The utilisation of transformative models, such as RoBERTa, has demonstrated promising results in the detection of fake news across diverse datasets, with RoBERTa exhibiting superior performance in terms of accuracy and training time compared to other models [38]. Similarly, Mina Schütz and colleagues [39] investigated the potential of pre-trained transformer models for the detection of fake news, using the FakeNewsNet dataset. The methodology entailed the application of transformer-based models to the text of news articles and a combination of the text of news articles and their titles, with an accuracy rate of up to 85%. In their study, Divyam Mehta et al., [40] examine the effectiveness of BERT. The authors fine-tune BERT on LIAR datasets and improve its performance by incorporating additional inputs such as human justification and metadata. Their methodology involves adapting BERT's deep contextual capabilities to classify news articles into both binary and six-label categories. The study shows that BERT significantly outperforms traditional models in binary classification, achieving 74% accuracy. While the improvement in the more complex six-label classification is modest, it remains important and highlights BERT's ability to capture and utilise nuanced textual and contextual information.

These advances underscore the significant potential of transformer-based models in enhancing the accuracy and efficiency of fake news detection on platforms like Twitter, pushing the boundaries of what is achievable in this critical area of research.

Table I provides a comprehensive overview of the different machine learning, deep learning and transformer models used for fake news detection.

This comprehensive analysis compares and contrasts traditional machine learning, deep learning, and transformer models to illustrate the advancements in the detection of fake news. The application of traditional machine learning techniques, including logistic regression, SVM and Naive Bayes, is contingent upon data pre-processing and feature extraction. However, these techniques often encounter difficulties in addressing complex relationships and handling large data sets. Deep learning models, such as convolutional neural networks (CNNs) and long short-term memory (LSTM) units have advanced the field by automatically extracting complex patterns and processing large amounts of data, achieving significant results in terms of accuracy and user behavior analysis.

TABLE I. LITERATURE REVIEW FINDINGS SUMMARY

References	**Literature review findings summary**			
	Category	***Model***	***Dataset***	***Results***
[10], [11]	Machine Learning	SVM	Twitter data	99.36%
[12]		SVM	Twitter	93%
[13]		MNB, RF, SVC, LR	Twitter	79.05%.
[15], [16], [17]		Logistic Regression, SVM	Twitter	_
[18]		Passive-Aggressive Classifier	-	93%
[20], [21]	Deep Learning	CNN, BiGRU	Twitter	90%.
[23], [24]		Geometric Deep Learning	Twitter	92.7%
[25], [26]		FND Net, DNN	Twitter	99.6%.
[28]		DNN with Contextual Features	FakeNewsNet dataset	98.7%.
[41]	Transformer	BERT,	LIAR	74%
[39]		BERT	FakeNewsNet	85%

III. Proposed Multi-Model Embedding Approach

Before presenting the approach and methodology of the proposed fusion framework, it is imperative to define the problem in order to establish the objective of the proposed model.

A. Problem Definition

In the context of the detection of fake news on Twitter, the problem is defined as a supervised learning task with the objective of determining whether a specific tweet is true or fake. The task is formulated as a binary classification problem with the objective of accurately classifying tweets based on both the textual content and supplementary features derived through the utilisation of multiple transformer models.

Given a collection of tweets T= {T1, T2… Tn}, each tweet Ti consists of textual data. The corresponding labels L indicate the authenticity of each tweet, where:

- 0 indicates that the tweet is real.
- 1 indicates that the tweet is fake.

The objective is to model a prediction function F that takes as input a comprehensive feature vector **X(T)** derived from a fusion of multiple transformer models (BERT, ELECTRA, and XLNet) and additional handcrafted features. The function F predicts the label of the tweet, i.e., $F(T) \rightarrow \{0,1\}$, where:

- F(T)=0 if the tweet T is predicted to be real.
- F(T)=1if the tweet TTT is predicted to be fake.

In this fusion approach, the feature vector **X(T)** is composed of:

- Transformer-Based Embeddings:
  - CLS token embeddings from BERT and ELECTRA.
  - Mean-pooled embeddings from XLNet.
- Additional features:
  - Sentiment label and sentiment score.
  - Tweet length and word count.

The prediction function **F** is implemented as a machine learning classifier (e.g. Random Forest, XGBoost) that is trained on the fused feature vector X (T). The objective is to classify tweets as true or false using the diverse linguistic and contextual information provided by the fusion of multiple models and additional features.

The following section will provide a comprehensive explanation of the proposed approach.

B. Approach and Methodology

The objective of this article is to introduce the Multi-Model Embedding Approach to Fake News Detection (MT-FND), which represents a novel approach specifically designed for detecting fake news. It works by leveraging advanced natural language processing (NLP) techniques and integrating multiple transformer models—BERT, ELECTRA and XLNET.

The transformer models are selected for their robustness in NLP tasks, particularly in generating contextual embeddings that capture semantic nuances and contextual understanding, as demonstrated by prior research.

To operationalize the selected models within MT-FND, specific functions are employed for tokenizing text entries and extracting embeddings. For BERT and ELECTRA, embeddings are derived from the [CLS] token, while XLNet calculates embeddings by averaging all token embeddings. These embeddings serve as fundamental representations of the textual content. Additionally, technical features such as tweet length, word count, sentiment tags, and sentiment scores obtained through sentiment analysis are incorporated into the feature vector. Furthermore, MT-FND employs graph representation techniques using NetworkX in order to extend the scope of its analysis beyond that of textual embeddings. In this context, each tweet within the dataset is treated as a node within a graph, and is subsequently enriched with a number of attributes including labels, sentiment metrics, tweet length, word count, and sentiment scores. A placeholder function is employed to convert these attributes into preliminary embeddings, thereby establishing the foundation for representing articles within the graph structure.

This approach enables MT-FND to identify relationships between tweets, potentially revealing patterns that may not be discernible through text analysis alone.

The next step in the MT-FND approach involves using these enriched feature vectors to train two different classifiers: the Random Forest Classifier and the XGBoost Classifier.

The Random Forest Classifier is employed to utilise the combined embeddings derived from BERT, ELECTRA, and XLNet, in conjunction with supplementary features, to ascertain the veracity of a given tweet. This combination exploits the diverse strengths of each transformer model, thereby facilitating predictions that are more accurate. In contrast, the XGBoost Classifier is configured to assess the discrete embeddings from BERT and XLNet individually, thereby enabling it to capitalise on the distinctive capabilities of each model in formulating its predictions. The dual-classifier approach guarantees that MT-FND will benefit from both ensemble learning and the distinctive capabilities of each transformer model.

In summary, MT-FND improves the detection and classification of fake news by combining multiple transformer models (BERT, ELECTRA, and XLNet) with cutting-edge natural language processing techniques. It offers a comprehensive view of text content by fusing text elements with technical features like sentiment scores and tweet length. By utilizing Random Forest and XGBoost classifiers, MT-FND seeks to increase the precision of disinformation identification. Subsequent enhancements could involve utilizing NetworkX for graph representations, which could lead to better comprehension of relational data models via sophisticated graph neural network methods. This methodical approach uses complementary machine learning techniques and sophisticated transformers to tackle the challenges of detecting fake news. Fig. 1 provides an overview of the approach proposed and tested in this article.

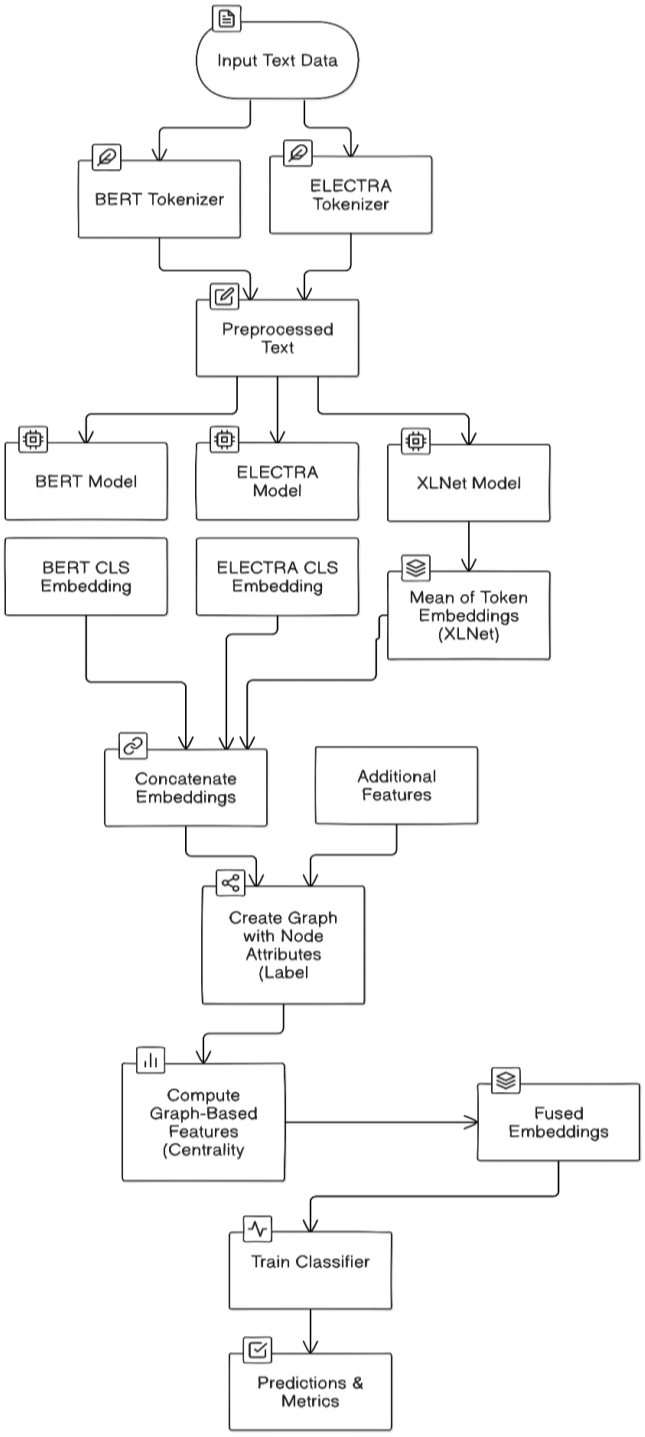


Fig. 1. Multi-model embedding approach to fake news detection (MT-FND).

IV. Experimental Set Up

In MT-FND approach, we opted for BERT, ELECTRA, and XLNet due to their established proficiency in natural language understanding tasks, each offering unique advantages such as contextual understanding, discriminative training efficiency, and bidirectional context capturing. By incorporating these diverse perspectives, we aimed to enhance the detection capabilities, effectively capturing the intricate nuances present in news articles, particularly within the constrained context of Twitter.

A. Dataset Description

For the purposes of this study, we utilized the FakeNewsNet dataset. The dataset is a comprehensive collection that was designed to help researchers study fake news detection and analysis. Two primary sources of data are included: BuzzFeed and PolitiFact. The dataset contains various features such as news content, social context, and spatiotemporal information. In the dataset, there are 11,510 news articles, with a proportionally balanced number of real and fake articles, with 5,755 articles labeled as real (label 1) and 5,755 articles labeled as fake (label 0). The balanced distribution of this dataset makes it possible for models trained on it to learn how to distinguish between real and fake news, which provides strong evaluation metrics and makes detection algorithms more reliable in real-world situations.

Fig. 2 shows the balance distribution of the dataset:

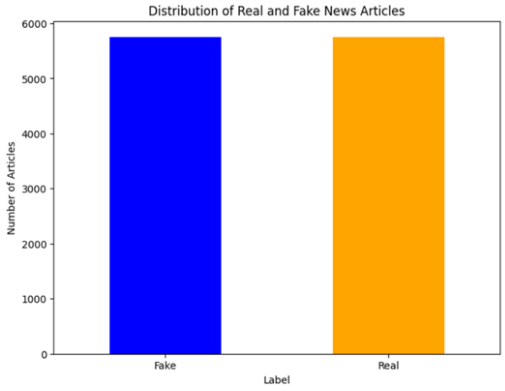


Fig. 2. Distribution of the balanced dataset.

B. Feature Engineering and Sentiment Analysis

The initial phase of feature engineering comprises the tokenization and embedding of news articles' titles, which is conducted through the utilisation of three distinct pre-trained transformer models. The models used are BERT, ELECTRA, and XLNet. Each model processes the text differently, thereby capturing various nuances of language. For example, BERT focuses on contextual relationships between words by employing bidirectional attention, while XLNet emphasize sequence order by considering permutations of input tokens. The output from each model is a high-dimensional vector representing the semantic content of the article's title. These embeddings serve as the foundational features for the classification model. In addition to the text embeddings, a sentiment analysis pipeline is utilised for the evaluation of the titles in question. This phase of the process classifies each ‘tweet’ as either "positive" or "negative" and assigns a corresponding sentiment score.

In order to represent the relationships between the tweets in a graph, the proposed approach involves constructing a graph using the NetworkX library. In this graph, each node represents a tweet, and edges are created based on cosine similarity between the text embeddings. Should the similarity between two articles exceed a specified threshold, an edge is formed, with the strength of this connection quantified by the similarity score.

Subsequently, centrality measures (such as degree centrality) and clustering coefficients are calculated for each node, thereby providing insights into the article's importance and its tendency to cluster with similar articles. These graph-based features are of paramount importance for understanding the broader context in which an article exists.

In the final stage, the extracted features are integrated into a unified representation for each article. The embeddings from BERT, ELECTRA, and XLNet are concatenated with sentiment analysis features and graph-based features such as centrality and clustering coefficients. This fusion creates a comprehensive feature set that captures both the textual and relational properties of each article.

C. Evaluation Criteria

The evaluation criteria used in this study are:

- Accuracy: The purpose of the accuracy metric is to ascertain the proportion of correctly predicted labels out of the total number of predictions. It is a simple yet effective metric for assessing the overall performance of the MT-FND model.

Metric: the accuracy metric is calculated during the testing phase by comparing the predicted labels (test_predictions) with the true labels (test_labels). The accuracy is computed using the accuracy_score function from the sklearn.metrics library, which divides the number of correct predictions by the total number of predictions. This metric provides a clear indication of the model's effectiveness in distinguishing between real and fake news articles.

- Precision: The purpose of the precision metric is to measure the accuracy of the positive predictions made by the MT-FND model. Specifically, it assesses the proportion of true positive predictions out of all positive predictions made by the model.

Metric: precision is calculated during the testing phase by comparing the predicted labels (test_predictions) with the true labels (test_labels). The precision is computed using the precision_score function from the sklearn.metrics library, which divides the number of true positive predictions by the total number of positive predictions. This metric is particularly important in scenarios where the cost of false positives is high, as it reflects the model's ability to avoid incorrect positive classifications.

- Recall: The purpose of the recall metric is to measure the model's ability to correctly identify all actual positive instances. It calculates the proportion of true positive predictions out of all actual positive cases.

Metric: recall is calculated during the testing phase by comparing the predicted labels (test_predictions) with the true labels (test_labels). The recall is computed using the recall_score function from the sklearn.metrics library, which divides the number of true positive predictions by the total number of actual positive instances. This metric is crucial in scenarios where missing a positive instance (false negatives) is costly, as it reflects the model's effectiveness in capturing all relevant instances.

- F1 Score: The purpose of the F1 score is to provide a balance between precision and recall, offering a single metric that accounts for both false positives and false negatives.

Metric: the F1 score is calculated during the testing phase by comparing the predicted labels (test_predictions) with the true labels (test_labels). The F1 score is computed using the f1_score function from the sklearn.metrics library, which combines precision and recall into a single metric.

V. Results and Discussions

A. Performance of MT-FND

In our evaluation of various classification models on a balanced dataset of real and false news articles, we introduced the Multi-Model Embedding Approach to Fake News Detection (MTFND), which integrates BERT, ELECTRA, and XLNet embeddings with graph features. This approach significantly enhances the accuracy and robustness of misinformation detection. The results clearly demonstrate the superiority of MTFND over models utilizing individual embeddings.

The performance of the proposed MTFND framework, which integrates fusion embeddings with graph features, demonstrates a strong enhancement in predictive capabilities across various classifiers. Among the models evaluated, the XGBoost classifier within the MTFND framework stands out, achieving an impressive accuracy of 87.28%. This model also shows a high precision of 85.56%, a recall of 89.53%, and an F1-score of 87.50%. These results highlight the ability of the MTFND framework to effectively capture and leverage complex features, providing a significant improvement over individual embedding approaches.

Table II presents a detailed comparison of model performance across various approaches, including the proposed MTFND framework and individual embeddings from BERT, XLNet, and ELECTRA.

TABLE II. Findings of the Experiments

Approach (Embedding type)	Model	Results by metrics type			
		Accuracy	*Precision*	*Recall*	*F1-score*
Proposed framework MTFND (Fusion embeddings + Graph Features)	RF	85.55	87.65	82.56	85.03
	SVM	84.97	84.09	86.05	85.06
	XGB	87.28	85.56	89.53	87.50
Individual BERT embeddings:	RF	84.97	84.09	85.06	84.97
	SVM	86.71	84.62	87.01	86.71
	XGB	84.97	84.09	85.06	84.97
Individual ELECTRA embeddings:	RF	79.19	82.89	73.26	77.78
	SVM	77.46	78.31	75.58	76.92
	XGB	81.50	83.75	77.91	80.72
Individual XLNET embeddings:	RF	78.03	79.27	75.58	77.38
	SVM	75.72	72.92	81.40	76.92
	XGB	78.61	77.53	80.23	78.86

When we evaluated models without the MTFND enhancements—relying solely on BERT embeddings or other contextual embeddings—the performance metrics were generally lower, highlighting the benefit of our approach.

For instance, the Random Forest classifier, when using only BERT embeddings without additional graph features, achieved an accuracy of 87.86%, a precision of 86.52%, and an F1 score of 88.00%. Although these figures are strong, they demonstrate that the inclusion of graph features in MTFND helps to provide a more nuanced understanding of the data, leading to predictions that are more accurate.

Similarly, the SVM classifier without the MTFND approach achieved an accuracy of 87.86%, a precision of 84.95%, and an F1 score of 88.27%. Again, while these results are respectable, they fall short of those achieved using the MTFND-enhanced models, particularly in terms of precision and F1 score, where the nuanced data representation provided by fusion embeddings and graph features shows its value.

The XGBoost classifier, even though robust, showed a noticeable drop in performance without the MTFND approach, recording an accuracy of 83.02%, a precision of 81.81%, and an F1 score of 79.27%. This further underscores the efficacy of the MTFND approach in enhancing the model's capability to detect and classify misinformation.

In conclusion, the MTFND approach significantly improves the performance of traditional machine learning models by enhancing their ability to analyze and interpret complex data patterns in news articles. The results clearly indicate that incorporating multimodal features, such as those in MTFND, is a powerful strategy for advancing the state of misinformation detection.

B. State-of-the-Art Comparison

Table III presents a comparative analysis of the proposed MTFND framework with several recent models from the literature. The MTFND framework achieved an accuracy of 87.28% using the XGB model, thereby demonstrating a significant improvement over the approaches described in study [39]. In particular, the BERT model proposed in study [39] achieved an accuracy of 85.0%. This comparison demonstrates that the MTFND framework, which integrates fusion embeddings (BERT, ELECTRA and XLNET) with graph features, exhibits superior performance in the detection of fake news in comparison to these transformer-based approaches. The XGB model within MTFND not only outperforms BERT and ALBERT in terms of accuracy but also demonstrates the efficacy of integrating diverse features and techniques. This suggests that the MTFND approach effectively harnesses both global text semantics and advanced features, resulting in a notable enhancement in performance. Further optimization and refinement of the MTFND framework may potentially yield even greater improvements.

TABLE III. COMPARATIVE MT-FND WITH OTHER WORKS

References	**Comparative MT-FND with other works**		
	Dataset	***Technique***	***Accuracy***
[39]	FakeNewsNet	BERT	85.0%
[40]	LIAR	BERT	74%
[13]	MNB, RF, SVC, LR	Twitter	79.05%.
Proposed framework MTFND	FakeNewsNet	Fusion embeddings (BERT,ELECTRA,XLNET)+ Graph Features with XGB classifier	87.28%

VI. CONCLUSION AND FUTURE WORK

This study presents a comprehensive approach to the detection of fake news, which integrates transformer-based embeddings with graph-based features. The integration of embeddings derived from BERT, ELECTRA, and XLNet with graph-based metrics has yielded a promising enhancement in performance, with notable improvements in accuracy, precision, recall, and F1-score values.

The fusion of transformer models' embeddings with additional features, such as sentiment scores, tweet length, and word count, provided a robust feature set that captures nuanced patterns in the data. By incorporating graph-based features like centrality and clustering coefficients, the model further enriched the representation of the data, enabling it to leverage the interrelationships between data points for improved classification. This hybrid approach outperforms traditional methods that rely solely on textual or structural features, underscoring the effectiveness of combining diverse sources of information.

By integrating the embeddings derived from three language models (BERT, ELECTRA, and XLNet) with graph-based features, the framework attains an exceptional accuracy of 87.28% on the FakeNewsNet dataset. This fusion approach exploits the distinctive capabilities of each model. The combination of BERT's contextual embeddings, ELECTRA's efficient training, and XLNet's permutation-based learning allows for the capture of a comprehensive representation of the text. Furthermore, the incorporation of graph features derived from cosine similarity and centrality measures enhances the feature set by capturing relational dynamics and structural patterns among articles.

This comprehensive approach not only enhances the model's capacity to discern subtle distinctions between authentic and fabricated news items but also improves its resilience and generalisability. The fusion of diverse embeddings with graph-based insights represents a significant advancement in the field of fake news detection, offering a more nuanced and effective solution.

Further research could enhanced by the incorporation of additional graph-based metrics, such as those pertaining to community detection and influence propagation. These advanced graph features have the potential to capture more intricate relationships and patterns within the data, thereby enhancing the model's ability to understand and detect nuanced misinformation. Furthermore, the investigation of novel or domain-specific transformer models, such as GPT-3, may facilitate enhancements in feature extraction and classification accuracy. By leveraging the latest advancements in model architecture and integrating these with graph-based insights, future models could achieve enhanced performance in the detection of fake news.

Finally, the expansion of the dataset through augmentation and enrichment represents another promising avenue of enquiry. The incorporation of diverse sources and the generation of synthetic data could improve the model's generalisation and robustness, addressing class imbalances and simulating various news scenarios. Furthermore, the development of real-time detection capabilities and the integration of the system with social media platforms would facilitate the timely and effective detection of fake news.

REFERENCES

[1] Lahrou, Y. (2019). Automatic detection of fake news on online platforms: A survey. In Proceedings of the 1st International Conference on Smart Systems and Data Science (ICSSD).

[2] Shu, K., Wang, S., & Liu, H. (2017). Fake news detection on social media: A data mining perspective. ACM SIGKDD Explorations Newsletter, 19(1), 22–36.

[3] Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. In Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics (NAACL).

[4] Clark, K., Luong, M. T., Le, Q. V., & Manning, C. D. (2020). ELECTRA: Pre-training text encoders as discriminators rather than generators. In Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP).

[5] Yang, Z., Dai, Z., Yang, Y., Cohen, W. W., & Salakhutdinov, R. (2019). XLNet: Generalized autoregressive pretraining for language understanding. In Proceedings of NeurIPS 2019: Advances in Neural Information Processing Systems.

[6] Liu, Y., Ott, M., Goyal, N., Du, J., & Clark, K. (2019). RoBERTa: A robustly optimized BERT pretraining approach. arXiv preprint arXiv:1907.11692.

[7] Lahrou, Y. (2021). Automatic detection of fake news on Twitter by using a new feature: User credibility. In Proceedings of the 5th International Conference on Big Data Cloud and Internet of Things (BDIoT).

[8] Yang, Z., Dai, Z., Yang, Y., Cohen, W. W., & Salakhutdinov, R. (2019). XLNet: Generalized autoregressive pretraining for language understanding. In Proceedings of NeurIPS 2019.

[9] Liu, Y., Ott, M., Goyal, N., Du, J., & Clark, K. (2019). RoBERTa: A robustly optimized BERT pretraining approach. arXiv preprint arXiv:1907.11692.

[10] Shabbir, A. S., & Khan, M. I. (2023). Fake Twitter followers detection using machine learning approach. In Proceedings of the International Conference on Business Analytics for Technology and Security (ICBATS).

[11] Hisham, M. (2023). An innovative approach for fake news detection using machine learning. University Research Journal of Engineering and Technology, 13.

[12] Raja, L. R. (2022). Fake news detection on social networks using machine learning techniques. Materials Today: Proceedings.

[13] Srinivas, J. S., & Pal, R. J. (2021). Automatic fake news detector in social media using machine learning and natural language processing approaches. In Proceedings of Smart Computing Techniques.

[14] Yadav, S. (2023). Machine learning based approach to disinformation detection using Twitter data. In Proceedings of the International Conference for Advancement in Technology (ICONAT).

[15] Zhou, X. (2018). Fake news: A survey of research, detection methods, and opportunities. arXiv preprint arXiv:1812.00315.

[16] Shu, K., Wang, S., & Liu, H. (2017). Fake news detection on social media: A data mining perspective. arXiv preprint arXiv:1708.01967.

[17] Shetty, M. S., & Swamy, A. S. (2023). Automated method for fake news detection using machine learning. In Proceedings of the International Conference on Network, Multimedia and Information Technology (NMITCON).

[18] Bhogi, A. D., & Nair, S. K. (2023). Machine learning for fake news detection on social media text. In Proceedings of the International Conference on Advances in Computation, Communication and Information Technology.

[19] Nguyen, V. S., & Do, T. H. (2020). FND: A framework for fake news detection on social media platforms. Computers in Human Behavior.

[20] Bhatia, T. M., & Kumar, A. (2023). Detecting fake news sources on Twitter using deep neural network. In Proceedings of the 11th International Conference on Information and Education Technology (ICIET) (pp. 508-512).

[21] Alghamdi, J. L., & Aljohani, S. (2023). Does context matter? Effective deep learning approaches to curb fake news dissemination on social media. Applied Sciences: Multidisciplinary Digital Publishing Institute, 13.

[22] Patel, U., & Gupta, P. (2023). Fake news detection using neural network. In Proceedings of the IEEE International Conference on Integrated Circuits and Communication Systems (ICICACS).

[23] Monti, F., & Defferrard, M. (2019). Fake news detection on social media using geometric deep learning. Computer Science.

[24] Sahoo, S. R., & Bhunia, A. (2021). Multiple features based approach for automatic fake news detection on social networks using deep learning. Applied Soft Computing.

[25] Kaliyar, R. K., & Sahu, M. (2020). FNDNet – A deep convolutional neural network for fake news detection. Cognitive Systems Research.

[26] Sedik, A. M., & Alshamrani, S. (2022). Deep fake news detection system based on concatenated and recurrent modalities. Expert Systems with Applications.

[27] Mouratidis, D., & Papatheodorou, C. (2021). Deep learning for fake news detection in a pairwise textual input schema. De Computis.

[28] Bhatia, T. M., & Sharma, A. (2023). Detecting fake news sources on Twitter using deep neural network. In Proceedings of the International Conference on Innovation Engineering and Technology.

[29] Rajakumaran, K. A. R., & Kumar, K. (2021). Fake news detection in Twitter datasets using deep learning techniques. Computer Science.

[30] LeCun, Y., & Bengio, Y. (2015). Deep learning. Nature, 521(7553), 436-444.

[31] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., & Jones, L. (2023). Attention is all you need. arXiv preprint arXiv:1706.03762.

[32] Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). BERT: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805.

[33] Radford, A., Narasimhan, K., & Salimans, T. (2019). Language models are unsupervised multitask learners. In Proceedings of the 2019 OpenAI.

[34] Raffel, C., Shinn, C., & Roberts, A. (2023). Exploring the limits of transfer learning with a unified text-to-text transformer. arXiv preprint arXiv:1910.10683.

[35] Luo, Y., Wu, X., & Yang, Z. (2023). Social media fake news detection algorithm based on multiple feature groups. In Proceedings of the IEEE 3rd International Conference on Information Technology, Big Data and Artificial Intelligence (ICIBA).

[36] Raza, S. (2022). Fake news detection based on news content and social contexts: A transformer-based approach. Computer Science.

[37] Do, T. H., & Tran, T. D. (2021). Context-aware deep Markov random fields for fake news detection. Computer Science.

[38] Babu, N. R., & Kumar, A. (2023). Performance evaluation of transformer-based NLP models on fake news detection datasets. Engineering, Electrical and Computer.

[39] Schütz, M., & Mehta, A. (2020). Automatic fake news detection with pre-trained transformer models. In Proceedings of the ICPR Workshops.

[40] Mehta, D., & Patel, R. (2021). A transformer-based architecture for fake news classification. Social Network Analysis and Mining.

[41] Kalkatawi, M. M., & Alotaibi, M. (2023). The detection of fake news in Arabic tweets using deep learning. Applied Science.

Protein-Coding sORFs Prediction Based on U-Net and Coordinate Attention with Hybrid Encoding

Ziling Wang[1], Wenxi Yang[2], Zhijian Qu[3]
School of Computer Science and Technology, Shandong University of Technology, SDUT, Country Zibo, China[1]
School of Computer Science and Technology, Shandong University of Technology, Country Zibo, China[2, 3]

***Abstract*—Small proteins encoded by small open reading frames (sORFs) exhibit significant biological activity in crucial biological processes such as embryonic development and metabolism. Accurately predicting whether sORFs encode small proteins is a key challenge in current research. To address this challenge, many methods have been proposed, however, existing methods rely solely on biological features as the sequence encoding scheme, which results in high feature extraction complexity and limited applicability across species. To tackle this issue, we proposed a deep learning architecture UAsORFs based on hybrid coding of sORFs sequences. In contrast to mainstream prediction methods, this framework processes sORF sequences using a mixed encoding approach, including both one-hot and gapped k-mer encodings, which effectively captures global and local sequence information. Additionally, it autonomously learns to extract features of sORFs and captures both long-range and short-range interactions between sequences through U-Net and coordinate attention mechanisms. Our research demonstrates significant progress in predicting encoded peptides from eukaryotic and prokaryotic sORFs, particularly in improving the cross-species predictive MCC index on the eukaryotic dataset.**

***Keywords*—Small open reading frames; deep learning; hybrid coding; U-Net; coordinate attention**

I. INTRODUCTION

With the rapid development of transcriptomic and proteomic technologies, researchers have come to better understand the potential coding regions of the genome [1]. Open reading frames (ORFs) are widely recognized as important sequence regions for protein coding [2], but short open reading frames (sORFs), as a group of ORFs up to 300 nucleotides in length, which were previously considered unlikely to code due to their short length [3, 4]. However, recent studies have shown that sORFs have a wide range of biological functions and can be directly transcribed and translated into biologically active small proteins [5-8]. These small proteins are involved in a variety of biological processes, including embryonic development [9], muscle function [5, 10-12], the regulation of cell growth and development [13-16], and the control of metabolic pathways [17]. Researchers have identified several sORFs with ribosomal activity by using versatile histological sequencing techniques, such as mass spectrometry and ribosome profiling [7, 18-20].

A prerequisite for the search for new protein-coding sORFs is their correct identification. Due to the short length, low expression levels, and lack of experimental validation for their functionality, sORFs have long been insufficiently annotated and studied. Investigating the coding potential of sORFs for small proteins is complex. Therefore, there is an urgent need for accurate and rapid methods to predict the coding ability of sORFs.

One of the main problems in predicting the microproteins-coding sORFs is to design an effective biological sequences coding scheme. The coding features are also crucial for distinguishing between coding and non-coding sORFs. Biological sequence coding schemes can be mainly divided into two types: sequential models and discrete models. Sequential models assign numerical values to each nucleotide in the biological sequence while preserving the order of the bases [1]. A prominent example of this is one-hot encoding (also known as C4 coding) [21], where each of the four nucleotides is represented by a unique four-bit binary vector (A-[1,0,0,0], C-[0,1,0,0], etc.). Each nucleotide's binary number is orthogonal to each other and has the same Hamming distance. In contrast, discrete models aim to design a set of features based on knowledge from the biological sequence. Some commonly used biological features include the codons usage [22], codon prototype [23], hexamer usage [24] and Z curves [25].

The sequential and discrete encoding models each present distinct advantages and limitations. While the sequential model preserves the global sequences order information [26], but this approach cannot fully capture biological features. Neural networks are not easily able to learn higher-order correlations from very low-level input [27]. Additionally, one-hot coding is unable capture frequency domain features such as k-mer [28]. On the other hand, taking 3-mer as an example, it is a discrete model of biological sequences and has become one of the features used to distinguish small proteins from non-encoding ones. Although the 3-mer is effective, it can only incorporate the local sequences order information between neighboring nucleotides and cannot reflect the global sequences order information [26].

Therefore, we designed a coding-protein prediction tool, named UAsORFs, utilizing a hybrid encoding strategy. This tool incorporates U-Net and Coordinate Attention (CA) mechanisms within its deep learning framework. This method effectively utilizes global sequence information, non-overlapping gapped k-mers and deep learning to autonomously learn sORF sequence features. By employing hybrid encoding, the method effectively extracts global and local sequence information of sORFs. Additionally, neural networks are employed to automatically differentiate between encoding and non-encoding sORFs.

The main contributions of this article are summarized as follows:

- A hybrid coding scheme combining sequence model and discrete model is designed. Unlike previous prediction tools that only use discrete models to extract sequence features, UAsORFs introduces sequence coding, effectively capturing global sequence information with one-hot coding, thereby enriching the encoding scheme and expression level.
- By exclusively using gkm as the discrete biological feature, excessive manual feature extraction is reduced, and gkm features significantly improve predictive performance on prokaryotic datasets.
- U-Net is used for the first time in the prediction of coding-protein sORF, facilitating the extraction of multi-scale, long-range and short-range interaction features of sORF sequences. A combined model (UCA) of U-Net and CA is constructed, leading to improved cross-species prediction results.

II. Related Work

In numerous biological prediction tasks, combining sequence information with biological features can lead to significant performance enhancements in specific applications. For instance, iTIS-PseTNC [26] introduced a sequence-based predictor for identifying translation initiation sites in human genes and claims that using k-mer representation in DNA sequences only reflects local sequence order information while losing global sequence order information. To overcome this limitation, the approach leveraged collaborative representations known as pseudo trinucleotide assemblies, integrating physicochemical properties into DNA sequences alongside k-mer features [26]. MHCDG [29] is a hybrid sequence-based deep learning model that integrates MeDIP-seq data with histone information to predict DNA methylation CpG status. By incorporating multiple biologically relevant features and sequence information, it outperformed other methods achieves more satisfactory promoter prediction performance. These works demonstrate the importance of hybrid coding.

In the prediction work of protein-coding sORFs, many prediction tools solely utilize various biological features for coding schemes. Among them, MiPepid [30] identified protein-coding sORFs using a logistic regression model and tetramer (4-mer) features from sequences. CPPred [31] is an SVM-base classifier that uses 38-dimensional biological features such as ORF coverage, Fickett score and CTD, etc., to predict the coding potential in both regular ORFs and sORFs. CPPred-sORF [32], based on CPPred coding, incorporates additional features such as GC count and mRNN-11 codons, and evaluates sORFs using non-AUG start codons. PsORFs [33] predict protein-coding sORFs in other species using 64 codon frequencies based on a random forest model trained on sORFs from prokaryotes. CodingCapacity [34] predicts the coding potential of sORFs using the Z-curve, codon frequencies, k-mer, and all features included in CPPred-sORF. DeepCPP [35] use a 589-dimensional feature set composed of nucleotide bias information and minimal distribution similarity. Notably, DeepCPP employs a convolutional neural network based on deep learning for prediction.

All the aforementioned work gives us a strong intuition that combining global sequence order information with biological features (such as gkm [36]) can enhance the prediction of protein-coding sORFs.

III. Materials and Methods

The prediction problem of protein-coding sORFs aims to determine whether an sORF has the ability to be transcribed and translated into a small protein. Given an sORF sequence $s = s_1 s_2 ... s_n$, the label of the sequence is $y = i, i \in \{1, 0\}$, i represents coding (1) or noncoding (0). Our goal is to convert the original sequences into a computer-recognizable format and predict it as either coding or noncoding sORFs using a deep learning framework.

A. Data Description

Our study aimed to establish a model for predicting the coding potential of sORFs in multiple species, covering both prokaryotes and eukaryotes. We utilized the standard dataset from PsORFs, which was generated based on a random order strategy. The same prokaryotic training dataset Pro-1282 and five test datasets (Hum-7111, Mou-7385, Ara-2125, Pro-6318 and Bac-150) from PsORFs were employed. Prokaryotic sORFs were selected from the Ref-Seq database [37]. Whereas human and mouse sORFs were downloaded from the sORFs.org database [38], while Arabidopsis thaliana sORFs were obtained from the TAIR database [39]. An experimental validation dataset (Bac-150) published by Hemm et al [40], included 150 positive sORFs and 53 negative sORFs detected from the E. coli genome. The detailed information of the datasets is presented in Table I, where coding sORFs refer to sequences capable of being translated into small proteins.

TABLE I. Number of Datasets for Each Species

Dataset	Species	Number of coding sORFs	Number of non-coding sORFs	Number of sORFs
Hum-7111	Prokaryotic genomes	7111	7111	14222
Mou-7385	Mouse	7385	7385	14770
Ara-2125	Arabidopsis thalian	2125	2125	4150
Pro-6318	Prokaryotic genomes	6318	6318	12645
Pro-1228	Prokaryotic genomes	1228	1327	2556
Bac-150	E.coli genome	150	53	203

The length distributions of sORFs across six datasets are illustrated in Fig. 1. The length distributions of sORFs in the Hum-7111 and Mou-7385 dataset are very similar, predominantly concentrated within the range of 60 to 140 nucleotides. In contrast to mammals, Arabidopsis, which is also a eukaryotic organism, exhibits a different distribution, with sORF lengths mainly distributed between 200 and 300 nucleotides. The length distributions of sORFs in the prokaryotic datasets Pro-6318 and Pro-1282 was similar, primarily concentrated between 150 and 300 nucleotides. Thus, it is evident that there are significant differences in sORF lengths among different species.

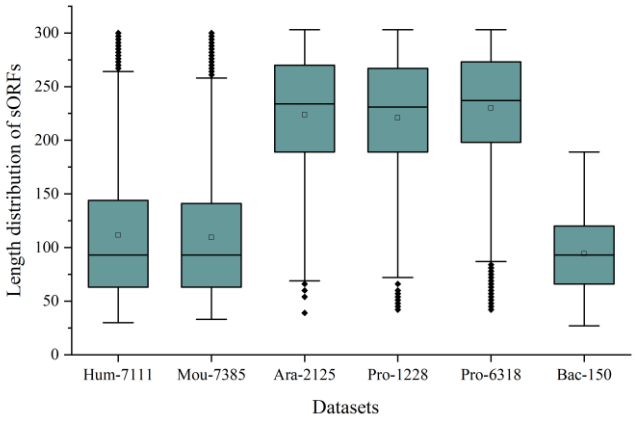


Fig. 1. Distribution of sORFs length of six datasets.

The differences in sORFs lengths among different species may pose challenges in predicting protein-coding sORFs in multiple species. sORFs with short lengths may be overlooked or misinterpreted as noise in certain species, thereby increasing the difficulty of prediction. Additionally, significant differences may exist in the sequence characteristics and preferences of sORFs across different species, making it challenging to identify universal features and patterns for predicting protein-coding sORFs across species. To address these challenges, the consideration of deep learning models is warranted, as deep learning offers advantages such as automatic feature learning, strong generalization capabilities, and flexibility.

B. Overview of the Designed Framework

As illustrated in Fig. 2, the overall workflow of UAsORFs comprises three stages. Firstly, the sORF sequences in the fasta file are preprocessed and encoded into one-hot and gkm coding formats. Subsequently, the one-hot coding is fed into the neural network UCA, which is composed of a U-Net and CA that can learn the importance of each nucleotide in the sequence. The resulting one-hot coding processed by the UCA model is concatenated with gkm coding to form a hybrid encoding representation. Finally, the hybrid encoding undergoes processing through a Multilayer Perceptron (MLP) and SoftMax activation function to obtain the final prediction results.

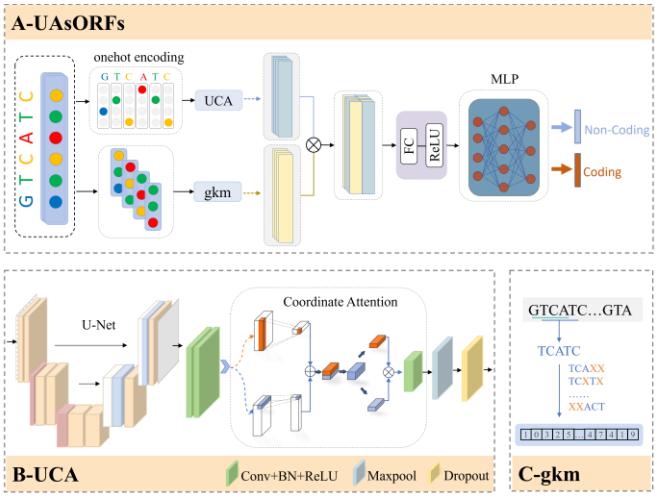


Fig. 2. The overall workflow of UAsORFs.

C. Hybrid Coding

To overcome the limitations of sequential models and discrete models, we propose a hybrid encoding scheme that combines global sequence models and biological features. The aim is to fully utilize the advantages of both, thereby enhancing the prediction of protein-coding sORFs. As illustrated in Fig. 3, For a given sequence $_S$, the hybrid encoding scheme can be formulated as follows:

$$E=[E_1,E_2] \tag{1}$$

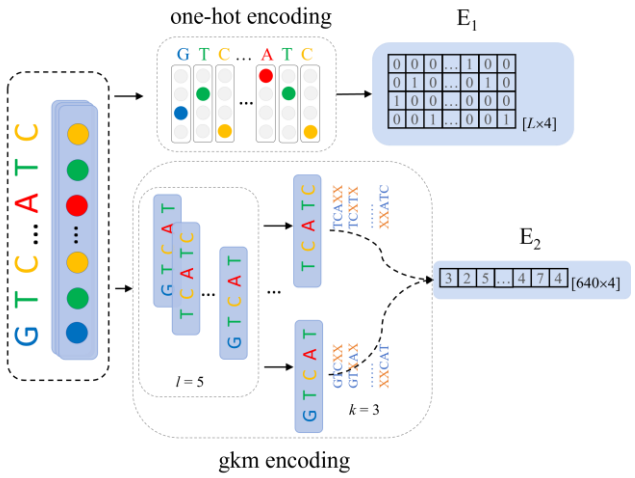


Fig. 3. Hybrid encoding scheme. E_1 is one-hot sequential model, E_2 is gkm discrete model.

Here, we utilize the one-hot sequential model [41] to capture the global sequence order information, Therefore, E_1 can be expressed as:

$$E_1=[x_1,x_2,...,x_i] \tag{2}$$

$$x_i=\begin{cases}[1,0,0,0], if\ i='A',\\ [0,1,0,0], if\ i='T',\\ [0,0,1,0], if\ i='G',\\ [0,0,0,1], if\ i='C',\\ [0,0,0,0], otherwise.\end{cases} \tag{3}$$

During one-hot encoding, we employ binary vectors of length 4 to encode the four nucleotides in biological sequences. Specifically, the position corresponding to a base is represented as 1, while the other positions are represented as 0. Additionally, when dealing with shorter sequences, we use the encoding [0,0,0,0] for sequence padding to maintain consistency in sequence length. Once a small protein sequence N of length i is inputted, a feature vector matrix is obtained to be fed into the model for training.

For the discrete model, we employ non-overlapping gkm [36] to capture local sequences order information effectively. K-mer, as a classic and effective feature representation, have been widely used in the field of bioinformatics, notably in the prediction of protein coding regions [42-44], coding potentials [31, 45], and identification of regulatory elements [29, 41].

Nevertheless, traditional k-mer methodologies are constrained by a pivotal issue that the increase of k leads to a very long and sparse feature vector [36]. To overcome this issue, we introduce the concept of gaps, which allows for certain mismatch exist in the k-mer sequence [36]. Gkm not only effectively reduce the dimensionality and sparsity of the feature vector but also demonstrates outstanding predictive prowess over conventional k-mer approaches, as evidenced by multiple studies in the biological domain [36, 46].

The gkm [36] method has two parameters: the length of the whole word l and the number of informative positions k. Therefore, the gap count is $l-k$. Combining previous work [1] and the dimension range of one-hot feature vectors, which is [100, 1200], we set $l=5$ and $k=3$. This not only effectively reduces the feature vector's dimensionality from 451,024 to $C_5^2 4^3 = 640$, which is close to the dimension of the one-hot feature vectors, but also allows using both together to enhance the ability to learn relevant patterns. This benefits the deep learning model by improving its expressive power and predictive accuracy. But also encompasses non-overlapping 3-mers information (e.g., AAAXX, ..., TTTXX). As shown in Fig. 3, when $l=5$ the length of each sORF subsequence is 5. For $k=3$, calculate the frequency of occurrence of each subsequence with three consecutive nucleotides. Thus, E_2 can be expressed as:

$$E_2 = [f(XXAAA), f(XAXAA), ..., f(TTTXX)] \quad (4)$$

Where $f(XXAAA)$ calculates the frequency of non-overlapping gapped trinucleotides (XXAAA) occurring in biological sequences. By introducing two gaps XX, the two words $GTACA$ and $CTACA$ of length 5 have the same gapped trinucleotides $XTXCA$.

D. UCA Model

Considering the hybrid coding approach used for sORF sequences, effectively integrating global sequence order information and gkm through deep learning, and autonomously learn features of sORFs of different lengths in different species is a problem that needs to be addressed. To this end, as illustrated in Fig. 2, we propose a UAsORFs architecture aimed at addressing this issue. In the UAsORFs architecture, sORF sequences are first encoded into one-hot coding and gkm feature representations. These are then processed through the UCA module and concatenated with gkm encoding.

While the issue of hybrid encoding has been discussed in previous sections, the focus of this section is to provide a detailed explanation of the UCA network module. The UCA module mainly consists of two key components: the U-Net [47-49] and the CA [50] based on convolutional neural network (CNN).

We are the first apply U-Net to the prediction of protein-coding sORFs in order to extract multi-scale, long-range and short-range interaction features from input sequences. This design is intended to ensure that the network can capture sufficient contextual information for longer sequences and maintains effective representation capabilities for shorter sequences.

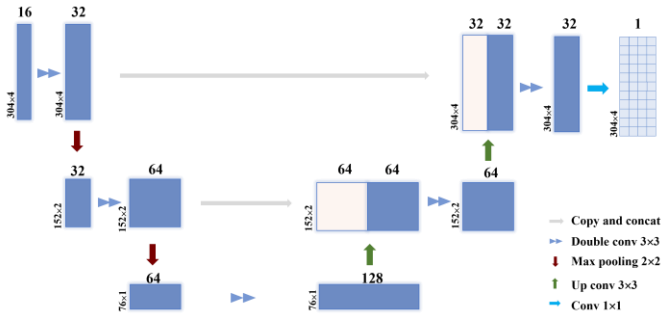


Fig. 4. U-Net model.

As shown in Fig. 4, the input of the U-Net network is denoted as E_1 after a convolution operation, while the output is represented as Y. The numbers at the top of the image represent the number of channels, while the lower-left corner shows dimensions as length × width. The kernels for Double conv and Up conv are 3 × 3, Conv uses a 1 × 1 kernel, and the max-pooling layer uses a 2 × 2 kernel. Through convolutional operations, $E_1 \in \{0,1\}^{304\times4}$ is transformed into $\in \{0,1\}^{16\times304\times4}$, where K can be regarded as an image that contains 16 color channels with a size of 304×4. Each channel signifies one of 16 possible nucleotide combinations, enabling the model to explicitly consider long-range interactions among nucleotides. Following processing by the U-Net, the output feature layer Y, $Y \in R^{16\times304\times4}$ forms a matrix of size $16 \times 304 \times 4$. Compared with the original one-hot coding, feature map Y can capture more abundant sORF sequence feature information, extract local patterns, context relations and higher-level semantic information in the sequence after multi-layer convolution and pooling layer processing of U-Net. It can provide powerful support for the next prediction work.

Next, for CA-based CNN, in image processing, CNN plays a crucial role in image processing, enabling the learning and extraction of effective image features. In the feature extraction phase, we employ a series of layer structures including Convolutional Layer, Batch Normalization (BN), Revised Linear Unit Activation Function (ReLU), Coordinate Attention, Max-Pooling and Dropout operation. The collaboration between these layer structures helps to achieve effective feature extraction and characterization of the input data.

In the traditional convolutional pooling process, applying channel attention mechanisms like Squeeze-and-Excitation (SE) attention [51] can assess the importance of each channel to learn the weights of different channel features [50]. However, the SE attention only considers encoding inter-channel information but neglects the importance of positional information, especially in cases of global sequential encoding (such as one-hot coding). The CA mechanism considers both inter-channel relationships and positional information within the feature space. Such a mechanism effectively focuses on different spatial locations of the input feature maps, enhancing the model's perception of key features and aiding in the better learning of useful features by the network model.

The CA Block is divided into two processes: Coordinate Information Embedding and Coordinate Attention Generation, as illustrated in Fig. 5.

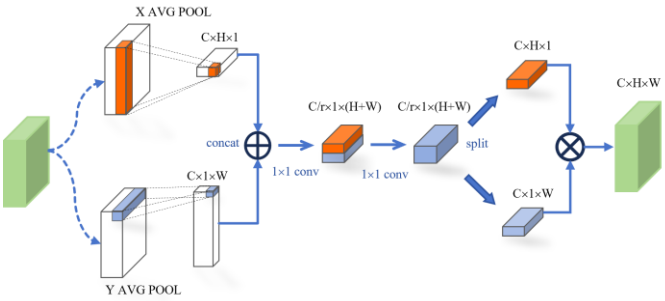


Fig. 5. Coordinate attention model.

$$y = \sigma(W_c \times f(W_g \times x)) \times x \tag{5}$$

Where x denotes the input feature map, W_c and W_g are the weight matrices of generation attention and channel attention, × denotes element-by-element multiplication, σ denotes the sigmoid function, and f denotes the weight coefficients of channel attention. Through the generative attention mechanism, a weight coefficient is learned for each spatial location. This weight coefficient is multiplied element-by-element with the input feature map to obtain a weighted feature map. Then, a weight coefficient is learned for each channel through the channel attention mechanism.

This weight coefficient is multiplied element-by-element with the weighted feature map and is used to weight the different channels of the feature map. This allows the network to pay more attention to the important channel information and suppress the unimportant channels to extract more effective feature representations and get the final output feature map.

IV. Training and Evaluation

A. Loss Function

To enhance the generalization ability and robustness against noise ability of the model, we use Label Smoothing (LS) loss as the loss function. The Label Smoothing loss function reduces the risk of overfitting and overconfidence by introducing a certain degree of smoothness. Its formula is as follows:

$$L = (1-\varepsilon) \times CE(y, y') + \varepsilon \times CE(\mu, y') \tag{6}$$

$$CE(p,q) = -\sum(p_i \times \log(q_i)) \tag{7}$$

Where y is the true label, y' is the output label probability distribution of the model, u is the smoothed label, ε is the smoothing factor, and CE is the cross-entropy loss function. In the loss function, we multiply the loss of the true labels by $(1-\varepsilon)$, multiply the loss of the smoothed labels by ε, and then weight and sum the two portions to get the final loss value. $\varepsilon > 0$ the loss portion of the smoothed labels will play a certain role of regularization, which helps to reduce the risk of overfitting.

B. Evaluation Indicators

We adopted four evaluation metrics, including Sensitivity (SN), Specificity (SP), Accuracy (ACC), and Matthews Correlation Coefficient (MCC), to evaluate the robustness of the model and its predictive performance for encoding sORFs. The formulas are as follows:

$$SN = \frac{TP}{TP+FP} \tag{8}$$

$$SP = \frac{FP}{TP+FP} \tag{9}$$

$$ACC = \frac{TP+TN}{TP+TN+FP+FN} \tag{10}$$

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP+FP)}\sqrt{(TP+FN)}\sqrt{(TN+FP)}\sqrt{(TN+FN)}} \tag{11}$$

Among them, TP and TN represent the number of correctly predicted coding and coding sORFs. FP and FN denote the number of incorrectly predicted coding and non-coding sORFs. SN and SP measure the model's ability to identify coding and non-coding sORFs. ACC reflects the proportion of correct predictions among all predictions. On the other hand, MCC comprehensively considers the relationships among TP, TN, FP, and FN, evaluating the correlation between predictions and annotations within the range of [-1, 1]. This metric system provides a comprehensive assessment of the model performance.

C. Training Parameter Settings

During training, we employed a learning rate decay strategy to prevent overfitting and accelerate the convergence of the learning algorithm. The initial learning rate was set to 4e-3, step size = 10 and gamma = 0.7. The Adam optimizer was chosen for parameter adjustment and optimization. In addition, set the batch size to 256, the number of epochs to 20 and the random seed to 42. Table II provides the UCA network architecture and parameters, with the U-Net and CA network architectures illustrated in Fig.4 and Fig.5. Our experimental environment consists of a CPU: AMD Ryzen 7 5800H and a GPU: NVIDIA GeForce RTX 3060.

TABLE II. UCA Network Architecture and Parameterisation

Layer	Size
Input	16×304×4
U-Net	16×304×4
CA	16×304×4
Conv	16×(3,3)
Maxpool	(2,2)
CA	16×152×2
Dropout	0.1
Conv	6×(3,3)
Maxpool	(2,1)
Dropout	0.1
Flatten	912
Concat	1096(912+640)
Linear	50
Linear	2
Softmax	2

V. RESULTS

In this section, we conducted four experiments on six datasets. The first demonstrated the importance of hybrid coding. The second and third experiments compared our approach with current popular methods on multiple species datasets. The fourth experiment demonstrated the importance of each module of UAsORFs.

A. Significance of Hybrid Coding

To validate the effectiveness of hybrid coding, we conducted ablation experiments aimed at separating hybrid coding and observing the performance of single coding in sORFs prediction. Specifically, as shown in Table III and Fig. 6, the hybrid coding was divided into one-hot and gkm encoding, which were then fed into MLP and CNN models (e.g., one-hot + CNN and gkm + MLP). We uesd a training and evaluation strategy trained on the Pro-6318 dataset and tested on the Ara-2125 (Fig. 6A) and Pro-1228 (Fig. 6B) datasets to evaluate the performance of hybrid coding on both eukaryotic and prokaryotic datasets.

TABLE III. DESIGN AND RESULTS OF ABLATION EXPERIMENTS FOR HYBRID CODING

Dataset	Methods	SN	SP	ACC	MCC
Ara-2125	PsORFs	0.4706	0.8613	0.6974	0.443
	one-hot+CNN	0.4815	0.9562	0.7188	0.4973
	k-mer+MLP	0.6918	0.8216	0.7267	0.5078
	gkm+MLP	0.5153	0.9501	0.7327	0.5168
	one-hot+k-mer+CNN	0.5976	0.9082	0.7529	0.5322
	one-hot+gkm+CNN	0.5962	0.9205	0.7584	0.5462
	one-hot+gkm+UAsORFs	0.6316	0.9031	0.7682	0.5571
Pro-1228	PsORFs	0.8698	0.8997	0.8908	0.7814
	one-hot+CNN	0.6458	0.9525	0.8051	0.6333
	k-mer+MLP	0.798	0.8432	0.8215	0.6424
	gkm+MLP	0.8739	0.9546	0.9142	0.8311
	one-hot+k-mer+CNN	0.8278	0.8964	0.8705	0.7418
	one-hot+gkm+CNN	0.8772	0.9584	0.923	0.8482
	one-hot+gkm+UAsORFs	0.9137	0.9435	0.9292	0.8582

According to the data presented in Table III and Fig. 6D, one-hot+gmk+CNN and one-hot+k-mer+CNN outperform one-hot+CNN, gmk+MLP, and k-mer+MLP on both the prokaryotic and eukaryotic datasets, which indicates that the hybrid coding scheme has a better prediction than the single coding model. It is worth noting that gmk+MLP achieved better predictive performance than k-mer+MLP, especially on the prokaryotic dataset, where ACC and MCC are improved by 9.27% (0.9142-0.8215) and 18.87% (0.8311-0.6424) , demonstrating the effectiveness of gkm (l=5, k=3) features in distinguishing coding and non-coding fields regions. Meanwhile, one-hot+CNN has outperformed PsORFs in Ara-2125, which proves the effectiveness of global order information.

In conclusion, our results suggest that there is a complementary relationship between one-hot coding and gkm features, and their combination helps deep learning methods to capture coding features more comprehensively.

B. Multi Species Predictions Results

To evaluate the performance of the different models, we conducted two experiments: (a) Training on the prokaryotic dataset Pro-1282 and testing on the remaining five datasets (Hum-7111, Ara-2125, Mou-7385 and Pro-6318, Bac-150). (b) Training on the prokaryotic dataset Pro-6318 and testing on the remaining five datasets.

Since there is no overlap of sequences in the test and training datasets, the multi-species validation is considered as independent dataset testing. With these two multi-species experiments, we evaluate the generalization performance of the model in multi-species prediction. These experimental designs help to validate the model's ability to generalize across different biological species and provide an important reference for further model improvement.

In our study, we evaluated seven different computational algorithms, some of which have been tested in the original literature on sORF [33]. As shown in Table IV，tools such as codingCapacity, PsORFs, CPPred-sORF, and DeepCPP employ discrete encoding schemes based solely on biological features. In contrast, UAsORFs enhances this discrete encoding framework by incorporating one-hot encoding, thereby increasing the representational capacity of sORFs. Furthermore, U-Net can integrate one-hot encoded features with spatial features, improving the accuracy of identifying specific regions or categories within sORFs data and enhancing segmentation performance, thus capturing details that other tools might overlook. The CA mechanism can emphasize crucial channels within the one-hot encoded data, ensuring that key classification information is prioritized and enabling the capture of significant features that might be missed by other tools.

As can be seen in Fig. 7 and 8, UAsORFs showed improvements across various independent datasets, particularly in terms of ACC and MCC metrics. On eukaryotic datasets, UAsORF achieved the highest ACC and MCC, surpassing the best-performing tool codingCapacity. In the Mou-7385 dataset, our method exhibited increases in ACC and MCC by 2.38% (0.5935-0.5697) and 7.57% (0.2398-0.1641). Similar improvements were observed in the Hum-7111 and Ara-2125 datasets. In the Pro-6318 dataset, UAsORF outperformed PsORFs but slightly lagged behind codingCapacity. On the Bac-150 dataset, UAsORF saw improvements in ACC and MCC by 2.15% (0.79-0.7685) and 8.96% (0.4715-0.3819).

From Fig. 8, it is evident that experiment (b) achieved better predictive performance compared to the training evaluation strategy of experiment (a). On the Bac-150 dataset, compared to PsORFs, UAsORFs showed approximately 5.12% increase in ACC (0.8-0.7488) and approximately 27.09% increase in MCC (0.5764-0.3055). With an increase in the number of training samples, UAsORFs more effectively captured sORF sequence features, resulting in better predictive performance. This also underscores the importance of constructing high-quality training and evaluation data.

In summary, our study demonstrates that UAsORFs exhibit strong generalization capabilities, showcasing excellent cross-

species predictive ability, and demonstrating their capacity to distinguish coding sORFs from non-coding ones. Furthermore, by altering training and evaluation strategies, we further validate the outstanding performance of the UAsORFs model in cross-species prediction.

C. UAsORFs Ablation Experiments

To investigate the robustness and reliability of the UAsORFs model, we conducted a series of ablation experiments. As shown in Table V, we systematically remove various components from the UAsORFs model, including the U-Net, CA and other modules. We utilized the Pro-1228 dataset for training and compared their predictive performance on both the eukaryotic (Hum-7111) and the prokaryotic (Pro-6318) dataset.

Fig. 9A and Fig. 9B show a performance comparison of the ACC and MCC across three eukaryotic test datasets (Hum-7111, Ara-2125, and Mou-7385) and two prokaryotic test datasets (Pro-6318 and Bac-150) under different methods. Fig. 9C and Fig. 9D show a performance comparison of ACC and MCC between Base, CA+LS and joining U-Net block on multi-species test datasets. Performance comparison of ACC index and MCC index on multi-species test dataset by adding CA(Fig. 9E), LS(Fig. 9F) block with Base et al.

From Table V and Fig. 9, it is evident that compared with the Base version, after adding U-Net, CA, and LS modules, the UAsORFs model improves the ACC(Fig. 9A) and MCC (Fig. 9B)to 62.39% and 29.59% on the eukaryotic dataset, and 91.7% and 83.67% on the prokaryotic dataset respectively. Specifically, in Fig. 9C and Fig. 9D, compared Base and CA+LS, the inclusion of the U-Net module in UAsORFs lead to significant improvements in ACC and MCC, increasing by 2.19% and 3.36% on the Hum-7111 test dataset. Fig. 9E and Fig. 9F demonstrate the enhancement in predictive performance after incorporating the CA and LS modules. The experimental results demonstrate the effectiveness of using U-Net, CA, and LS modules for extracting sORFs features.

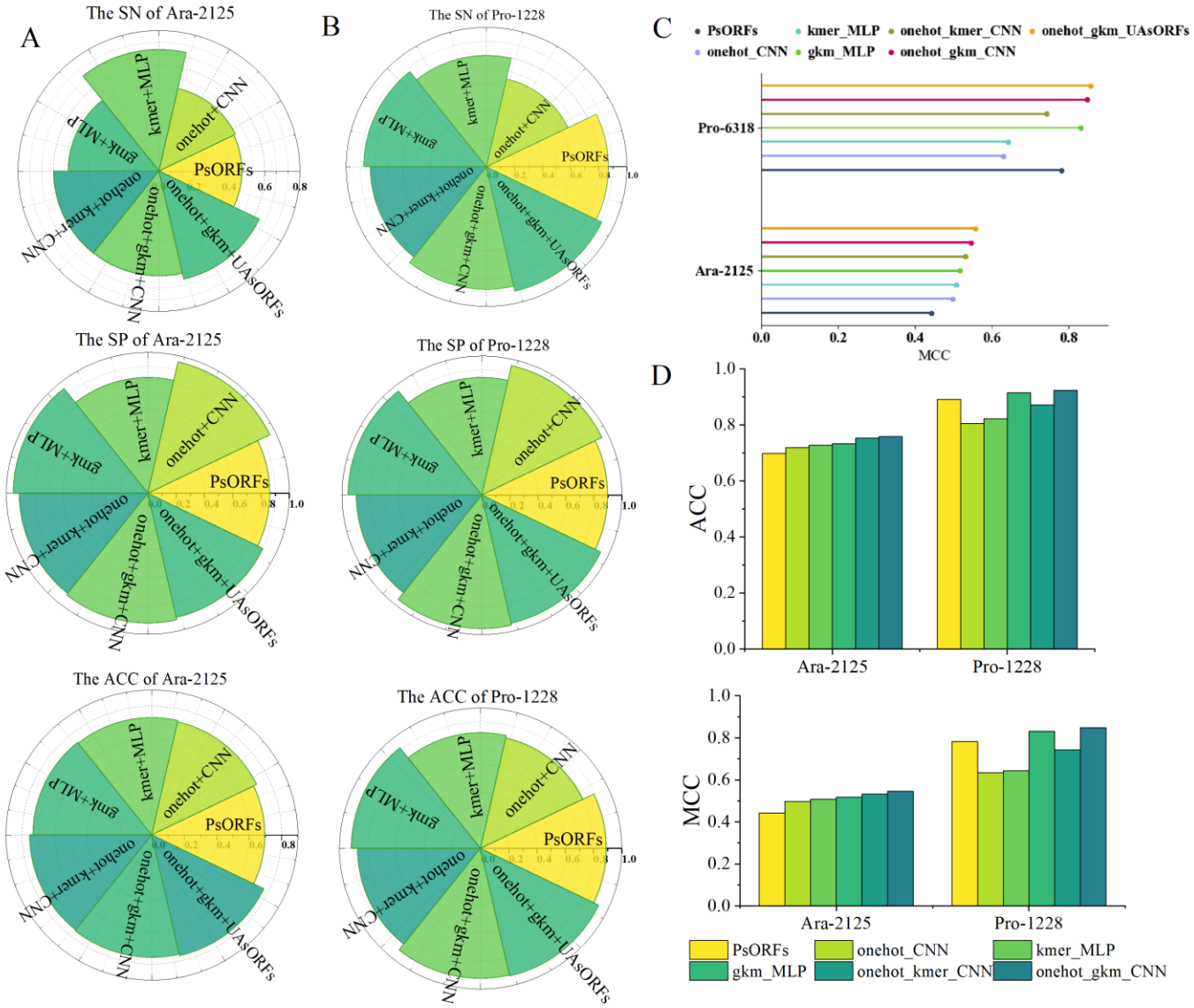


Fig. 6. Performance comparison of hybrid coding.

TABLE IV. COMPARISON OF UASORFS WITH SEVERAL OTHER PREDICTION TOOLS

Method	Year	Feature	Model
CPPred	2019	ORF length, ORF coverage, ORF integrity, Fickett score, Hexamer score, PI, Gravy, instability, CTD features	SVM classifier
MiPepid	2019	4-kmer	logistic regression
CPPred-sORF	2020	GCcount, mRNN-11codons and all features used by CPPred	SVM classifier
DeepCPP	2020	maximum ORF length, mean hexamer score, k-mer, ORF coverage, Fickett score, g-gap and nucleotide bias	CNN
PsORFs	2021	Codon frequency	Random forest
codingCapacity	2023	z-curve, codon frequency, k-mer and all features used by CPPred-sORF	Random forest
Our-method	2024	gmk	U-Net,CA

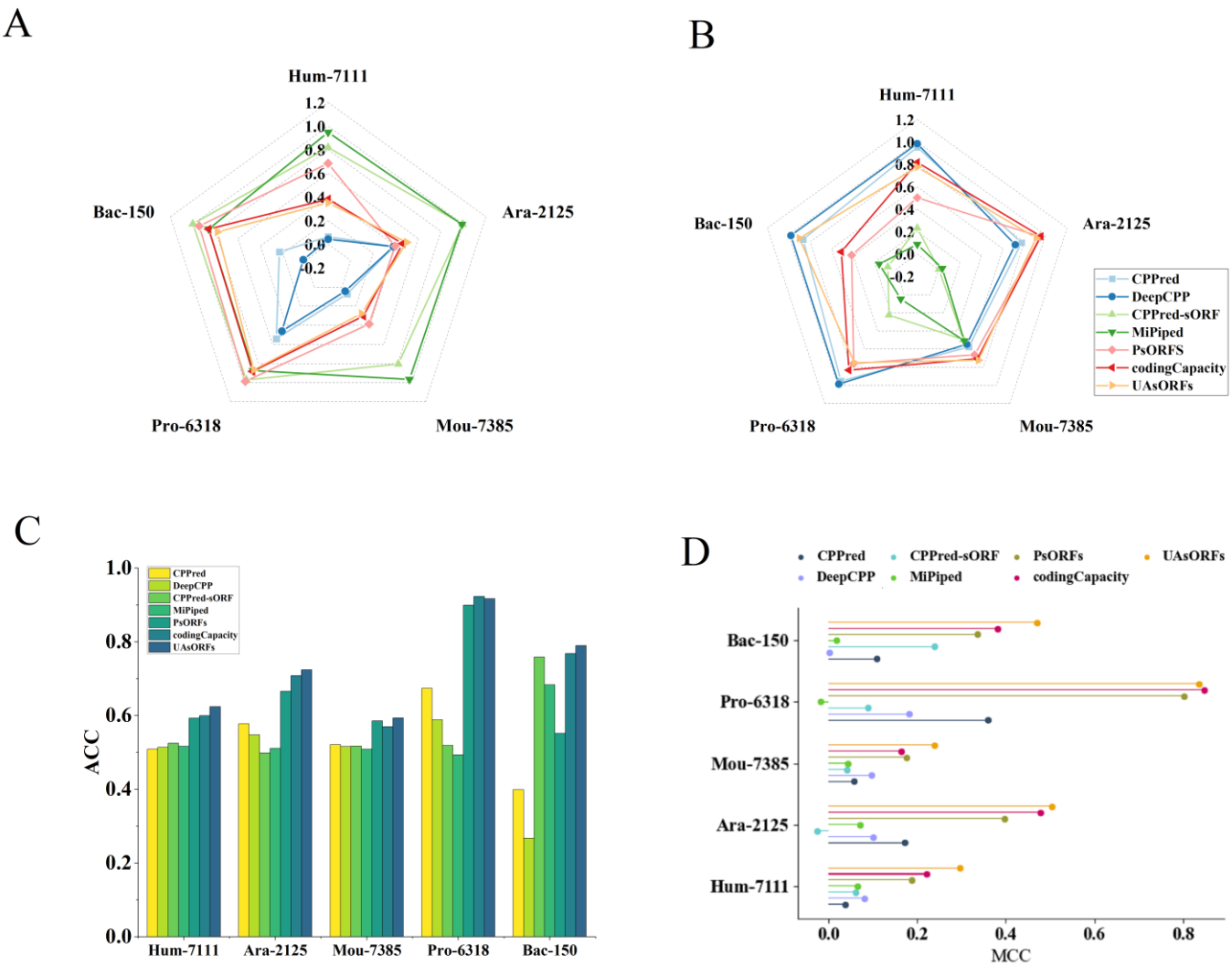


Fig. 7. The result of SN(A), SP(B), ACC(C)and MCC(D)for experiment (a).

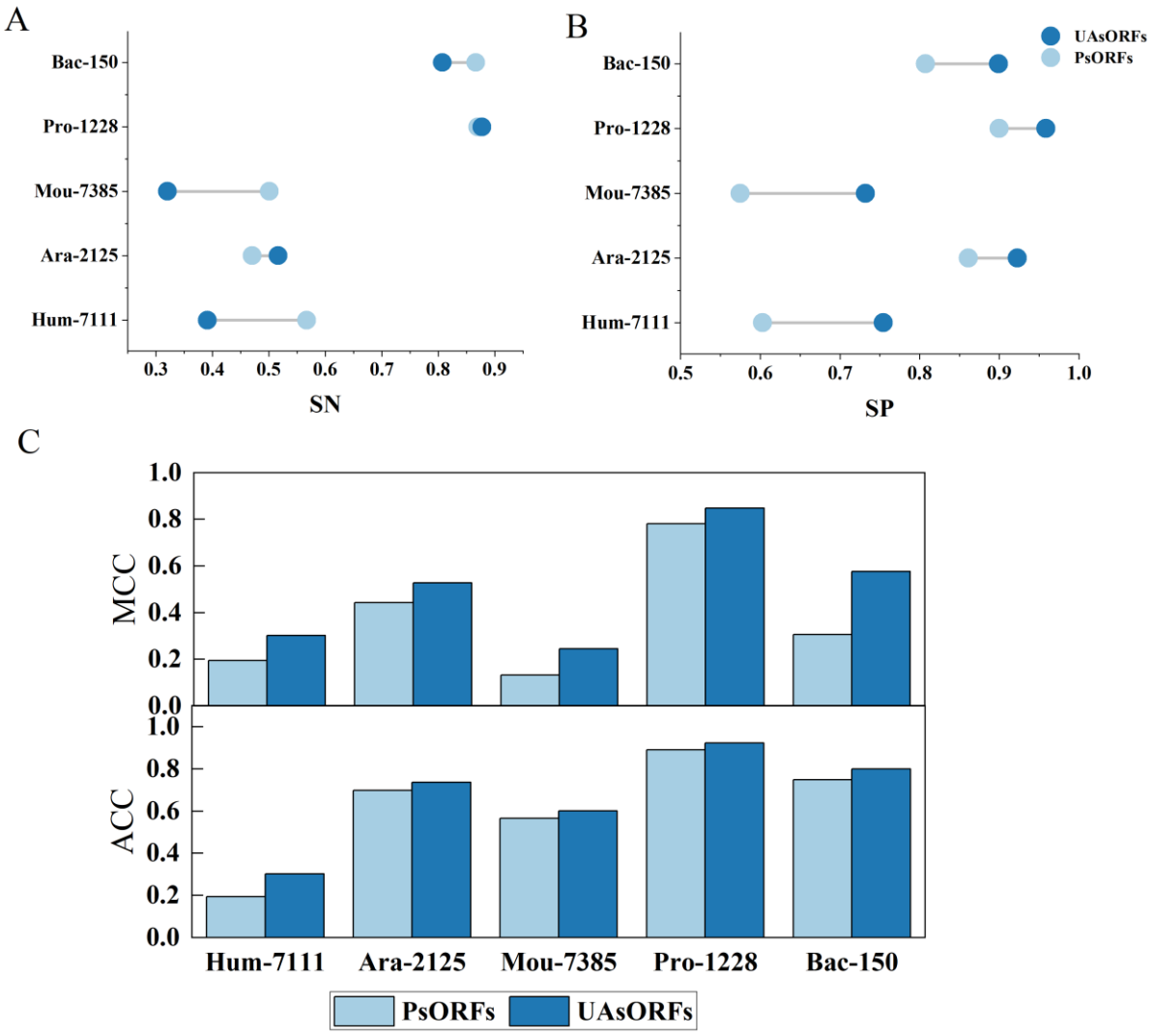


Fig. 8. The results of SN(A), SP(B), ACC and MCC(C) for PsORFs and UAsORFs using Pro-6318 train dataset.

TABLE V. DESIGN AND RESULTS OF ABLATION EXPERIMENTS FOR HYBRID CODING

Dataset	Method	U-Net	CA	LS	ACC	MCC
Hum-7111	Base	-	-	-	0.5719	0.1719
	CA+LS	-	√	√	0.6020	0.2623
	U-Net+LS	√	-	√	0.5679	0.1773
	U-Net+CA	√	√	-	0.5945	0.2301
	U-Net+CA+LS	√	√	√	0.6239	0.2959
Pro-6318	Base	-	-	-	0.8900	0.7864
	CA+LS	-	√	√	0.9094	0.8033
	U-Net+LS	√	-	√	0.8960	0.7983
	U-Net+CA	√	√	-	0.9156	0.8322
	U-Net+CA+LS	√	√	√	0.9170	0.8367

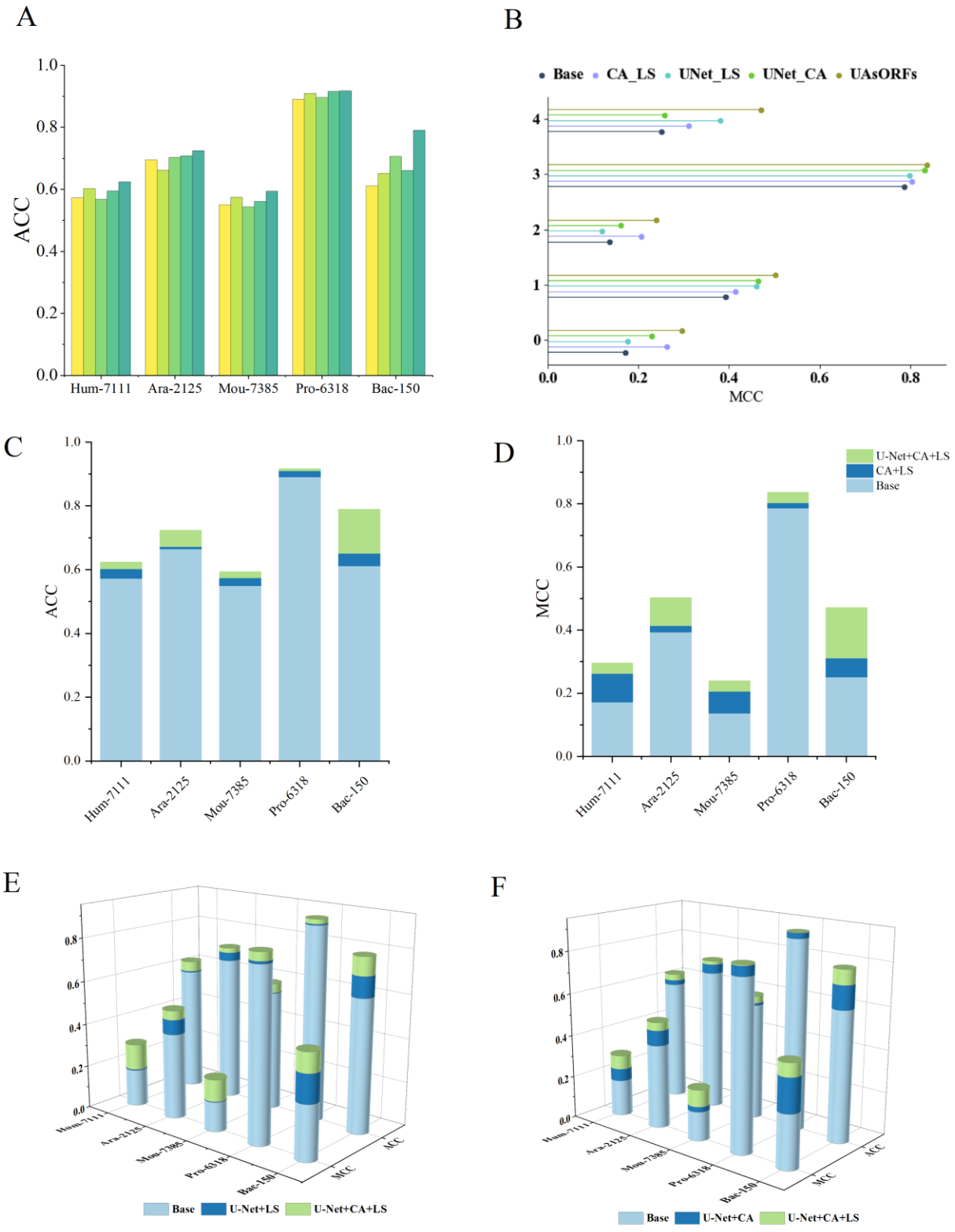


Fig. 9. Comparison of multi-species prediction performance of U-Net, CA and LS blocks of UAsORFs on independent test datasets.

D. Summarize

Through cross-species experiments and ablation studies on datasets from different species, we comprehensively evaluated the performance of the UAsORFs model, and verified the effectiveness of hybrid coding, U-Net, CA, and LS in the UAsORFs, which can effectively distinguish coding sORFs from non-coding ones. There are two main reasons for the outperformance of UAsORFs over the other methods. Firstly, the hybrid coding strategy can access and combine the global and local sequence information of sORFs to enhance the sORF representation. Secondly, deep learning effectively and autonomously learns to extract the features of sORFs, the feature fusion mechanism of up-sampling and intermediate variables in the U-Net module contribute to deep feature mining of sORFs. The CA attention mechanism is able to better capture the complex dependencies between nucleotides, thereby improving the understanding of interactions in sORF sequences. The channel attention mechanism can adaptively learn the importance of each channel, which enhances the model's representation of nucleotide pairing features.

VI. Conclusions

In this work, our study uses a hybrid encoding of one-hot and gkm coding, which retains both the global sequence order information and captures biological features. This approach fully utilizes the advantages of both methods, enhancing the encoding capability of the sequences and greatly avoiding the shortcomings such as insufficient sequence features and human intervention caused by a single encoding method. Additionally, we propose a deep learning architecture called UAsORFs, the deep learning framework distinguish between coding sORFs and non-coding sORFs through autonomous learning. The framework used in this study does not require extensive manual extraction of features, effectively learns essential sORFs features across multiple species and achieve remarkable predictive performance for multi-species sORFs. Additionally, the UAsORFs is a new, novel and efficient method for the prediction of protein-coding sORFs.

The study has seversal limitations. Firstly, the smaller sORFs dataset restricts the ability of the model to learn sORFs features. This is supported by the Multi-species predictions experiment (b), which demonstrates that increasing training samples allows UAsORFs to better capture sORFs sequence features, resulting in better predictive performance. Additionally, models trained on prokaryotic species exhibit suboptimal performance on eukaryotic protein-coding sORFs, suggesting that the current prokaryotic models may not capture certain features present in eukaryotic organisms. Future research should focus on expanding the sORF dataset and constructing a large-scale multi-species dataset to enhance capture features of eukaryotic sORFs and improve the applicability for sORF prediction. Furthermore, exploring species-specific characteristics and integrating other types of biological data (e.g., epigenetic marks, RNA-Seq data) could lead to the development of new biological sequence encoding schemes and further enhance prediction accuracy.

Acknowledgment

All the data mentioned in this article can be available at http://guolab.whu.edu.cn/codingCapacity/download.html.

We would like to thank the research team that provided the dataset.

Funding

This work was supported by the Outstanding Youth Innovation Teams in Higher Education of Shandong Province (2019KJN048).

Author Contributions

Ziling Wang: writing-original draft, methodology, analysis of results, data curation, validation. Wenxi Yang: analysis of the results, supervision, validation. Zhijian Qu: corresponding author, supervision, writing–review & editing, project administration, methodology, funding acquisition.

References

[1] WEI C, ZHANG J, YUAN X. Enhancing the prediction of protein coding regions in biological sequence via a deep learning framework with hybrid encoding. Digital Signal Processing, 2022, 123: 103430.

[2] SIEBER P, PLATZER M, SCHUSTER S. The Definition of Open Reading Frame Revisited. Trends Genet, 2018, 34(3): 167-70.

[3] WRIGHT B W, YI Z, WEISSMAN J S, et al. The dark proteome: translation from noncanonical open reading frames. Trends Cell Biol, 2022, 32(3): 243-58.

[4] FUCHS S, KUCKLICK M, LEHMANN E, et al. Towards the characterization of the hidden world of small proteins in Staphylococcus aureus, a proteogenomics approach. PLOS Genetics, 2021, 17(6): e1009585.

[5] ATAKAN M M, TüRKEL İ, ÖZERKLIĞ B, et al. Small peptides: could they have a big role in metabolism and the response to exercise? J Physiol, 2024, 602(4): 545-68.

[6] SANDMANN C-L, SCHULZ J F, RUIZ-ORERA J, et al. Evolutionary origins and interactomes of human, young microproteins and small peptides translated from short open reading frames. Molecular Cell, 2023, 83(6): 994-1011.e18.

[7] PRENSNER J R, ABELIN J G, KOK L W, et al. What Can Ribo-Seq, Immunopeptidomics, and Proteomics Tell Us About the Noncanonical Proteome? Molecular & Cellular Proteomics, 2023, 22(9): 100631.

[8] VAKIRLIS N, VANCE Z, DUGGAN K M, et al. De novo birth of functional microproteins in the human lineage. Cell Reports, 2022, 41(12): 111808.

[9] PAULI A, NORRIS M L, VALEN E, et al. Toddler: an embryonic signal that promotes cell movement via Apelin receptors. Science, 2014, 343(6172): 1248636.

[10] MATSUMOTO A, PASUT A, MATSUMOTO M, et al. mTORC1 and muscle regeneration are regulated by the LINC00961-encoded SPAR polypeptide. Nature, 2017, 541(7636): 228-32.

[11] NELSON B R, MAKAREWICH C A, ANDERSON D M, et al. A peptide encoded by a transcript annotated as long noncoding RNA enhances SERCA activity in muscle. Science, 2016, 351(6270): 271-5.

[12] ANDERSON D M, MAKAREWICH C A, ANDERSON K M, et al. Widespread control of calcium signaling by a family of SERCA-inhibiting micropeptides. Sci Signal, 2016, 9(457): ra119.

[13] ZHENG X, WANG M, LIU S, et al. A lncRNA-encoded mitochondrial micropeptide exacerbates microglia-mediated neuroinflammation in retinal ischemia/reperfusion injury. Cell Death Dis, 2023, 14(2): 126.

[14] ZHENG C, WEI Y, ZHANG P, et al. CRISPR/Cas9 screen uncovers functional translation of cryptic lncRNA-encoded open reading frames in human cancer. J Clin Invest, 2023, 133(5).

[15] PRENSNER J R, ENACHE O M, LURIA V, et al. Noncanonical open reading frames encode functional proteins essential for cancer cell survival. Nat Biotechnol, 2021, 39(6): 697-704.

[16] LAURESSERGUES D, COUZIGOU J M, CLEMENTE H S, et al. Primary transcripts of microRNAs encode regulatory peptides. Nature, 2015, 520(7545): 90-3.

[17] LEE C, ZENG J, DREW B G, et al. The mitochondrial-derived peptide MOTS-c promotes metabolic homeostasis and reduces obesity and insulin resistance. Cell Metab, 2015, 21(3): 443-54.

[18] MARTINEZ T F, LYONS-ABBOTT S, BOOKOUT A L, et al. Profiling mouse brown and white adipocytes to identify metabolically relevant small ORFs and functional microproteins. Cell Metabolism, 2023, 35(1): 166-83.e11.

[19] LI J, SMITH L S, ZHU H-J. Data-independent acquisition (DIA): An emerging proteomics technology for analysis of drug-metabolizing enzymes and transporters. Drug Discovery Today: Technologies, 2021, 39: 49-56.

[20] PALAZZO A F, KOONIN E V. Functional Long Non-coding RNAs Evolve from Junk Transcripts. Cell, 2020, 183(5): 1151-61.

[21] VOSS R F. Evolution of long-range fractal correlations and 1/f noise in DNA base sequences. Phys Rev Lett, 1992, 68(25): 3805-8.

[22] STADEN R, MCLACHIAN A D. Codon preference and its use in identifying protein coding regions in long DNA sequences. Nucleic Acids Research, 1982, 10(1): 141-56.

[23] SHEPHERD J C. Method to determine the reading frame of a protein from the purine/pyrimidine genome sequence and its possible evolutionary justification. Proc Natl Acad Sci U S A, 1981, 78(3): 1596-600.

[24] CLAVERIE J M, SAUVAGET I, BOUGUELERET L. K-tuple frequency analysis: from intron/exon discrimination to T-cell epitope mapping. Methods Enzymol, 1990, 183: 237-52.

[25] ZHANG C-T, ZHANG R. Analysis of distribution of bases in the coding sequences by a diagrammatic technique. Nucleic acids research, 1991, 19 22: 6313-7.

[26] CHEN W, FENG P M, DENG E Z, et al. iTIS-PseTNC: a sequence-based predictor for identifying translation initiation site in human genes using pseudo trinucleotide composition. Anal Biochem, 2014, 462: 76-83.

[27] RAJAPAKSE J C, LOI SY H. Markov encoding for detecting signals in genomic sequences. IEEE/ACM Transactions on Computational Biology and Bioinformatics, 2005, 2(2): 131-42.

[28] CHOONG A C H, LEE N K. Evaluation of convolutionary neural networks modeling of DNA sequences using ordinal versus one-hot encoding method; proceedings of the 2017 International Conference on Computer and Drone Applications (IConDA), F 9-11 Nov. 2017, 2017 .

[29] FU L, PENG Q, CHAI L. Predicting DNA Methylation States with Hybrid Information Based Deep-Learning Model. IEEE/ACM Trans Comput Biol Bioinform, 2020, 17(5): 1721-8.

[30] ZHU M, GRIBSKOV M. MiPepid: MicroPeptide identification tool using machine learning. BMC Bioinformatics, 2019, 20(1): 559.

[31] TONG X, LIU S. CPPred: coding potential prediction based on the global description of RNA sequence. Nucleic Acids Res, 2019, 47(8): e43.

[32] TONG X, HONG X, XIE J, et al. CPPred-sORF: Coding Potential Prediction of sORF based on non-AUG. bioRxiv, 2020.

[33] YU J, GUO L, DOU X, et al. Comprehensive evaluation of protein-coding sORFs prediction based on a random sequence strategy. Front Biosci (Landmark Ed), 2021, 26(8): 272-8.

[34] YU J, JIANG W, S.-B. Z, et al. Prediction of protein-coding small ORFs in multi-species using integrated sequence-derived features and the random forest model. Methods: A Companion to Methods in Enzymology, 2023.

[35] ZHANG Y, JIA C, FULLWOOD M J, et al. DeepCPP: a deep neural network based on nucleotide bias information and minimum distribution similarity feature selection for RNA coding potential prediction. Brief Bioinform, 2021, 22(2): 2073-84.

[36] GHANDI M, LEE D, MOHAMMAD-NOORI M, et al. Enhanced Regulatory Sequence Prediction Using Gapped k-mer Features. PLoS Computational Biology,10,7(2014-7-17), 2014, 10(7): e1003711.

[37] HAFT D H, DICUCCIO M, BADRETDIN A, et al. RefSeq: an update on prokaryotic genome annotation and curation. Nucleic Acids Res, 2018, 46(D1): D851-d60.

[38] OLEXIOUK V, MENSCHAERT G. Using the sORFs.Org Database. Curr Protoc Bioinformatics, 2019, 65(1): e68.

[39] BERARDINI T Z, REISER L, LI D, et al. The Arabidopsis information resource: Making and mining the "gold standard" annotated reference plant genome. Genesis, 2015, 53(8): 474-85.

[40] HEMM M R, WEAVER J, STORZ G. Escherichia coli Small Proteome. EcoSal Plus, 2020, 9(1).

[41] ARNIKER S B, KWAN H K, LAW N F, et al. DNA numerical representation and neural network based human promoter prediction system; proceedings of the 2011 Annual IEEE India Conference, F 16-18 Dec. 2011, 2011.

[42] WEI C, YE Z, ZHANG J, et al. CPPVec: an accurate coding potential predictor based on a distributed representation of protein sequence. BMC Genomics, 2023, 24(1): 264.

[43] BERNARD G, GREENFIELD P, RAGAN M A, et al. k-mer Similarity, Networks of Microbial Genomes, and Taxonomic Rank . mSystems, 2018, 3(6).

[44] HATZIGEORGIOU A G. Translation initiation start prediction in human cDNAs with high accuracy. Bioinformatics, 2002, 18(2): 343-50.

[45] WEN J, LIU Y, SHI Y, et al. A classification model for lncRNA and mRNA based on k-mers and a convolutional neural network. BMC Bioinformatics, 2019, 20(1): 469.

[46] LAFFERTY J D, MCCALLUM A, PEREIRA F. Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data; proceedings of the International Conference on Machine Learning, F, 2001 .

[47] ZHOU H, HU B, YI N, et al. Balancing High-performance and Lightweight: HL-UNet for 3D Cardiac Medical Image Segmentation. Academic Radiology, 2024.

[48] WANG B, QIN J, LV L, et al. DSML-UNet: Depthwise separable convolution network with multiscale large kernel for medical image segmentation. Biomedical Signal Processing and Control, 2024, 97: 106731.

[49] LUO K, TU F, LIANG C, et al. RPA-UNet: A robust approach for arteriovenous fistula ultrasound image segmentation. Biomedical Signal Processing and Control, 2024, 95: 106453.

[50] HOU Q, ZHOU D, FENG J. Coordinate Attention for Efficient Mobile Network Design; proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), F 20-25 June 2021, 2021.

[51] HU J, SHEN L, SUN G. Squeeze-and-Excitation Networks; proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, F 18-23 June 2018, 2018 .

An Improved YOLOv8 Method for Measuring the Body Size of Xinjiang Bactrian Camels

Yue Peng, Alifu Kurban, Mengmei Sang
Department School of Software, Xinjiang University, Urumqi, China

***Abstract*—Camel body size measurement has initially been applied in livestock production. However, current methods suffer from low measurement accuracy due to detection box localization loss and occlusions. This study proposes an effective algorithm, Camel-YOLOv8, specifically designed for detecting Xinjiang Bactrian camels and calculating their body sizes. By integrating the Selective Kernel Networks (SKAttention) mechanism and an enhanced Asymptotic Feature Pyramid Network structure (AFPN-beta), the algorithm successfully captures the body characteristics of Bactrian camels in natural environments and converts these into precise size data. We have developed a Xinjiang Bactrian camel body size measurement dataset and applied the enhanced YOLOv8 model for accurate classification and detection. By extracting key point pixel values and applying Zhang Zhengyou's calibration method, the coordinate system data is converted into accurate body size measurements. The Camel-YOLOv8 achieves a detection accuracy of 76.4% on the Xinjiang Bactrian camel dataset, marking a 3.7% improvement over the baseline model. In terms of body size calculation, the average relative errors for height and chest circumference are -3.39% and 4.1%, respectively, demonstrating high measurement precision. The algorithm not only maintains high detection accuracy but also achieves a reasonable balance between detection speed and efficiency, providing an effective solution for rapid acquisition of camel body size information.**

Keywords*—YOLOv8; Asymptotic Feature Pyramid Network; SKAttention; Bactrian camel body size measurement

I. INTRODUCTION

The scaling and precision enhancement in modern camel farming have led breeders to recognize the critical importance of selecting, cultivating, and fostering the healthy development of superior camel breeds. The physical development and size of camels are reflected through body measurement indicators, closely linked to the camels' adaptability to their environment, productivity, reproductive performance, and economic value [1]. Accurate determination of camel body size and morphological assessments are vital for monitoring camel growth, genetic selection within herds, improving reproductive capabilities, enhancing productivity, and standardizing rearing practices [2]. Traditional livestock measurement and condition assessment methods, which typically involve contact tools like tape measures, can induce stress in animals, impacting their growth and development. With the evolution of technologies such as image processing [3], pattern recognition [4], and artificial intelligence [5], machine learning [6] has increasingly been applied in China's livestock industry, yielding significant results. Leveraging machine vision for measuring camel body size and performing morphological evaluations facilitates a comprehensive analysis of camel development, quickly assessing their growth conditions and preparing for potential health crises.

The application of computer vision [7] in livestock management is diverse and widespread, enhancing management efficiency, productivity, and animal welfare through advanced image analysis techniques. Computer vision systems can autonomously monitor animal activities, behavioral patterns, and body language, identifying behaviors such as eating, resting, and socializing. These insights are crucial for assessing animal health, welfare, and productive performance. By analyzing appearances and behaviors, computer vision technologies can aid in the early identification of disease signs or health issues, such as limping, weight changes, or decreased appetite, enabling timely intervention [8]. Computer vision can also measure animal body dimensions, such as length and height, to estimate weight and monitor growth, aiding in nutritional management and breeding program improvements. By integrating computer vision with automation technologies, feeding quantities and nutritional ratios can be automatically adjusted based on an animal's weight, health condition, and growth needs, supporting individual animal management, population structure analysis, and breeding and selection decisions [9].

The advancement and implementation of computer vision technology have significantly raised the productivity and management levels in the livestock industry, also providing strong support for enhancing animal welfare and ensuring food safety. Through target detection technologies, camel body size data can be calculated using detection boxes. However, ensuring the accuracy of these calculations is crucial, as the precise positioning of detection boxes is paramount. Detection box localization loss refers to inaccuracies in positioning detection boxes during the object detection process due to various factors, impacting the accuracy of body size calculations. Additionally, occlusion among camels, a common issue due to their large size and tendencies to move closely within groups, can cause some parts to be obscured by other camels or objects, affecting the accuracy of detection box positioning. This occlusion can lead to inaccuracies in the detection box area, thereby affecting the results of body size calculations. The primary contributions of this paper include:

1) Establishing a rich dataset containing images of camel postures in various environments, providing necessary data support for algorithm training and evaluation.

2) Current camel body size detection methods suffer from detection box localization loss; in YOLOv8 [10], the

SKAttention mechanism is added to reduce this loss, thereby minimizing errors in calculating camel body size.

3) Using the Asymptotic Feature Pyramid Network (AFPN-beta) structure effectively resolves issues of camel occlusion, ensuring that the body size information of obscured camels can be more accurately calculated.

4) Employing the camera calibration method from Zhang Zhengyou's calibration technique to obtain camera model parameters, such as internal, external, and distortion parameters. Based on the single-camera imaging principle and the transformation relationships between coordinate systems, a measurement model for the body size of Xinjiang Bactrian camels is established, facilitating the measurement of body size information for Bactrian camels.

In summary, Section II of this paper reviews the existing research in the field of camel body measurement and analyzes the current state of computer vision and object detection technologies in livestock management. Section III provides a detailed description of the improved YOLOv8 method proposed in this paper, including feature selection, integration of the SKAttention mechanism, application of the AFPN-beta structure, and the computation methods for camel body metrics. In Section IV, we present the experimental design, dataset preparation, training parameter settings, and evaluation metrics of the model, and provide an in-depth analysis of the experimental results. Section V summarizes the main research findings of this paper and discusses future research directions.

II. Related Work

The use of computer technology for animal biometric measurement through image processing and feature extraction facilitates efficient and accurate measurements, widely applied in biological research and ethology. Federico Pallottino [11] and colleagues compared manual measurements with stereovision measurements, finding a high overall correlation and lower variability, with an average error below 3%. The variability in error magnitude depends on the specific traits, but Pallottino's study did not delve into the theoretical basis of the measurement algorithm. Additionally, the research focused mainly on correlation analysis without involving the design and optimization of specific algorithms, limiting its potential for technical deepening and application expansion. Qin [12] and others introduced deep learning, proposing a unique method for measuring livestock body dimensions using Mask R-CNN [13] to measure cattle and goats of various sizes against different backgrounds. This study initially utilized the idea of human joint location to accurately position livestock body feature points and perform precise measurements. However, due to the unique physiological structure of camels, especially the presence of humps, the use of deep learning models like Mask R-CNN and the search for measurement points can cause significant interference. This issue limits the suitability and accuracy of this technology in camel measurement. Mahdi Khojastehkey [14] and colleagues used SPSS software and Pearson correlation coefficients to select features more relevant to single-humped camel measurement, discovering the mathematical relationships between digital image-extracted features and camel body dimensions. This finding provides an important theoretical basis for developing more accurate camel body dimension calculation models, further supporting the feasibility and effectiveness of using digital image technology for camel body measurement.

Computer vision, a critical component of the field of artificial intelligence, plays a vital role across various industries. With the development of machine learning, especially deep learning, the application of object detection [15], [16] has expanded beyond emerging industries. In traditional sectors, such as livestock, object detection algorithms are gradually showcasing their significant functions. AlNujaidi [17] and others researched a camel-vehicle collision mitigation system through computer vision. Although progress was made in camel detection, the study did not extend to specific calculations of camel body dimensions, limiting its completeness and practicality in biometric applications. Wang [18] and others proposed a portable and automated Xtion measurement system for assessing pig body sizes, which showed significant advantages in measuring pig body dimensions and confirmed the importance of these dimensions in predicting weight. However, this system needs more detailed measurement methods for body dimensions to ensure accuracy when adapting to the more complex morphology of camels. Finally, Li and Teng [19] studied deep learning-based body dimension measurement methods for goats and cattle against different backgrounds. While this method performs well in measuring cattle and goat body sizes, the complexity and specificity of camel body dimensions pose significant challenges for this technology in camel measurement. Wang Yusha [20] and others developed a computer vision-based device for measuring the body dimensions and body mass traits of large yellow croaker, performing well in measuring organisms with smooth body surfaces. However, the same measurement point positioning strategy may not be suitable for camels, which have more structurally complex body surfaces. Munir Ahmad et al. [21] developed a deep transfer learning-based animal face identification model, demonstrating the significant potential of deep learning in the field of biometric identification, particularly for recognizing complex body structures. However, the study primarily focuses on facial recognition and does not fully explore the application of this technology to the measurement of more complex body parts, such as the humps of camels. Additionally, the model's scalability to handle the complex morphological structures of different animal species still requires further validation. Dhivya Mohanavel and Muthu Ishwarya [22] developed a deep learning and computer vision-based animal detection warning system for agricultural environments, highlighting the growing importance of these technologies in agriculture. However, the limitation of this study lies in its primary focus on animal detection and warning functions, without addressing the collection and analysis of specific biometric data. While it provides valuable insights into the application of deep learning in animal management within agricultural settings, the potential for precise animal body measurement and biometric analysis has not been fully explored.

III. Research Method

The Camel-YOLOv8 model primarily consists of three components: the Backbone, the Neck, and the Head networks.

Initially, the input image is processed through multiple convolutional layers (Conv) for preliminary feature extraction. This is followed by several c2f modules that refine the features further. During this process, the SK Attention module enhances the attention mechanism in feature extraction, enabling the model to focus more on important feature regions. Subsequently, features undergo multiscale processing via the SPPF (Spatial Pyramid Pooling - Fast) module. The Neck network then integrates multiple layers of features using convolutional layers and the AFPN-beta (Asymptotic Feature Pyramid Network - beta) module, with further processing by c2f modules. In the Head network, a series of convolutional layers transform the feature map into the format required by the output layer, calculating both the Bounding Box Loss (BBox Loss) and the Classification Loss (Cls Loss). Ultimately, the model outputs the detection results in the image, including bounding boxes and class labels. Overall, the YOLOv8 model, through its multi-layer feature extraction and integration mechanisms, effectively enhances the accuracy and efficiency of object detection while maintaining computational efficiency. The proposed model structure is illustrated in Fig. 1.

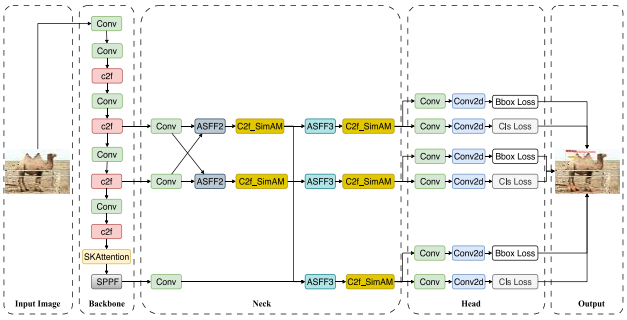


Fig. 1. Structure diagram of camel-YOLOv8 model.

A. Feature Selection

Many research findings indicate that chest girth, body length, pelvic width, and shoulder height are the most suitable and reliable parameters for estimating the live weight of animals. Fig. 2 shows diagram of camel body measurement points.

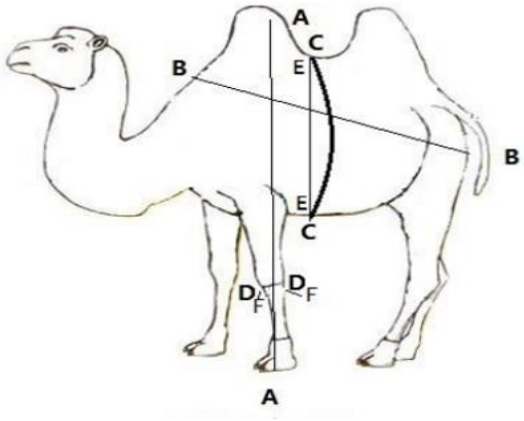


Fig. 2. Diagram of camel body measurement points.

In recent studies, in some cases, features from digital images have been used to estimate the body size of livestock [23]. The camel body measurement points are shown in the diagram. For the Xinjiang Bactrian camel, the height (A) refers to the vertical distance from the base of the camel's front hump to the ground, the body length (B) refers to the distance from the shoulder end to the hip end, the chest girth (C) is the vertical circumference measured from the base of the front hump down through the center of the horny pad at the chest bottom around the body, the cannon circumference (D) refers to the horizontal circumference measured around the cannon at the upper third of the left forelimb, chest diameter (E), and cannon diameter (F).

Using SPSS software, this paper analyzes the correlation between camel weight and other body metrics from existing camel data in Fuyun Town, Fukang City, Xinjiang Uighur Autonomous Region. We have chosen to use the Pearson correlation coefficient [24] to measure the linear relationship between them. The Pearson correlation coefficient is a statistic used to measure the degree of linear correlation between two variables. It is employed to assess the linear relationship between two variables, with values ranging from -1 to +1. A correlation coefficient of 1 indicates a perfect positive correlation; a coefficient of -1 indicates a perfect negative correlation; and a coefficient of 0 indicates no linear correlation between the variables. The formula for calculating the Pearson correlation coefficient is as follows:

$$r = \frac{\mathrm{cov}(X,Y)}{\sigma_X \sigma_Y} \quad (1)$$

Here, Cov represents covariance, and σ represents standard deviation. Specifically, the Pearson correlation coefficient is calculated as the covariance of the two variables divided by the product of their respective standard deviations. Below are the correlation coefficients between camel characteristic information and body weight in the camel data from Ziniquanzi Town, Fukang City, Xinjiang Uighur Autonomous Region:

From the Table I, it can be observed that there is a positive correlation between body weight and body measurement indicators, with the height, chest girth, and chest diameter all showing correlations above 0.5. This indicates that in this dataset, these features may have a strong linear relationship with body weight and can be used to predict it. Therefore, by improving the YOLOv8 algorithm for classifying the body parts of Xinjiang Bactrian camels, the indicators of body length and chest diameter are selected as the main subjects of study in this paper.

TABLE I. PEARSON CORRELATION COEFFICIENTS BETWEEN CAMEL BODY MEASUREMENT FEATURES AND BODY WEIGHT

Feature	Correlation Coefficient
Body Height	0.538
Body Length	0.138
Body Diagonal Length	0.252
Chest Girth	0.793
Chest Diameter	0.594
Teat Diameter	0.256
Teat Length	0.255

B. SKAttention

Different sizes of receptive fields have varying effects on targets of different scales. In the process of classifying different parts of a camel, such as the camel's hump, body, and foot, the appropriate receptive field varies due to the varying sizes of these parts. To address the issue of the camel's body being obscured, we propose the SKAttention mechanism [25], which allows the network to automatically utilize information captured by effective receptive fields for classification. The structure of the Selective Kernel Attention is shown in Fig. 3. And consists of three parts: Split, Fuse, and Select. We input a feature map with dimensions C×H×W, and in the Split part, the input image undergoes convolution operations with 3×3 and 5×5 kernels, resulting in two feature maps, U_1 and U_2. Fuse involves calculating the weights of the two convolution kernels, summing the feature maps element-wise, and then averaging along the H and W dimensions to obtain a one-dimensional vector of size C×1×1. This weight information represents the importance of each channel's information. The formula is as follows:

$$U = U_1 + U_2 \tag{2}$$

In this process, U is generated through global average pooling (AvgPool) to provide channel statistics. A linear transformation is then applied to map the original C-dimensional information to Z-dimensional information. Following this, two linear transformations are used to convert from Z dimensions back to the original C dimensions, thereby completing the extraction of channel dimensions. In the Select part, a softmax operation is performed on the channel dimension to merge the soft attention vectors of each branch, allowing for the selection of multiple branches with different kernel sizes. An SKNet composed of multiple SK units can capture objects of various scales through its neurons.

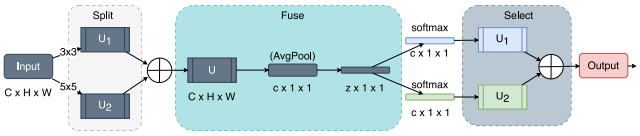


Fig. 3. Selective kernel attention structure diagram.

In the backbone part of YOLOv8, we set the input feature channel number (Channel parameter) to 512, the size list of convolution kernels (kernels parameter) to [1, 3, 5, 7], the dimension reduction ratio (reduction parameter) to 16, the number of convolution groups (group parameter) to 1, and the dimension (L parameter) to 32. Notably, in this process, adopting an adaptive receptive field size allows the model to better adapt to targets of varying scales, thereby enhancing accuracy and efficiency.

C. Asymptotic Feature Pyramid Network AFPN-beta Structure

Inspired by the Asymptotic Feature Pyramid Network (AFPN) [26], we propose a target detection method, AFPN-beta, for detecting various body parts of the Xinjiang Bactrian camel. This method helps to address the issues of target box loss and occlusion among camels, thereby enhancing detection accuracy. The AFPN-beta mainly consists of two key components: the Feature Pyramid Network and the Adaptive Spatial Fusion operation.

The Feature Pyramid Network allows for direct interaction between different layers, preventing the loss of feature information. By extracting features at various scales and integrating them, the model can better capture target features across different scales and semantic levels. This direct interaction helps to improve detection accuracy, especially for targets with complex shapes and varying sizes. The Adaptive Spatial Fusion operation is another crucial component within AFPN-beta. It helps to resolve conflicts of information during the feature fusion process. By dynamically adjusting the fusion weights according to the spatial distribution of features at different layers, the Adaptive Spatial Fusion operation effectively merges features from different levels, reduces information conflicts, and enhances the accuracy of target detection. The specific structure of the AFPN-beta module is illustrated in Fig. 4 [24].

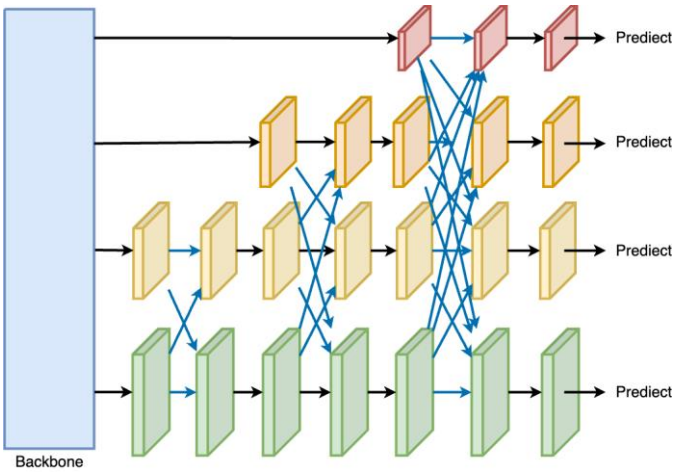


Fig. 4. AFPN-beta module structure diagram.

D. Camel Body Measurement Calculation Algorithm

During the data collection process for Xinjiang Bactrian camels, a set of calibration images using Zhang Zhengyou's calibration board [27] is essential. These images should include a calibration board whose corners have precisely known three-dimensional coordinates. Throughout this process, computer vision techniques such as Harris or Shi-Tomasi corner detection algorithms are utilized to accurately detect these corners on the calibration board. By aligning these detected corners with their established three-dimensional points and employing the camera's projection model, we can accurately estimate the camera's internal parameters—such as focal length and principal point—as well as external parameters, including rotation and translation vectors. Utilizing Zhang Zhengyou's calibration method ensures the precision of the measurement setup, which is crucial for accurate data collection. The calibration process and its results are depicted in Fig. 5, showing the setup used specifically in the data collection of Xinjiang Bactrian camels.

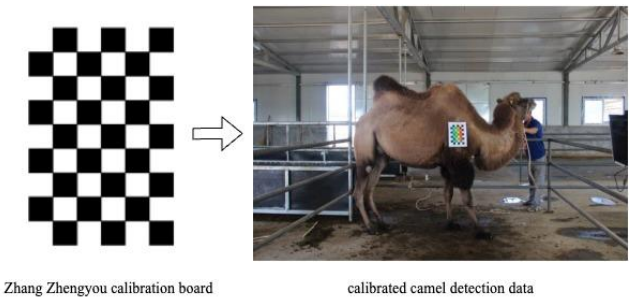


Fig. 5. Apply the Zhang Zhengyou calibration board on the left to the camel data collection on the right.

After the classification is complete, in order to facilitate the extension of the calculation method to three-dimensional space, the Minkowski distance is chosen to calculate the pixel information of various camel parts. The Minkowski distance [28] is a method for measuring the distance between two points in multidimensional space. It is a very common method for measuring the distance between numerical points, with the coordinates of points P and Q assumed to be as follows:

$$P = (x_1, x_2, \ldots, x_n) \text{and } Q = (y_1, y_2, \ldots, y_n) \in R^n \quad (3)$$

Thus, the Minkowski distance is defined as:

$$\left(\sum_{i=1}^{n} |x_i - y_i|^p\right)^{\frac{1}{p}} \quad (4)$$

Once the pixel information is calculated, we can use the similar triangles formula to determine the distance to the object. This process typically requires camera calibration information, including focal length and pixel size parameters. This can be achieved through the internal and external parameters obtained from camera calibration. Below is a basic formula for this conversion:

$$D_real = \frac{P \times S}{f} \quad (5)$$

Where:

D_real is the actual distance.

f is the focal length of the camera, P= camera_matrix[1,1].

S is the actual distance between the object and the camera.

P is the width of the object in the image.

After completing the calculation of camel body measurements, the mean relative error (MRE) [29] is used to represent the degree of deviation of the measurement error relative to the actual measurement values. The calculation method is shown in the following formula.

$$MRE = \frac{\sum \frac{\text{Actual Value} - \text{Predicted Value}}{\text{Actual Value}}}{n} \quad (6)$$

where, n is the number of samples, Σ represents the sum over all samples, and |x| denotes the absolute value of x.

IV. EXPERIMENT

A. Dataset Preparation

Images of Xinjiang Bactrian camels were taken in Ziniquanzi Town, Fukang City, Xinjiang Uygur Autonomous Region, using a Canon 60D camera and an 18-135 lens. Images were collected in December 2020, March to April 2021, and July 2021, capturing 146 Bactrian camels under different lighting conditions, angles, distances, and various occlusions, resulting in 389 valid images. Due to the small dataset size, additional data was collected in September 2023 in Keping County, Aksu Prefecture, Xinjiang Uygur Autonomous Region, amounting to approximately 400 valid images.

A total of 789 images were used to create the dataset for calculating the body measurements of Xinjiang Bactrian camels. The dataset was annotated using the LabelImg software [30].

1) Image selection: Poor quality images were discarded. According to the requirements of the network model, images where key parts of the camel's body were obscured were excluded. Such images were discarded due to overlapping or large occlusions caused by the presence of too many camels in the same frame, which severely affected the training quality of the model. This resulted in 789 valid images.

2) Image annotation: The open-source software LabelImg was used for manual annotation of the targets. SPSS software analysis revealed that four indicators—body size, chest diameter, chest girth, and pipe diameter—are highly correlated with the weight of Bactrian camels. Therefore, the annotated categories include "Camel," "Hump," "Body," "Bottle," and "Foot." The annotations were saved in .txt files with the same names as the images.

3) Data augmentation [31]: The statistical distribution of the sample numbers in the training set showed an imbalance in the number of images of different Xinjiang Bactrian camels. To enhance the model's robustness and adapt to different weather conditions farmers might encounter during camel body detection, techniques such as weather changes, noise addition, and occlusion were randomly applied to the training set images for augmentation.

4) Dataset formatting: The dataset was randomly sampled from images collected at different times and divided into training, validation, and test sets. The training and validation sets together comprise 80% of the dataset, with a 9:1 ratio, while the test set comprises 20%. To verify the model's generalization performance, the collection times of the images in the test set were as different as possible from those in the training set.

B. Training Parameter Settings

The experiments were conducted using an NVIDIA GeForce RTX 3060 GPU, with the operating system being Ubuntu 16.04. The model was built using the Pytorch deep learning framework, with Python version 3.8 and CUDA version 10.2. The improved YOLOv8 model used input images of size 640×640, with an epoch set to 400 and a batch size of 32.

C. Model Evaluation Metrics

The performance of the model was evaluated using precision (P), recall (R), mean average precision (mAP), frame

per second (FPS), and memory usage. FPS measures the number of image frames detected per second.

D. Results and Analysis

1) Analysis of body part classification results of xinjiang bactrian camels: In the Xinjiang Bactrian camel dataset, the original training set contained 721 images. Through the application of image augmentation techniques such as weather effect simulation, color jittering, noise addition, and blurring, the training set was expanded to 2518 images for model training. The dataset also included a validation set of 629 images and a test set of 99 images. The model was trained over 400 epochs, and the results of camel body part recognition are shown in Table II.

TABLE II. RESULTS OF COMPARATIVE EXPERIMENTS ON CAMEL PART RECOGNITION

	Epoch	*loss*	*P*	*R*	*mAP 50*	*mAP 50-95*	*FPS*
YOLOv3-tiny	400	0.622	0.745	0.95	0.854	0.69	303.0
YOLOv5n	400	0.648	0.687	0.934	0.804	0.677	285.7
YOLOv5m	400	0.485	0.690	0.943	0.831	0.741	149.2
YOLOv5s	400	0.696	0.730	0.952	0.87	0.717	357.1
YOLOv6n	400	0.569	0.691	0.948	0.833	0.726	357.1
YOLOv6s	400	0.628	0.686	0.943	0.833	0.714	333.3
YOLOv8n	400	0.619	0.691	0.959	0.836	0.719	344.8
YOLOv8s	400	0.611	0.735	0.962	0.879	0.738	303.0
YOLOv10s	400	0.533	0.708	0.923	0.836	0.729	204.1
Camel-YOLOv8	400	0.588	0.745	0.964	0.888	0.764	303.0

From the table above, it is evident that the improved YOLOv8 model performs better in measuring the body dimensions of Xinjiang Bactrian camels compared to YOLOv3, YOLOv5, YOLOv6, YOLOv8, and YOLOv10. The improved YOLOv8 model achieves a precision rate of 74.5%, a recall rate of 96.4%, a mAP50 of 88.88%, a mAP50-90 of 76.4%, and an FPS of 303.03 frames per second. Although the FPS has decreased slightly and the model's memory usage has slightly increased compared to the original YOLOv8 model, other metrics have improved, with the mAP50-95 increasing by 3.7 percentage points. Compared to YOLOv3-tiny, YOLOv5n, YOLOv5s, YOLOv6n, YOLOv6s, YOLOv8n, and YOLOv8s models, the improved model shows an increase in detection accuracy, with the mAP50-95 improving by 7.4, 8.7, 4.7, 3.8, 5.0, 4.5 and 2.6 percentage points, respectively. Additionally, compared to the YOLOv5m and YOLOv10s models, although the bounding box errors have significantly decreased, there has also been a substantial increase in FPS, resulting in a decrease in detection speed.

Therefore, this study successfully enhances the accuracy of body measurement for Xinjiang Bactrian camels while ensuring detection speed, validating the effectiveness of the proposed method. After training for 400 epochs, the training results of the Camel-YOLOv8 model are shown in Fig. 6.

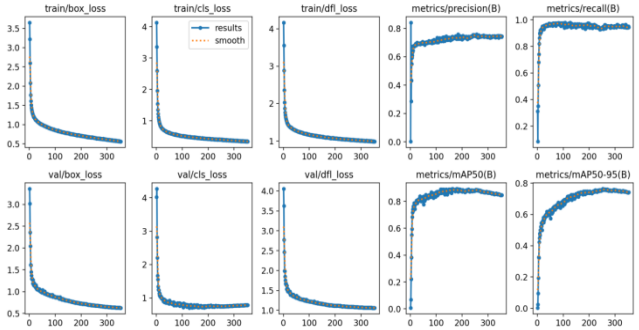


Fig. 6. Training results of the camel-YOLO v8 model.

2) Ablation study: To evaluate the enhancement effects of the proposed improved modules, we designed four ablation experiments. These experiments include the standard YOLOv8 model, YOLOv8 with the SKAttention mechanism, YOLOv8 with the AFPN-beta structure, and a combination of these improvements. Each experimental setup was trained and tested on the same dataset, and consistent evaluation metrics were used to measure changes in model performance. The purpose of these ablation studies is to isolate and quantify the impact of each individual module on the overall performance of the YOLOv8 model. These experiments provide a comprehensive understanding of how each modification contributes to the model's efficacy. The experimental results are presented in Table III.

3) Analysis of Bactrian camel body measurements: The pixel length of the camel's body is determined from the detection bounding box results, and then the actual length of the camel's body is calculated using Formula (6). Partial results of the estimated camel body lengths are shown in Table IV.

Table V presents the error results of this experiment, evaluated using the mean relative error as the metric. The findings indicate that the proposed object detection model demonstrates a high level of accuracy in estimating the pixel-based body measurements of camels.

TABLE III. RESULTS OF ABLATION EXPERIMENTS ON CAMEL PART RECOGNITION

YOLOv8	*SKAttention*	*AFPN-beta*	*Camel-YOLOv8*	*loss*	*P*	*R*	*mAP50*	*mAP50-95*	*FPS*
✓	-	-	-	0.5962	0.720	0.955	0.867	0.734	333.3
✓	✓	-	-	0.5866	0.735	0.952	0.886	0.758	312.5
✓	-	✓	-	0.6111	0.735	0.962	0.879	0.738	303.0
✓	✓	✓	✓	0.5885	0.745	0.964	0.888	0.764	303.0

TABLE IV. PARTIAL RESULTS OF CAMEL BODY LENGTH ESTIMATION

Image Number	*Estimated Camel Body Length (mm)*	*Actual Camel Body Length (mm)*	*Error*	*Estimated Camel Chest Diameter (mm)*	*Actual Camel Chest Diameter (mm)*	*Error*
1	16537	17500	-0.0550	8584.22	9000	-0.0462
2	15817	17100	-0.0750	9890.22	9000	0.0989
3	15045	17000	-0.1150	9849.13	10000	-0.0151
4	16012	17500	-0.0850	11089.13	10000	0.1089
5	15214	16100	-0.0550	9159.29	9300	-0.0151
6	15910	17200	-0.0750	9044.22	9000	0.0049
7	20362	18100	0.1250	9890.22	9000	0.0989
8	18879	17400	0.0850	10837.36	9600	0.1289
9	15970	16900	-0.0550	9847.75	9800	0.0049

TABLE V. RESULTS OF ERRORS IN CAMEL BODY DIMENSION INFORMATION

	MRE
Camel Body Height	-3.39%
Camel Chest Diameter	4.1%

V. CONCLUSION AND FUTURE WORK

A. Conclusion

This research significantly enhances the model's ability to recognize the features of different parts of Bactrian camels, raising the precision, recall, and mean average precision to 74.5%, 96.4%, and 88.8%, respectively. These performance metrics surpass those of the existing YOLOv3, YOLOv5, YOLOv6, YOLOv10, and the original YOLOv8 models. Moreover, the improved model detects at a speed of 303.03 frames per second (fps), demonstrating good real-time performance, and occupies only 14.45MB of memory with fewer parameters, facilitating quick deployment and portability across various devices. Compared to the actual size dimensions of camels, the model's computational results are highly accurate, with an average relative error of -3.39% for height and 4.1% for girth.

In the current context of scaled and refined breeding environments, real-time detection, monitoring, and management of camel populations are becoming increasingly important [32]. Therefore, the outcomes of this study hold significant appeal for camel breeders. By utilizing this model, breeders can more rapidly and precisely assess the growth conditions of camels and calculate their body measurements and weight [33], which is crucial for optimizing selection, improving breeding strategies, and enhancing economic benefits. Additionally, the low error in calculating the body size of Xinjiang Bactrian camels can aid breeders in promptly identifying health issues [34], thereby enabling early intervention measures to reduce the impact of diseases and other health risks on camel breeding.

B. Future Work

Although this study has made significant progress in calculating the body dimensions of camels, there are still some limitations:

1) The model primarily focuses on algorithm optimization for Xinjiang Bactrian camels, and its applicability to other camel breeds or camels from different geographical areas requires further validation.

2) While some technical issues have been addressed through algorithm improvements, the results have not met expectations, particularly in handling occlusions and feature extraction. Therefore, future efforts could explore more effective improvement methods to further enhance the model's accuracy and reliability [35].

Looking ahead, applying this method to other fields will present certain technical challenges. For example, when applied to other animal species, the model may need significant adjustments to accommodate different body features and behavioral patterns. The universality of the model in detecting other livestock or animal body types still needs to be validated and optimized through more experiments. Future research directions include exploring more advanced attention mechanisms and feature fusion techniques to further improve the model's performance in various application scenarios [36][37]. These efforts will help enhance the model's practicality and provide more comprehensive and reliable technical support for intelligent breeding [38].

ACKNOWLEDGMENT

This work was supported by the Natural Science Foundation of Xinjiang Uygur Autonomous Region (No. 2021D01C082).

REFERENCES

[1] Yao Huaibing, Li Na, Liang Xiaorui, Hao Zelin, Yan Hui, Zhao Zhongkai, Liu Ying, Ma Qiang, Wang Yuzhuo, Chen Gangliang, Yang Jie. "Determination and Correlation of Milk Production Traits with Body Size, Body Weight, and Blood Physicochemical Indicators in Camelus bactrianus in Xinjiang." *China Animal Husbandry & Veterinary Medicine*, 2023, Vol. 50, Issue (8): 3210-3220.

[2] N.S.N. Abd Aziz, S.M. Daud, R.A. Dziyauddin, M.Z. Adam and A. Azizan, "A review on computer vision technology for monitoring poultry Farm—Application, hardware, and software," IEEE access, vol. 9, 2020, pp. 12431-12445.

[3] Yuheng S, Hao Y. "Image segmentation algorithms overview." *arXiv preprint arXiv:1707.02051*, 2017.

[4] Md Golam Morshed, Tangina Sultana, Aftab Alam, and Young-Koo Lee. "Human Action Recognition: A Taxonomy-Based Survey, Updates, and Opportunities." *Sensors*, 2023, Vol. 23, Issue 4, 2182. doi: 10.3390/s23042182.

[5] Suresh Neethirajan. "Artificial Intelligence and Sensor Technologies in Dairy Livestock Export: Charting a Digital Transformation." *Sensors*, 2023, Vol. 23, Issue 16, 7045. doi: 10.3390/s23167045.

[6] Bishop, C. M. "Pattern Recognition and Machine Learning." *Springer*, 2006.

[7] L.O. Chua, "CNN: A vision of complexity," International Journal of Bifurcation and Chaos, vol. 7, no. 10, 1997, pp. 2219-2425.

[8] B. Koger, A. Deshpande, J.T. Kerby, J.M. Graving, B.R. Costelloe and I.D. Couzin, "Quantifying the movement, behaviour and environmental context of group-living animals using drones and computer vision," Journal of Animal Ecology, vol. 92, no. 7, 2023, pp. 1357-1371.

[9] M. Madi, Y. Basha, Y. Albadersawi, F. Alenezi, S.A. Mahmoud, et al., “Camel detection and monitoring using image processing and IoT,” Proc. 2023 15th International Conference on Developments in eSystems Engineering (DeSE), IEEE, 2023, pp. 305-308.

[10] M. Hussain, “YOLO-v1 to YOLO-v8, the rise of YOLO and its complementary nature toward digital manufacturing and industrial defect detection,” Machines, vol. 11, no. 7, 2023, pp. 677.

[11] F. Pallottino, R. Steri, P. Menesatti, F. Antonucci, C. Costa, S. Figorilli and G. Catillo, “Comparison between manual and stereovision body traits measurements of Lipizzan horses,” Computers and electronics in agriculture, vol. 118, 2015, pp. 408-413.

[12] Q. Qin, D. Dai, C. Zhang, C. Zhao, Z. Liu, X. Xu, M. Lan, Z. Wang, Y. Zhang and R. Su, “Identification of body size characteristic points based on the Mask R-CNN and correlation with body weight in Ujumqin sheep,” Frontiers in Veterinary Science, vol. 9, 2022, pp. 995724.

[13] K. He, G. Gkioxari, P. Dollár and R. Girshick, “Mask r-cnn,” Proc. Proceedings of the IEEE international conference on computer vision, 2017, pp. 2961-2969.

[14] M. Khojastehkey, M. Yeganehparast, A. Jafari Arvari, N. Asadzadeh and M. Khaki, “Biometric measurement of one-humped camels using machine vision technology,” Journal of Ruminant Research, vol. 7, no. 1, 2019, pp. 19-32.

[15] Zou, Z., Shi, Z., Guo, Y., & Ye, J. "Object Detection in 20 Years: A Survey." *arXiv preprint arXiv:1807.05511*, 2019.

[16] Zhao, Zhong-Qiu, Peng Zheng, Shou-Tao Xu, and Xindong Wu. "Object detection with deep learning: A review." *IEEE Transactions on Neural Networks and Learning Systems*, 2019, Vol. 30, No. 11, pp. 3212-3232.

[17] K. AlNujaidi, G. Alhabib and A. AlOdhieb, “Spot-the-camel: Computer vision for safer roads,” arXiv preprint arXiv:2304.00757, 2023.

[18] K. Wang, H. Guo, Q. Ma, W. Su, L. Chen and D. Zhu, “A portable and automatic Xtion-based measurement system for pig body size,” Computers and electronics in agriculture, vol. 148, 2018, pp. 291-298.

[19] 1. K. Li and G. Teng, “Study on body size measurement method of goat and cattle under different background based on deep learning,” Electronics, vol. 11, no. 7, 2022, pp. 993.

[20] Wang Yusha, Wang Jiaying, Xin Rui, Ke Qiaozhen, Jiang Pengxin, Zhou Tao, Xu Peng. "Application of computer vision in morphological and body weight measurements of large yellow croaker (Larimichthys crocea)." Journal of Fishery Sciences of China, 2023.

[21] M. Ahmad, S. Abbas, A. Fatima, G. F. Issa, T. M. Ghazal, and M. A. Khan, “Deep transfer learning-based animal face identification model empowered with vision-based hybrid approach,” Appl. Sci., vol. 13, no. 2, p. 1178, 2023, DOI: 10.3390/app13021178.

[22] D. Mohanavel and A. M. Ishwarya, “Deep learning and computer vision based warning system for animal disruption in farming environments,” IEEE, 2024, DOI: 10.1109/AIIoT58432.2024.10574724.

[23] Chu Mengyuan, Si Yongsheng, Li Qian, Liu Gang. "Research progress on automatic measurement technology of livestock body size." Transactions of the Chinese Society of Agricultural Engineering, 2022, Vol. 38, No. 13.

[24] Benesty, J., Chen, J., Huang, Y., & Cohen, I. "Pearson Correlation Coefficient." *Noise Reduction in Speech Processing*, Springer, 2009, pp. 1-4.

[25] X. Li, W. Wang, X. Hu and J. Yang, “Selective kernel networks,” Proc. Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2019, pp. 510-519.

[26] G. Yang, J. Lei, Z. Zhu, S. Cheng, Z. Feng and R. Liang, “AFPN: Asymptotic feature pyramid network for object detection,” Proc. 2023 IEEE International Conference on Systems, Man, and Cybernetics (SMC), IEEE, 2023, pp. 2184-2189.

[27] Z. Zhang, “A flexible new technique for camera calibration,” IEEE Transactions on pattern analysis and machine intelligence, vol. 22, no. 11, 2000, pp. 1330-1334.

[28] A.A. Thant, S.M. Aye and M. Mandalay, “Euclidean, manhattan and minkowski distance methods for clustering algorithms,” International Journal of Scientific Research in Science, Engineering and Technology, vol. 7, no. 3, 2020, pp. 553-559.

[29] Liu Tong, Teng Guojun, Fu Weidong, Li Zhi. "Extraction algorithms and applications of pig body size measurement points based on computer vision." *Transactions of the Chinese Society of Agricultural Engineering*, 2013, Vol. 29, No. 2, pp. 161-168.

[30] Ali Alameer. "Facial Emotion Recognition Datasets for YOLOv8 Annotation." *University of Salford*, 2023.

[31] Mingle Xu, Sook Yoon, Alvaro Fuentes, and Dong Sun Park. A comprehensive survey of image augmentation techniques for deep learning. Pattern Recognition, page 109347, 2023.

[32] A. A. Samsu Aliar, J. Yesudhasan, M. Alagarsamy, K. Anbalagan, J. Sakkarai, and K. Suriyan, “A comprehensive analysis on IoT based smart farming solutions using machine learning algorithms,” Bull. Electr. Eng. Informatics, vol. 11, no. 3.

[33] C. Tırınk, E. Eyduran, A. Faraz, A. Waheed, N. A. Tauqir, M. Nabeel, M. Tariq, and I. Sheikh, “Use of multivariate adaptive regression splines for prediction of body weight from body measurements in Marecha (Camelus dromedaries) camels in Pakistan,” Trop. Anim. Health Prod., 2021, DOI: 10.1007/s11250-021-02788-y.

[34] M. R. Islam, M. M. Kabir, M. F. Mridha, S. Alfarhood, M. Safran, and D. Che, “Deep learning-based IoT system for remote monitoring and early detection of health issues in real-time,” Sensors, vol. 23, no. 11, p. 5204, 2023.

[35] F. M. J. M. Shamrat, A. Hossain, and T. Roy, “IoT based smart automated agriculture and real time monitoring system,” IEEE.

[36] Y. Hu, X. Deng, Y. Lan, X. Chen, Y. Long, and C. Liu, “Detection of rice pests based on self-attention mechanism and multi-scale feature fusion,” Insects, vol. 14, no. 3, p. 280, 2023.

[37] G. Koc, C. Koc, H. E. Polat, et al., “Artificial intelligence-based camel face identification system for sustainable livestock farming,” Neural Comput. Appl., vol. 36, no. 6, pp. 3107–3124, 2024.

[38] S. Neethirajan, “Recent advances in wearable sensors for animal health management,” Sens. Bio-Sens. Res., vol. 12, pp. 15–29, 2017.

Towards Secure Internet of Things-Enabled Healthcare: Integrating Elliptic Curve Digital Signatures and Rivest Cipher Encryption

Longyang Du*, Tian Xie
School of Artificial Intelligence, Jiaozuo University, Jiaozuo 454000, Henan, China

***Abstract*—The expansion of Internet of Things (IoT) applications, such as wireless sensor networks, intelligent devices, Internet technologies, and machine-to-machine interaction, has changed current information technology in recent decades. The IoT enables the exchange of information and communication between items via an internal network. Nevertheless, the advancement of technology raises the urgent issue of ensuring data privacy and security, particularly in critical sectors like healthcare. This study aims to address the problem by developing a hybrid security scheme that combines the Secure Hash Algorithm (SHA-256), Rivest Cipher 4 (RC4), and Elliptic Curve Digital Signature Scheme (ECDSS) to ensure the confidentiality and integrity of medical data transmitted by IoT-enabled healthcare systems. This hybrid model employs the Elliptic Curve Digital Signature Scheme (ECDSS) to perform exclusive OR (XOR) operations inside the RC4 encryption algorithm. This enhances the RC4 encryption process by manipulating the encryption key. Moreover, SHA-256 is used to convert incoming data in order to guarantee data security. An empirical investigation validates the superiority of the suggested model. This framework attains a data transfer rate of 11.67 megabytes per millisecond, accompanied by an encryption duration of 846 milliseconds and a decryption duration of 627 milliseconds.**

***Keywords*—IoT-enabled healthcare; data privacy; security; hybrid security framework; SHA-256; RC4; encryption; data integrity**

I. Introduction

A. Background

The conventional healthcare system faces challenges in meeting the demands of an extensive population due to its constraints in terms of cost and accessibility [1, 2]. The emerging concept of smart healthcare empowers individuals regarding their health conditions, enabling them to manage certain medical situations and enhance the quality of care [3]. This technology permits remote patient monitoring, reducing healthcare expenses and allowing medical practitioners to extend their services across geographical boundaries [4]. An operationally intelligent healthcare system corresponds with the development of innovative urban areas, offering inhabitants a better lifestyle [5]. The National Sanitation Foundation (NSF) examined how nanotechnology and Information and Communication Technology (ICT) could enhance human well-being. This convergence enables the interconnection of items via nanotechnology, embedded systems, sensors, and wireless networks, giving each Internet-connected object its own unique identity [6]. The Internet of Things (IoT) covers a broad spectrum of technologies providing connectivity between different objects, and it has been successfully applied in several industries, particularly healthcare facilities [7]. IoT-enabled healthcare is a sophisticated process involving computer science, medical technology, medicine, microelectronics, and other related fields [8].

B. Challenges

According to current projections, the industrial IoT domain is anticipated to have a market worth $110.6 billion by 2025 after a substantial growth trend in recent times [9]. Projections suggest that by 2030, the quantity of IoT objects and devices in operation will exceed 50 billion, forming an extensive interconnected system that encompasses smartphones and household appliances [10]. The widespread adoption of IoT, combined with the declining costs of electronic devices and networking, has significantly facilitated the proliferation of its use in the healthcare industry. The introduction of IoT in the healthcare field holds immense potential. The utilization of IoT for remote health monitoring is expected to have a substantial influence on both healthcare establishments and people's homes [11]. The technology offers considerable potential to augment healthcare quality and save costs by enabling early identification and avoidance of illnesses and other hazardous situations [12]. The potential uses of this technology include managing long-term medical conditions, providing elderly care, facilitating physical fitness endeavors, and several other domains. Using technology to enable remote patient monitoring can significantly decrease hospitalization expenses by communicating up-to-date health information to healthcare professionals and promptly identifying and managing health conditions [13].

The IoT combines health sensors, imaging, and diagnostic devices to provide healthcare services that improve efficiency and prolong patient lives. The IoT allows healthcare professionals to remotely manage equipment, effectively allocate resources, and support cost-efficient interaction through safe, real-time communication among medical facilities and patients [14]. This helps in minimizing equipment downtime and improving overall efficiency in healthcare operations. Moreover, healthcare networks enabled by the IoT are positioned to assist in timely identifying diseases, managing long-term medical conditions, handling medical crises, and providing healthcare services as needed, aided by database systems, entry points, and medical servers [15].

Recent advancements in machine learning, such as the multi-expert large language model architecture for Verilog code

generation, have shown promise in automating complex design tasks, potentially offering new avenues for secure and efficient hardware design in IoT applications [16]. The automatic synthesis of models from communication traces in System-on-Chip (SoC) designs has been shown to streamline the development of secure and efficient systems, which is critical for managing the complex data flows in IoT-enabled healthcare environments [17].

C. Problem Statement

Security is a significant concern in large-scale network configurations, particularly in healthcare IoT deployments that handle sensitive patient information. The wireless nature of most devices and their communications in IoT-enabled healthcare systems raises considerable privacy and security concerns [18, 19]. For example, under a Medical Sensor Network (MSN), an IP-enabled sensor might send health information to distant healthcare services. Nevertheless, patient privacy becomes a significant concern when medical data is transmitted through potentially untrustworthy network infrastructures, such as the Internet [20]. Therefore, it is imperative to guarantee the confidentiality and security of health information in healthcare IoT applications. Ensuring the identity verification and permission granting of distant healthcare facilities or caregivers and safeguarding data during its transmission are crucial requirements in healthcare IoT to avoid unauthorized access to confidential medical information or intentional disruption of particular operations. Given individuals' ongoing engagement with these applications, ensuring secure and reliable data communication across healthcare sensors, patients, caregivers, and actuators is vital. Concerns about misuse or privacy issues might deter people from embracing IoT-based healthcare applications.

D. Contribution

This paper introduces a hybrid model that combines the Secure Hash Algorithm (SHA-256), Rivest Cipher 4 (RC4), and Elliptic Curve Digital Signature Scheme (ECDSS) to enhance the security of patient data collected through IoT and various health devices. SHA-256+RC4+ECDSS employs RC4 encryption using ECDSS and introduces shift-right operation to improve the Pseudo-Random Number Generation Mechanism (PRGM) stage. Through the use of shift-right, the key is modified, and subsequently, the plaintext is encrypted. ECDSS encrypts the resulting ciphertext, which is further processed by SHA-256. This combination of algorithms prevents third-party access to data and ensures complete privacy assurance. The following questions guide the study:

- How can a hybrid security framework improve the confidentiality and integrity of data in IoT-enabled healthcare systems?
- What are the computational trade-offs involved in integrating SHA-256, RC4, and ECDSS in a single security framework?
- How does the proposed framework perform in comparison to existing security solutions?

The paper is formatted in the following fashion. Section II comprehensively analyzes previous research on the same topic. Section III outlines the suggested approach. Section IV presents a comprehensive overview of the simulation parameters and presents the findings. Section V discusses the obtained results in detail. Finally, Section VI concludes and discusses potential future enhancements for this research.

II. Related Work

Research efforts in recent years have shown significant progress in securing IoT-enabled healthcare systems. Security, privacy, and efficient data management have been addressed in numerous studies. Table I compares the methodologies, key contributions, and performance metrics of several research studies on data encryption, attack detection, blockchain integration, energy efficiency, anomaly identification, and medical image security.

Al Shahrani, et al. [21] proposed an efficient hashing technique that utilizes digital certificates to boost safety measures. At first, medical data is collected and filtered through normalization before being saved on the IoT device. In this process, digital certificates play a role in authentication. Their approach, known as the Discrete Decision Tree Hashing Algorithm (DDTHA), incorporates the Ant Colony Optimization (ACO) algorithm to hash the unsigned certificates. Encryption is carried out using the Blowfish algorithm, resulting in signed digital certificates for authentication. The proposed method underwent analysis and comparative evaluation against existing approaches, demonstrating superior performance for crucial factors such as energy consumption, avalanche effect, execution time, decryption time, and encryption time compared to other existing methods.

Aruna Santhi and Vijaya Saradhi [22] proposed a method to identify attacks on healthcare IoT devices by using an improved deep learning framework to support the Bring Your Own Device (BYOD) concept. Their approach involves modeling a simulated hospital environment where numerous IoT devices and medical equipment communicate. The datasets on malware assessment in medical IoT devices are collected from every node and treated as features. The processing of these characteristics is carried out using a Deep Belief Network (DBN), which is a constituent of the deep learning algorithm. To optimize the DBN's performance, they fine-tune the number of hidden neurons by leveraging a hybrid meta-heuristic algorithm, a combination of the Spider Monkey Optimization (SMO) and Grasshopper Optimization Algorithm (GOA), called Local Leader Phase-based GOA (LLP-GOA). The DBN trains the nodes by constructing an extensive data store, including attack specifics, allowing precise detection throughout testing. The analysis shows that the suggested LLP-GOA-based DBN model achieved a higher accuracy of 0.25% compared to Particle Swarm Optimization (PSO)-DBN, 0.15% compared to Grey Wolf Algorithm (GWO)-DBN, 0.26% compared to SMO-DBN, and 0.43% compared to GOA-DBN. In addition, the LLP-GOA-DBN model demonstrated a 13% improvement in accuracy compared to the Support Vector Machine (SVM), a 5.4% improvement over the K-Nearest Neighbor (KNN), an 8.7% improvement over the Neural Network (NN), and a 3.5% improvement over a regular DBN.

TABLE I. IOT-ENABLED HEALTHCARE SYSTEMS

Reference	Methodology/Approach	Key contributions	Comparative performance/results
[21]	Optimized hashing algorithm using digital certificates for security. Encryption via the blowfish algorithm.	Using the discrete decision tree hashing algorithm (DDTHA) with ant colony optimization (ACO) for unsigned digital certificates' hashing.	Superior performance compared to existing methods in encryption/decryption time, execution, avalanche effect, and energy consumption.
[22]	Attack detection in medical IoT devices using a deep learning architecture enhanced with a hybrid meta-heuristic algorithm (LLP-GOA).	Utilization of Deep Belief Network (DBN) with LLP-GOA for attack detection.	Improved accuracy (0.25% - 13%) compared to PSO-DBN, GWO-DBN, SMO-DBN, GOA-DBN, SVM, KNN, NN, and standard DBN.
[23]	Addressing security concerns in exchanging patients' records using blockchain and NuCypher threshold re-encryption mechanism.	Introduction of a secure architecture using blockchain for E-healthcare data security, redesigning medical WSN lifecycle, and employing NuCypher for data encryption. Implementation of customized lightweight blockchain PoW/PoS with digital signatures.	Improved security, reduced storage load, and improved transaction processing for E-healthcare compared to centralized systems.
[24]	Utilizing permissioned blockchain, MEC, and DRL-enabled IoT for secure and energy-efficient healthcare services.	Integration of permissioned blockchain and MEC to enhance security and energy efficiency. The application of DRL is to optimize system security and energy consumption.	Balanced security and energy efficiency to combat COVID-19-related challenges.
[25]	Introduction of PRISM, an edge-centric system for intelligent healthcare tools assessment using IoT trials.	A systematic approach for IoT-based trials in domestic healthcare settings. Achieved high precision in anomaly identification.	High precision in models trained on individual patients, decline in accuracy observed on diverse patients.
[26]	Exploration of medical image security within IoT-based healthcare systems using cryptography-based networks.	Use of ResNet-50 architecture for encryption and decryption of medical images. Implementing reconstructive network for decryption and Return on Investment (ROI) framework.	Strong security outcomes in medical image encryption/decryption and potential for precise therapy assessments.

Khan, et al. [23] have proposed a comprehensive strategy to address security concerns in exchanging patients' records across centralized server-based systems. Their solutions address node connectivity rates, parallel data-sharing failures, and delivery complications. The strategy comprises three main components. Initially, it presents a new and reliable framework for ensuring the security of electronic healthcare data by utilizing blockchain-distributed ledger architecture. Furthermore, it improves how medical Wireless Sensor Networks (WSNs) operate by implementing a distributed tiered structure, improving network capacity, and promoting confidence in the blockchain-enabled Peer-to-Peer (P2P) environment. Also, it utilizes the NuCypher threshold re-encryption process to encrypt data, guaranteeing the security of shared resources stored in blocks inside an immutable blockchain. The system utilizes chain codes to automate the processes of verification, logging, distribution of index information, and transaction traceability to prevent illegal actions in the e-healthcare distributed application. In addition, it implements tailored, efficient blockchain systems that utilize both multi-proof-of-work (PoW) and multi-proof-of-stake (PoS) mechanisms, together with digital signatures. These enhancements optimize resource usage, minimize storage requirements, and streamline the transaction process specifically for e-healthcare.

Liu and Li [24] proposed a novel system that combines a permissioned blockchain with Deep Reinforcement Learning (DRL)-enabled IoT to address privacy and power constraints. The proposed mechanism aims to provide healthcare services in real time, ensuring security and energy efficiency. Its primary focus is on assisting in the management of the COVID-19 pandemic. To deal with issues regarding security, a technique based on permissioned blockchain has been developed to guarantee the system's security. The strategy incorporates mobile edge computing (MEC) to disperse computing tasks to address energy restrictions. This helps decrease the proposed H-IoT system's computational load and energy consumption. Additionally, the system employs energy harvesting techniques to enhance performance. Moreover, using a DRL technique simultaneously improves the system's energy efficiency and security aspects. The simulation findings demonstrate that the suggested method successfully achieves a harmonious equilibrium between security and energy efficiency, providing a solid and effective response to the issues brought about by the COVID-19 pandemic.

Hadjixenophontos, et al. [25] introduced PRISM, an edge-centric system designed to assess innovative healthcare tools within domestic settings. They established a systematic approach rooted in automated IoT trials. Leveraging an extensive real-world dataset from in-home patient surveillance across 44 residences of People Living with Dementia (PLWD) spanning a two-year duration. Findings revealed anomaly identification with precision reaching 99%, alongside an average training duration as brief as 0.88 seconds. Despite the high precision observed in models trained on the same individual, a decline in accuracy occurred when assessed on diverse patients.

In healthcare, preserving the confidentiality, integrity, and availability of medical images is critical for precise diagnoses, treatment planning, and patient well-being. Consequently, Nadhan and Jacob [26] explored medical image security within IoT-based healthcare systems. Their study focused on using cryptography-based networks to encrypt and decrypt medical images, especially in secure image transmission through deep learning. The critical network was built upon the ResNet-50 architecture to establish the correlation between various image representations, enabling the incorporation of these intricate elements into the learning model for fine-tuning the encryption technique in specific domains. A reconstructive network was

then used in the decryption process to convert the encrypted image to its original "plaintext" form. Upon revealing the concealed components, an accessible Return on Investment (ROI) framework was established, streamlining data mining by accessing information directly from the user's local environment. The proposed system presented highly reliable imaging tools for assessing therapy outcomes. The utilization of two distinct publicly available datasets facilitated the accomplishment of their research objectives. The robust empirical setup and security analysis outcomes strongly suggest that the proposed approach can offer unparalleled security and yield powerful outcomes in medical image encryption and decryption.

Researchers have focused on efficient hashing techniques and digital certificates but may overlook the importance of integrating multiple cryptographic methods to enhance both security and efficiency. The work of Aruna Santhi and Vijaya Saradhi on attack detection using deep learning frameworks is innovative but does not specifically address encryption and integrity of medical data transmission. In the same way, Khan et al.'s blockchain-based approach enhances security in centralized systems but may introduce complexities in terms of data management and computational overhead. A hybrid security framework combining RC4, ECDSS, and SHA-256 is proposed in our research to address these gaps, providing a balanced solution to the challenges of encryption, authentication, and data transmission in IoT-enabled healthcare systems.

III. Proposed Approach

Securing the privacy and integrity of healthcare data transmission via networks is an ongoing and dynamic problem in IoT technologies. Conventional approaches to network access control are often ineffective and susceptible to unauthorized access or replication. Novel techniques like encryption algorithms have been developed to provide a highly efficient data invisibility framework. However, encryption algorithms encounter difficulties as a result of the strong relationship between the original message and the encrypted message, which facilitates the extraction of encryption keys and the retrieval of the original message by attackers. This study addresses these concerns, proposing a hybrid algorithm called RC4+ECDSS+SHA-256 to enhance the privacy and security of incoming data originating from healthcare and IoT equipment.

The RC4 technique has two primary stages: the Key Scheduling Process (KSP) and PRGM. The KSP populates the S-Box, a 256-byte table, by distributing the elements based on a key. The PRGM then generates a pseudo-random key, which is then used for XOR encryption to form the ciphertext. For more extensive plaintext data, the encryption time may be longer. Moreover, the ECDSS produces a cryptographic key for the RC4 method. ECDSS relies on the Discrete Logarithmic Problem (DLP) and operates within a group of cycles known as EC(Fp), which has a designated generator (G) and a co-factor (h) of 1. ECDSS utilizes key sets comprising a public key (PuK) and a private key (PrK). The SHA-256 is employed to guarantee data integrity. SHA-256 is a cryptographic hash algorithm used to validate data integrity. The suggested hybrid technique enhances privacy and security for medical data exchanged over IoT networks by integrating the RC4, ECDSS, and SHA-256 algorithms.

In the proposed technique, the efficiency and security of the system are improved by leveraging ECDSS for faster key generation during the PRGM process in RC4. KSP in RC4 modifies the key, which can be 40 to 256 bits depending on specific requirements. The suggested approach enhances security resilience versus various threats and improves storage effectiveness by incorporating modifications of ECDSS into RC4. The XOR operation encrypts the plaintext by combining it with the pseudo-random bits generated via the PRGM procedure. This process occurs after the critical encryption is performed by ECDSS and the permutation operation is performed in the KSP section.

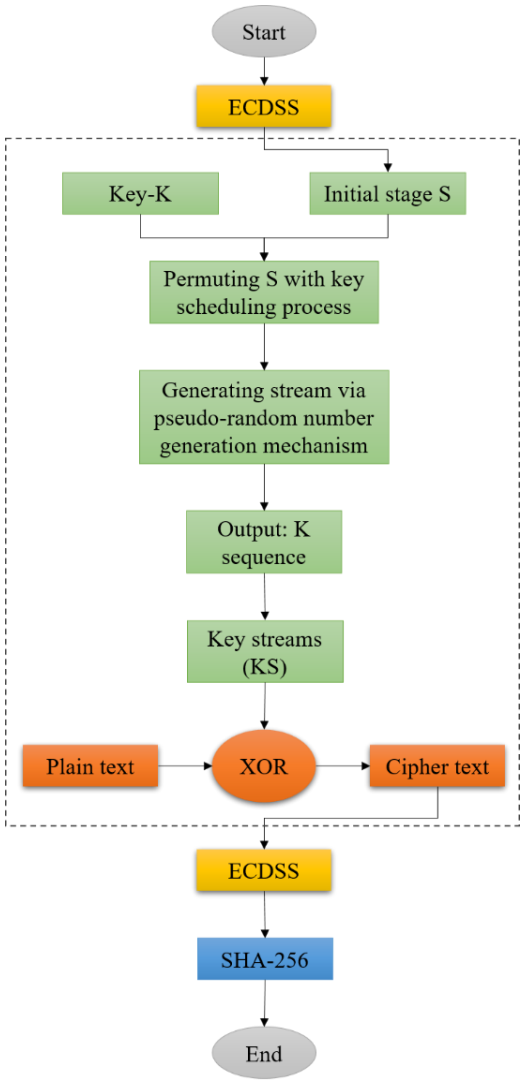


Fig. 1. The workflow of SHA-256+RC4+ECDSS.

To further increase security, a shift-right function is employed to choose a value prior to executing the XOR function. The plaintext is encrypted using the RC4 technique, and the entire process is further encrypted using the ECDSS certification. In addition, to ensure data integrity, the SHA-256 algorithm is used to encrypt the ECDSS signature. This ensures

that only authenticated users with access rights can access the ciphertext. Fig. 1 illustrates the process of the proposed SHA-256+RC4+ECDSS architecture. It provides a comprehensive overview of the various steps of the encryption process, including key generation, encryption with RC4, and data integrity verification with SHA-256. Overall, this hybrid approach combines the strengths of ECDSS, RC4, and SHA-256 to ensure security, efficiency, and data integrity in transmission and improve medical data via IoT networks.

The encryption process involves the following steps. Firstly, an elliptic curve (EC) is selected, and a generator (G) is chosen from its cyclic group. Next, the public-private key pair (*PuK*, *PrK*) is generated using ECDSS. This is achieved by randomly selecting a private key (*PrK*) within a specific range and computing the corresponding public key ($PuK = PrK * G$). The public key is then used as input for the KSP within the RC4 algorithm, modifying the *S* array (the key schedule) and determining its final state. The modified *S* array is employed in the PRGM to generate a pseudo-random key for encryption. Shift-right operations are performed on the *PuK* vector values to create a new key. The plaintext is encrypted using XOR encryption with the generated pseudo-random key. Additionally, a signature is generated for the Encrypted Plaintext (*EP*) using ECDSS with the private key and other parameters. The signature is then hashed using the SHA-256 algorithm. The encrypted message and the signature are the final output of the encryption process.

On the other hand, the decryption process involves the following steps. Given the encrypted message (EM) or ciphertext, the message is decrypted using the SHA-256 decryption process. The signature is verified to determine its acceptance or rejection. Signature verification involves the computation of various values using the ECDSS parameters and the received signature. The *R.y* value is calculated using the square root method. To verify the signature, the computation of *RwPrK + uPuK*EM (FPrK)* is performed. The signature is accepted if *x(R)* is congruent to *k (mod n)*, where *x(R)* represents the x-coordinate of *R*. Finally, the ciphertext is decrypted using the RC4 algorithm, and the resulting plaintext is the output of the decryption process.

Algorithm 1 outlines the pseudocode of the proposed technique, which incorporates the use of ECDSS, RC4, and SHA-256 algorithms in the workflow. In this situation, users transmit medical data from Internet of Things (IoT) devices to cloud storage. At this point, the initial task is to create the procedure for each workflow utilizing the SHA-256, ECDSS, and RC4 algorithms. ECDSS is a public key cryptography algorithm used for generating digital signatures, ensuring the authenticity and integrity of the data. RC4 is a widely used stream cipher algorithm for encryption and decryption, providing confidentiality to the transmitted data. SHA-256 is a cryptographic hash function that generates a unique hash value to ensure data integrity and detect modifications. Subsequently, the S array is altered for the data arriving utilizing the ECDSA public key. This phase guarantees that the received data is thoroughly validated and certified securely. Ultimately, users execute the encryption and decryption processes according to the instructions outlined in Algorithm 1. This allows for the secure transmission and retrieval of medical data, ensuring confidentiality and privacy.

Algorithm 1. Pseudocode of the proposed method

Input: Elliptic Curve Parameters (EC)
Output: Key Pairs (Public Key PuK, Private Key PrK)
Encryption process:
- Iterate over each EC(Fp) to choose a point G from order n.
- Generate the Elliptic Curve over Fp (GEEC(Fp)).
- Select a random integer PrK within the range $2 < PrK < n-2$.
- Compute $PuK = PrK * G$.
- Apply PuK as input for the KSA function.
- Generate the state S = KSA(PuK, S).
- Define the key length for S, modify the S array using PuK, and execute the PRGA function for S.
- Assign S as PuK, shift-right the PuK vector values, and create a new key based on shift-right operations.
- Encrypt the plaintext using the XOR operation, defining the encrypted plaintext as EP, private key PrK, and ECD parameters.
- Choose a random integer r within the range $2 < r < n-2$.
- Calculate R = rG.
- Determine $k = x(R) \pmod n$.
- Compute $S = r^{-1} * (h(M) + dk) \pmod n$.
- Hash(S) using SHA-256.
- Generate the Encrypted Message EM.

Decryption Process: Input the Encrypted Message EM or Cipher Text.
- Decrypt the message using SHA-256-decryption.
- Accept or reject the signature.
- Compute $v = S^{-1} \pmod n$ for verification.
- Compute $w = h(M) * v \pmod n$.
- Calculate $u = k * v \pmod n$.
- Calculate R.y value.
- Retrieve point R by k value using the square root method.
- Calculate $R = wPrK + uPuKeEM(F_{PrK})$.
- Accept the signature If $x(R) = k \pmod n$.
- Decrypt the cipher text using RC4.

Output: Plaintext

The input submitted for encryption is converted into an array using ECDSS, which is altered using the public key of the ECDSS encryption scheme. To illustrate, the original key "abcd" is converted into array K by the ECDSS process. The strings are then converted into bytes. Encryption and decryption are done using the RC4 algorithm based on the byte arrays. For example, array *A* is obtained as [97, 98, 99, 100] from abcd, and the encrypted RC4 key is generated using ECDSS, resulting in array *K* as [65, 121, 53, 47, 104, 87, 86].

The proposed model aims to improve data security for healthcare infrastructure and applications. In these systems, diverse health information is collected from IoT devices, archived in the cloud, and transmitted to medical facilities, healthcare institutions, and private doctors for assessment. However, the focus is on unauthorized access by third parties, which poses a significant risk to the confidentiality of the data. In the suggested algorithm, the collected IoT data is encoded and decoded by a combined approach of RC4+ECDSS+SHA-256. This ensures that only trusted parties, such as doctors, health centers, and research centers, can access the data, lowering the risk of unauthorized access.

In this proposed hybrid algorithm, RC4, ECDSS, and SHA-256 are individually complex cryptographic components, and their sequential application determines the algorithm's computational complexity. As a stream encryption algorithm, RC4 has a time complexity of *O(n)*, where *n* is the length of the input data. However, KSP and PRGM in RC4 incur some overhead due to their iterative nature. For key generation and signature generation, the ECDSS, which uses elliptic curve cryptography, requires *O(log(n))* and *O(n)*, respectively. SHA-256 is a cryptographic hash function that operates with a fixed complexity of *O(n)*. These algorithms, particularly the repetition of key generation and transformation steps, result in a complexity of *O(n)* for encryption and decryption, but with additional overhead from elliptic curve operations and hash computation, making the hybrid approach computationally intensive but robust against various security threats at the same time.

IV. Simulation Results

We evaluated the presented security framework using various performance metrics, including throughput, decryption time, and encryption time. A summary of the components and parameters considered in the system is outlined in Table II. The CPU is an Intel i5 processor running at 3.2 GHz, while the operating system is Windows 10. The system boasts 4 GB of RAM and operates on a 32-bit architecture. The simulation leverages Python with the Cryptography class for its configuration. Moreover, various cryptographic models and their respective specifications, including key size, block size, and other specific parameters like S-box size for RC4, are outlined in Table II, providing a comprehensive overview of the system used for the study.

A ratio of original data size to encryption time determines encryption throughput. An increase in encryption throughput indicates higher algorithm efficiency. As shown in Table III, we adopted a file size of 10 MB for the test, which resulted in a throughput time of 11.67 ms. Eq. (1) calculates encryption throughput. The analysis of encryption throughput compared to other techniques is shown in Fig. 2.

$$E_t(KB/ms) = \frac{\sum input\ file}{\sum encryption\ time} \quad (1)$$

Likewise, the decryption throughput is calculated by dividing the input file size by the decryption time. For a file size of 10 MB, the analysis resulted in a decryption throughput of 12.79 ms, as shown in Table IV. The calculation of the decryption throughput follows Eq. (2). Fig. 3 illustrates the analysis of the decryption throughput.

$$D_t(KB/ms) = \frac{\sum input\ file}{\sum decryption\ time} \quad (2)$$

TABLE II. Simulation Parameters

Cryptographic model	Block size (bits)	Key size (bits)
RC2	128	128
AES	128	256
RC4	S-box (256 bytes)	256
ECDSS	128	256
3DES	64	256

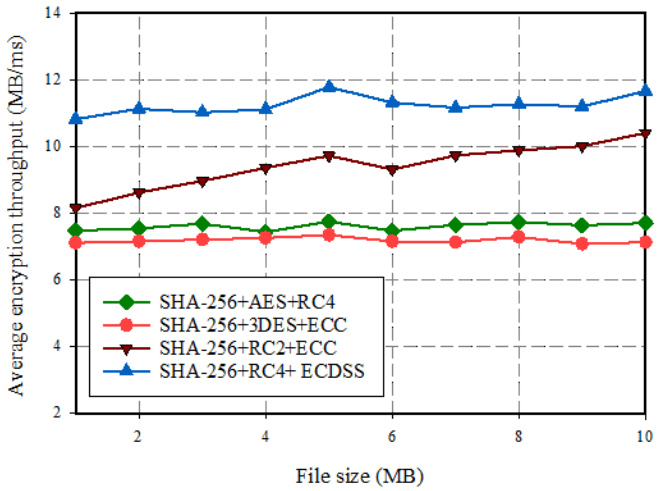


Fig. 2. Encryption throughput comparison.

TABLE III. Encryption Throughput Analysis

Size of the original file (MB)	Average duration taken for encryption throughput			
	SHA-256+AES+RC4	SHA-256+3DES+ECC	SHA-256+RC2+ECC	SHA-256+RC4+ ECDSS
1	7.48	7.11	8.16	10.83
2	7.55	7.16	8.63	11.14
3	7.69	7.21	8.97	11.03
4	7.44	7.26	9.37	11.12
5	7.75	7.35	9.73	11.79
6	7.48	7.16	9.32	11.33
7	7.65	7.13	9.74	11.17
8	7.74	7.29	9.89	11.28
9	7.64	7.08	10.02	11.21
10	7.71	7.13	10.41	11.67
Average time	7.613	7.188	9.424	11.257
Total time	83.743	79.068	103.664	123.827

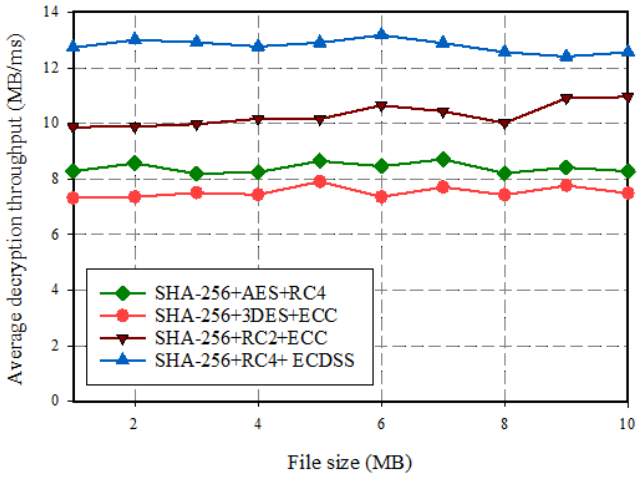


Fig. 3. Decryption throughput comparison.

Table V compares the proposed hybrid algorithm with other algorithms regarding the encryption time. All files with the suggested algorithm were encrypted in 331.5 ms on average, indicating a faster encryption process. Fig. 4 provides a visualization of the encryption time analysis. Furthermore, In Table VI, the developed method and alternatives are compared in terms of decryption times. Our method has a shorter decryption time of 312.1 ms than any other method. Fig. 5 provides a graphical representation of the decryption time analysis and illustrates the performance of the proposed and existing methods.

The results show that the proposed hybrid security framework has superior performance in terms of encryption and decryption efficiency. The observed encryption throughput of 11.67 KB/ms and decryption throughput of 12.79 KB/ms with a file size of 10 MB indicates a highly efficient process that outperforms many existing methods. The average encryption time of 331.5 ms and the decryption time of 312.1 ms further underline the effectiveness of the model, as these times are significantly lower than those of other algorithms. These metrics justify the conclusion that the proposed framework increases both security and efficiency, making it suitable for real-time applications in IoT-enabled healthcare systems. The reduced processing times mean the framework can process large volumes of sensitive data quickly, ensuring robust security without compromising performance, which is critical in healthcare environments where timely data processing is essential.

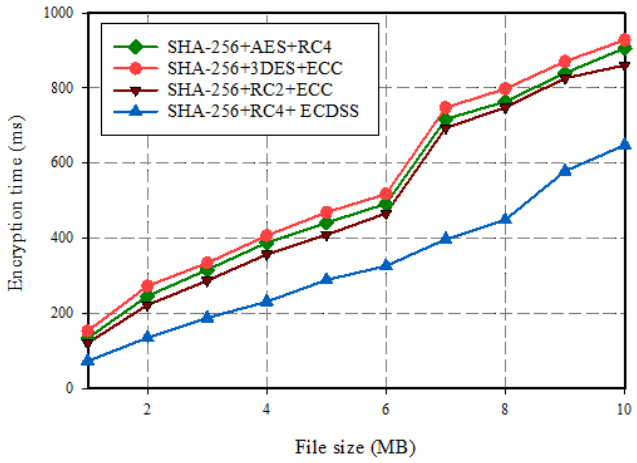


Fig. 4. Encryption time comparison.

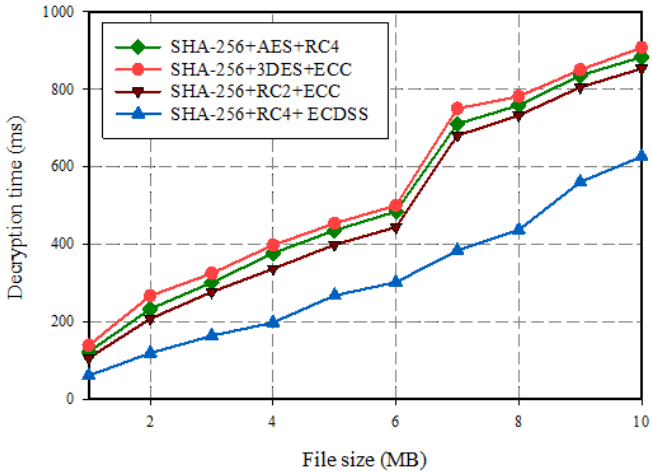


Fig. 5. Decryption time comparison.

TABLE IV. DECRYPTION THROUGHPUT ANALYSIS

Size of the original file (MB)	Average duration taken for decryption throughput			
	SHA-256+AES+RC4	SHA-256+3DES+ECC	SHA-256+RC2+ECC	SHA-256+RC4+ ECDSS
1	8.28	7.32	9.87	12.73
2	8.57	7.36	9.89	13.01
3	8.19	7.51	9.98	12.92
4	8.25	7.44	10.17	12.77
5	8.65	7.91	10.16	12.91
6	8.46	7.36	10.65	13.18
7	8.71	7.71	10.44	12.89
8	8.21	7.43	10.02	12.57
9	8.42	7.77	10.91	12.41
10	8.28	7.49	10.96	12.57
Average time	8.4	7.53	10.3	12.79
Total time	92.42	82.83	113.35	140.75

TABLE V. ENCRYPTION TIME ANALYSIS

Size of the original file (MB)	Required duration for encryption			
	SHA-256+AES+RC4	SHA-256+3DES+ECC	SHA-256+RC2+ECC	SHA-256+RC4+ ECDSS
1	134	153	121	73
2	246	273	222	135
3	316	334	287	188
4	388	407	357	231
5	441	469	409	289
6	492	518	466	326
7	717	748	694	397
8	764	798	748	449
9	840	871	826	579
10	906	928	861	648
Average time	524.4	549.9	499.1	331.5
Total time	5768.4	6048.9	5490.1	3646.5

TABLE VI. DECRYPTION TIME ANALYSIS

Size of the original file (MB)	Required duration for decryption			
	SHA-256+AES+RC4	SHA-256+3DES+ECC	SHA-256+RC2+ECC	SHA-256+RC4+ ECDSS
1	122	139	106	61
2	233	267	208	119
3	301	325	277	164
4	377	398	337	198
5	436	454	399	268
6	485	501	444	302
7	711	751	681	384
8	759	782	733	437
9	836	851	806	561
10	884	908	854	627
Average time	514.4	537.6	484.5	312.1
Total time	5658.4	5913.6	5329.5	3433.1

V. DISCUSSION

The proposed hybrid security framework combining RC4, ECDSS, and SHA-256 has demonstrated significant improvements in healthcare systems' efficiency and security. The encryption and decryption throughput results show that the framework can process large amounts of data quickly, which is critical in real-time healthcare environments where timely access to patient data can be life-saving. Integrating ECDSS with RC4 not only increases key management efficiency but also strengthens encryption against common cryptographic attacks. Using SHA-256 ensures data integrity and provides an additional layer of security that is essential to maintaining patient confidentiality and complying with healthcare regulatory standards.

Furthermore, the empirical analysis confirms that the proposed framework outperforms existing security models in terms of encryption and decryption times, indicating its suitability for resource-constrained IoT devices commonly used in healthcare. This performance improvement is critical for practical use as it minimizes computational overhead while maximizing security. The hybrid approach also addresses the identified gaps in previous research by providing a more comprehensive solution that integrates multiple cryptographic techniques, providing robust protection against unauthorized access and ensuring the integrity and confidentiality of medical data during transmission. This makes the proposed framework a viable option for improving the security of IoT-enabled healthcare systems in an increasingly connected and vulnerable digital landscape.

VI. CONCLUSION

This study addresses the critical challenge of data privacy and security in IoT-based healthcare systems by proposing a hybrid security framework. The framework combined ECDSS, RC4, and SHA-256 to preserve the integrity and confidentiality of data sent. The experimental analysis demonstrated the superiority of the proposed model, particularly when encrypting

data of 10 MB. The framework achieved a high throughput of 11.67 MB per millisecond, with an encryption time of 846 milliseconds and a decryption time of 627 milliseconds. These results highlighted the efficiency and effectiveness of the proposed hybrid security framework. Compared to existing techniques such as SHA-256+3DES+ECC, SHA-256+ AES+RC4, and SHA-256+ RC2+ECC, the proposed model outperformed for encryption time, decryption time, and overall security. It provided a robust solution for protecting sensitive healthcare data transmitted through IoT devices.

Further investigation in healthcare IoT security could concentrate on using advanced cryptographic techniques, such as homomorphic encryption and quantum-resistant algorithms, to enhance data protection. Exploring decentralized identity management and blockchain applications has the potential to improve identity verification in healthcare systems. In addition, creating machine learning models to identify anomalies in healthcare data collected by IoT devices could facilitate proactive security measures. Furthermore, the combination of edge computing with federated learning has the potential to resolve privacy issues by locally processing confidential data. These instructions guarantee the enhancement of the durability of healthcare IoT systems against advancing cyber dangers while promoting innovation in patient data confidentiality and protection.

REFERENCES

[1] S. K. Dezfuli, "Targeted killings and the erosion of international norm against assassination," Defense & Security Analysis, vol. 39, no. 2, pp. 191-206, 2023, doi: https://doi.org/10.1080/14751798.2023.2185947.

[2] S. Abdidizaji, A. K. Yalabadi, M. Yazdani-Jahromi, O. O. Garibay, and I. Garibay, "Agent-Based Modeling of C. Difficile Spread in Hospitals: Assessing Contribution of High-Touch vs. Low-Touch Surfaces and Inoculations' Containment Impact," arXiv preprint arXiv:2401.11656, 2024, doi: https://doi.org/10.48550/arXiv.2401.11656.

[3] F. Kamalov, B. Pourghebleh, M. Gheisari, Y. Liu, and S. Moussa, "Internet of medical things privacy and security: Challenges, solutions, and future trends from a new perspective," Sustainability, vol. 15, no. 4, p. 3317, 2023.

[4] T. Arpitha, D. Chouhan, and J. Shreyas, "Anonymous and robust biometric authentication scheme for secure social IoT healthcare applications," Journal of Engineering and Applied Science, vol. 71, no. 1, pp. 1-23, 2024.

[5] B. Pourghebleh, N. Hekmati, Z. Davoudnia, and M. Sadeghi, "A roadmap towards energy - efficient data fusion methods in the Internet of Things," Concurrency and Computation: Practice and Experience, vol. 34, no. 15, p. e6959, 2022.

[6] Z. N. Aghdam, A. M. Rahmani, and M. Hosseinzadeh, "The role of the Internet of Things in healthcare: Future trends and challenges," Computer methods and programs in biomedicine, vol. 199, p. 105903, 2021.

[7] B. Pourghebleh and N. J. Navimipour, "Data aggregation mechanisms in the Internet of things: A systematic review of the literature and recommendations for future research," Journal of Network and Computer Applications, vol. 97, pp. 23-34, 2017.

[8] B. Pourghebleh, V. Hayyolalam, and A. A. Anvigh, "Service discovery in the Internet of Things: review of current trends and research challenges," Wireless Networks, vol. 26, no. 7, pp. 5371-5391, 2020.

[9] D. Hao and C. JianHua, "A Survey of Structural Health Monitoring Advances Based on Internet of Things (IoT) Sensors," International Journal of Advanced Computer Science and Applications, vol. 14, no. 10, 2023.

[10] A. SRHIR, T. MAZRI, and M. BENBRAHIM, "Security in the IoT: State-of-the-art, issues, solutions, and challenges," International Journal of Advanced Computer Science and Applications, vol. 14, no. 5, 2023.

[11] R. Zgheib, S. Kristiansen, E. Conchon, T. Plageman, V. Goebel, and R. Bastide, "A scalable semantic framework for IoT healthcare applications," Journal of Ambient Intelligence and Humanized Computing, vol. 14, no. 5, pp. 4883-4901, 2023.

[12] S. Vairachilai, A. Bostani, A. Mehbodniya, J. L. Webber, O. Hemakesavulu, and P. Vijayakumar, "Body sensor 5 G networks utilising deep learning architectures for emotion detection based on EEG signal processing," Optik, p. 170469, 2022.

[13] A. Rejeb et al., "The Internet of Things (IoT) in healthcare: Taking stock and moving forward," Internet of Things, p. 100721, 2023.

[14] S. S. Sefati, B. Arasteh, S. Halunga, O. Fratu, and A. Bouyer, "Meet User's Service Requirements in Smart Cities Using Recurrent Neural Networks and Optimization Algorithm," IEEE Internet of Things Journal, 2023.

[15] S. Yazdanpanah, S. S. Chaeikar, and A. Jolfaei, "Monitoring the security of audio biomedical signals communications in wearable IoT healthcare," Digital Communications and Networks, vol. 9, no. 2, pp. 393-399, 2023.

[16] B. Nadimi and H. Zheng, "A Multi-Expert Large Language Model Architecture for Verilog Code Generation," arXiv preprint arXiv:2404.08029, 2024.

[17] M. R. Ahmed, B. Nadimi, and H. Zheng, "AutoModel: Automatic Synthesis of Models from Communication Traces of SoC Designs," IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems, 2024.

[18] S. Boopathi, "Securing Healthcare Systems Integrated With IoT: Fundamentals, Applications, and Future Trends," in Dynamics of Swarm Intelligence Health Analysis for the Next Generation: IGI Global, 2023, pp. 186-209.

[19] M. A. Tofighi, B. Ousat, J. Zandi, E. Schafir, and A. Kharraz, "Constructs of Deceit: Exploring Nuances in Modern Social Engineering Attacks," in International Conference on Detection of Intrusions and Malware, and Vulnerability Assessment, 2024: Springer, pp. 107-127, doi: https://doi.org/10.1007/978-3-031-64171-8_6

[20] H. Verma, N. Chauhan, and L. K. Awasthi, "A Comprehensive review of 'Internet of Healthcare Things': Networking aspects, technologies, services, applications, challenges, and security concerns," Computer Science Review, vol. 50, p. 100591, 2023.

[21] A. M. Al Shahrani, A. Rizwan, M. Sánchez-Chero, C. E. Rosas-Prado, E. B. Salazar, and N. A. Awad, "An internet of things (IoT)-based optimization to enhance security in healthcare applications," Mathematical Problems in Engineering, vol. 2022, 2022.

[22] J. Aruna Santhi and T. Vijaya Saradhi, "Attack detection in medical Internet of things using optimized deep learning: Enhanced security in healthcare sector," Data Technologies and Applications, vol. 55, no. 5, pp. 682-714, 2021.

[23] A. A. Khan et al., "Data Security in Healthcare Industrial Internet of Things with Blockchain," IEEE Sensors Journal, 2023.

[24] L. Liu and Z. Li, "Permissioned blockchain and deep reinforcement learning enabled security and energy efficient Healthcare Internet of Things," Ieee Access, vol. 10, pp. 53640-53651, 2022.

[25] S. Hadjixenophontos, A. M. Mandalari, Y. Zhao, and H. Haddadi, "PRISM: Privacy Preserving Healthcare Internet of Things Security Management," in 2023 IEEE Symposium on Computers and Communications (ISCC), 2023: IEEE, pp. 1-5.

[26] A. S. Nadhan and I. J. Jacob, "Enhancing healthcare security in the digital era: Safeguarding medical images with lightweight cryptographic techniques in IoT healthcare applications," Biomedical Signal Processing and Control, vol. 88, p. 105511, 2024.

Dose Archiving and Communication System in Moroccan Healthcare: A Unified Approach to X-Ray Dose Management and Analysis

Lhoucine Ben Youssef[1], Abdelmajid Bybi[2], Hilal Drissi[3], El Ayachi Chater[4]
Mohammed V University in Rabat, Higher School of Technology of Salé, LASTIMI, Salé, Morocco[1, 3, 4]
Mohammed V University in Rabat, Higher School of Technology of Salé, MEAT, Salé, Morocco[2]

***Abstract*—This study explores the implementation of a Dose Archiving and Communication System (DACS) in Moroccan healthcare, highlighting the importance of X-ray dose management in modern radiology. It emphasizes patient safety and the ALARA principle to minimize radiation exposure while maintaining diagnostic accuracy. The research discusses advancements in imaging technologies, such as dose-reduction algorithms and real-time monitoring systems. A survey of 1000 healthcare professionals reveals significant challenges in X-ray dose management, including poor dose tracking, regulatory non-compliance, and inadequate radiation protection training. Noteworthy findings reveal that 10% of patients received doses exceeding 5 Gray, underscoring the exigency for robust dose management systems. The article delineates a strategic implementation approach for DACS in Moroccan hospitals, comprising meticulous needs assessment, infrastructure fortification, and stakeholder engagement. By harnessing cloud-based storage, blockchain technology, and industry-standard encryption protocols, the envisioned DACS endeavors to furnisha secure, scalable, and efficient framework for radiation dose management. This holistic approach, underpinned by empirical statistics regarding training in radiation protection, network infrastructure, and DACS implementation strategies, aims to elevate patient outcomes and ensure stringent regulatory compliance.**

***Keywords*—DACS; Real-time monitoring; radiation protection; radiology practice; healthcare professionals; x-ray doses; regulatory compliance; patient safety; Moroccan healthcare**

I. Introduction

X-ray dose management represents a fundamental aspect of contemporary radiology practice, emphasizing the ethical imperative to prioritize patient safety while harnessing the diagnostic potential of X-ray imaging. By fostering a culture of radiation stewardship and embracing evidence-based strategies, healthcare institutions can uphold the principles of ALARA (As Low As Reasonably Achievable) and deliver high-quality care that maximizes clinical efficacy while minimizing radiation-related risks [1].

In the realm of modern medical imaging, X-ray technology stands as a cornerstone for diagnostic and therapeutic purposes, facilitating crucial insights into the human body's intricate structures. However, alongside its undeniable benefits, the utilization of X-ray radiation carries inherent risks, particularly concerning cumulative radiation exposure and its potential adverse effects on patients and healthcare professionals alike.

Effective X-ray dose management encompasses various dimensions, including dose monitoring, optimization of imaging protocols, equipment calibration, personnel training, and patient education. Through meticulous monitoring and analysis of radiation doses delivered during diagnostic and interventional procedures, healthcare providers can gain valuable insights into radiation utilization patterns, identify potential areas for improvement, and tailor interventions to mitigate unnecessary exposure.

Within this framework, Garba et al. [2] conducted research focusing on the development of a manual radiation dose management system to monitor and track radiation doses and scan parameters for brain CT scans. The system monitored CTDI vol and DLP, using notification values to identify procedures requiring optimization. Data analysis was conducted to compare with national and international diagnostic reference levels (DRLs) to ensure compliance and enhance patient safety. Heron et al. [3] examined the impact of X-ray-based medical imaging on staff safety and explored how new technologies affect medical staff's exposure to X-rays. They highlighted the crucial importance of using protective measures and ongoing training. Polizzi-et al. [4] carried out a study to standardize X-ray cabinet irradiator dose, geometry, and calibration reporting, focusing on a dual X-ray source cabinet irradiator (CIXD, Xstrahl Limited, UK). They assessed dose distribution under various experimental conditions using methods such as half-value layer (HVL) measurement, profile measurements, and output calibration with an ion chamber, alongside two weeks of constancy measurements. Film measurements evaluated percent depth dose and homogeneity. The X-ray tubes showed an output of 1.27 Gy/min with an HVL of 1.7 mm Cu. Simultaneous operation of the tubes reduced the heel effect observed in individual tubes. Despite a 15% dose inhomogeneity within the tray area, film measurements indicated only minor nonuniformities. Additionally, Silva et al. [5] undertook a study to investigate and evaluate the radiation doses received by professionals during chest CT scans and to assess the effectiveness of personal protective equipment (PPE). Computational scenarios were simulated using pediatric (1 and 10 years old) and adult virtual anthropomorphic phantoms to represent patients and professionals. The MCNP 6.2 Monte Carlo code was employed to determine conversion coefficients for equivalent (CC[HT]) and effective (CC[E]) doses. Another noteworthy contribution to radiation protection research was offered by Kawauchi et al. [6]. Their study

emphasized the importance of protecting lenses during cerebral angiography examinations to ensure patient lens safety. To assess lens doses, they employed both a phantom and a real-time dosimeter. Additionally, they computed an artifact index for evaluating image quality through pixel and noise analysis.

Healthcare professionals and patients face increased health risks from radiation overdoses due to inadequate use of protective measures and passive dosimeters during imaging exams. Developing effective X-ray dose management systems is essential to mitigate these risks by ensuring rigorous dose monitoring and adherence to safety protocols. In this context, Choi et al. [7] conducted a study focusing on implementing a Dose Archiving and Communication System (DACS) in a large healthcare system to manage radiation doses. Liu et al. [8] provide a comprehensive review of DACS in radiation oncology, focusing on its capabilities, challenges, and future prospects. Furthermore, Faggioni et al. [9] describe the implementation and evaluation of a DACS tailored for pediatric cardiac catheterization procedures. The study evaluates how DACS improves radiation dose monitoring and management in pediatric settings. Rehani et al. [10] explore the comprehensive landscape of radiation dose management systems, including DACS, discussing challenges in monitoring doses across various medical imaging modalities. Wang et al. [11] examine the advancements in DACS technology and its role in enhancing radiation dose management in clinical settings. Moreover, Martin et al. [12] explore how the implementation of DACS affects clinical workflow and patient care outcomes, highlighting both benefits and challenges.

We have designed an innovative Dose Archiving and Communication System (DACS) that offers several significant advantages. This system operates independently and adapts seamlessly to any IT infrastructure installed in the relevant departments. Unlike other solutions, our DACS is engineered to meet the constraints of network infrastructure and limited access, providing exceptional flexibility and compatibility. It can be used with both new equipment technologies and older technologies, making it a durable and adaptable solution. Furthermore, our system allows for real-time dose monitoring, with alerts and notifications to ensure optimal management of prescribed examination protocols for patients and the doses absorbed by healthcare personnel during their daily tasks. This level of control and responsiveness helps improve the safety and efficiency of radiological practices.

This paper is organized into seven sections. The first section is Introduction. Section II describes the general context. Section III explores the study's foundational aspects and methodology. Section IV is devoted to elucidating the dose system management. Following that, Section V sheds light on the results and fosters discussions. Then, Section VI concentrates on outlining an approach for implementing dose archiving and communication systems in healthcare institutions. Finally, Section VII presents the conclusion.

II. General Context

X-ray dose management emerges as a pivotal discipline within radiology and healthcare at large, aiming to optimize the utilization of X-ray imaging while minimizing radiation exposure risks. By implementing comprehensive strategies, protocols, and technologies, X-ray dose management endeavors to strike a delicate balance between acquiring diagnostically valuable images and safeguarding patient safety. Moreover, advancements in imaging technology, such as dose-reduction algorithms, dose-tracking software, and real-time dose monitoring systems, offer invaluable tools for enhancing X-ray dose management practices. These innovations empower healthcare professionals to optimize imaging parameters, adjust radiation doses based on patient characteristics and clinical indications, and ensure that diagnostic goals are achieved with the lowest feasible radiation exposure.

In the domain of X-ray exposure, a nuanced understanding of various dose metrics is indispensable for assessing radiation risks and ensuring patient safety. These metrics provide crucial insights into the amount of radiation absorbed by individuals during imaging procedures. Among the key types of X-ray doses, the Entrance Skin Dose (ESD) stands out as it quantifies the radiation absorbed by the skin at the point of entry of the X-ray beam, serving as a critical indicator of potential skin effects. Additionally, Organ Dose evaluation allows for a targeted assessment of radiation absorbed by specific organs or tissues, recognizing their differing sensitivities and facilitating a more accurate estimation of associated health risks. The Effective Dose (ED) metric goes further by synthesizing the varying sensitivities of different tissues and organs, offering a comprehensive assessment of the overall risk posed by a particular radiation exposure, measured in Sieverts (Sv). Dose Area Product (DAP) is instrumental in quantifying the total radiation delivered to a specific area during an X-ray procedure, factoring in both radiation intensity and exposed area. Peak Skin Dose (PSD) is particularly pertinent in interventional radiology procedures, identifying the maximum skin dose reached during a specific imaging intervention. Cumulative Dose accounts for the total radiation accumulated over time from multiple X-ray exposures, particularly crucial for individuals undergoing frequent medical imaging. Finally, Effective Dose Equivalent (EDE) provides a refined measure of radiation risk by considering the biological harm associated with different types of ionizing radiation, thereby guiding radiation protection strategies. For medical professionals, radiologists, and radiation safety experts, a comprehensive grasp of these dose metrics is essential in ensuring that X-ray procedures are conducted with utmost precision and minimal risk to patients, while still yielding valuable diagnostic insights [13-15].

The legal framework governing nuclear and radiological safety, security, and ionizing radiation control in Morocco is outlined in references [16-19]. These documents articulate three fundamental principles of radiation protection: justification, optimization, and limitation. Justification dictates that no activity involving exposure to ionizing radiation should occur unless it produces a positive economic, social, or other benefit that outweighs potential health risks. Optimization requires minimizing individuals' exposure to ionizing radiation to the fullest extent possible, taking into account economic and social considerations. Limitation mandates that cumulative doses from all activities must not exceed the dose limits specified by regulations.

To complement these regulatory efforts, several agencies have been established to serve specific purposes such as dosimetry monitoring, calibration, and metrology of ionizing radiation devices, as well as control and expertise in radiation protection, quality assurance in medical imaging, radiological environmental monitoring, and radiation protection training. These agencies include "The National Center for Radiation Protection of the Ministry of Public Health," "The National Commission for Radiation Protection," "The National Commission for Nuclear Safety,""The Department of Nuclear Energy of the Ministry of Energy and Mines," and "The National Center for Energy of Nuclear Sciences and Techniques" [16-19].

The integration of dosimetry control and monitoring is an increasingly prevalent trend within Moroccan healthcare establishments, influenced by various challenges: technological (quality, safety, and innovation), regulatory, and economic considerations. This research is crafted for the benefit of all practitioners involved in ionizing radiation. The primary objectives of this study revolve around the adoption of best practices in radiation protection. Additionally, it aims to promote the utilization of both individual and collective protective measures. Lastly, the research emphasizes the significance of monitoring the dosimetry of each individual to ensure personal protection and minimize absorbed radiation doses.

III. Study Background and Methodology

While technology continues to evolve and regulatory standards change, the field of X-ray dose management remains a major challenge for the Moroccan authorities, particularly in terms of the doses absorbed by healthcare professionals.

To understand users' needs in X-ray dose management, a survey was conducted among 1000 healthcare professionals. The results revealed that the main challenges encountered included: X-ray dose tracking, regulatory and normative compliance and lack of adequate training, underscoring the importance of an integrated dose management system to address users' diverse needs and enhance patient and worker safety.

The significance of network infrastructure in the context of radiology departments is paramount for enhancing radiation protection and dose management. A robust network infrastructure plays a crucial role in facilitating the integration of advanced technologies and tools that contribute to the overall safety and efficiency of x-rays practices [20-22]. Several key points highlighting the importance of network infrastructure in this regard:

1) Data management and storage: Network infrastructure enables seamless data management and storage of radiological images and patient information. This centralized approach allows for efficient access, retrieval, and secure archival of critical data, supporting comprehensive dose monitoring and analysis.

2) Integration of imaging systems: A well-developed network allows for the integration of various imaging systems and devices within the radiology department. This integration enhances the coordination and interoperability of equipment, ensuring a streamlined approach to dose control and radiation protection protocols.

3) Real-time monitoring and analysis: Network connectivity facilitates real-time monitoring of imaging procedures and radiation dose levels. With instant access to this information, healthcare professionals can make informed decisions promptly, adjusting protocols as needed to optimize radiation exposure for patients and staff.

4) Telemedicine and remote consultations: Network infrastructure supports telemedicine initiatives and remote consultations, enabling radiologists to collaborate and share expertise regardless of physical location. This capability contributes to more extensive knowledge exchange, fostering best practices in radiation protection.

5) Implementation of dose tracking systems: Dose tracking systems, crucial for monitoring and managing radiation exposure, rely on a robust network infrastructure. The connectivity provided by the network allows for the seamless collection, analysis, and reporting of dose data, aiding in the implementation of dose optimization strategies.

6) Security and compliance: A secure network infrastructure is essential to protect sensitive patient data and ensure compliance with regulatory standards. Compliance with security protocols is integral to maintaining the integrity of radiation protection measures and preventing unauthorized access to patient information.

7) Education and training programs: Network connectivity supports online education and training programs for healthcare professionals involved in radiology. This allows for continuous learning, ensuring that the staff stays updated on the latest advancements in radiation protection practices.

Dose Archiving and Communication System (DACS) aims to enhance dose monitoring and management practices in medical imaging by providing comprehensive solutions for capturing, storing, and analyzing dose data. Through advanced tracking and reporting functionalities, DACS facilitates the optimization of radiation dose levels, ensuring patient safety, regulatory compliance, and the quality of diagnostic imaging procedures. By centralizing dose data and promoting data-driven decision-making, DACS empowers healthcare providers to improve patient care and enhance operational efficiency in medical imaging facilities.

In the context of analyzing the needs for managing and analyzing X-ray doses, understanding users (Biomedical, Radiology, Oncology and Surgery staff) specific requirements and the current challenges faced by professionals in this field is paramount. Our system is a comprehensive solution designed to manage and monitor radiation doses administered to patients during radiological procedures. Upon launching the system, users are greeted with a secure authentication screen, ensuring that only authorized personnel can access sensitive patient and dose information. Once logged in, users are directed to the main dashboard, which serves as the central hub of the system. The interface is intuitively organized into several key sections, providing seamless navigation and functionality.

The settings section allows administrators to configure system parameters, user roles, and access permissions, ensuring that the system operates according to institutional protocols and user needs. The statistics display provides real-time analytics and visualizations of dose data, including charts and graphs that illustrate trends, compliance with safety standards, and individual patient dose histories. This feature empowers healthcare professionals to make informed decisions and optimize patient safety. Additionally, the dashboard includes a comprehensive dose archive section, where detailed records of all administered doses are stored and easily retrievable. The interface also supports advanced search capabilities, enabling users to quickly locate specific patient records or procedure details. Overall, the system interface is designed to be used friendly, efficient, and secure, facilitating x-ray dose management and communication within radiology departments.

Our solution is designed with four primary objectives (Fig. 1) to ensure optimal radiation dose management and patient safety. First, the collect objective focuses on the systematic collection and secure storage of dosimetric data. This involves gathering detailed records of all administered doses, ensuring that data integrity and confidentiality are maintained. Second, the monitor objective enables continuous, real-time tracking of patient radiation doses. This feature provides healthcare professionals with immediate feedback on radiation exposure levels, allowing for timely adjustments and interventions. Third, the evaluate objective involves comprehensive statistical analysis of dosimetric information. By utilizing advanced analytics tools, the system can identify trends, generate reports, and support evidence-based decision-making to improve radiation safety protocols. Finally, the optimize objective aims at the assessment and enhancement of radiation dose safety and efficacy. This includes evaluating current practices, implementing optimization strategies, and ensuring compliance with regulatory standards to minimize patient exposure while maintaining diagnostic image quality. Together, these objectives create a robust framework for x-ray dose management in radiological practices.

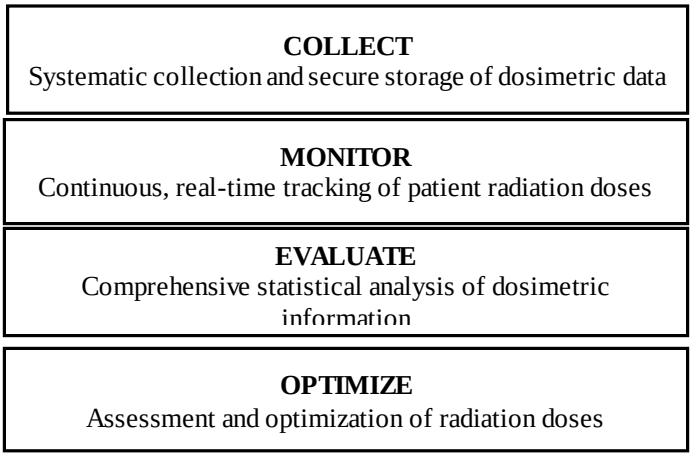


Fig. 1. Dose archinving and communication system objectives.

In conclusion, a robust network infrastructure is indispensable for creating a connected and technologically advanced radiology department. By enabling seamless data management, integrating imaging systems, facilitating real-time monitoring, supporting telemedicine, implementing dose tracking systems, ensuring security and compliance, and fostering education programs, the network plays a pivotal role in enhancing radiation protection and dose control in medical imaging based on X-rays.

IV. X-Ray Dose Management

According to our survey, users' specific needs vary depending on their roles and usage contexts. Radiologists and medical imaging technicians require effective tools to monitor and evaluate x-ray doses administered to patients while optimizing diagnostic image quality and controlling the doses absorbed by them. Similarly, radiation protection and safety professionals need dose management systems to ensure compliance with regulatory standards and worker safety. Current challenges encountered by professionals in this domain include the increasing complexity of imaging equipment, which makes dose management more challenging, as well as time and resource constraints that limit the ability to implement effective dose management practices. Additionally, compliance with regulatory standards and safeguarding patients' health data are critical considerations.

By conducting a comprehensive analysis of users' needs and the challenges faced by professionals in managing and analyzing X-ray doses, we have developed effective solutions tailored to their requirements. Fig. 2 illustrates an overall structure of our system application area.

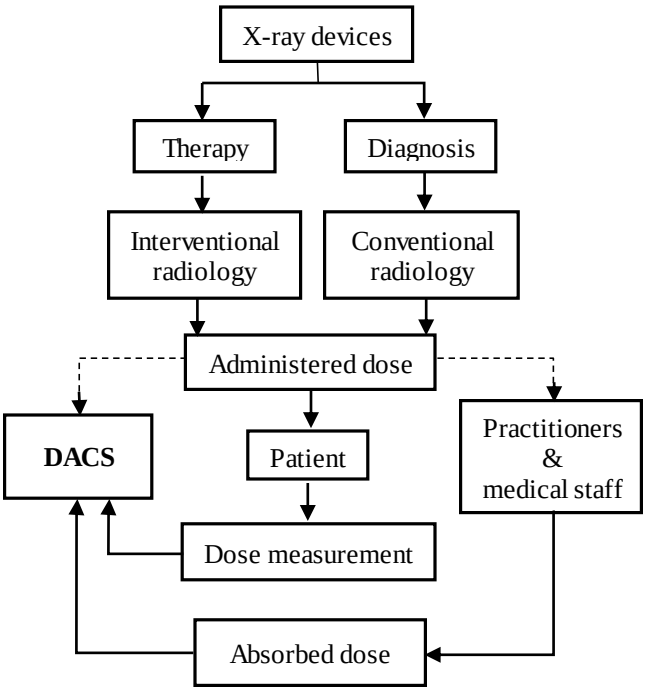


Fig. 2. General diagram and application area.

In the field of X-ray based medical imaging, two primary types of equipment are utilized: those employed in conventional radiology for diagnostic purposes and those utilized in interventional radiology for therapeutic interventions. Prior to conducting an examination, technicians program the dose of X-

rays to be administered to the patient based on various parameters such as the type of exam and patient anatomical characteristics. Following exposure, the patient absorbs a dose of X-rays, which is measured during image detection using radiographic or fluoroscopic imaging. Simultaneously, medical personnel present during the examination are also exposed to X-ray doses, which vary depending on their proximity to the radiation source and the protective measures employed, suchas lead aprons. All these doses, both those administered to patients and those absorbed by medical personnel, are then transmitted to the Dose Archiving and Communication System for storage and analysis. Additionally, this system conducts equipment checks based on these doses and the utilized examination protocols, ensuring adherence to safety standards and the quality of radiological exams performed.

To understand the framework and components of our Dose and Archiving Communication System, we present a modeling overview (Fig. 3) including:

- Data Acquisition and Integration: This component is responsible for collecting dose data from x-ray devices, integrating it into a standardized format, and transmitting it to the central storage.
- Central Dose Data Storage and Management: This component stores and manages the dose data in a centralized database, facilitating efficient archiving, retrieval, and analysis.
- User Interface: The user interface provides a dashboard for visualizing dose data, and reporting functionalities.
- Security and Compliance: This component ensures data security through encryption, access control mechanisms, and maintains compliance with regulatory requirements.

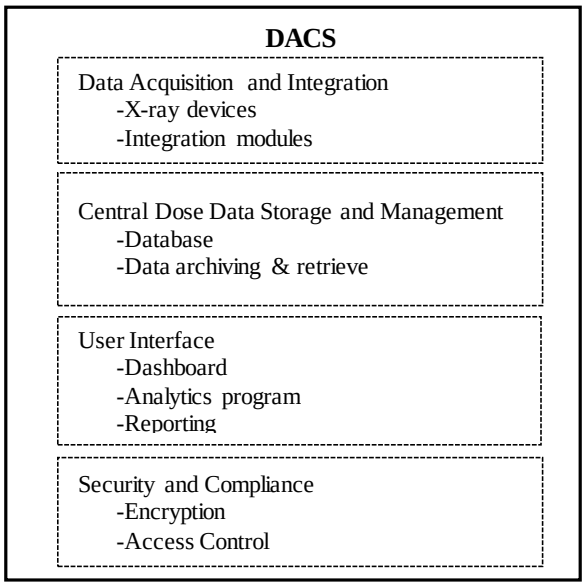


Fig. 3. Overview of DACS architecture.

Designing an effective architecture for a Dose Archiving and Communication System requires careful consideration to ensure both flexibility and security. An optimal approach involves the integration of modular components that can adapt to evolving requirements while maintaining robust security measures. Technologies such as cloud-based storage solutions offer scalability and accessibility, allowing for seamless data archiving and retrieval [23]. Additionally, blockchain technology [24-25] provides a secure and transparent framework for data communication and authentication, ensuring the integrity and traceability of dose data. Implementation of industry standard encryption protocols further fortifies data security, safeguarding sensitive information against unauthorized access. By leveraging these technologies synergistically, a DACS architecture can achieve the delicate balance between flexibility and security, laying the foundation for efficient dose management and analysis.

In the diverse landscape of medical imaging, a range of Dose Archiving and Communication Systems are available, each designed to address specific needs and requirements within healthcare facilities. Integrated Picture Archiving and Communication Systems (PACS) with dose monitoring capabilities offer a consolidated approach, allowing for the storage and management of both medical images and dose data within a single platform. While these systems provide basic dose tracking functionalities, they may lack the advanced analytics and reporting features found in standalone solutions. Standalone DACS, on the other hand, are dedicated platforms focused solely on dose monitoring and management. These systems offer comprehensive dose tracking, analysis, and reporting capabilities, enabling healthcare providers to perform in-depth assessments and trend analyses. Cloud based DACS leverage the scalability and accessibility of cloud computing technology, providing healthcare facilities with flexible and remotely accessible dose management solutions. By centralizing dose data in the cloud, these systems enable seamless collaboration and data sharing across multiple locations. Additionally, some medical imaging equipment vendors offer proprietary solutions tailored to their specific imaging systems. These vendor specific DACS are seamlessly integrated into the equipment's software and workflow, providing healthcare providers with specialized features optimized for their imaging modalities. However, they may lack the flexibility and interoperability of standalone or cloud based DACS. Healthcare providers must carefully evaluate the features, scalability, interoperability, and cost effectiveness of each solution to determine the most suitable option for their organization, considering factors such as existing infrastructure, budget constraints, and regulatory compliance requirements.

Our dose management system is a specialized platform designed to effectively manage, archive, and communicate dose data from medical imaging procedures and it provide several features as below (Fig. 4):

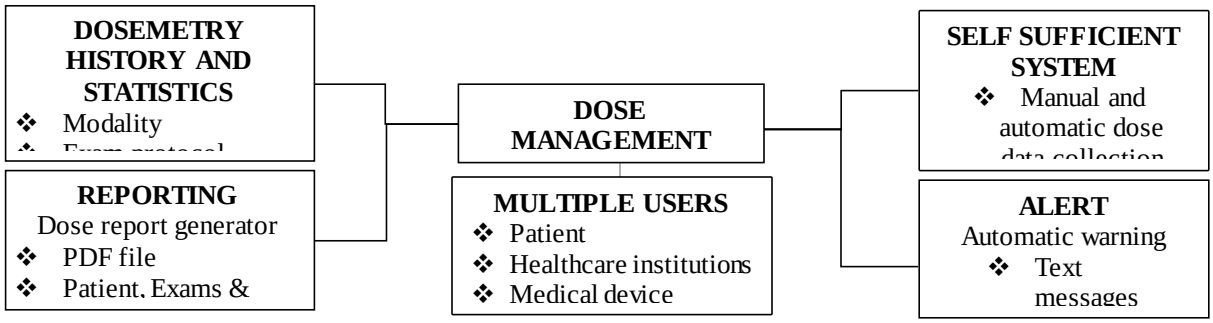


Fig. 4. Features of the DACS system.

A. Dosimetry History and Statistics

DACS facilitates the integration and comprehensive analysis of dose data across various imaging modalities. This section outlines its capabilities in documenting exam protocols and providing customizable statistical analysis over different time periods.

1) Modality: DACS seamlessly integrates with various imaging modalities, capturing dose data from x-ray, CT scans, and more.

2) Exam protocol: Detailed documentation of exam protocols used during imaging procedures, ensuring comprehensive data capture.

3) Period: Provides a comprehensive dose history and statistical analysis over customizable time periods, facilitating trend identification and dose optimization efforts.

B. Self Sufficient System

This section highlights the DACS features that ensure system autonomy and flexibility in dose data management.

1) Manual and automatic dose data collection: DACS supports both manual entry and automated collection of dose data, ensuring accuracy and flexibility.

2) Remote access: Enables secure remote access to dose data and system functionalities, allowing users to monitor and manage dose information from anywhere, at any time.

C. Multiple Users

This part describes how DACS accommodates various users, from patients to healthcare institutions and medical device suppliers, enhancing transparency, collaboration, and centralized dose management.

1) Patient access: Empowers patients to access their own dose history and reports, promoting transparency and patient engagement in their healthcare journey.

2) Healthcare institutions: Facilitates centralized dose management across healthcare facilities, allowing institutions to monitor dose levels, compliance, and performance across departments and modalities.

3) Medical device suppliers: Provides access to dose data for medical device suppliers, fostering collaboration and optimization of imaging equipment performance.

D. Reporting

This segment outlines DACS's capabilities in generating detailed and customizable dose reports for both healthcare providers and patients.

1) Dose report generator: Generates detailed and customizable reports in PDF format, incorporating patient information, exam details, and dose data for comprehensive documentation.

2) Patient centric reports: Delivers personalized dose reports to patients, enhancing communication and understanding of dose exposure and associated risks.

E. Alert System

This category details DACS's alert features, ensuring timely intervention and communication through automatic alerts and notifications.

1) Automatic alerts: DACS triggers automatic alerts for dose threshold breaches or abnormal dose patterns, ensuring timely intervention and dose optimization.

2) Text messages and email notifications: Alerts are disseminated via text messages or email to designated recipients, facilitating prompt action and communication among relevant stakeholders.

The Dose Archiving and Communication System streamlines dose management processes, enhances patient safety, and promotes collaboration among healthcare providers and stakeholders. With its advanced features and user-friendly interface, our solution empowers healthcare organizations to effectively manage dose data, optimize imaging practices, and deliver high-quality patient care.

V. Results and Discussions

The dashboard of our Dose Archiving and Communication System is meticulously designed to offer the users a comprehensive snapshot of key statistics and metrics crucial for understanding the system's performance and ensuring patient safety. Upon accessing the system through secure login protocols, users are greeted with the main dashboard, where they can quickly grasp essential information. This includes the total number of patients, providing an overview of the volume of dosimetric data within the system, and the average dose per patient, offering insight into the typical radiation exposure experienced by individuals undergoing procedures. Additionally, a dose distribution chart visually represents the

spread of doses across different ranges, aiding in identifying any anomalies or trends.

The procedure statistics section presents users with valuable insights into the total number of procedures performed, segmented by procedure type, along with the average doses administered for each procedure category. A timeline graph depicting procedure volumes over time facilitates the tracking of procedural trends, enabling proactive decision-making and resource allocation. Real-time monitoring capabilities allow users to stay updated on current active procedures and receive prompt alert notifications for doses that exceed predefined safety thresholds, ensuring timely intervention and adherence to safety protocols.

The dashboard's analytical section offers in-depth analysis and historical perspectives, with features such as dose trends over time and patient specific dose histories presented inauser-friendly format. Compliance reports provide transparency regarding adherence to safety standards, while benchmarking facilitates performance comparison against industry norms, fostering a culture of continuous improvement and excellence. Finally, the optimization insights section leverages data driven recommendations derived from statistical analysis to enhance dose safety and efficacy, empowering users to make informed decisions aimed at improving patient outcomes and optimizing resource utilization. Through its intuitive interface and comprehensive feature set, our DACS dashboard facilitates seamless access to critical dosimetric data, empowering stakeholders with the insights needed to drive informed decision-making and deliver high-quality patient care.

The patient dose summary collected from different healthcare institutions provides a comprehensive overview of radiation exposure among of 350 patients over a year (Fig. 5). On average, each patient received a dose of 4.5 Gray (Gy), a unit measuring the absorbed radiation energy. The distribution of doses reveals that 40% of the patients fell into the low dose category, receiving between 0 and 2 Gy. The majority (about 50%) received a medium dose ranging from 2 to 5 Gy, while 10% were exposed to high doses exceeding 5 Gy. This summary highlights the variation in radiation exposure levels among the patients during this period and underscores the need for careful monitoring and management of radiation doses in clinical settings.

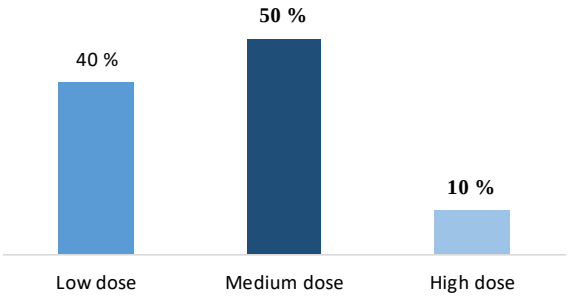


Fig. 5. Patient dose summary and distribution.

The Table I provides a detailed breakdown of the dose absorbed by ten patients across various medical procedures. Each row corresponds to a patient, while the columns represent different procedures undertaken by them. The values within the table denote the dose of radiation absorbed by each patient during the respective procedures. This comprehensive overview allows for a comparative analysis of radiation exposure among patients undergoing different medical interventions. By examining the data within the table, trends in radiation dosage across procedures and variations among patients become apparent. Such insights are crucial for ensuring the optimization of radiation doses, minimizing unnecessary exposure, and enhancing patient safety in clinical settings. Furthermore, this information facilitates informed decision-making by healthcare professionals regarding the appropriate dosage levels for specific procedures based on individual patient characteristics and medical requirements.

To calculate the risk associated with X-ray doses, the first step is to determine the effective dose, which is measured in Sieverts (Sv). The effective dose accounts for the type of radiation and the varying sensitivity of different tissues and organs. This measure provides an overall risk estimate from radiation exposure [26]. The effective dose is calculated using the formula:

$$E=\sum(Dt\cdot Wr\cdot Wt) \quad (1)$$

TABLE I. DOSE ABSORBED FOR DIFFERENT PROCEDURES ACROSS 10 PATIENTS

Procedure / Patient	Chest CT	Abdomen CT	Brain CT	Cardiac CT	Lumbar Standard	Extremity Standard	Chest Standard	Mammography	Dental
Patient 1	6.1	*	*	*	1.4	*	0.2	0.21	*
Patient 2	*	5.3	1.6	*	*	*	*	*	0.25
Patient 3	6	5.2	*	*	*	0.2	0.1	0.20	
Patient 4	*	*	1.5	1.6	*	*	0.2	*	0.23
Patient 5	6.2	*	*	1.5	*	0.1	*	*	0.3
Patient 6	*	5.4	1.4	*	1.3	*	*	0.22	*
Patient 7	*	*	*	*	1.2	*	0.2	*	0.31
Patient 8	*	5	*	1.4	*	0.3	0.1	*	*
Patient 9	6	4.8	*	*	*	*	*	*	0.2
Patient 10	7	*	1.8	*	1	*	0.12	0.3	*

Where:

E is the effective dose. Dt is the absorbed dose in tissue *t*, Wr is the radiation weighting factor (depends on the type of radiation, usually 1 for X-rays) and Wt is the tissue weighting factor (varies for different tissues).

Tissue weighting factors (*Wt*) are used to account for the varying sensitivity of different tissues to radiation. These factors, provided by the International Commission on Radiological Protection (ICRP) [27], help in calculating the contribution of each tissue or organ to the effective dose. By applying these factors (Table II), the effective dose calculation becomes more accurate, reflecting the differential risk of radiation exposure to various tissues.

TABLE II. TISSUE WEIGHTING FACTORS

Tissue or Organ	ICRP 60	ICRP 103
Gonads	0.20	0.08
Red bone marrow	0.12	0.12
Lung	0.12	0.12
Colon	0.12	0.12
Stomach	0.12	0.12
Breast	0.05	0.12
Bladder	0.05	0.04
Liver	0.05	0.04
Esophagus	0.05	0.04
Thyroid	0.05	0.04
Skin	0.01	0.01
Bone surface	0.01	0.01
Brain		0.01
Salivary glands		0.01

A comprehensive empirical investigation was conducted within healthcare institutions in Morocco to assess the implementation of radiation protection measures in both diagnostic and therapeutic procedures. The study's objectives encompassed evaluating adherence to the three fundamental principles of radiation protection and examining infrastructure, notably computer network connectivity, across services employing ionizing radiation sources. To accomplish this, a questionnaire was administered to over 1000 healthcare professionals, comprising radiology technicians, radiologists, surgeons, biomedical technicians, and biomedical engineers, representing diverse healthcare facilities across different regions of the kingdom. Key areas of inquiry included training in radiation protection, utilization of personal protective equipment, availability of individual dosimeters, effectiveness of control and monitoring systems, and comprehensiveness of dosimetry reports. Survey results encompass a broad spectrum of hospital structures spanning various regions in Morocco, with contributions from multiple healthcare institutions as delineated in Table III.

Professionals from various hospital services participated in the survey, with representation as follows: the biomedical department accounted for 13% (Fig. 6), where technicians and engineers, serving as both internal and external interfaces, manage device-user interaction and liaise with suppliers. The radiology department accounted for 63% (Fig. 6), including radiology technicians and managers in the medical imaging department. The oncology department represented 7% (Fig. 6), comprised of radiation therapists and radiologists, while the surgery department contributed 17% (Fig. 6), involving traumatologists, neurologists, and other medical personnel.

TABLE III. HEALTHCARE INSTITUTIONS PARTICIPATING IN THE STUDY

Healthcare institutions	Number
University Hospital Center	5
Regional Hospital Center	14
Provincial Hospital Center	60
Military Hospital	4
Clinic	10
Radiology Center	15

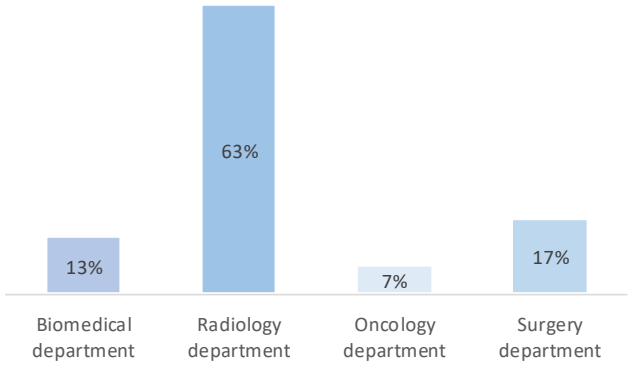


Fig. 6. Participating departments.

In the array of healthcare institutions, a significant portion of healthcare professionals, totaling 58%, lack training in radiation protection and are unaware of the relevant standards (Fig. 7). This gap primarily stems from inadequacies in the academic curriculum of universities and higher education institutions, which fail to sufficiently cover radiation protection. Additionally, there is a dearth of coverage on this topic in the continuing education courses pursued by these professionals, exacerbating the issue. The study reveals that only 42% have received training on radiation protection and its critical implications for both patient well-being and healthcare practitioners (Fig. 7).

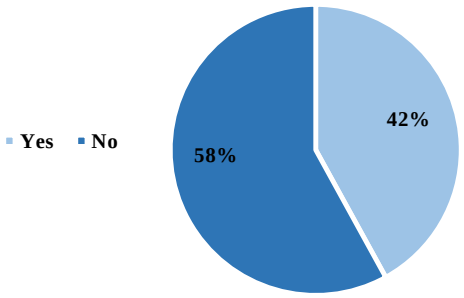


Fig. 7. Radiation safety-trained professionals.

Adhering to dose limits ensures that the radiation risk from all controllable sources of ionizing radiation remains sufficiently low to pose no concern to individuals. The emphasis lies not in solely controlling the radiological risk from one specific source, but in restricting individual risk arising from exposure to all sources. This underscores the necessity of utilizing both individual protective gear such as lead skirts, lead gowns, thyroid covers, X-ray gloves, and X-ray glasses, as well as collective protective measures like lead walls and X-ray screens [28-35]. According to the findings, 26% of practitioners utilize lead gowns (Fig. 8), 10% use both thyroid covers and lead gowns (Fig. 3), while a notable 64% do not employ any protective measures (Fig. 8).

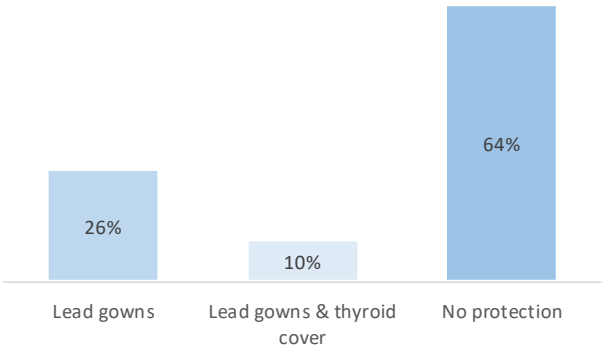


Fig. 8. Individual protection items.

Regulations stipulate the monitoring of external exposure to ionizing radiation in areas where individuals may encounter such risks, necessitating the use of individual dosimeters. This practice evaluates the radiation dose received by each person during their professional activities, ensuring ongoing safety measures and facilitating the early detection of any anomalies for timely intervention. It's essential to clarify that while this monitoring doesn't provide direct protection, its purpose is to maintain safety conditions. The associated risk with a specific radiation dose is determined by the likelihood of an individual experiencing particular radiation-induced effects upon exposure.

Within the purview of the National Center for Radiation Protection (NCRP), the dosimetry-monitoring department assumes responsibility for the provision and management of individual dosimeters used in medical settings, both private and public. Under the authority of the Ministry of Health, the NCRP oversees the monitoring of external exposure for individuals engaged in tasks involving ionizing radiation. Typically, this monitoring is reserved for medical and paramedical personnel operating within controlled areas, particularly those classified as category A, who are anticipated to be occupationally exposed to an effective dose exceeding 6 millisieverts over a 12-month period and directly involved in radiation-related tasks.

The primary objectives of individual dosimetry monitoring encompass several key aspects:

1) Quantification of ionizing radiation levels: The foremost purpose is to measure and quantify the doses of ionizing radiation accumulated by an individual during their occupational activities. This data provides crucial information regarding the extent of radiation exposure experienced by workers.

2) Empowerment of occupational physicians: Individual dosimetry monitoring empowers occupational physicians to take necessary actions based on the radiation exposure levels detected. This may include implementing supplementary medical examinations or temporarily relocating individuals from high-risk areas to mitigate potential health risks associated with radiation exposure.

3) Effective monitoring and regulation of working conditions: By utilizing individual dosimetry data, employers can effectively monitor and regulate working conditions to ensure compliance with radiation safety standards and guidelines. This proactive approach aids in maintaining a safe and healthy work environment for employees exposed to ionizing radiation.

For all personnel exposed to ionizing radiation falling within categories A and B, the utilization of a passive dosimeter is obligatory. This dosimeter must be worn either at chest level or, if not feasible, on the belt beneath protective clothing (such as a lead gown) for the entire duration of work. At the end of the workday, the dosimeter is securely stored in a designated area, ensuring distance from any sources of radiation, heat, or humidity. Subsequently, it is dispatched to the relevant agency responsible for passive dosimetry either monthly or, exclusively for category B workers, on a quarterly basis. The measured results are quantified in millisieverts (mSv), and the outcome report is conveyed in an individualized and nominative manner to the occupational physician, who then communicates them to the healthcare professionals. The survey's results pertaining to this protocol are depicted in Fig. 9. The statistics indicating that 54% of healthcare professionals use passive dosimeters while 46% do not underscore the varying levels of adherence to radiation safety practices within healthcare settings. Those utilizing passive dosimeters benefit from continuous monitoring of radiation exposure, enabling them to proactively manage risks and adhere to regulatory requirements such as the ALARA principle. In contrast, those not using dosimeters may potentially overlook the cumulative effects of radiation exposure, highlighting a need for enhanced awareness and adherence to safety protocols across the healthcare field.

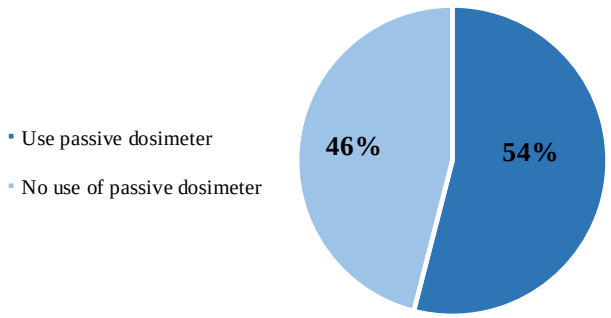


Fig. 9. Monitoring radiation exposure.

When X-rays interact with matter, they induce various effects primarily because of their capacity to ionize atoms and molecules [36-38]. These effects include ionization, where X-rays possess adequate energy to expel electrons from atoms, resulting in the formation of positively charged ions. This ionization process can disrupt chemical bonds, thereby impacting the structure and functionality of molecules. Furthermore, the ionizing nature of X-rays is central to their utility in medical imaging, but it can also inflict cellular damage by disrupting DNA strands within cells. This mechanism underlies their application in radiation therapy, where X-rays are employed to target and eradicate cancer cells. Additionally, high doses of X-rays can lead to radiation burns on the skin and underlying tissues, necessitating careful control of radiation doses in medical environments. Moreover, prolonged or repeated exposure to X-rays, particularly at elevated doses, heightens the risk of cancer, emphasizing the critical importance of radiation protection measures in both medical and industrial settings. Furthermore, X-rays can induce fluorescence in certain materials, causing them to emit light of characteristic wavelengths, a property exploited in X-ray fluorescence spectroscopy for elemental analysis. Finally, X-rays are extensively utilized in medical diagnostics, including diagnostic radiography, computed tomography (CT) scans, and fluoroscopy, facilitating the visualization of internal bodily structures for the detection and diagnosis of various medical conditions. It is crucial to acknowledge that while X-rays offer significant benefits in medical diagnostics and treatment, adherence to proper precautions and safety protocols is imperative to mitigate potential risks associated with their ionizing nature.

The importance of network infrastructure within radiology departments cannot be overstated, as it is fundamental for improving radiation protection and dose control measures. A resilient network infrastructure is instrumental in enabling the seamless integration of cutting-edge technologies and tools, thereby enhancing the overall safety and efficiency of radiological practices. However, the study findings (Fig. 10) reveal a concerning statistic: only 11% of diverse radiology department spaces have opted to incorporate a computer network, leaving a substantial 89% without such integration. This deficiency has emerged as a significant barrier to the adoption of new technologies that rely on a computer network for essential dose monitoring functionalities.

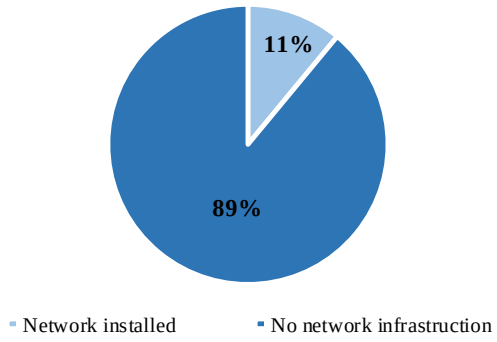


Fig. 10. Network infrastructure implementation.

The effective management of radiation exposure is paramount for both X-ray practitioners and patients alike. Implementing key practices such as utilizing personal dosimetry, maintaining safe distances from X-ray sources, proper collimation, emphasizing accurate positioning and technique, using shielding equipment, and adjusting imaging protocols based on patient parameters are essential steps in minimizing radiation doses while ensuring diagnostic quality. By adhering to these recommendations, X-ray practitioners can safeguard themselves from excessive radiation exposure and contribute to the overarching goal of optimizing patient care. Moreover, ongoing training, open communication, and stringent quality assurance protocols are vital components for maintaining a safe and effective radiological practice. By prioritizing radiation protection measures, X-ray practitioners play a crucial role in promoting both occupational safety and patient well-being within radiology departments.

VI. DACS Implementation Approach

Implementing a Dose Archiving and Communication System in Moroccan hospitals involves several key steps to ensure accurate radiation dose tracking, storage, and communication. This system will enhance patient safety, optimize radiology practices, and comply with regulatory standards. To initiate the implementation, it is crucial to conduct a thorough needs assessment and planning phase. Engage key stakeholders, including radiologists, medical physicists, IT specialists, hospital administrators, and regulatory authorities, to gather comprehensive requirements and understand existing workflows and challenges. Organize meetings and workshops to ensure all perspectives are considered. Evaluate the current radiology and IT infrastructure to identify gaps in radiation dose tracking, archiving, and communication processes. Once the needs assessment is complete, define the technical and functional requirements for the DACS. These requirements should include integration capabilities with existing Radiology Information Systems (RIS) and Picture Archiving and Communication Systems (PACS).

Developing the necessary infrastructure is a key step in implementing DACS. Upgrade or procure the required hardware, such as servers, storage systems, and network infrastructure, to support the new system. Ensure high-speed and secure network connectivity within and between hospitals to facilitate seamless data transfer. Install the DACS software on designated servers and configure it for integration with existing RIS and PACS systems. Proper infrastructure development ensures the system's reliability and performance, supporting efficient radiation dose management. With the infrastructure in place, focus on migrating existing dose records from legacy systems to the new DACS. This process involves validating the accuracy and completeness of the migrated data to ensure no critical information is lost. Integrate the DACS with RIS and PACS for automatic dose data capture and sharing, ensuring interoperability with other hospital systems. Successful data migration and integration are essential for maintaining continuity and accuracy in radiation dose tracking and management.

Implementing a new system requires comprehensive training and effective change management strategies. Develop

training programs for radiologists, technologists, IT staff, and administrative personnel, including hands-on training sessions and the provision of user manuals and support materials. Address potential resistance and concerns through regular communication, highlighting the benefits of the new system. By facilitating a smooth transition and ensuring all users are proficient with the new system, hospitals can maximize the benefits of the DACS. Before full-scale deployment, conduct a pilot implementation in a few selected hospitals to test the DACS in a real-world setting. Monitor system performance, user feedback, and data accuracy during this pilot phase. Validate the system's functionality, reliability, and compliance with regulatory standards, making necessary adjustments based on the pilot testing results. This step helps identify and resolve any issues, ensuring a smoother rollout across all hospitals.

Following a successful pilot, proceed with the full-scale deployment of the DACS to all targeted hospitals across Morocco. Provide continuous support and troubleshooting during the initial deployment phase to address any challenges promptly. Establish a monitoring system to track the performance and usage of the DACS, ensuring it meets operational needs and regulatory requirements. Regular monitoring and maintenance are vital to ensure the system operates optimally and continues to benefit the hospital's radiology practices. Post-deployment, regularly evaluate the system's impact on radiation dose management, patient safety, and operational efficiency. Collect and analyze feedback from users to identify areas for improvement. Implement an ongoing process for system enhancements based on evolving needs and technological advancements. Stay updated with international best practices and regulatory changes to ensure the system remains compliant and effective. Continuous evaluation and improvement will help maintain the DACS's relevance and utility in enhancing patient care and radiology workflows.

VII. Conclusion

In Morocco, legislation mandates regular dosimetry monitoring for medical personnel exposed to ionizing radiation, yet the current implementation of such measures remains inadequate, with a significant portion of healthcare professionals lacking training in radiation protection. To address this gap, comprehensive training and education programs must be prioritized, encompassing all personnel working with ionizing radiation. Concurrently, consistent utilization and maintenance of Personal Protective Equipment (PPE) are essential, alongside optimization of imaging procedures to minimize radiation doses while maintaining diagnostic quality. Furthermore, fostering a radiation safety culture within healthcare institutions is crucial, promoting open communication and awareness about safety concerns among healthcare professionals and patients alike.

Implementing a robust dose archiving and communication system offers numerous benefits in enhancing radiation protection efforts. Such a system enables systematic storage and retrieval of radiation dose data, facilitating comprehensive dose monitoring and timely identification of potential overexposure incidents. Additionally, it supports informed decision-making regarding patient care and radiation protection measures. Despite potential obstacles such as technological limitations and concerns about data security and privacy, the benefits of a dose archiving and communication system justify continued efforts to promote its widespread adoption, ultimately ensuring the safety and well-being of both patients and healthcare professionals within radiology departments. However, several obstacles may hinder the successful implementation of a dose archiving and communication system. These may include technological limitations, such as compatibility issues with existing healthcare infrastructure and electronic health record systems. Additionally, concerns regarding data security and patient privacy may pose challenges in establishing trust and compliance with regulatory requirements. Furthermore, resource constraints, including financial limitations and staffing shortages, may impact the feasibility of implementing and maintaining such a system on a large scale. Despite these challenges, the benefits of a dose archiving and communication system in enhancing radiation protection efforts and ensuring the safety of both patients and healthcare professionals justify continued efforts to overcome obstacles and promote its widespread adoption in healthcare settings.

References

[1] World Health Organization. https://www.who.int.

[2] Garba, I., Penelope, E. H., Davidson, F., & Ismail, A. (2024). Prospective dose monitoring using a manual dose management system: experience in brain computed tomography from a tertiary hospital in Nigeria. Radiation Protection Dosimetry, 200(7), 648–658. https://doi.org/10.1093/RPD/NCAE094.

[3] J. L. Heron, R. Padovani, l. Smith, R. Czarwinski, "Radiation protection of medical staff", Volume 76, Issue 1, pp 20-23, October 2010.

[4] Polizzi, M., Valerie, K., & Kim, S. (2024). Commissioning and Assessment of Radiation Field and Dose Inhomogeneity for a dual x-ray tube cabinet irradiator: to ensure accurate dosimetry in radiation biology experiments. Advances in Radiation Oncology, 101486. https://doi.org/10.1016/J.ADRO.2024.101486.

[5] Silva, M. F., Caixeta, A. L. O., Souza, S. P., Tavares, O. J., Costa, P. R., Santos, W. S., P. Neves, L., & Perini, A. P. (2024). A dosimetric study of occupational exposure during computed tomography procedures. Radiation Physics and Chemistry, 218. https://doi.org/10.1016/J.RADPHYSCHEM.2024.111564.

[6] S. Kawauchi, K. Chida, T. Moritake, Y. Hamada, W. Tsuruta, "Radioprotection of eye lens using protective material in neuro cone-beam computed tomography: Estimation of dose reduction rate and image quality", Volume 82, pp 192-199, February 2021.

[7] Choi, Y., & Lee, Y. (2022). Implementation of a dose archiving and communication system (DACS) for radiation dose management in a large healthcare system. Radiation Oncology Journal, 40(1), 58-64. https://doi.org/10.3857/roj.2021.00913.

[8] Liu, C., et al. (2021). Dose archiving and communication system (DACS) in radiation oncology: A comprehensive review and future perspectives. Medical Physics, 48(10), e888-e900. https://doi.org/10.1002/mp.15193.

[9] Faggioni, L., et al. (2019). Implementation and evaluation of a dose archiving and communication system (DACS) for pediatric cardiac catheterization procedures. European Radiology, 29(10), 5714-5722. https://doi.org/10.1007/s00330-019-06092-2.

[10] Rehani, M. M., & Ortiz-Lopez, P. (2018). Radiation dose management systems: Needs and opportunities. Radiation Protection Dosimetry, 182(1), 98-104. https://doi.org/10.1093/rpd/ncy061.

[11] Wang, Y., & Thompson, L. (2023). Enhancing radiation dose management with advanced DACS technology. Radiology Management, 45(2), 87-94.

[12] Martin, K., & Rivera, J. (2022). Impact of dose archiving and communication systems on clinical workflow and patient care. Journal of Digital Imaging, 35(1), 123-130. doi:10.1007/s10278-021-00433-9.

[13] D'Alessio, A., Matheoud, R., Cannillo, B., Guzzardi, G., Galbani, F., Galbiati, A., Spinetta, M., Stanca, C., Tettoni, S. M., Carriero, A., & Brambilla, M. (2023). Evaluation of operator eye exposure and eye protective devices in interventional radiology: Results on clinical staff and phantom. Physica Medica, 110. https://doi.org/10.1016/J.EJMP.2023.102603.

[14] Healthdirect. https://www.healthdirect.gov.au/x-rays.

[15] International Commission on Radiological Protection. https://www.icrp.org/index.asp.

[16] Law No. 142-12 on nuclear and radiological safety and security, Morocco. https://www.mem.gov.ma/.

[17] Ministry of Health and Social Protection - Morocco. https://www.sante.gov.ma/.

[18] Decree No. 2-20-131 on the authorization and declaration regime for activities, installations, and associated sources of ionizing radiation in category II. https://www.mem.gov.ma/.

[19] Decree No. 2-97-30 of 25 Joumada II 1418 (October 28,1997) taken for the application of Law No. 005-71 of 21 Chaabane 1391 (October 12, 1971) relating to the protection against radiations.

[20] Langlotz, C. P., Allen, B., Erickson, B. J., et al. (2022). The future of radiology: Augmented workflow, artificial intelligence, and the networked hospital. Radiology.

[21] Patel, V., Kumar, S., & Gupta, R. (2023). Implementing advanced network infrastructures to support tele-radiology services. Telemedicine and e-Health, 29(2), 189-198.

[22] Chen, Y., Wang, X., & Sun, Y. (2023). Network security in radiology: Protecting patient data in the digital age. Insights into Imaging, 14, 32.

[23] Huang, J., Smith, S. E., & Liu, B. (2021). The role of cloud-based solutions in modern radiology departments. Journal of Digital Imaging, 34(5), 1234-1245.

[24] Kuo, T. T., Kim, H. E., & Ohno-Machado, L. (2019). Applications of blockchain technology in medicine and healthcare: Challenges and future perspectives. Computational and Structural Biotechnology Journal, 17, 164-170.

[25] Zhang, P., White, J., Schmidt, D. C., et al. (2020). Blockchain technology in healthcare: A comprehensive review and directions for future research. IEEE Transactions on Engineering Management.

[26] Sherer, M. A., Visconti, P. J., & Ritenour, E. R. (2017). Radiation protection in medical radiography. Mosby.

[27] International Commission on Radiological Protection (ICRP). (2007). The 2007 Recommendations of the International Commission on Radiological Protection. Annals of the ICRP, 37(2-4), 1-332.

[28] D. Sirinelli, H. Ducou, P. Roch, J.F. Chateil, "Principles and implementation of radiation protection". Bordeaux University.

[29] "Follow-up of patient dosimetry and evaluation of practices"-Institute for Radiation Protection and Nuclear Safety, https://c2isante.fr/.

[30] A. J. Gonzalez, "Basic standards of radiation protection", General direction for nuclear safety and radiation protection Paris.

[31] J. -M. Bodry, "External dosimetry of ionizing radiation from the national reference to users in radiotherapy and radiation protection" Health and safety at work – INRS, https://www.inrs.fr/.

[32] Radioactive hazard and radiation protection - 2nd edition – January 2018, https://www.jstor.org/.

[33] Radiation_Protection_Manual_v June 2018, https://www.jstor.org/.

[34] Radiation Protection and Safety of Radiation sources-IAEA, https://www.iaea.org/fr.

[35] Training in radiation protection for workers- F. Durandn Physics Unit, https://www.who.int/.

[36] Rahman, T., Khandakar, A., Qiblawey, Y., Tahir, A., Kiranyaz, S., Kashem, S. B. A., ... & Chowdhury, M. E. (2021). Exploring the effect of image enhancement techniques on COVID-19 detection using chest X-ray images. Computers in biology and medicine, 132, 104319.

[37] Bliznyuk, U., Avdyukhina, V., Borshchegovskaya, P., Bolotnik, T., Ipatova, V., Nikitina, Z., ... & Yurov, D. (2022). Effect of electron and X-ray irradiation on microbiological and chemical parameters of chilled turkey. Scientific reports, 12(1), 750.

[38] Liu, H., Zhao, J., Xue, Y., Zhang, J., Bai, H., Pan, S., ... & Voelcker, N. H. (2023). X-Ray-Induced Drug Release for Cancer Therapy. Angewandte Chemie International Edition, 62(39), e202306100.

UTAUT Model for Digital Mental Health Interventions: Factors Influencing User Adoption

Mohammed Alojail
Management Information System Department, College of Business Administration, King Saud University, Riyadh, Saudi Arabia

***Abstract*—The impact of digital revolution on mental health therapies is examined in this research. As explained in the paper, the delivery of mental healthcare is being revolutionized by digital transformation, which is providing creative answers to the problems associated with mental health illnesses. But knowing and managing user approval is crucial to the effective integration of digital transformation approaches into mental health therapies. To investigate user's acceptance regarding digital transformation in Mental Health therapies, this study outlines a modeling-based method based on a well-established Unified Theory of acceptability and Use of Technology theory, abbreviated as UTAUT. This study delves into the base constructs of Expected Performance, Expected Effort, Social Influence, Conditions facilitating the use of proposed solution, Hedonic Motivation, and Value for Money, utilizing the UTAUT model as a framework. By employing Structural Equation Modelling (SEM) in a thorough study, the goal of this research is to identify statistical correlations that impact user acceptance dynamics. To offer context-specific insights, this article also delves into digital mental health solutions, including teletherapy platforms, mood monitoring smartphone applications, and virtual reality-based exposure therapy. This study enhances accessibility, engagement, and results for people seeking mental health care by providing a deeper knowledge of user acceptability, which aids in the creation and roll-out of digital mental health solutions.**

Keywords*—Digital transformation; mental health interventions; UTAUT model; TAM; SmartPLS; user acceptance; hypothesis testing

I. Introduction

The integration of digital technologies in healthcare represents a transformative shift, offering profound opportunities for enhanced patient care, improved clinical outcomes, and optimized healthcare delivery processes [1][2]. Within this realm of transformation, mental health interventions are at the forefront, leveraging various digital solutions to address the multifaceted challenges associated with mental health disorders [4]. Digital transformation in mental healthcare includes innovations such as teletherapy platforms, mobile applications for mood monitoring, virtual reality-based exposure therapy, and online support communities. These advancements hold significant promise in extending the accessibility of mental health services, reducing associated stigma, and empowering individuals to actively manage their mental health [4]. A central component of the success of digital mental health interventions is user acceptance, which reflects the degree to which individuals perceive a technology as useful, user-friendly, and worth adopting [3]. Navigating user acceptance dynamics is crucial for the effective implementation and uptake of these interventions across diverse user populations. Unified Theory of Acceptance and Use of Technology (UTAUT) model offers a robust framework for testing factors which effect user acceptance within the context of digital transformation. The UTAUT model, proposed by Venkatesh et al. [6], integrates key constructs from various other new technology acceptance theories, like Technology Acceptance Model (TAM), Theory of Reasoned Action (TRA), and Innovation Diffusion Theory (IDT) [3]. According to this model, user acceptance is influenced by core constructs such as Performance expectation, Expected Effort, Influence of Social Life, Conditions facilitating the use of proposed solution, Hedonic Motivation, and Value for Money. These constructs encompass the usefulness through perception, ease of use, social norms, organizational support, and cost-effectiveness linked to adopting a technology.

While the UTAUT model has seen widespread application in various contexts, its utility within the domain of digital mental health interventions remains underexplored. This paper is written with an objective to bridge this gap by employing a UTAUT modeling-based approach to investigate user acceptance dynamics specific to digital transformation in mental healthcare. Utilizing Structural Equation Modeling (SEM), this study examines the connection between UTAUT constructs and user acceptance. The insights derived aim to inform the design, implementation, and evaluation of effective digital mental health strategies, ultimately advancing the delivery of evidence-based care to those in need.

A. Motivations and Contributions

Since mental health is a major subject of the current world. It is also seen that the world is moving towards the involvement of more digital transformation driven interventions in various facets of life. Therefore, it was felt prudent that, while the paper is developed diligently towards the development of the solutions – how well these solutions will be accepted by the user group. It was done, because ultimately, it is the users who decide the success or failure of the specific solution. This paper makes an endeavour to analyze the responses from a varied set of people to understand the following parameters:

1) How eager are people to employ a digital transformation driven mental health care intervention in their lives?

2) Is cost a driving factor in this?

3) Will people use the functionality under the peer pressure of friends, family or society?

4) How significant is the user friendliness of the solution to drive its usage?

The paper is divided into the following sections. Section I Lists the Introduction and Section II presents the Literature review followed by Proposed Research Model in Section III.

Section IV describes the Model Validation and Section V Presents the Results and last Section Concludes.

II. Literature Review

At the very onset of this paper, the paper reviews the previous works done on the premise of digital transformation implementation, then it delves into digital transformation acceptance and then into implementation of digital transformation methodologies in mental health scenario.

Digital transformation refers to the amalgamation of digital technologies, strategies, and capabilities into various aspects of an organization or industry in order to fundamentally alter its operations and value delivery to stakeholders [5]. It is not merely a matter of adopting new technologies, but rather a holistic reimagining and restructuring of an organization's culture, processes, and customer experiences to align with the evolving demands of the digital age. This phenomenon of digital transformation is not confined to any specific sector or industry. It has had a widespread impact, transforming diverse fields such as finance, healthcare, retail, and manufacturing [6]. While the potential benefits of digital transformation can be substantial, including enhanced competitiveness, customer satisfaction, and long-term sustainability, it also presents significant challenges. These challenges include the need for substantial investment, overcoming resistance to change, and addressing ethical and regulatory issues related to data use.

Conceptualizing Digital Transformation in Organizations the concept of digital transformation transcends the mere adoption of new technologies, encompassing a more holistic and strategic approach to fundamentally reshape how organizations operate and deliver value to their stakeholders [5]. This transformative process involves the integration of digital technologies, capabilities, and innovative strategies across various aspects of the business, from culture and processes to customer experiences [6]. By embracing digital transformation, organizations can reinvent themselves to better align with the evolving demands of the digital age.

While digital transformation has impacted a wide range of industries, from finance and healthcare to retail and manufacturing, the extent of its adoption and the challenges faced can vary significantly across sectors [6]. Organizations that navigate this transformation successfully can reap substantial benefits, including enhanced competitiveness, improved customer satisfaction, and long-term sustainability. However, the path to digital transformation is not without its obstacles, requiring substantial investment, overcoming resistance to change, and addressing complex ethical and regulatory issues related to data utilization [7].

One significant implementation case study of Digital Transformation is the Real Estate Sector.

Realizing the Full Potential of Digital Transformation in Real Estate The real estate sector has historically been slower to embrace the digital revolution compared to other industries, but there is growing recognition of the immense potential for growth and innovation through digital transformation [8]. This discrepancy can be attributed to the deeply entrenched processes and systems that have long defined the upstream and downstream segments of the real estate industry, as well as the disruptive impact of digital technologies on conventional business practices.

Despite this initial hesitation, the real estate sector has witnessed a surge in the adoption of cutting-edge technologies, becoming a topic of increasing interest in relevant research studies [5]. Over the past decade, the pace and scope of innovation have accelerated, driven by a wave of technological advancements, such as cloud and mobile computing, as well as the rising prominence of platform-based business models that have attracted substantial venture capital investment [9].

The mental healthcare sector is undergoing a significant transformation driven by the integration of digital technologies. Digital transformation in this context refers to the adoption and implementation of innovative digital solutions to address the complex challenges associated with mental health disorders [10]. These digital interventions have the potential to revolutionize mental healthcare delivery, enhancing accessibility, engagement, and treatment outcomes for individuals seeking support.

However, the successful integration of digital technologies into mental health interventions is contingent upon understanding and navigating user acceptance. UTAUT model has emerged as a valuable framework for exploring the issues that influence user behaviour towards new technologies [11]. The UTAUT model identifies core constructs, such as Expected Performance, Expected Effort, Social Influence and Conditions facilitating the use of proposed solution, as key determinants of user acceptance and usage behavior.

Recent studies have further expanded the UTAUT model to include additional factors relevant to the context of digital mental health interventions. For instance, Hedonic Motivation, which refers to the intrinsic enjoyment and pleasure derived from using a technology, and Price Value, which considers the perceived cost-benefit ratio, have been incorporated to develop an understanding towards user acceptance dynamics [12].

Structural Equation Modeling (SEM) has emerged as a powerful analytical technique to explain the complex relationships between these UTAUT determinants and user acceptance. By employing SEM, researchers can uncover the intricate interplay between factors such as Expected Performance, Expected Effort, Social Influence and Conditions facilitating the use of proposed solution, and their collective impact on user acceptance and usage behavior.

The existing literature has highlighted the importance of context-specific insights when studying user acceptance of digital mental health interventions. Researchers have delved into the adoption and usage patterns of various digital mental health solutions, including teletherapy platforms, mobile applications for mood monitoring, and virtual reality-based exposure therapy [13]. These context-specific investigations provide valuable insights into the unique challenges and opportunities associated with the integration of digital technologies in mental healthcare.

By offering a nuanced understanding of user acceptance, this line of research contributes to the design, development and roll-out of effective digital mental health related strategies. Leveraging the insights gained from UTAUT-based studies can help mental healthcare providers and digital health innovators to

design and deploy digital interventions that are more aligned with user needs and preferences, ultimately enhancing accessibility, engagement, and treatment outcomes for individuals seeking mental health support.

Ensuring User Adoption is Key to Successful Digital Transformation. The success of digital transformation initiatives is fundamentally contingent upon user adoption and acceptance of the implemented technological solutions [12]. Recognizing the central role of end-users, organizations must prioritize understanding the factors that drive technology acceptance in order to effectively promote and sustain the usage of their digital innovations. To this end, researchers have developed several conceptual models that provide valuable frameworks for exploring the nuances of user acceptance behavior. These include the extended expectation-confirmation model (EECM), UTAUT and TAM [14]. The EECM is specifically insightful in relation of post-adoption technology usage. This model extends the TAM by integrating expectation-confirmation theory, suggesting that users' ongoing satisfaction and continued usage of a technology are influenced not only by their initial expectations, but also by the degree to which those expectations are confirmed through actual system usage.

Complementing the EECM, the UTAUT model provides a comprehensive perspective on the key indicators of acceptance of technology, including expected performance, expected effort, influence of social life, and conditions that facilitate the situation [6]. Similarly, the TAM focuses on the role of perceived usefulness and perceived ease of use in shaping user attitudes and behavioural intentions towards technology [17]. These frameworks can help organizations develop a more holistic realization of the multifaceted behaviour that influence user acceptance and adoption of mental health related digital innovations.

By leveraging these well-established theoretical models, organizations can design and deploy their digital transformation strategies in a manner that is more closely aligned with user needs, preferences, and ongoing experiences. This user-centric approach is crucial for fostering sustained engagement, driving long-term adoption, and ultimately, realizing the full potential of digital transformation initiatives. While the EECM, UTAUT, and TAM have similarities in studying user acceptance of technology, they also have distinct differences. After reviewing the literature on user acceptance of the system, it became evident that the TAM, UTAUT, and EECM are all relevant frameworks for understanding user acceptance of technology.

Building upon the well-established theory of reasoned action, TAM has emerged as a seminal framework for understanding and predicting user adoption of new technologies is demonstrated [15]. At the core of this model are two key constructs that shape an individual's attitudes, intentions, and ultimately, their usage behavior. The first construct, perceived ease of use (PEOU), reflects the degree to which a user expects the target system to be free of effort and user-friendly [17]. This factor speaks to the intrinsic motivations of the user, as individuals are more likely to embrace technologies that they perceive as intuitive and effortless to navigate. The second construct, perceived usefulness (PU), captures the user's subjective assessment of the likelihood that using a particular system will enhance their performance or productivity within a given context [16]. This extrinsic motivation is a critical determinant of user acceptance, as individuals are more inclined to adopt technologies that they believe will tangibly improve their outcomes or experiences.

By integrating these two core determinants - PEOU and PU - the TAM provides a robust theoretical foundation for analyzing the complex interplay of cognitive, attitudinal, and behavioral factors that shape an individual's technology adoption decisions [15]. This model, grounded in the broader theory of reasoned action, acknowledges that user behavior is not solely driven by personal beliefs and attitudes, but is also influenced by subjective norms and social influences.

The versatility and predictive power of the TAM have made it a widely adopted framework for exploring user acceptance across a diverse range of technological innovations, from enterprise software and mobile applications to e-commerce platforms and smart home devices [6]. By leveraging the TAM, researchers and practitioners can gain valuable insights into the nuanced factors that drive technology adoption, enabling the design and deployment of more user-centric digital solutions.

It has been observes that the previous works done encircling the key words have done only a partial level of analysis, where either TAM, PU or any one acceptance factor is analyzed solely without any cohesion amongst other factors, this has deeply restricted the performance of the solutions in models shown in available literature review.

III. Proposed Research Model

In this paper, examination of the total acceptance model along with some modifications to handle mental health intervention specific tests in the system. In the extant scenario through the proposed models, it has been proposes to contribute a holistic approach towards UTAUT where various critical factors are analyzed in conjunction with other critical factors of the model, such as Societal Conditions, Attitude Towards Use etc. to come to a more effective conclusion at the end of the study.

In the instant case, the following is considered:

Behavioral Intention is taken as the base measure of TAM, which is a concept derived from base UTAUT architecture [6], System Quality (SQ) plays zero role in this solution, as the proposed research paper plays no role in this.

Therefore, a system is designed where the following are considered:

H1: Perceived Usefulness (PU)

H2: Perceived Ease of Use (PEOU)

H3: Social Influence/ Social Impact (SI)

H4: Facilitating Conditions (FC)

H5: Attitude Towards Use (ATT)

TAM is centered around two main constructs that shape an individual's acceptance and usage of technology, which are Perceived Usefulness (PU) which refers to "the degree to which a person believes that using a particular system would enhance

their job performance" and Perceived Ease of Use (PEOU) which is defined as "the degree to which a person believes that using a particular system would be free of effort" [17]. These two main units, PU and PEOU, are influenced by external variables and in turn affect the user's "attitude toward using (ATT)" and "behavioral intention to use (BI)". The model presumes that the perceived ease of use and usefulness mediate the relationship between these independent variables and the actual system usage behavior. The validity of the TAM model has been extensively demonstrated across a wide range of fields and disciplines, including healthcare, technology related to assistive care, social networking, e-shopping, internet, computer, online public services, and entertainment [17].

Notably, the TAM has been widely applied in the context of technology acceptance and adoption among the population. Researchers have explored the use of TAM in various technological contexts relevant to the youth, such as social network related works & digital well-being related interventions and more [18].

The versatility and predictive power of the TAM have made it a leading framework for understanding the factors that drive technology acceptance, enabling researchers and practitioners to design and deploy more user-centric digital solutions.

Fig. 1 shows a graphical representation of the proposed hypothesis.

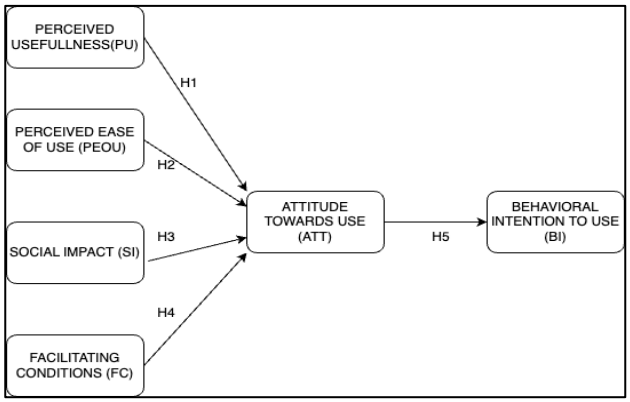


Fig. 1. Research model.

Most crucial parameter for our proposed research model is Attitude to Use (ATT) and its subsequent factorem Behavioral Intention to Use (BI). These factors help in influencing user experience with information technology, encompassing aspects like ease of use, accessibility, and reliability [19]. This study integrates BI as an independent variable within the TAM framework to explore its influence on perceived ease of use (PEOU) and perceived usefulness (PU) of the use case (U) of "Using Digital Transformation based interventions in Mental Health". Understanding this relationship is vital for assessing the impact of ATT and BI on both PEOU and PU, ultimately influencing user behavioral intentions towards user acceptance.

A. Hypothesis Development

The hypothesis formulated for this subject research model are as follows:

1) H1. Perceived usefulness of the use case "Using Digital Transformation based interventions in Mental Health" has a positive effect on the ATT & BI.

2) H2. Perceived Ease of Use of the use case "Using Digital Transformation based interventions in Mental Health" has a positive effect on the ATT & BI.

The degree to which an individual feels that utilizing a certain technology will enhance their quality of life is known as Perceived Usefulness (PU). Perceived Ease of Use (PEOU) is the degree to which a person believes that using a technology involves no effort.

Most studies employing the Technology Adoption Model (TAM) approach make the assumption that there is a relationship between a product's adoption and usage patterns and consumers' evaluations of its utility. This suggests that people are more likely to adopt new technology when they believe it to be beneficial or has clear advantages.

The Technology Acceptance Model (TAM) states that an individual's opinion of a technology's usefulness influences their assessment of its value. A thorough evaluation of the variables impacting healthcare professionals' adoption of information and communication technology (ICT) revealed that the PEOU and PU of the system were the two most crucial components. These two components made up the majority of the original TAM [20].

H3. Social Influence/Social Impact of the use case "Using Digital Transformation based interventions in Mental Health" has a positive effect on the ATT & BI.

The degree to which friends, family, peers, and carers feel something will affect someone's decision to use new healthcare technology, either favourably or unfavourably, is known as social influence [6].

Previous studies ([21] & [22]) have shown how important social impact is when deciding whether or not to utilise technology for ageing in place. Peers, relatives, kids, and professional caretakers for the elderly are all potential sources of social influence.

H4. Facilitating Conditions of the use case "Using Digital Transformation based interventions in Mental Health" has a positive effect on the ATT and BI.

The "facilitating conditions" are the extent to which people believe that an organisational and technological infrastructure is in place to support the information system and motivate users to use it. Previous studies have shown that encouraging surroundings, such as training programmes, technical support, and financial aid, have a beneficial influence on people's intents to use smart wearables and assist them in overcoming their reservations about utilising cutting-edge technology.

Using the effort expectation construct in the same model, researchers discovered that while facilitating conditions by themselves—that is, in the absence of any moderator—do not significantly predict intention to use the system, they do so with a strong effect on older workers with more experience when they are moderated by age and experience. Enabling aspects that are considered important for young people include accessibility, price, and the availability of technical assistance.

H5. Attitude towards use of the use case “Using Digital Transformation based interventions in Mental Health” has a positive effect on the BI.

"An individual's positive or negative feelings or appraisal about using a technology" is the definition of attitude factors (ATT) [6]. Prior research indicated that older adults with more favorable attitudes towards technology are more likely to utilize it. Though they may have a good attitude towards the relevant technology, older people may not want to utilize healthcare technology, according to a study, where the attitude towards technology utilization was not statistically significant.

Behavioural intention (BI) is a key component of technology acceptance models (TAM). It refers to the choice to adopt or utilise a technology based on one's attitude towards using it. Research indicates that a person's behavioural desire to utilise technology positively effects how they actually use it. The first TAM referred to BI as "the degree to which an individual has formulated conscious plans to perform or not perform some specified behaviour in future". Therefore, behavioural intention to use technology has been the focus of various research studies in attempt to predict actual technology usage and adoption.

B. Research Methodology

In the healthcare sector, TAM is the most often used and well-liked model [23], and certain correlations found in TAM have consistently been shown to be significant.

TABLE I. MEASUREMENT ITEMS

Construct	Item	Measurements	Reference
	PU1	Medical Care Managing my health will be easier for me if I use a smart wearable.	
Perceived Usefulness	PU2	I believe that using an innovative wearable to track my mental wellbeing willmake my daily life safer.	[6] & [17]
	PU3	By using a mental health tracker, I will live a better life.	
	PEOU1	I think using a smart mental health monitor will be easy to use.	
Perceived Ease of Use	PEOU2	Learning involved to use this app/technology will be easy.	[6] & [17]
	PEOU3	This app/technology will be convenient to use.	
	SI1	Family will approve of my use of a smart mental health monitor.	
Social impact/influence	SI2	My friends will recommend that I use a mental health monitor.	[6] & [17]
	SI3	My friends will approve that I use a mental health monitor.	
	FC1	I will know how to use the app/technology.	
Facilitating Conditions	FC2	Someone will always be there for help, if I encounter any problem during using the app/technology.	[6] & [17]
	FC3	I have sufficient financial means to use this app/technology.	
	ATT1	Using this will have a positive impact on my lifestyle	
Attitude towards use	ATT2	This app/technology will benefit my family life.	[6] & [17]
	ATT3	I feel positive about this app/technology	
	BI1	I would be happy to use a digital health monitor, if I get the opportunity	
Behavioural Intention to Use	BI2	I will use this app/technology for the betterment of my mental health	[6] & [17]
	BI3	I will use this app/technology to increase my quality of life	

Several research changed and added variables to improve the model's predicting ability in light of the pertinent circumstances for optimal results.

The TAM has a weakness in that it disregards social influence and subjective standards because it was designed based on individual ideas. As a result, the original Technology Acceptance Model (TAM), which included just the two predictive variables PU and PEOU, would not be sufficient to explain why older people embrace technology in a hospital setting.

Based on the suggested model, a structured survey was created to collect information from users who have been exposed to Digital medical/consultation for their medical needs on the variables impacting platform adoption. In order to evaluate important dimensions including perceived utility, perceived ease of use, social impact and facilitating conditions, the survey made use of well-established metrics from earlier research (see Table I).

The concept measurement section uses a 5-point Likert scale, with 1 representing strongly disagree and 5 representing strongly agree. Demographic questions regarding age, gender, qualification, nationality, and user type are also included. Nineteen questions comprise the measure of independent and dependent variables, each specifically designed for the use case scenario.

In this section, all of the questions pertaining to the six variables that make up our suggested model were also accessible. Three measures each, derived from measuring items published ([6] and [17]), were used to test variables including perceived utility, perceived ease of use, and intention to utilise digital health wearables. Three questions each from the Venkatesh et al. measuring items were used to measure social influence and facilitating circumstances. To measure each

component of the research model, a five-point Likert scale ranging from strongly disagree to strongly agree was utilized.

Data was gathered using an online survey tool. The poll was sent to 500 people at random in Saudi Arabia, and 239 of them answered. Participation was completely optional. Three days later, a follow-up email reminder was issued in an attempt to increase participation. 56 valid replies—or a 23.3% response rate—were obtained after the data was cleaned to remove incomplete or incorrect responses as well as to detect and exclude biased respondents. Therefore, then total sample size is 295.

Male and female participants in the research sample showed comparable levels of involvement, according to Table II's findings. The age distribution of the participants was found to be variable, with 42.37% of the individuals lying between the 30- to 39-year-old age ranges. Moreover, 15.25% of the population was under 30, 10.17% was over 50, and 21.2% of the population was between the ages of 40 and 49.

TABLE II. DEMOGRAPHIC DATA

		N=295	%
Gender	Male	166	56.27%
	Female	129	43.73%
Age	Younger than 30	95	32.20%
	30 – 39	125	42.37%
	40 – 49	45	15.25%
	Above 50	30	10.17%
Occupation	Teaching	95	32.20%
	Engineering	121	41.02%
	Medical	52	17.63%
	Govt.	25	8.47%
	Others	2	0.68%
Education	Below High School	2	3.20%
	High School	16	10.30%
	Diploma	85	7.10%
	Bachelor’s	121	63.50%
	Postgraduate	71	16.00%
Income	> 1000$ & < 5000 $	101	34.24%
	> 5000$ & < 10000$	129	43.73%
	> 10000$	65	22.03%
Expense	> 1000$ & < 5000 $	48	16.27%
	> 5000$ & < 10000$	212	71.86%
	> 10000$	35	11.86%
Existing ailment	Yes	114	38.64%
	No	145	49.15%
	Cannot disclose	36	12.20%

The study model was validated using the partial least squares (PLS) method, which is based on structural equation modelling. First, the paper uses measurement analysis, which comprised factor loading, the average variance extracted (AVE), Cronbach alpha, and path coefficient, to assess the internal consistency and validity of our study model.

This paper also confirmed the association between various variables. This paper uses SmartPLS 3 in our investigation to examine the information gathered.

The validity of the convergence and accuracy of the measurement model were evaluated. Internal reliability is verified using Cronbach's alpha test, and consistency of internal validity indication of >0.7 is deemed suitable. A summary of the loadings, SMC, composition reliability, AVE, and Cronbach's alpha can be seen in Table III.

IV. MODEL VALIDATION

The converging validity hypothesis was supported by the composite reliability coefficients (CR) and average variance extracted (AVE) values, both of which were more than 0.65. The loading, variance inflation factors (VIF) and AVE values of our variables meet the criteria for convergent validity, and Cronbach’s alpha (CA) values show that they are internally consistent. A resampling technique called cross-validation is used to get almost unbiased estimates of model performance without compromising sample quantity.

The model validation findings at each of the measurement locations are shown in Tables III-V.

TABLE III. FACTOR LOADING AND RELIABILITY TEST

Construct	Item	Loading	VIF	SMC	CR	AVE	CA
Perceived Usefulness	PU1	0.871	2.37	0.8			
	PU2	0.912	2.62	0.82	0.93	0.81	0.88
	PU3	0.821	2.4	0.8			
Perceived Ease of Use	PEOU1	0.921	3.33	0.87	0.94	0.84	0.9
	PEOU2	0.932	3.76	0.89			
	PEOU3	0.881	2.35	0.79			
Social impact/influence	SI1	0.882	2.38	5.37	0.92	0.8	0.87
	SI2	0.873	2.17	4.79			
	SI3	0.917	2.76	8			
Facilitating Conditions	FC1	0.891	2.14	0.79	0.92	0.79	0.87
	FC2	0.899	2.47	0.89			
	FC3	0.873	2.19	0.91			
Attitude towards use	ATT1	0.894	2.29	0.79			
	ATT2	0.893	2.38	0.79	0.92	0.79	0.87
	ATT3	0.893	2.36	0.75			
Behavioral Intention to Use	BI1	0.899	2.43	0.89			
	BI2	0.922	3.01	0.81	0.93	0.82	0.89
	BI3	0.891	2.79	0.79			

TABLE IV. DISCRIMINANT VALIDITY (FORNELL-LARCKER)

Item	PU	PEOU	SI	FC	ATT	BI
PU	**0.869**					
PEOU	0.614	**0.925**				
SI	0.768	0.593	**0.883**			
FC	0.673	0.794	0.627	**0.898**		
ATT	0.798	0.731	0.654	0.714	**0.883**	
BI	0.662	0.612	0.624	0.649	0.718	**0.926**

TABLE V. DISCRIMINANT VALIDITY (CROSS LOADINGS)

Item	**PU**	**PEOU**	**SI**	**FC**	**ATT**	**BI**
PU1	**0.896**	0.574	0.711	0.61	0.705	0.647
PU2	**0.904**	0.629	0.675	0.606	0.67	0.618
PU3	**0.896**	0.588	0.657	0.572	0.693	0.601
PEOU1	0.641	**0.924**	0.572	0.757	0.666	0.638
PEOU2	0.592	**0.934**	0.519	0.713	0.629	0.65
PEOU3	0.588	**0.887**	0.508	0.681	0.629	0.562
SI1	0.689	0.518	**0.892**	0.562	0.608	0.543
SI2	0.647	0.489	**0.872**	0.52	0.555	0.619
SI3	0.695	0.554	**0.916**	0.571	0.609	0.576
FC1	0.621	0.724	0.533	**0.89**	0.667	0.629
FC2	0.561	0.656	0.537	**0.899**	0.618	0.528
FC3	0.582	0.706	0.576	**0.874**	0.586	0.562
ATT1	0.717	0.628	0.621	0.62	**0.894**	0.666
ATT2	0.654	0.604	0.585	0.65	**0.893**	0.624
ATT3	0.684	0.646	0.566	0.617	**0.891**	0.627
BI1	0.644	0.659	0.627	0.621	0.655	**0.899**
BI2	0.642	0.612	0.588	0.586	0.657	**0.922**
BI3	0.589	0.577	0.529	0.547	0.629	**0.89**

One of the most often used methods for examining the discriminant validity of measurement models is the Fronell-Larcker criteria. This criteria states that a construct's square root of the average variance it extracts must be larger than the correlation it has with any other construct. Discriminant validity is proven once this requirement is met.

The evaluation of a scale's validity involves determining whether or not it captures the idea of what it is meant to capture. Convergent and discriminant validity are established in order to evaluate construct validity. Convergent and discriminant validity are proven in constructs that are reflectively assessed.

When items in a given measure converge to represent the underlying construct, this is known as convergent validity. The mean of the squared loadings of each indicator connected to a construct is how the AVE is computed. In terms of statistics, convergent validity is proven when the Average Variance Extracted (AVE) value is greater than 0.50.

Discriminant Validity: The purpose of discriminant validity is to determine how unique the study's constructs are. It demonstrates that each research construct has a distinct identity and is not too connected with other study constructs. Three methods are used to establish discriminant validity in SMART-PLS.

Fornell and Larcker Criterion: When discriminant validity is proved, it means that the Sq. When it comes to a certain construct, the root of AVE is larger than its association with every other construct.

Cross Loadings: When compared to other research constructs, an item should have larger loadings on its own parent construct, according to cross loadings.

There are problems with discriminant validity when an item loads more favourably onto a different construct than it does onto its own parent construct. The item may be endangering discriminant validity if the difference in loading smaller than.10 also suggests that it is cross-loading onto the other construct.

The correlations between the constructs are shown by numbers outside of the diagonal, while values in bold indicate the square root of the variance extracted (AVE). PEOU: perceived usability; PU stands for perceived utility, SI for social impact, and FC for facilitating condition. ATT: Usage-related attitude; BI: Intention to utilize behavior.

"Discriminant validity is shown when each measurement item correlates weakly with another construct, except for the ones to which it is theoretically associated,". When analyzing cross-loadings, the researcher looks at each item to determine which ones load heavily on numerous constructions and which ones have high loadings on the same construct [24]. Consequently, to demonstrate discriminant validity at the item level, there must be a significant correlation between items that belong to the same concept and a relatively weak correlation between items that belong to different constructs. Despite its simplicity, this technique lacks empirical evidence and theoretical support [25].

A factor's square root of average variance for each latent variable needs to be greater than the other correlation coefficient in order for it to be deemed significant in terms of discriminant validity. Table V displays the detailed cross-loading data. In Table V, the construct that received the highest weight relative to all other constructs is denoted by bolded numerals.

The results of the investigation show that the conditions for discriminant validity are satisfied. The validity and reliability outcomes of our model are displayed in Tables III–V.

A. Hypothesis Testing and Evaluation

This paper used the Smart PLS 3's bootstrapping module (5000 epochs) to get the path coefficients and t values in order to evaluate our hypothesis. Using the bootstrap technique, subsamples from the observed data were selected at random to verify the data's stability. Perceived usefulness, perceived ease of use, social influence, and enabling condition are the four factors that predict 51.3% of the behaviour intention (R2 = 0.513) and 67.6% of the attitude (R2 = 0.676). The route co-efficient (β) and t statistics were used to examine the association between the variables. Table VI displays the PLS findings for the hypotheses.

TABLE VI. HYPOTHESIS TEST

H	Route	β	t-Value	Comments
Hl	PU to ATT	0.425	4.296**	Supported
H2	PEOU to ATT	0.204	2.136 *	Supported
H3	SI to ATT	0.095	1.285	Not supported
H4	FC to ATT	0.204	2.276*	Supported
HS	ATT to BI	0.716	18.047**	Supported

* $p < 0.05$
** $p < 0.001$

From the above, it is observed that the impact of Social Impact/Influence on the Attitude to Use of Digital Transformation driven interventions in mental health treatment is not supported, therefore, doesn't play any role.

V. RESULTS

This study's primary objective was to look at the elements that affect people' behavioural intention and acceptance of digital transformation-driven mental health therapies, such as wearables or mobile apps.

The two fundamental TAM components of perceived utility and perceived ease of use were supplemented by two additional dimensions in our study: social influence and enabling situation. In this study, which included a rather flexible targeting method for population selection, it is discovered that attitudes towards utilising digital mental health monitors are favourably impacted by perceived usefulness, ease of use, and enabling conditions.

It was demonstrated that attitudes regarding the adoption of digital mental health monitoring were positively impacted by the core TAM qualities of perceived usefulness and perceived ease of use.

In this study, the sample population's attitude towards digital mental health monitoring was strongly impacted by perceived usefulness ($\beta = 0.425$). Similarly, the sample set members' attitude towards utilising digital mental health monitors was positively impacted by their perception of the devices' ease of use ($\beta = 0.200$).

According to the findings, people's attitudes about using digital mental health monitors were favourably correlated with enabling conditions ($\beta = 0.200$), but social influence ($\beta = 0.090$) did not appear to play a significant role in fostering the population's favourable attitudes about using them.

In summary, the research outlines the role that perceived usefulness, perceived ease of use, and facilitating conditions have in supporting attitudes toward mental health support online. While social influence is less prominent, the success of digital transformation of mental health rests on how well these platforms demonstrate alignment with the end-user needs, as well as the ease of access and perceived ease of use associated with these digital technologies. This knowledge can effectively inform the future design and implementation of digital mental health interventions, by focusing on user-centered design, and provide the requisite supporting technical factors to allow for sustained use.

VI. CONCLUSION

Digital mental health monitors must live up to public expectations for usability and practicality in daily life in order to foster good attitudes toward these technologies and promote their usage, taking into account the first two results on perceived utility and perceived ease of use. Moreover, the third result suggests that the population's positive sentiments towards the use of digital wearables may not be significantly influenced by social influence. Friends, relatives, and acquaintances won't have a significant impact on the population's desire to utilise these digital mental health monitors unless they are willing to use the technology themselves.

The public is likely to utilise digital mental health monitors if they obtain technical support and recommendations whenever they require assistance utilising the application or technology, according to the fourth result about enabling conditions.

Finally, it is discovered that the sample population's behavioural intention to utilise digital mental health monitoring was significantly influenced by their opinions. These demonstrated that the population's behavioural intention to utilise digital mental health monitoring is higher among those with a more positive attitude towards using these devices.

Like any other research work, this study is also not free from flaws and is not without restrictions.

The data were gathered from Saudi Arabia, therefore the implementation of this study will require the survey to be done on a global scale. In addition, the sample size was modest given the size of the response received from our survey.

In the future, gathering data is advised should involve a bigger sample size. Furthermore, the term "mental health" is used significantly in the paper and in the survey conducted, therefore it may have acted as a deterrent while filling information as the same may have been a taboo subject. In addition to this, it is also necessary to bind the available principles of UTAUT with modern methodologies that revolve around AI based Surveying and Machin learning processes.

Future researchers are urged to consider other factors in order to present a more thorough and definitive view on the behavioural intention of Saudi Arabian youth to use digital mental healthcare wearables, including age-related characteristics, physical changes, digital literacy, technically savvy—experience with technology—and adoption. This will contribute to a more thorough comprehension of this research. In addition to the above, it is to be noted that, like any other research work, this study also has its own limitations, one major limitation is the usage of relatively old methodology for implementation of UTAUT and other is the geo-specific dataset used for the study which limits the overall effectiveness of the study.

VII. REFERENCES

[1] H. Ameera, and L. Anaya, "Digital Technologies in the Healthcare Industry: Literature review," Procedia Computer Science 237 , 2024, pp. 363-370.

[2] H. Dhawan, "Revolutionizing Patient Care: The Impact of Digital Transformation in Healthcare," Available online: https://www.neuronimbus.com/blog/revolutionizing-patient-care-the-impact-of-digital-transformation-in-healthcare/.

[3] K. Hameed, R. Naha and F. Hameed, "Digital Transformation for Sustainable Health and Well-Being: A Review and Future Research Directions," Discover Sustainability 2024, 5, doi:10.1007/s43621-024-00273-8.

[4] V. Rentrop, M. Damerau, A. Schweda, J. Steinbach, and L.C. Schüren, "Predicting Acceptance of e–Mental Health Interventions in Patients With Obesity by Using an Extended Unified Theory of Acceptance Model: Cross-Sectional Study," JMIR Formative Research 2022, 6, e31229,.

[5] D. L. Rogers, "The Digital Transformation Playbook: rethink your business for the digital age," New York: Columbia University Press, 2016.

[6] V. Viswanath, Y.L. James K.Y. Frank K. Y. C.P.j.H. Hu, and S.A. Brown, "Extending the Two-Stage Information Systems Continuance Model: Incorporating UTAUT Predictors and the Role of Context," Information Systems Journal, 21, 2011, pp. 527–55.

[7] P. Langley and A. Leyshon, " Platform Capitalism: The Intermediation and Capitalisation of Digital Economic Circulation. Finance and Society," 2017, 3, pp. 11–31.

[8] S. Barns,"Platform Urbanism", 2020;.

[9] S. Kraus, S. Durst, J.J. Ferreira, P.Veiga, N. Kailer, and A. Weinmann, "Digital Transformation in Business and Management Research: An Overview of the Current Status Quo. International Journal of Information Management", 2022, 63, 102466.

[10] Addressing the Mental Health Impact of COVID-19 through Digital… Available online: https://www.kcl.ac.uk/addressing-the-mental-health-impact-of-covid-19-through-digital-therapies-1.

[11] M. Alojail, J. Alshehri and S.B. Khan, "Critical Success Factors and Challenges in Adopting Digital Transformation in the Saudi Ministry of Education," Sustainability, 2023, 15(21), 15492.

[12] M. Al Moteri, S.B. Khan and M. Alojail, "Economic growth forecast model urban supply chain logistics distribution path decision using an improved genetic algorithm", Malaysian Journal of Computer Science,2023, pp.76-89.

[13] P.K. Chopdar and V.J. Sivakumar, "Understanding Continuance Usage of Mobile Shopping Applications in India: The Role of Espoused Cultural Values and Perceived Risk," Behaviour & Information Technology 2018, 38, pp. 42–64,.

[14] S. Lal, CE. Adair, "E-mental health: a rapid review of the literatur," Psychiatr Serv. 2014 Jan 1;65(1), pp. 24-32.

[15] S. Singh, "An Integrated Model Combining ECM and UTAUT to Explain Users' Post-Adoption Behaviour towards Mobile Payment Systems," AJIS. Australasian Journal of Information Systems/AJIS. Australian Journal of Information Systems/Australian Journal of Information Systems, 2020, 24.

[16] S. Bhattarai and S. Maharjan, " Determining the Factors Affecting on Digital Learning Adoption among the Students in Kathmandu Valley: An Application of Technology Acceptance Model (TAM)", Available online: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3650176.

[17] F.D. Davis, "Perceived Usefulness, "Perceived Ease of Use, and User Acceptance of Information Technology", Available online: https://www.semanticscholar.org/paper/Perceived-Usefulness%2C-Perceived-Ease-of-Use%2C-and-of Davis/ea349162d97873d4493502e205968ffccb23fcf2.

[18] T.H, Ki and H.S. Kim, "Delivery App Understanding and Acceptance among Food Tech Customers Using the Modified Technology Acceptance Model," 2016, 40, pp. 127–144,.

[19] Y. Yann, S. Hooi Tan and S. Wei Choon, "Elderly's Intention to Use Technologies: A Systematic Literature Review," Heliyon 8: 2022, e08765.

[20] N. Saif, S.U. Khan, I. Shaheen, F.A. ALotaibi, M.M. Alnfiai, and M. Arif, " Validating Technology Acceptance Model (TAM) in Education Sector via Ubiquitous Learning Mechanism," Computers in Human Behavior, 2024, 154, 108097.

[21] Gagnon, Marie-Pierre, Marie Desmartis, Michel Labrecque, Josip Car, Claudia Pagliari, Pierre Pluye, Pierre Frémont, Johanne Gagnon, Nadine Tremblay, and France Légaré. 2012. Systematic review of factors influencing the adoption of information and communication technologies by healthcare professionals. Journal of Medical Systems 36: 241–77.

[22] L. Huber, Lesa, M. Boutain, L. Jean Camp, K. Shankar, and K. H. Connelly. " Privacy, Technology, and Aging: A Proposed Framework." Ageing International 36, 2011, pp. 232–52.

[23] G. Demiris, K. Brian K. Hensel, M. Skubic, and M. Rantz. " Senior Residents' Perceived Need of and Preferences for 'Smart Home'," Sensor Technologies. International Journal of Technology Assessment in Health Care 2, 2008, pp. 120–24.

[24] Holden, J. Richard and B.Tzion Karsh, "The Technology Acceptance Model: Its Past and Its Future in Health Care," Journal of Biomedical Informatics, 43, 2010, pp. 159–72.

[25] M. Alojail and S.B. Khan, " Impact of digital transformation toward sustainable development,", 2023, Sustainability, 15(20), 14697.

ResNet50 and GRU: A Synergistic Model for Accurate Facial Emotion Recognition

Shanimol. A[1], J Charles[2]
Research Scholar, Department of Computer Applications Engineering,
Noorul Islam Centre for Higher Education, Tamil Nadu, India[1]
Associate Professor, Department of Computer Applications Engineering,
Noorul Islam Centre for Higher Education, Tamil Nadu, India[2]

***Abstract*—Humans use voice, gestures, and emotions to communicate with one another. It improves oral communication effectiveness and facilitates concept of understanding. Majority of people are able to identify facial emotions with ease, regardless of gender, nationality, culture, or ethnicity. The recognition of facial expressions is becoming more and more significant in a variety of newly developed computing applications. Facial expression detection is a hot topic in almost every industry, including marketing, artificial intelligence, gaming, and healthcare. This study proposes a novel hybrid model combining ResNet-50 and Gated Recurrent Unit (GRU) for enhanced Facial emotion recognition (FER) accuracy. The dataset for the study is taken from Kaggle repository. ResNet-50, a deep convolutional neural network, excels in feature extraction by capturing intricate spatial hierarchies in facial images. GRU, effectively processes sequential data, capturing temporal dependencies crucial for emotion recognition. The integration of ResNet-50 and GRU leverages the strengths of both architectures, enabling robust and accurate emotion detection. Experimental result on CK+ dataset demonstrate that the proposed hybrid model outperforms current methods, achieving a remarkable accuracy of 95.56%. This superior performance underscores the model's potential for real-world applications in diverse domains such as security, healthcare, and interactive systems.**

***Keywords*—Deep convolutional neural network; ResNet-50; Facial Emotion Recognition; Gated Recurrent Unit**

I. INTRODUCTION

Emotions play a major role during communication. A person's mental condition is one of the most important things that can reveal their facial expression. Humans are able to communicate nearly forty five percent of their information verbally and about fifty-five percent nonverbally [1]. Psychologist has defined seven basic emotions such as Disgust, Fear, Surprise Angry, Neutral, Unhappy and Happy [2]. In a broader sense, there are three emotional states of the person. First, neutral emotions, second, positive emotion comprising Happy and Surprise expressions and third, negative emotions comprising Fear, Disgust, Angry and Unhappy expressions. These are basic expressions and are independent of gender and ethnicity [3]. Humans also recognizes other emotions such as contempt, confusion, excitement, stress and desire. Darwin suggested that emotional facial expressions have evolved for a reason.

Nonverbal communication heavily relies on the understanding of facial emotions. Right now automatic face expression recognition is the hardest task thus, there is a strong need for systems that can recognize the same in many different sectors. FER has several uses outside of analyzing behavior and keep track on emotions and mental health of people. It has applications in a variety of domains, including data-driven animation, medical diagnostics, human–robot communication, human–computer interfaces [4], education, robotics, entertainment, holography, smart healthcare systems, security systems, criminology [5], and stress detection [6–7]. Facial expressions are becoming increasingly significant in the medical sciences, especially for bipolar patients whose mood swings are frequent.

FER is also beneficial for applications like smart card readers, social robots, e-learning, criminal justice systems, and customer satisfaction identification [8-9]. The classic emotion identification system consists of three key blocks namely feature extraction, face detection, and emotion classification. Conventional methods of emotion detection have the disadvantage of mutually optimizing feature extraction and categorization. The automation of face emotion recognition and classification is a difficult task. A few fundamental emotions are used by the research community, including fear, anger, upset, and pleasure. However, machines find it extremely difficult to distinguish between a wide ranges of emotions. The major contributions of the proposed research work are follows:

- Developed an efficient Facial Emotion Recognition (FER) method utilizing a deep learning model.
- Implemented a hybrid deep learning approach by integrating ResNet 50 with Gated Recurrent Unit (GRU).
- Achieved enhanced accuracy in Facial Emotion Recognition.

The rest of the paper is organized as follows: In Section II, a summary of literature is provided, highlighting areas that indicate a need for more investigation. In Section III, the methodology is explained in depth. Section IV goes into great detail about the results that the suggested strategy produced. A discussion is provided in Section V and finally, a summary of the findings is included in Section VI, which gives a conclusion to the paper.

II. Literature Review

A framework that used a BiLSTM fusion network and simultaneously learned temporal dynamics and spatial information for FER was presented by Liang et al. [10]. Three benchmark databases—namely MMI, CK+, Oulu-CASIA, were used in the experiment. The technique learned discriminative spatial features and short-term dynamic features using two separate CNNs, and then combined them at the feature level. A comparison of the model's performance using the Oulu-CASIA dataset revealed an accuracy of 91.07%. Because there were only few training samples available for the study, the method's generalizability was constrained. A spatio-temporal feature representation learning method for FER that was resistant to changes in expression intensity was presented by Kim et al. [11]. Regardless of the intensity of the expression, the approach made use of representative expression described in face sequences. Using a CNN, spatial properties were learned. Long short-term memory of the face expression was used to learn the temporal property of the spatial feature representation. The studies were carried out using two datasets namely one for spontaneous micro-expression (CASME II) and the other for purposeful expression (MMI). The accuracy of the approach was found to be 72.83% and 78.61% in the intra- and inter-dataset evaluations.

A Transfer learning approach to emotion recognition was proposed by Chowdary et al. [12]. Pre-trained vgg19, Resnet50, and Mobile Net, Inception V3, networks were used in this work. The CK + database was used in the experiment and 94.2% accuracy was attained with the help of MobileNet. Using a CNN, Debnath et al.'s new facial emotional recognition model [13] identified seven distinct emotions from image data. The model attained an accuracy of 92.05% using the generalization strategy on the JAFFE dataset. A modular framework was presented by Alreshidi et al. [14] for the classification of facial emotions into seven distinct states. The authors failed to utilize geometric elements that could enhance the performance. The approach achieved 59.0% accuracy and might be used to treat and diagnose patients with emotional problems.

A Custom CNN Architecture was presented by Borgalli et al. [15] to do fundamental FER in static images. The three datasets utilized in the methodology to evaluate the model are FER13, JAFFE, and CK+. The CK+ datasets achieved an accuracy rate of 92.27% on fundamental emotions. The method's significant recognition error amount was one of its limitations. In order to circumvent the conventional feature extraction procedure, Bukhari et al. [16] created a CNN model that was utilized as a feature extractor for emotion identification using facial expression. Three pre-trained models were employed in this work by the authors: VGG-16, ResNet-50, and Inception-V3. The accuracy rates for CNN 92.91% on ck+ dataset, according to the experimental results.

CNN, which predict and assign probabilities to each emotion, were the basis of an efficient facial emotion identification system for the seven basic human emotions presented by Ghaffar et al. [17]. In order to improve prediction, the system applied a variety of preprocessing procedures to each image as deep learning models learn from data. To include each image in the training dataset, the face detection algorithm was run on each one first. With the combined dataset, the method's maximum accuracy was 78.1%. In order to overcome the facial expression recognition (FER) problem, Kandhro et al. [18] proposed a CNN; in recent years, more and more significant efforts have been done in this area. This FER technique can be used to obtain facial expressions based on regularization settings, activations, and optimizations from databases such as CK+ and JAFFE. Numerous techniques, such as regularization, optimization, and activation, in addition to additional hyperparameters, were used to assess the model's performance. The authors obtained 71% test accuracy and 97% training accuracy using the FER2013 dataset. By utilizing several facial features with appropriate dimensions space reduction and applying a kernel filter as part of the preprocessing technique, Kumar Arora et al. [19] suggested the facial feature for emotional recognition using a deep learning algorithm.

A. Research Gap

Facial expression recognition heavily depends on facial landmarks in image-based approaches, which can be prone to errors in varying conditions. Model-based approaches, while accurate, rely on intense numerical computations due to the need for complex mapping functions, making them resource-intensive and time-consuming when training on large datasets. Current models exhibit good performance but require further improvement to handle real-world variations such as different lighting conditions, occlusions, and diverse facial expressions across various ethnicities and age groups. Existing methods often fall short in robustness and generalization, partly due to training on relatively small or biased datasets. Additionally, many FER systems are not optimized for real-time operation on edge devices with limited computational resources. Fourier transform techniques, commonly used in these systems, may miss important spatial features crucial for emotion detection as they focus primarily on frequency domain information. Addressing these gaps is essential for developing FER systems that are both efficient and applicable in diverse, real-world scenarios.

III. Materials and Methods

The proposed methodology as shown in Fig. 1 comprises two main components: a pre-trained ResNet50 [20] for image feature extraction and a GRU [21] for capturing sequential patterns. ResNet-50 was chosen for its exceptional feature extraction capabilities, which are crucial for capturing the intricate spatial details of facial expressions, while GRU was integrated to leverage its strength in sequential pattern recognition, enabling the model to effectively analyze the temporal dynamics of emotions. Existing methods, such as those relying solely on Convolutional Neural Networks (CNNs) or Long Short-Term Memory (LSTM) networks, often fall short in either spatial feature extraction or temporal sequence modeling, making them less effective for FER tasks that require a holistic approach. The proposed method overcomes these limitations by combining the strengths of both architectures, making it more suitable for the complex pattern recognition required in accurately classifying emotions.

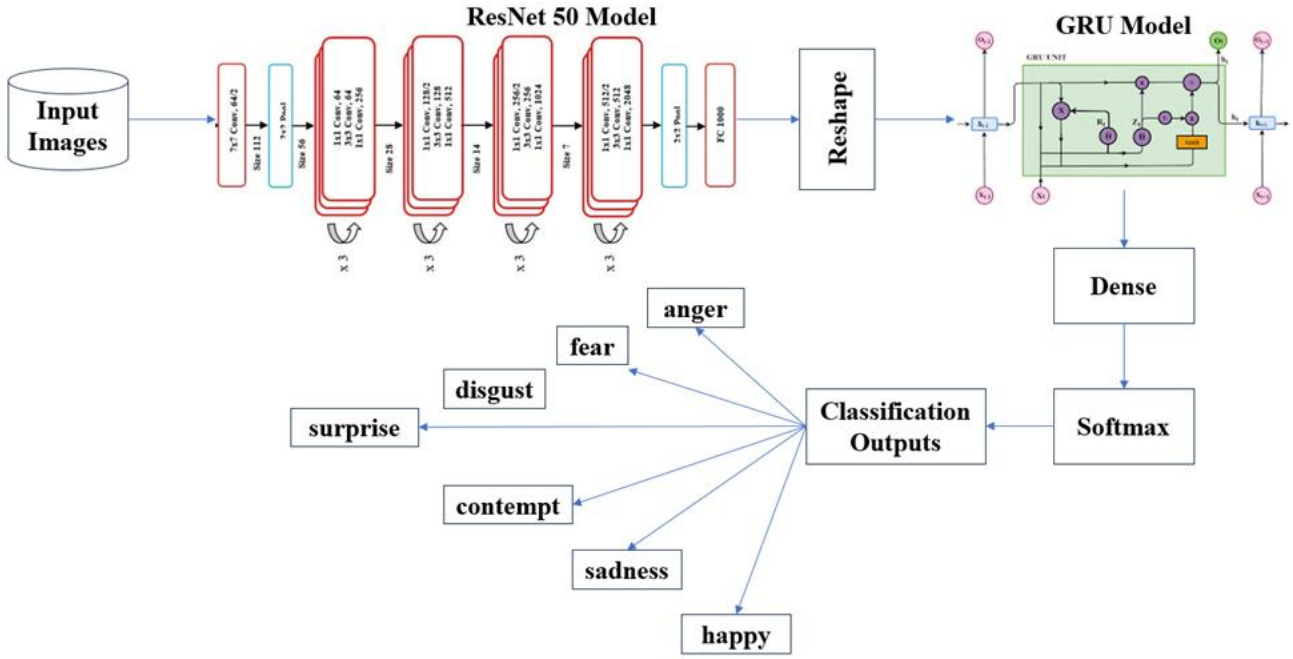


Fig. 1. Block diagram of the proposed model.

Following dataset collection, pre-processing and augmentation techniques are applied. The data generator resizes images to a target size of 224x224 pixels, rescale pixel value to the range [0, 1], and apply shear transformations, zoom in/out, horizontal and vertical flips, and rotations up to 30 degrees. The images are batched into groups of 32, and class labels are encoded in categorical format. This setup is crucial for training and evaluating the model on the CK+ dataset, enhancing generalization through data augmentation. The ResNet50 and GRU networks are integrated into a hybrid model by concatenating their outputs. The ResNet and GRU components are connected sequentially, with the GRU input reshaped to match its expected input shape. The performance of the model is evaluated on the several performance measures.

A. Dataset Description

Data is sourced from Kaggle repository https://www.kaggle.com/datasets/davilsena/ckdataset/. The sample images are shown in Fig. 2. Dataset Contains modified data up to 920 images from 920 original CK+ dataset. Data is already reshaped to 48x48 pixels, in grayscale format and face cropped using haarcascade_frontalface_default. Noisy images were adapted to be clearly identified using Haar classifier.

B. Data Pre-processing and Augmentation

The preprocessing and data augmentation steps for the model involve several techniques to enhance training and evaluation on the CK+ dataset. Initially, the dataset is preprocessed by rescaling pixel values to the range [0, 1]. The data generator is then configured to apply various augmentation techniques, including shear transformations, zooming in and out, horizontal and vertical flips, and rotations up to 30 degrees. Images are resized to a target size of 224x224 pixels and batched into groups of 32. Additionally, class labels are encoded in categorical format. This comprehensive setup is crucial for improving the model's generalization capabilities through effective data augmentation.

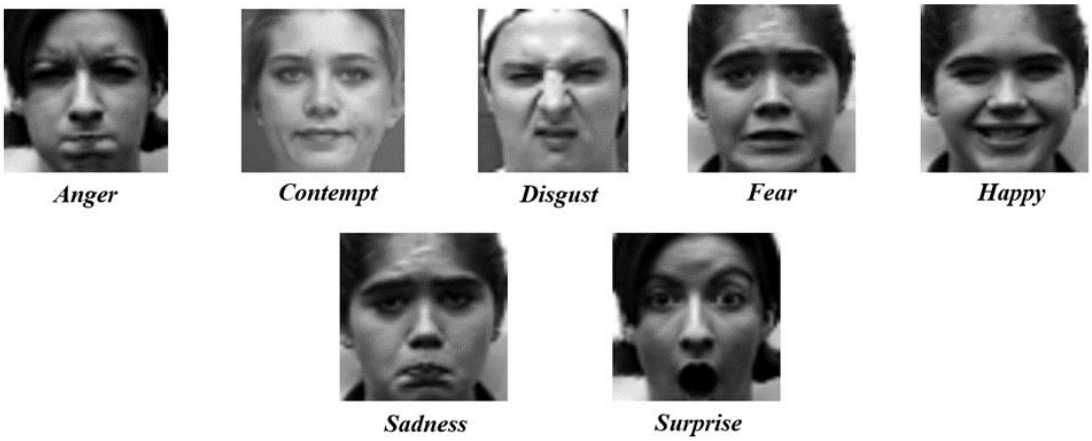


Fig. 2. Sample images from the dataset.

C. ResNet 50 Architecture

Four main parts of the ResNet50 architecture are the identity block, fully connected layer, convolutional block, and convolutional layer. Fig. 3 shows the architecture of the ResNet 50 model. The features that the convolutional layers have extracted from the input image are being processed and transformed by the identity block and convolutional block. The identity block is a straightforward block that adds the input back to the output after passing it through several convolutional layers. The network is able to learn residual functions, which convert input into desired output. In the convolutional layers of ResNet50, batch normalization and ReLU activation come after many convolutional layers. These layers are in charge of taking characteristics like edges, textures, and forms out of the input image. Max pooling layers, which minimize the spatial dimensions of the feature maps while maintaining the most crucial properties, come after convolutional layers. The fully connected layers make up the last section of ResNet50. The last classification is determined by these layers. The output of the final fully connected layer is fed into a softmax activation function to get the final class probabilities.

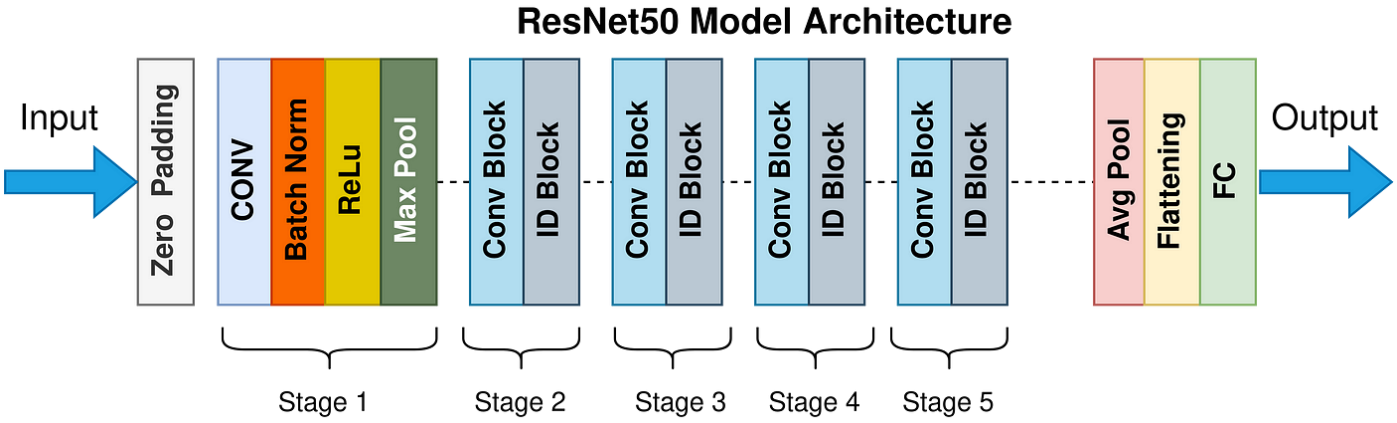


Fig. 3. Architecture of ResNet50 model.

When deep neural networks are trained, an issue known as "vanishing gradients" might arise. This is when the parameter gradients in the deeper layer get very small, which makes it harder for such layers to learn. The deeper the network, the more severe this issue gets. By enabling data to move straight from the network's input to its output and omitting one or more tiers, skip connections solve this issue. Instead of having to learn the complete mapping from scratch, the network can learn residual functions that convert input into the intended output. The residual block is depicted in Fig. 4. The output of the residual block is represented by Eq. (1).

$$Y = F(X, \{W_m\} + X) \quad (1)$$

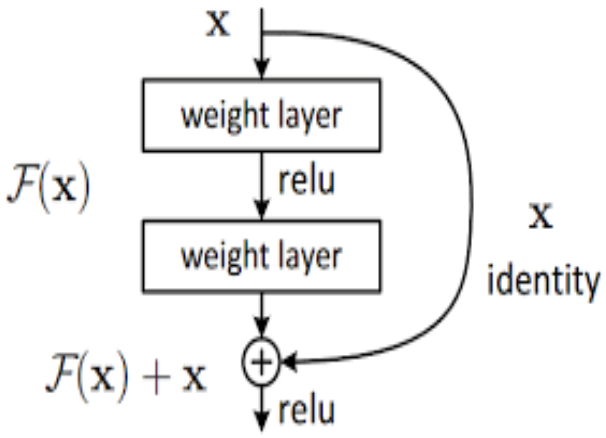


Fig. 4. Block diagram of the residual block.

The input to the residual block is denoted by X, Y is the output of the block, W_m represents the weights of the convolutional layers within the block, and F is the residual function. In a residual block, the residual function F typically consists of two or three convolutional layers. For a three-layer residual block, the residual function can be expressed as in Eq. (2).

$$F(X, \{W_m\} = W3\sigma(W2\sigma(W1X))F(X, \{Wm\}) = W3\sigma(W2\sigma(W1X)) \quad (2)$$

σ denotes the ReLU activation function, W1,W2,W3 are the weights of the convolutional layers. The identity mapping X sometimes be transformed using a linear projection to match the dimensions of $F(X, W_m)$ when necessary. This can be done using a 1x1 convolution as depicted in Eq. (3).

$$Y = F(X, \{W_m\}) + W_s X \quad (3)$$

where W_s is the weight matrix of the 1x1 convolution used for matching dimensions. The activation function used is typically the Rectified Linear Unit (ReLU), expressed as in Eq. (4).

$$\sigma(x) = max(0, x) \quad (4)$$

Batch normalization as expressed in Eq. (5) is applied after each convolution and before the activation function:

$$BN(x) = \frac{x-\mu}{\sigma^2+\epsilon} * \gamma + \beta \quad (5)$$

Where μ and σ2 are the batch mean and variance, γ and β are learnable parameters, and ϵ is a small constant to prevent division by zero.

D. Gated Recurrent Unit

By allowing information to be selectively retained or lost over time, GRU is intended to mimic sequential data. Because GRU has fewer parameters and a simpler architecture, it may be easier to train and use less computing power. The update gate and the reset gate are the two distinct gates that are part of the GRU architecture as shown in Fig. 5. The distinct functions of each gate greatly add to the high efficiency of the GRU. Long-term connections are recognized by the update gate, whereas short-term ties are identified by the reset gate.

The computation steps of a GRU in the Reset Gate, update gate, candidate hidden state, as are the following.

$$R_t = \sigma(A_{x,z}X_t + A_{H,z}H_{t-1} + B_z) \quad (6)$$

$$Z_t = \sigma(A_{x,z}X_t + A_{H,z}H_{t-1} + B_z) \quad (7)$$

$$\acute{H}_t = \tanh(A_{H,H}R_tH_{t-1}) + A_{x,H}X_t + B_H \quad (8)$$

$$H_t = (1 - Z_t)H_{t-1} + Z_t\acute{H}_t \quad (9)$$

where, $\acute{H}_t$ is the candidate hidden state that is incorporated proportionately to the hidden state, R_t is the reset gate value, and Z_t is the update gate.

E. Proposed ResNet 50-GRU Hybrid Model

The proposed hybrid model for FER integrates a pre-trained ResNet-50 and a GRU. The model architecture of the proposed hybrid model is depicted in Fig. 6. The model takes an image input of shape (224, 224, 3), with the ResNet and GRU components connected sequentially, reshaping the GRU input to match its expected input shape. The ResNet-50, configured as a feature extractor by removing its top layers and applying global average pooling, reduces spatial dimensions and is followed by a dense layer and ReLU activation to enhance feature representation. Concurrently, a GRU network with 128 units is constructed, also followed by a dense layer with 256 units and ReLU activation. These two networks are then integrated by concatenating their outputs, and the final output layer is a dense layer with softmax activation, which produces probability distributions over seven emotion classes.

F. Hardware and Software Setup

The computational setup for this research utilized a machine with robust specifications, featuring an Intel Core i7 processor. 32GB of RAM, and the formidable NVIDIA GeForce GTX 1080Ti GPU. Model implementation was seamlessly carried out through the Keras library, functioning as a prototype built upon the Tensorflow framework and executed using the versatile Python language. Keras, known for its user-friendly interface and powerful capabilities, proved instrumental in crafting intricate Neural Network architectures. This framework ensures efficient utilization of computing resources, seamlessly accommodating CPU, GPU, and TPU environments. To leverage extensive computational capabilities and streamline model training, the deployment was orchestrated on Google Colab. This cloud-based Python notebook environment not only provides complimentary access to robust computational resources but also facilitates collaborative development, making it an optimal choice for training models.

Hyper parameters are essential configuration settings that define the behaviour and characteristics of a machine learning framework throughout the training process. Unlike the parameters of the model, which are learned from the data itself, hyper parameters are set by the user before training begins. The neural network model uses the Adam optimizer. The training process is guided by the Categorical cross-entropy loss function. During training, the model processes input data in batches of 32 samples per iteration. The training is carried out over 50 epochs, signifying the number of times the entire training dataset is processed by the model. These hyper parameter choices, such as the optimizer, loss function, batch size, and number of epochs, collectively define the configuration for training the neural network model, aiming to optimize its performance on the proposed emotion detection. The model configuration of the suggested approach is tabulated in Table I.

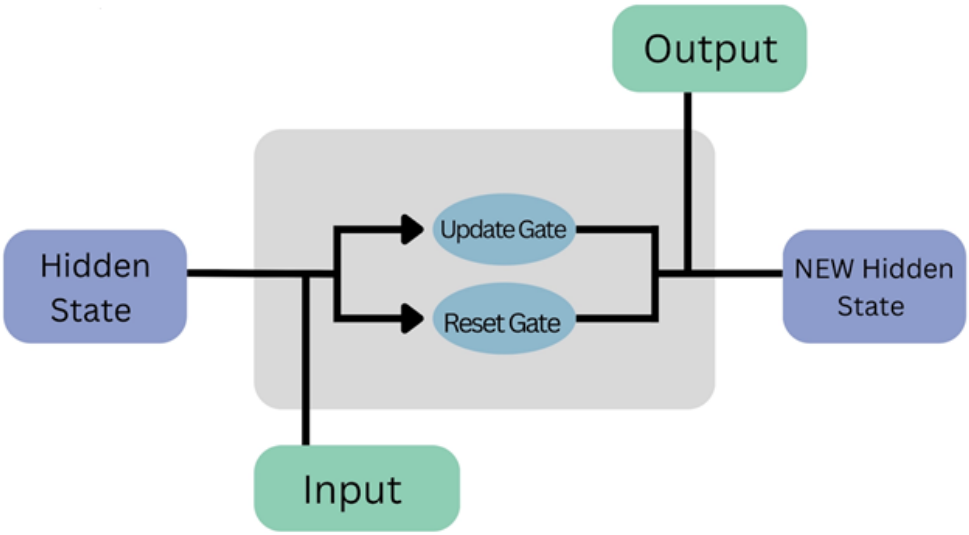


Fig. 5. Architecture of Gated Recurrent Unit.

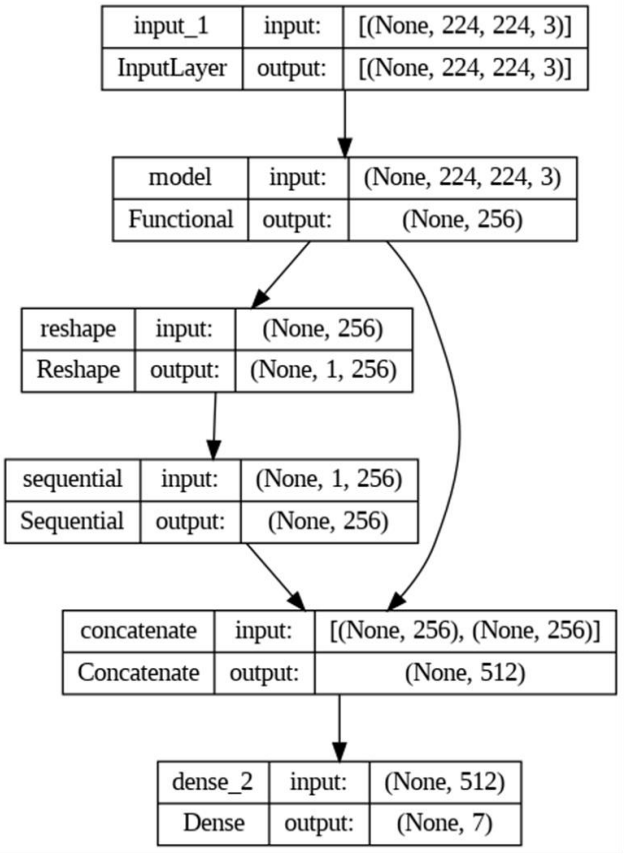


Fig. 6. Proposed model architecture.

TABLE I. MODEL CONFIGURATIONS

Hyper parameter	Values
Optimizer	Adam
Loss Function	Categorical crossentropy
No. of epochs	50
Batch Size	32
Activation Function	ReLu, Softmax

IV. EXPERIMENTAL RESULTS

The accuracy and loss plots are essential tools for assessing the performance and learning dynamics of the proposed emotion classification model. The accuracy plot visually depicts how accurately the model predicts the emotional labels of the data across training iterations for both the training and validation datasets. This plot tracks the consistency between the model's predictions and the actual emotional labels, serving as a crucial indicator of the model's performance throughout the training process. The accuracy plot highlights the model's ability to effectively distinguish between different facial emotions during training. Ideally, in the early epochs, both training and validation accuracies rise simultaneously, demonstrating the model's ability to generalize its knowledge beyond the training dataset.

The accuracy plot of the model is shown in Fig. 7. In the initial epochs of training, the proposed system demonstrates high accuracy, starting at 98.74% in epoch 1. This strong initial performance indicates the model's rapid learning capacity. As the training progresses, accuracy consistently remains high, reaching 99.20% by epoch 2 and continuing to show incremental improvements. However, some fluctuations in accuracy occur between epochs 13 and 16, where accuracy drops from 98.40% to 94.04%, reflecting a temporary period of instability. Despite these fluctuations, the model quickly recovers, achieving a perfect accuracy of 100% by epoch 10 and maintaining this level through the final epochs, from epoch 48 to epoch 50. These fluctuations are typical in deep learning training processes, often due to batch variance or learning rate adjustments.



Fig. 7. Accuracy plot of the proposed system.

The difference between the predicted emotions and the true labels is quantified as the model's loss, which is illustrated in the loss plot. Throughout the training process, the objective is for the loss to decrease steadily, reflecting that the model is improving its predictions and reducing errors with each iteration, as depicted in Fig. 8.

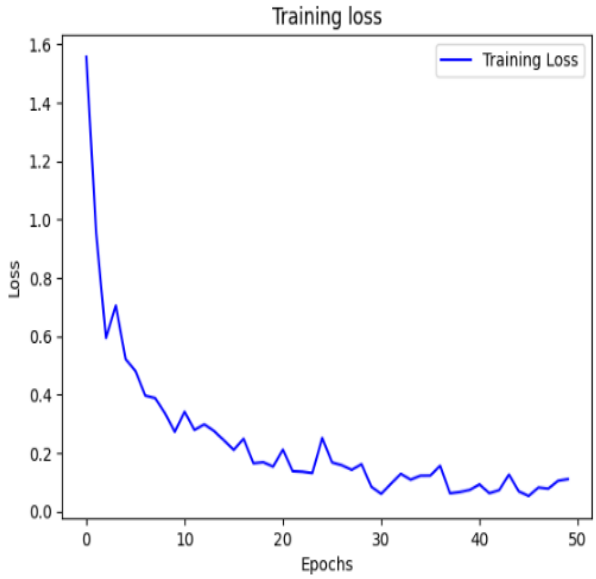


Fig. 8. Loss plot of the proposed system.

In the initial epochs, the loss decreases steadily from 0.0329 to 0.0010, and the accuracy increases from 0.9874 to 1.0000. By epoch 22, the loss decreases to 0.0911 with the same accuracy, but in epoch 23, the loss increases to 0.1567, and accuracy slightly drops to 95.53%, showing some initial fluctuation. In the final epochs, the loss significantly decreases, with epoch 48 showing a loss of 0.0021 and 100% accuracy, continuing to epoch 49 with a loss of 0.0011 and epoch 50 reaching the lowest loss of 0.0010, both maintaining 100% accuracy. This indicates that the model is learning well and improving its performance over time. However, there are fluctuations in the loss and accuracy throughout the epochs, such as a slight increase in loss during epochs 11 and 12, followed by a decrease. Overall, the fluctuations seem relatively minor, and the model achieves a very low loss and high accuracy by the final epoch, suggesting that it has effectively learned the patterns in the data and generalized well.

A valuable method for assessing the effectiveness of the proposed emotion classification system is the use of a confusion matrix. This matrix offers a structured overview of the model's performance by comparing its predicted emotional categories with the actual labels across various classes. It organizes the outcomes into a table format, where the rows correspond to the true emotional labels and the columns correspond to the predicted labels, as illustrated in Fig. 9. Each cell within the matrix displays the count of instances where the model's predictions either match or deviate from the true emotional labels. The confusion matrix is divided into four quadrants, with the diagonal elements representing correct predictions and the off-diagonal elements indicating misclassifications.

Performance metrics derived from the confusion matrix offer a thorough evaluation of the proposed model's efficacy in classifying emotions. The performance of the system is mainly evaluated on four parameters accuracy, precision, recall, F1-score. These measures, which are based on the concepts of False Positive (FP), False Negative (FN), True Negative (TN), and True Positive (TP), are essential for assessing the model's performance. The calculation of accuracy involves dividing the total number of predictions by the number of right predictions.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (10)$$

The exactness of a prediction is measured by its precision, or the number of true positives. Instead, recall quantifies completeness, or the number of real positives that were anticipated as positives.

$$Precision = \frac{TP}{TP+FP} \quad (11)$$

$$Recall = \frac{TP}{TP+FN} \quad (12)$$

$$F1-Score = 2 * \left(\frac{Precision*Recall}{Precision+Recall}\right) \quad (13)$$

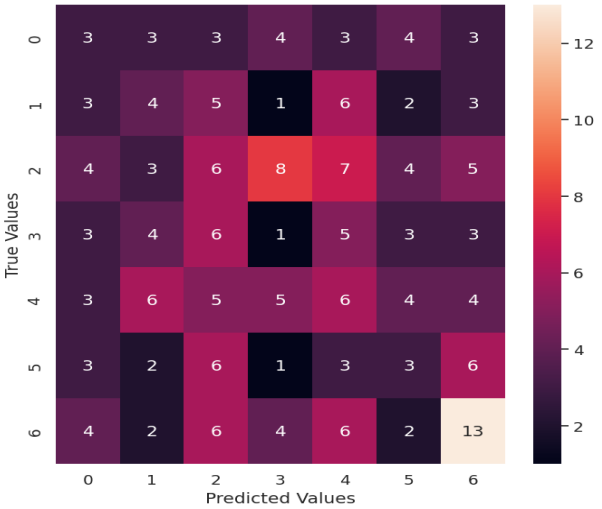


Fig. 9. Confusion matrix.

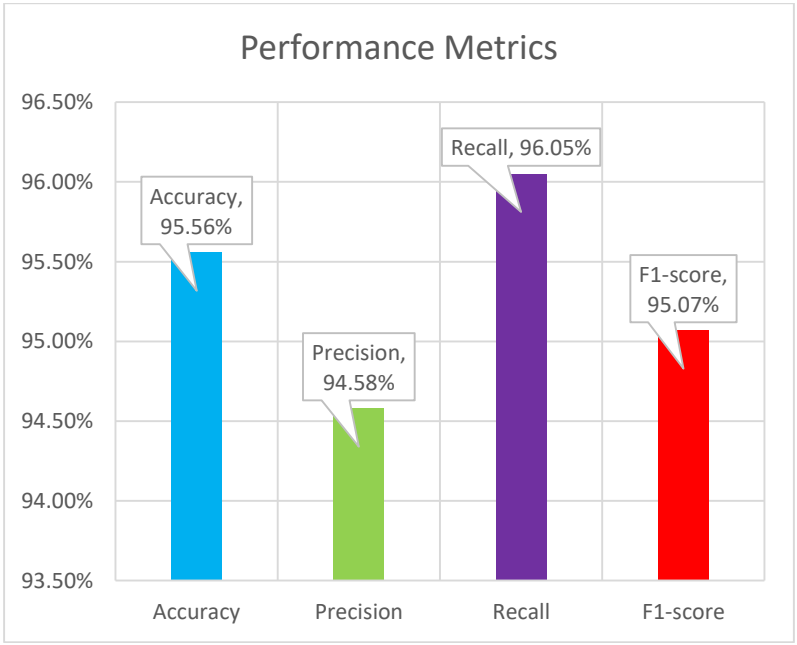


Fig. 10. Performance metrics.

The performance metrics of the proposed emotion classification system are detailed in Fig. 10, illustrating its effectiveness in accurately classifying emotions. The accuracy metric stands at 95.56%, indicating the overall correctness of the model's predictions compared to the total number of predictions made. Precision, measured at 94.58%, reflects the model's capability to correctly identify specific emotions among those predicted. Recall, which measures at 96.05%, signifies the model's ability to accurately retrieve all instances of a particular emotion from the dataset. The F1-score, calculated at 95.07%, harmonizes precision and recall into a single metric, offering a balanced assessment of the model's performance in emotion classification tasks. These metrics collectively demonstrate the system's high accuracy and reliability in recognizing and distinguishing between different emotional states, underscoring its potential for practical applications requiring nuanced emotion detection.

The prediction results of the proposed emotion classification system are illustrated in Fig. 11. This figure provides a comprehensive visual representation of the model's performance, showcasing how effectively it can identify and classify various emotional expressions. By examining the prediction results, one can assess the accuracy and reliability of the model in real-world scenarios. The figure highlights the model's capability to distinguish between different emotions, demonstrating its potential effectiveness and practical application in emotion recognition tasks.

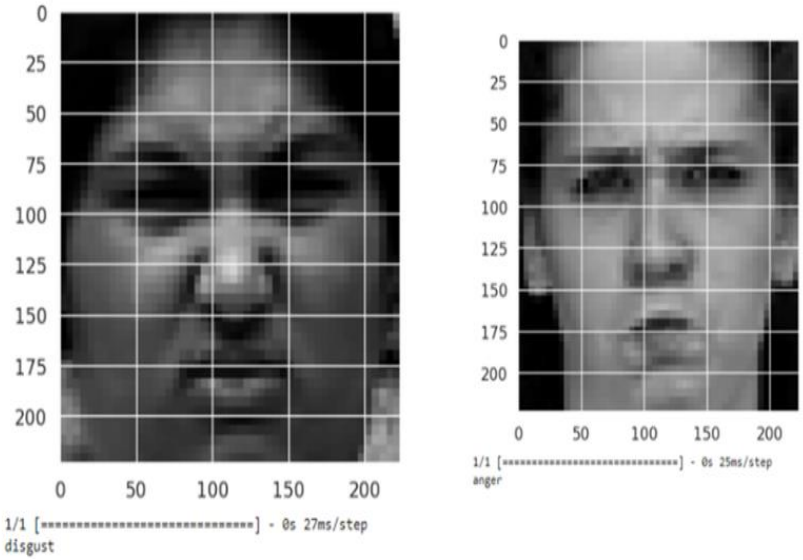


Fig. 11. Prediction output.

V. Discussion

Comparing the performance of the proposed hybrid emotion classification network against existing methods primarily based on machine learning and deep learning is a pivotal aspect of this study. Table II and Fig. 12 provides a comparative analysis that highlights the effectiveness of the hybrid model by carefully evaluating outcomes in contrast to established approaches. This model addresses the challenges of accurately recognizing and classifying emotions from facial expressions, which is critical in fields such as human-computer interaction, healthcare, and security. This evaluation rigorously examines a range of metrics and parameters to assess how well the proposed method performs compared to traditional methodologies used in emotion recognition. The aim is to demonstrate the superiority and robustness of the hybrid approach in accurately classifying emotions, showcasing its potential to outperform conventional techniques in real-world applications.

The comparison report presents various state-of-the-art models and their corresponding methodologies and results in a certain task, likely classification or prediction. Liang et al. employed a Bidirectional Long Short-Term Memory (BiLSTM) model achieving an accuracy of 91.07%. Kim et al. utilized a CNN-LSTM hybrid model, attaining a slightly lower accuracy of 78.61%. Debnath et al. deployed a CNN achieving a higher accuracy of 92.05%, while Borgalli et al. introduced a custom CNN with a slightly higher accuracy of 92.27%. Bukhari et al. implemented a ResNet-50 model, achieving an accuracy of 92.91%. Finally, the proposed model in the report, a hybrid of ResNet-50 and GRU, demonstrated the highest accuracy of 95.56%. This comparison highlights the effectiveness of different architectures and demonstrates the superiority of the proposed hybrid model, which integrates both convolutional and recurrent neural network components, achieving the highest accuracy among the compared methods.

TABLE II. Comparison of Proposed Model with Existing Methods

Authors	Methodology	Result
Liang et al [10]	Bi-LSTM	91.07%
Kim et al [11]	CNN-LSTM Hybrid model	78.61%
Debnath et al [13]	CNN	92.05%
Borgalli et al [15]	Custom CNN	92.27%
Bukhari et al [16]	ResNet -50	92.91%
Proposed model	**Hybrid model ResNet 50 and GRU**	**95.56%**

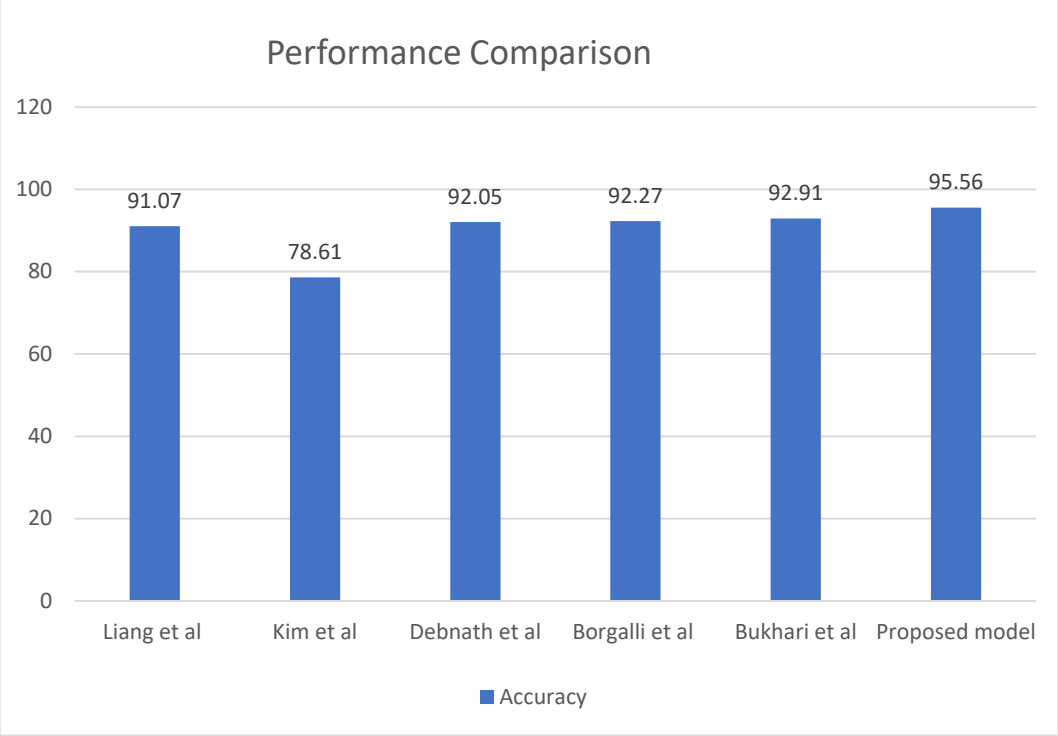


Fig. 12. Performance comparison.

Moreover, the robustness of the proposed model in handling diverse and complex emotional expressions demonstrates its potential for real-world applications where emotion recognition must be accurate and reliable. The comparison study underscores the model's capability to outperform conventional techniques, making it a promising solution for advancing the field of emotion recognition. This hybrid approach not only achieves higher accuracy but also offers a more nuanced understanding of emotional expressions, paving the way for more sophisticated and effective emotion recognition systems.

VI. Conclusion

Facial expressions are a vital tool for determining human emotions since they are reflections of those emotions. The majority of the time, a person's facial expressions are a nonverbal means of expressing their emotions. These emotions can be used as concrete evidence to determine whether or not someone is telling the truth. This study aimed to classify facial expressions into one of seven emotions using the CK+ dataset. The study successfully demonstrates the efficacy of a hybrid model combining a pre-trained ResNet50 and a GRU for FER. By leveraging the powerful feature extraction capabilities of ResNet50 and the sequential pattern recognition strengths of GRU, the proposed model achieves an accuracy of 95.56% in classifying emotions. When compared to cutting-edge outcomes, the developed model provides good accuracy. This hybrid approach not only validates the integration of ResNet 50 and GRU for complex pattern recognition tasks but also sets a robust framework for future research in FER and related fields.

While the proposed model has shown promising results, there are several areas for future work to further enhance the performance and applicability of FER systems. First, expanding the dataset to include more diverse and real-world scenarios could improve the model's generalizability, making it more robust in different environments and cultures. Additionally, exploring the integration of other advanced deep learning architectures, such as Transformer-based models, could potentially enhance the model's ability to capture complex dependencies in facial expressions. Moreover, incorporating multimodal data, such as voice and physiological signals, alongside facial expressions could lead to more comprehensive emotion recognition systems.

References

[1] Salim, M. S. (2023). Verbal and non-verbal communication in linguistics. International Journal of Innovative Technologies in Social Science, (2 (38)).

[2] Frenzel, A. C., Daniels, L., & Burić, I. (2021). Teacher emotions in the classroom and their implications for students. Educational Psychologist, 56(4), 250-264.

[3] Chen, Y., & Joo, J. (2021). Understanding and mitigating annotation bias in facial expression recognition. In Proceedings of the IEEE/CVF International Conference on Computer Vision (pp. 14980-14991).

[4] Rajesh Kumar, G., Srinivasa Rao, D., Rajasekhar, N., Ramesh Babu, C., Rohini, C., Ravi, T., & Mangathayaru, N. (2022, November). Emotion Detection Using Machine Learning and Deep Learning. In International Conference on Intelligent Computing and Communication (pp. 705-715). Singapore: Springer Nature Singapore.

[5] Channing, I., Churchill, D., & Yeomans, H. (2023). Renewing historical criminology: Scope, significance, and future directions. Annual Review of Criminology, 6, 339-361.

[6] Mansour, R. F., El Amraoui, A., Nouaouri, I., Díaz, V. G., Gupta, D., & Kumar, S. (2021). Artificial intelligence and internet of things enabled disease diagnosis model for smart healthcare systems. IEEE Access, 9, 45137-45146.

[7] Goel, R., & Gupta, P. (2020). Robotics and industry 4.0. A Roadmap to Industry 4.0: Smart Production, Sharp Business and Sustainable Development, 157-169.

[8] Hassouneh, A., Mutawa, A. M., & Murugappan, M. (2020). Development of a real-time emotion recognition system using facial expressions and EEG based on machine learning and deep neural network methods. Informatics in Medicine Unlocked, 20, 100372.

[9] Otamendi, F. J., & Sutil Martín, D. L. (2020). The emotional effectiveness of advertisement. Frontiers in Psychology, 11, 563695.

[10] Liang, D., Liang, H., Yu, Z., & Zhang, Y. (2020). Deep convolutional BiLSTM fusion network for facial expression recognition. The Visual Computer, 36, 499-508.

[11] Kim, D. H., Baddar, W. J., Jang, J., & Ro, Y. M. (2017). Multi-objective based spatio-temporal feature representation learning robust to expression intensity variations for facial expression recognition. IEEE Transactions on Affective Computing, 10(2), 223-236.

[12] Chowdary, M. K., Nguyen, T. N., & Hemanth, D. J. (2023). Deep learning-based facial emotion recognition for human–computer interaction applications. Neural Computing and Applications, 35(32), 23311-23328.

[13] Debnath, T., Reza, M. M., Rahman, A., Beheshti, A., Band, S. S., & Alinejad-Rokny, H. (2022). Four-layer ConvNet to facial emotion recognition with minimal epochs and the significance of data diversity. Scientific Reports, 12(1), 6991.

[14] Alreshidi, A., & Ullah, M. (2020, February). Facial emotion recognition using hybrid features. In Informatics (Vol. 7, No. 1, p. 6). MDPI.

[15] Borgalli, M. R. A., & Surve, S. (2022, March). Deep learning for facial emotion recognition using custom CNN architecture. In Journal of Physics: Conference Series (Vol. 2236, No. 1, p. 012004). IOP Publishing.

[16] Bukhari, N., Hussain, S., Ayoub, M., Yu, Y., & Khan, A. (2022). Deep learning based framework for emotion recognition using facial expression. Pakistan Journal of Engineering and Technology, 5(3), 51-57.

[17] Ghaffar, F. (2020). Facial emotions recognition using convolutional neural net. arXiv preprint arXiv:2001.01456.

[18] Kandhro, I. A., Uddin, M., Hussain, S., Chaudhery, T. J., Shorfuzzaman, M., Meshref, H., ... & Khalaf, O. I. (2022). Impact of activation, optimization, and regularization methods on the facial expression model using CNN. Computational Intelligence and Neuroscience, 2022.

[19] Kumar Arora, T., Kumar Chaubey, P., Shree Raman, M., Kumar, B., Nagesh, Y., Anjani, P. K., ... & Debtera, B. (2022). Optimal facial feature based emotional recognition using deep learning algorithm. Computational Intelligence and Neuroscience, 2022(1), 8379202.

[20] Koonce, B., & Koonce, B. (2021). ResNet 50. Convolutional neural networks with swift for tensorflow: image recognition and dataset categorization, 63-72.

[21] Dey, R., & Salem, F. M. (2017, August). Gate-variants of gated recurrent unit (GRU) neural networks. In 2017 IEEE 60th international midwest symposium on circuits and systems (MWSCAS) (pp. 1597-1600). IEEE.

Efficient Parallel Algorithm for Extracting Fuzzy-Crisp Formal Concepts

Ebtesam Shemis[1], Arabi Keshk[2], Ammar Mohammed[3], Gamal Elhady[4]
Department of Computer Science-Faculty of Computers and Information, Menoufia University, Shebin El-Kom City, Egypt[1, 2, 4]
Department of Computer Science-Faculty of Graduate Studies for Statistical Research, Cairo University, Giza, Egypt[3]

***Abstract*—Fuzzy Formal Concept Analysis (FFCA) is a robust mathematical tool for analyzing data, particularly where uncertainty or fuzziness is inherent. FFCA is utilized across various domains, including data mining, information retrieval, and knowledge representation. However, fuzzy concepts extraction is a crucial yet computationally intensive task. This paper addresses the challenge of time efficiency in extracting single-sided fuzzy concepts from large datasets. A parallel algorithm is proposed to reduce computational time and optimize resource utilization, thus enabling the scalable analysis of expanding datasets. By computing fuzzy concepts across multiple threads in parallel, each thread processes an attribute independently to extract fuzzy concepts, which are then merged in the final step. The proposed algorithm extracts fuzzy-crisp concepts, which are more concise than other types of fuzzy concepts. Experiments were conducted to evaluate the performance of the proposed parallel algorithm against existing sequential methods. Experimental results demonstrate significant gains in computational efficiency, with the algorithm achieving an average time reduction of 68% compared to the attribute-based algorithm and up to 83%-time reduction compared to the fuzzy CbO algorithm across various types of datasets, including binary, quantitative, and fuzzy.**

Keywords*—Fuzzy Formal Concept Analysis; single-sided fuzzy concept; fuzzy-crisp concepts; parallel algorithm; fuzzy concepts extraction; knowledge representation

I. Introduction

Formal Concept Analysis (FCA) is a mathematical framework that involves the systematic study of formal concepts and their interrelationships, providing a structured approach to analyzing complex data sets [1] [2]. Therefore, FCA has found diverse applications across various domains, including linguistics [3], information retrieval [4], bioinformatics [5], Text classification [36], and beyond [6][34]. Traditional FCA algorithms can process only binary datasets [15], namely formal context, to extract formal concepts, constituting logical clusters of related objects and attributes [7]. This framework is based on the principles of crisp set theory, which categorizes objects as either fully belonging to a set or completely outside of it, emphasizing clear boundaries in the conceptual space [35]. Besides, classical FCA algorithms use the crisp scaling method to handle quantitative or qualitative data (many-valued contexts MVC) [8]. As a result, the crisp boundary problem in crisp scaling introduces challenges in accurately defining the boundaries between sub-attributes, affecting the precision of the analysis [9].

Fuzzy Formal Concept Analysis (FFCA) effectively addresses the limitations of handling many-valued and fuzzy contexts by managing ambiguity and uncertainty similarly to human cognition[10]. It represents data on a spectrum rather than in binary terms, allowing for nuanced and accurate analysis of large datasets. FFCA captures the complexity of real-world data, leading to more meaningful insights and clear, interpretable results. Its flexibility allows adaptation to various data mining tasks, making it ideal for applications like customer feedback analysis, where understanding human sentiment is essential. Additionally, FFCA's ability to represent imprecise data aligns with subjective user views, making it suitable for everyday data handling [11] [28]. For example, in the education domain, a university using FFCA to analyze student course evaluations can handle nuanced feedback such as "somewhat satisfied", "mostly satisfied," and "highly satisfied", rather than just "satisfactory" or "unsatisfactory." This deeper understanding of student opinions allows for more targeted improvements in teaching methods and course content, enhancing the overall educational experience.

Extracting the entire set of fuzzy concepts from large datasets is a time-consuming task, recognized as a #P-complete problem [12]. However, if the relationship between object and attribute sets is sparse, even in large datasets, the complexity of this process can be reduced [11]. The generation time of classical formal concepts is often notably shorter than the extraction time of fuzzy concepts. This discrepancy arises due to the inherent complexity involved in computing fuzzy concepts, which typically requires additional computational resources and processing steps [13]. FCA primarily deals with crisp binary relations, leading to relatively faster computations, whereas FFCA involves handling fuzzy relations, which require more complex calculations to derive fuzzy concepts. Therefore, while FFCA handles different data types effectively, it requires extensive time than the traditional FCA that hinder its applicability across various domains. Addressing this research gap, the proposed algorithm significantly enhances the efficiency of generating fuzzy concepts by leveraging parallel processing techniques.

The proposed algorithm differs from In-Close4b [32], bit-close4 [30], and FPCbO [31] in its ability to effectively process a variety of data types. While these algorithms are limited to handling binary data, real-world applications typically involve heterogeneous datasets. In contrast, the proposed algorithm leverages fuzzy set theory, enabling it to process diverse data types, including crisp, vague, and quantitative.

FuzzyInClose4b algorithm [24] generates concepts where both the extent and intent are fuzzy, resulting in an excessive number of fuzzy concepts. This abundance hinders its practical application in data-intensive domains such as association rule extraction, semi-automatic ontology construction, and information retrieval [16]. Conversely, the proposed algorithm generates fuzzy concepts with fuzzy extents and crisp intents, reducing the number of extracted fuzzy concepts. Additionally, the proposed algorithm leverages parallel processing to significantly reduce execution time. These enhancements improve the algorithm's applicability and efficiency.

Through this research, a time-efficient algorithm is proposed for extracting fuzzy concepts from large datasets utilizing parallel processing. The main contributions of this work are summarized as follows:

1) The proposed algorithm utilizes fuzzy logic in extracting fuzzy-crisp concepts from various data types, unlike existing crisp algorithms such as In-Close4b [32], bit-close4 [30], and FPCbO [31], which only process binary data.

2) The proposed algorithm generates more concise fuzzy concepts compared to FuzzyInClose4b [24], facilitating the practical application of FFCA in real-world scenarios.

3) The proposed algorithm reduces the computational time of extracting fuzzy concepts by leveraging multithreading in processing multiple attributes simultaneously to extract fuzzy concepts in parallel, thereby enhancing efficiency and performance. Therefore, it utilizes the available resources for faster computation of fuzzy concepts.

The remainder of this paper is organized as follows: Section II offers a comprehensive review of fundamental notions and definitions concerning formal concept analysis (FCA) and its fuzzy variant. Subsequently, Section 0 takes a closer look at the related algorithms employed in the derivation of formal concepts. Next, Section IV presents the proposed parallel algorithm for generating fuzzy formal concepts, clarifying its operation. Section V introduces a practical case that utilizes the proposed algorithm to see how well it performs in practical in comparison to other algorithms. Section VI evaluates the proposed algorithm across different types of data, presenting a comparative study with other related algorithms. This section highlights the superior performance of the proposed algorithm through detailed analysis. Lastly, Section TABLE IX. provides a summary of the main findings, highlights the key contributions, and discusses avenues for future research and development.

II. PRELIMINARIES

This section provides an overview of the basic definitions and notions related to classical Formal Concept Analysis (FCA) and its fuzzy variant FFCA. Further definitions and foundations can be found in the references [14] for FCA and [17], [18] for FFCA [17].

A. Classical Formal Concept Analysis

Formal concept analysis primarily aims to identify groupings of entities and their associated attributes within a given dataset. Typically, FCA operates on binary formal contexts as its input, wherein a crisp relation I outlines the associations between entities G and their attributes M. Accordingly, a formal context can be concisely characterized as a tabular representation comprising rows denoting entities and columns representing attributes (or vice versa).

Definition 1: A formal context is delineated as a triple $\mathbb{K} = (\mathrm{G}, \mathrm{M}, \mathrm{I} \in \{0,1\})$, where G signifies the set of objects, M denotes the set of attributes, and I denotes a crisp relation among objects and attributes, where $\mathrm{I} \subseteq \mathrm{G} \times \mathrm{M}$. If an object $\mathrm{g} \in \mathrm{G}$ and an attribute $\mathrm{m} \in \mathrm{M}$ are related in I, it is expressed as the pair $(\mathrm{g}, \mathrm{m}) \in \mathrm{I}$ or $(\mathrm{g}\ \mathrm{I}\ \mathrm{m})$.

Definition 2: A formal concept is denoted as a pair (A, B) of the formal context $\mathbb{K} = (\mathrm{G}, \mathrm{M}, \mathrm{I})$ iff $\mathrm{A} \subseteq \mathrm{G}$, $\mathrm{B} \subseteq \mathrm{M}$, $\mathrm{A}\uparrow = \mathrm{B}$, and $\mathrm{B}\downarrow = \mathrm{A}$, where A and B the concept extent and intent, respectively. And A↑, B↓ are given by Eq. (1) and Eq. (2):

$$A\uparrow := \{ m \in M \mid (g,m) \in I \ \forall g \in A \} \quad (1)$$

Such that $A\uparrow$ represents the set of characteristics in M that are shared by all objects in A.

$$B\downarrow := \{ g \in G \mid (g,m) \in I \ \forall m \in B \} \quad (2)$$

Given that $B\downarrow$ is set of objects having all features in the set B.

Classical FCA is typically proficient in processing crisp binary contexts, a scenario that may not align with the inherent complexity of real-world datasets. One challenge encountered pertains to specifying sharp boundaries within scaled attribute intervals, thus posing limitations in effectively addressing the many valued contexts inherent in many datasets.

Fuzzy Formal Concept Analysis (FFCA), in contrast, offers a versatile framework capable of accommodating various types of data, thereby addressing the limitations associated with classical FCA. Unlike its crisp counterpart, fuzzy FCA embraces the inherent uncertainty and vagueness present in real-world datasets by allowing for graded membership degrees. FFCA can handle different types of data (continuous, discrete, or hybrid) types without needing strict boundaries between attribute intervals. Consequently, FFCA offers a more robust and adaptable approach for analyzing complex datasets characterized by imprecision and ambiguity.

B. Fuzzy Formal Concept Analysis

Fuzzy Formal Concept Analysis (FFCA) is a mathematical framework designed to analyze and extract meaningful patterns from datasets characterized by uncertainty and imprecision. Fuzzy sets [33] in FFCA replace the conventional binary representation of relationships between objects and attributes where a value in the range [0, 1] represents the degree of membership of an object to an attribute.

Definition 3: a fuzzy formal context is a triple $\hat{\mathbb{K}} = (\mathrm{G}, \mathrm{M}, \hat{\mathrm{I}} \in [0,1])$, where G is a set of objects, M is a set of attributes, $\hat{\mathrm{I}}$ is a fuzzy relation between objects and attributes each with membership μ that gives each pair (g, m) in $\mathrm{G} \times \mathrm{M}$ a degree of membership. This membership degree shows how strongly objects and attributes are linked.

FFCA extends traditional FCA by allowing for the representation of graded relationships, enabling a more flexible

and nuanced analysis of complex datasets. One-sided FFCA involves the fuzzification of either the extent or the intent of a formal concept while maintaining a crisp definition on the other side [10][17].

As depicted in Fig. 1, various viewpoints and formal definitions for fuzzy concepts exist. Fuzzy concepts can be categorized into single-sided fuzzy concepts and full-sided fuzzy concepts. In single-sided fuzzy concepts, only the intent or extent of the fuzzy concept is fuzzy, not both [26]. On the other hand, full-sided fuzzy concepts have both extent and extent represented as fuzzy sets [24]. This paper proposes a parallel algorithm for extracting single-sided fuzzy concepts in which extents are fuzzy and intents are crisp. The rest of this section presents the formal definitions and notions for this type of FFCA.

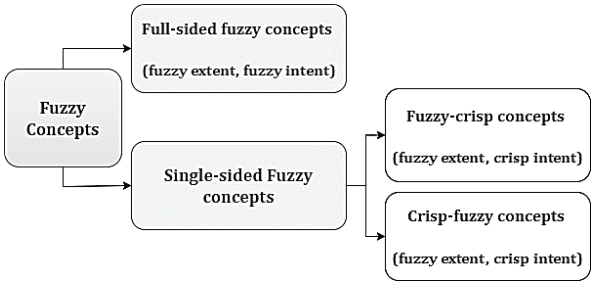


Fig. 1. Fuzzy concepts' taxonomy.

Definition 4: In a fuzzy context $\mathbb{\hat{K}} = (G, M, \hat{I})$, a fuzzy-crisp concept is represented by a pair (A_f, B), where $A_f \subseteq G$ denotes the fuzzy extent of the concept, $B \subseteq M$ represents the crisp intent, and $A_f \uparrow = B$ and $B \downarrow = A$. The definitions of $A_f \uparrow$ and $B \downarrow$ are provided by Eq. (3) and Eq. (4) [17]:

$$A_f \uparrow = \{ b \in B \mid \forall a \in A: \mu_I(a, b) \geq \mu_A(a) \} \quad (3)$$

$$B \downarrow := \{ a \mid \mu_A(a) = min_{\{b \in B\}} (\mu_I(a, b)) \} \quad (4)$$

In this context, $A_f \uparrow$, denoted as $A_f{}'$, delineates the crisp intent of the object set A_f, while $B \downarrow$, denoted as B', delineates the fuzzy extent of the attribute set B. A confidence threshold α is utilized where $\alpha < 1$. The user can adjust the threshold α according to their requirements for generating fuzzy concepts from a provided fuzzy context [20]. This threshold aids in filtering out fuzzy relations that fall outside the interval $[\alpha, 1]$ from the given fuzzy formal context, thereby facilitating knowledge discovery and representation [19] [21] [28]. Alternatively, the threshold interval is denoted as α-cut of the fuzzy context.

Example 1: Table I shows a fuzzy context example that refers to a part of the employee's dataset. According to Def. 3, the attribute set M is $\{A, B, C, D, E, F\}$, and each attribute corresponds to the properties: A: low salary, B: moderate salary, C: high salary, D: young, E: youth, and F: old. Moreover, the object set $G = \{O_0, O_1, O_2, O_3, O_4\}$ represents five different employee instances. Applying an $\alpha\text{-}cut$ of 0.4 to the fuzzy context in Table I produces the fuzzy context in Table II. Eliminated relationships are highlighted in gray background in Table II. Applying the threshold clearly produces a sparser context and can effectively eliminate unwanted relationships.

TABLE I. EXAMPLE FUZZY CONTEXT

Obj	A	B	C	D	E	F
O_0	1	0	0	1	0	0
O_1	0.22	0.78	0	1	0	0
O_2	0	0.33	0.67	0.73	0.27	0
O_3	0	0	1	0	0.36	0.64
O_4	0	0	1	0	0	1

TABLE II. FUZZY CONTEXT WITH $\alpha\text{-}cut = 0.4$

Obj	A	B	C	D	E	F
O_0	1	0	0	1	0	0
O_1	0	0.78	0	1	0	0
O_2	0	0	0.67	0.73	0	0
O_3	0	0	1	0	0	0.64
O_4	0	0	1	0	0	1

For more clarity, Table III illustrates the fuzzy concepts extracted from the fuzzy context in Table II after applying the threshold $\alpha = 0.4$. As shown in Table III, there are ten fuzzy concepts, each consisting of a fuzzy extent and a crisp intent.

TABLE III. FUZZY CONCEPTS GENERATED FROM THE FUZZY CONTEXT IN TABLE II

#	Fuzzy Extent	Crisp Intent
C_1	$\{O_0: 1, O_1: 1, O_2: 1, O_3: 1, O_4: 1\}$	{}
C_2	$\{O_0: 1\}$	{A, D}
C_3	{}	{A, B, C, D, E, F}
C_4	$\{O_1: 0.78, O_2: 0.33\}$	{B, D}
C_5	$\{O_2: 0.33\}$	{B, C, D}
C_6	$\{O_2: 0.67, O_3: 1, O_4: 1\}$	{C}
C_7	$\{O_2: 0.67\}$	{C, D}
C_8	$\{O_3: 0.36\}$	{C, E, F}
C_9	$\{O_3: 0.64, O_4: 1\}$	{C, F}
C_{10}	$\{O_0: 1, O_1: 1, O_2: 0.73\}$	{D}

A fuzzy extent consists of a set of objects and their corresponding memberships in the form of $\{O_i: \mu_{o_i}\}$, such that O_i refers to the object number i and μ_{o_i} is the membership of O_i to the corresponding crisp intent. For instance, fuzzy concept (4) is $(\{O_1: 0.78, O_2: 0.33\}, \{B, D\})$ where $(\{O_1: 0.78, O_2: 0.33\}$ is the fuzzy extent consisting of objects O_1 and O_2 with membership degrees of 0.78 and 0.33, respectively. These membership degrees determine to what extent the objects possess the attributes in the crisp intent $\{B, D\}$.

III. RELATED WORKS

Extracting formal concepts from formal contexts has gained a significant interest in classical algorithms that handle binary datasets using crisp set theory. In-Close4b [32], bit-close4 [30], and FPCbO [31] algorithms identify crisp concepts through depth-first search and pruning techniques. With a mix of arrays and bitsets, In-Close4b balances memory usage and speed,

making it suitable for large datasets. Bit-Close4 enhances performance for sparse data by optimizing memory efficiency with bitset data structures. FPCbO [31] has improved the classical CbO algorithm [23] by reducing computational redundancy and employing a canonical direct basis representation to generate only canonical concepts. Regarding optimizations, these algorithms differ in the following ways: In-Close4b balances resources; Bit-Close4 emphasizes memory efficiency; and FPCbO focuses on redundancy reduction.

Compared to the proposed algorithm, these crisp algorithms are less adaptable to diverse data types and cannot handle fuzzy or imprecise data, limiting their applicability in real-world scenarios where data uncertainty is common. While In-Close4b balances memory usage and speed, it struggles with extremely large or dense datasets, and its complexity makes it harder to implement. Bit-Close4's reliance on bitsets enhances performance for sparse matrices but can reduce flexibility and efficiency in dense datasets. FPCbO's strict canonicality minimizes redundancy but introduces significant computational overhead and increased memory consumption, making it less efficient for complex datasets. Overall, these limitations highlight the need for more versatile and scalable algorithms, offered by fuzzy concept extraction algorithms.

Fuzzy FCA algorithms gained less popularity due to its extensive time requirement despite its capability to process diverse data types, including fuzzy and imprecise data. This section reviews current fuzzy algorithms, identifying key limitations that the proposed algorithm aims to address.

Fuzzy CbO [22] presents an algorithm to extract fuzzy-crisp concepts. Therefore, fuzzy CbO can process fuzzy and vague data effectively. In addition, this recursive algorithm exploits canonicity tests to discover new fuzzy concepts, demonstrating superior efficacy in fuzzy concept generation when compared to the fuzzy NextClosure algorithm[25]. Despite its effectiveness in processing fuzzy data, the fuzzy CbO algorithm faces limitations such as significant computational overhead from canonicity tests, high memory consumption due to its recursive nature, and potential challenges in handling large and diverse datasets. These limitations are efficiently handled by the proposed parallel algorithm.

The attribute-based algorithm [27] excels with symmetrically correlated attributes, outperforming fuzzy CbO in execution time under these conditions. However, in other scenarios, both algorithms demonstrate similar performance, limiting the attribute-based algorithm's effectiveness in handling diverse datasets and asymmetric correlations. On the contrary, the proposed algorithm leverages parallel processing techniques to handle multiple attributes at the same time, which results in efficient processing of fuzzy concepts. Unlike sequential algorithms like Fuzzy CbO and attribute-based algorithm, the proposed algorithm excels in handling large, dense datasets and produces compact fuzzy-crisp concepts efficiently.

FuzzyInClose4b algorithm[24] utilizes incremental closure and matrix searching techniques to extract full-sided fuzzy concepts. It computes each closure incrementally, only once per concept, to prevent repeated closure calculations. Compared to the proposed algorithm, FuzzyInClose4b[24] generate more fuzzy concepts. Therefore, it requires plenty of time and storage space. But it may be more suitable when the intent's membership values are essential. But for Ontology construction applications and association rule mining, it can be sufficient to use fuzzy-crisp concepts algorithms like the proposed algorithm.

To the best of knowledge, all existing algorithms for generating fuzzy-crisp formal concepts operate sequentially, failing to leverage parallel computation. This limitation significantly impacts execution time, particularly with large datasets. This work addresses this gap by proposing a parallel algorithm to accelerate the generation of fuzzy concepts.

IV. Proposed Parallel Algorithm for Computing Fuzzy-Crisp Concepts

This section introduces the proposed parallel fuzzy concept (PFC) algorithm to generate fuzzy-crisp concepts from a fuzzy formal context in parallel, leveraging multi-threading to improve efficiency. The algorithm accesses the fuzzy context and eventually returns fuzzy concepts as pairs of fuzzy extents and crisp intents.

Algorithm: PFC( )

Input: A fuzzy formal context $\widehat{\mathbb{K}} = (\mathrm{G}, \mathrm{M}, \hat{\mathrm{I}} \in [0,1])$
Output: The set of FC of all fuzzy concepts of $\widehat{\mathbb{K}}$
Begin
1 Initialize FC ← $\{(G\uparrow, G), (M\downarrow, M)\}$
2 With ThreadPoolExecutor() as executor:
3 For each i from 0 to $|M|-1$ do
4 executor.map(ProcessAttribute, i)
5 End for
6 Return FC
Procedure: ProcessAttribute ($attr_i$)
1 F_extent ← $attr_i\downarrow$
2 intent ← F_extent ↑
3 If intent ⊈ FC.intents then:
4 FC← $FC \cup$ (F_extent , intent)
5 With threading.lock():
6 For c in FC \ {(F_extent, intent) ∪ (G, G↓)∪ ($M\downarrow$, M)}
7 Inters = Fextent ∧ c.extent
8 If Inters ↓ ⊈ FC.intents then:
9 $FC \leftarrow FC \cup (\text{Inters}, \text{Inters}\downarrow)$
10 End If
11 End For
12 End If
End Procedure

The Parallel Fuzzy Concept (PFC) algorithm begins by initializing the global fuzzy concepts set with the bottom and top concepts (line 1). The top concept is a set of all objects G and their shared intent $G\uparrow$ evaluated using Eq. (3). The bottom concept is a set of whole intent M and all objects that assess M, given by $M\downarrow$ provided using Eq. (4), and represents the fuzzy extent in the form $ext: \mu_{ext}$. Subsequently, a thread pool is employed to process attributes in parallel (Lines 2–5), where each thread is responsible for computing all fuzzy concepts discoverable using a particular attribute. Therefore, the algorithm ensures efficient computation of fuzzy concepts using multiple threads.

To identify all new fuzzy concepts given an attribute, the ProcessAttribute() procedure is invoked. First, lines $1-2$ calculate the fuzzy extent using Eq. (4) and the crisp intent using Eq. (5). Next, line 3 verifies that the concept intent is new and not present in the FC set, thereby storing the new fuzzy concept in FC (line 4). In line 5, a locking mechanism ensures thread safety during updates. Lines 6–9 then uncover all other fuzzy concepts that arise from intersections between the new fuzzy extent (F_extent) and the existing concepts' extents. Therefore, line 6 presents a loop that iterates over all existing concepts except all fuzzy concepts that cannot produce new concepts via the fuzzy intersection (i.e., the newly generated concept {F_extent, intent}, the top concept $(G, G\uparrow)$, and the bottom concept $(M\downarrow, M)$. In line 7, Zadah's intersection, as stated in definition (5), evaluates the fuzzy intersection between the extent of the new concept and the extent of the existing concept. If the intent of an intersection is not in FC, it adds the new concept to FC (lines 8 and 9).

Definition 5: The intersection of two fuzzy sets A and B with membership functions $\mu_A(x)$ and $\mu_B(x)$, respectively, results in a fuzzy set C, denoted as $C = A \cap B$. The membership function of C is defined as $\mu_C(x) = \mathrm{Min}[\mu_A(x), \mu_B(x)]$ for all $x \in X$ in the universe of discourse X. This can also be represented as $C = A \wedge B$.

The proposed algorithm significantly enhances efficiency by leveraging parallel processing technique (multi-threading), making it well-suited for large datasets. The combination of parallel attribute processing and intersection checking ensures the identification of all fuzzy concepts within the fuzzy context.

Unlike existing algorithms for extracting full-sided fuzzy concepts, such as FuzzyInClose4b [24], the proposed algorithm generates fuzzy-crisp concepts that are more compact and suitable for information retrieval, ontology construction and association rule mining. Additionally, the parallel nature of the proposed algorithm makes it more efficient in extracting fuzzy-crisp concepts from large dense datasets, a feature that is not present in existing sequential fuzzy-crisp concept generation algorithms like Fuzzy CbO [22], Fuzzy NextClosure [25], and attribute-based algorithm [27].

V. Case Study and Discussion

This section demonstrates the practical application of the proposed PFC algorithm in analyzing a fuzzy keyword-document context. This context involves eight keywords (k_1 to k_8) and four documents (d_1 to d_4), where fuzzy values represent the strength of relationships between keywords and documents. Keywords are treated as fuzzy extents, and documents as crisp intents, quantified by fuzzy membership values ranging from 0 to 1, as shown in Table IV.

Using the proposed PFC algorithm, this fuzzy context is systematically analyzed to extract fuzzy-crisp concepts, detailed in Table V. These concepts are characterized by fuzzy extents, indicating the strength of associations between keywords and documents, and crisp intents, describing documents based on keywords.

TABLE IV. Keyword-Document Fuzzy Context

	d_1	d_2	d_3	d_4
k_1	0	1	1	0
k_2	0.2	0.6	1	0.2
k_3	1	0	0	0
k_4	0.3	0.6	0.7	0.3
k_5	1	0	0	1
k_6	0	1	0	0
k_7	0.1	0.5	0.5	0.1
k_8	0.5	0.3	0	0.5

TABLE V. Fuzzy-Crisp Fuzzy Concepts Extracted by the Proposed Algorithm

#	Fuzzy Extents (keywords)	Crisp Intents (documents)
$C1$	$\{k_0: 1, k_1: 1, k_2: 1, k_3: 1, k_4: 1, k_5: 1, k_6: 1, k_7: 1\}$	{ }
$C2$	$\{k_1: 0.2, k_3: 0.3, k_6: 0.1\}$	$\{d_1, d_2, d_3, d_4\}$
$C3$	$\{k_1: 0.2, k_2: 1, k_3: 0.3, k_4: 1, k_6: 0.1, k_7: 0.5\}$	$\{d_1\}$
$C4$	$\{k_0: 1, k_1: 0.6, k_3: 0.6, k_5: 1.0, k_6: 0.5, k_7: 0.3\}$	$\{d_2\}$
$C5$	$\{k_1: 0.2, k_3: 0.3, k_6: 0.1, k_7: 0.3\}$	$\{d_1, d_2, d_4\}$
$C6$	$\{k_0: 1, k_1: 1, k_3: 0.7, k_6: 0.5\}$	$\{d_3\}$
$C7$	$\{k_0: 1, k_1: 0.6, k_3: 0.6, k_6: 0.5\}$	$\{d_2, d_3\}$
$C8$	$\{k_1: 0.2, k_3: 0.3, k_4: 1, k_6: 0.1, k_7: 0.5\}$	$\{d_1, d_4\}$

Fig. 2 illustrates the steps of the PFC algorithm. Initially, the top and bottom fuzzy concepts (C1 and C2) are added to the fuzzy concepts set (line 1 of the algorithm). According to line 2, four threads are utilized to process different attributes (d1, d2, d3, d4) in parallel. Each thread independently processes its assigned attribute by calculating the fuzzy extent and corresponding crisp intent and checks if the newly formed concept's intent is not in the set of fuzzy concepts (FC). If it's unique, it adds a new concept to FC. Additionally, each thread performs fuzzy intersections of extents with existing concepts and ensures that new intersection concepts, if unique, are also added to FC. For example, processing attribute d_1 results in a concept with the fuzzy extent $\{k_1: 0.2,\ k_2: 1,\ k_3: 0.3, k_4: 1,\ k_6: 0.1,\ k_7: 0.5\}$ and the crisp intent $\{d_1\}$.

This process is repeated for all attributes, ensuring systematic extraction of all fuzzy concepts. This tracing highlights the algorithm's parallel behavior, where multiple threads execute concurrently to process different attributes, significantly speeding up the concept generation process by leveraging parallel computation. By analyzing these fuzzy concepts, the varying strengths of relationships between keywords and documents can easily be demonstrated.

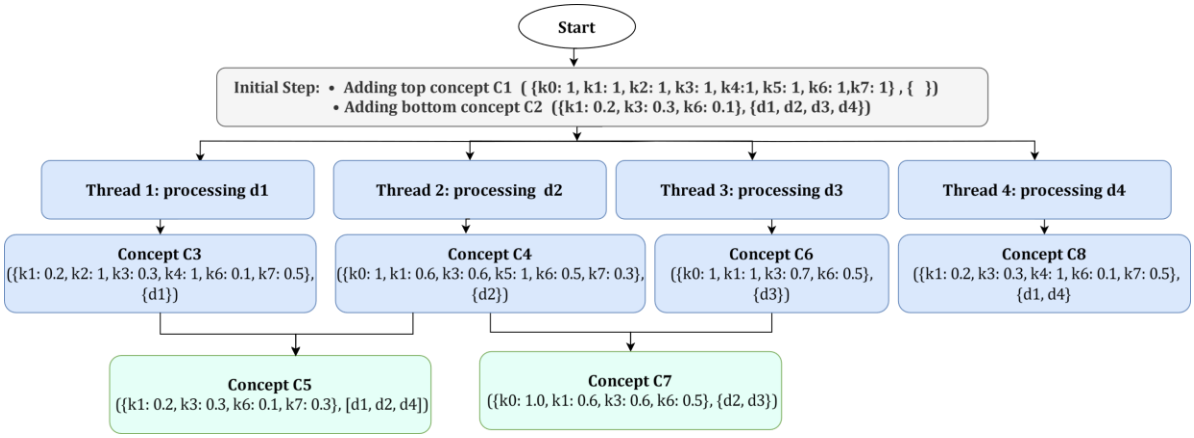


Fig. 2. Steps by the proposed PFC algorithm for extracting fuzzy concepts from fuzzy context in Table IV.

This analysis helps identify documents with strong, weak, or similar associations to multiple keywords. For instance, fuzzy concept C8 highlight that document d_1 and d_4 have a strong association with keyword k_4 and moderate to weak associations with other keywords. These keywords serve as a linking set for these two documents. Information from fuzzy-crisp concepts is valuable for tasks like information retrieval, as it improves search accuracy by highlighting documents strongly associated with the user query. The FuzzyInClose4b [24] algorithm generated 43 fuzzy concepts from Table IV in which both extent and intent are fuzzy. Example fuzzy concepts are: $(\{k_2, k_4, k_7, k_8: 0.3\},\ \{d_1: 0.1, d_2: 0.5, d_4: 0.1\})$ and $(\{k_2, k_4, k_7: 0.1, k_8: 0.3\},\ \{d_1: 0.2,\ d_2: 0.6,\ d_4: 0.2\})$ which are very similar concepts.

In this case study, fuzzy-crisp concepts extracted by the proposed PFC algorithm, which use fuzzy extents for keyword relevance and crisp intents for document associations, provide an efficient and nuanced approach. This method captures keyword importance precisely while maintaining clear document associations, leading to effective and context-aware search results. For example, in a search for "renewable energy sources," relevant keywords can be weighted, and documents are crisply linked, enhancing search efficiency. In contrast, full sided fuzzy concepts extracted by FuzzyInClose4b [24] offer more detailed relationships by representing both keywords and documents with varying degrees, but they introduce higher computational complexity and interpretation challenges with the existence of very similar fuzzy concepts. Thus, the fuzzy-crisp concepts extracted by the proposed PFC algorithm balance complexity and precision, delivering relevant and personalized search outcomes efficiently.

VI. Results and Discussion

This section shows the efficiency and scalability of the proposed PFC algorithm through experiments on synthetic and benchmark datasets. The aim of these experiments is to show that the proposed PFC algorithm significantly reduces computation time, making FFCA feasible for large-scale data analysis. Table VI provides an overview of the datasets utilized in the experiments. The Iris and red-wine datasets are benchmarks from the UCI Repository [29]. To further assess the robustness of the proposed algorithm, we synthesized three variations of a fuzzy dataset with different densities (20%, 30% and 40%) and a size of (20,000 × 15), so the performance of the proposed algorithm under various conditions can be assessed.

TABLE VI. Datasets Used in the Experiments

Dataset	$\|G\|$	$\|M\|$	Density	Description
Fuzzy Red Wine	1600	36	66.2%	Multivariate (fuzzified)
Fuzzy Iris	150	15	58%	Multivariate (fuzzified)
Synthetic Fuzzy Dataset	20,000	15	20%, 30%, 40%	Synthetic Fuzzy
Car	1,728	25	28%	Crisp

To maintain consistency with the fuzzy setting, all quantitative attributes are fuzzified using three linguistic labels per attribute: low, moderate, and high. Trapezoidal membership functions are employed to represent the low and high linguistic labels, while a triangular membership function is used for the moderate label. The proposed PFC algorithm is implemented in Python, and the experiments are carried out on a Windows machine equipped with an Intel Core i7 processor running at 2.60 GHz and 32 GB of RAM.

Before exploring experiments, Table VII presents feature comparisons between the proposed PFC algorithm and the related algorithms: In-Close4b [32], bit-close4 [30], FPCbO [31], FuzzyInClose4b [24], fuzzy CbO [22], attribute-based algorithm [27].

TABLE VII. Features Comparison

Algorithms	Data Types	Parallel	Concept Extent	Concept Intent
In-Close4b [32]	Binary	x	Crisp	Crisp
bit-close4 [30]		✓		
FPCbO [31]		✓		
FuzzyInClose4b[24]	Fuzzy	x	Fuzzy	Fuzzy
fuzzy CbO [22]		x	Fuzzy	Crisp
attribute-based [27]		x		
PFC algorithm	Fuzzy	✓		

The algorithms In-Close4b [32], bit-close4 [30], and FPCbO [31] can only extract formal concepts from binary data, such as Car dataset. They can't process other types of data, such as multivariate and fuzzy datasets. In contrast, the proposed PFC algorithm can process these data by utilizing fuzzy set theory in computing fuzzy concepts.

The FuzzyInClose4b algorithm [24] extracts full-sided fuzzy concepts, which are more numerous than fuzzy-crisp concepts, extracted by the proposed PFC algorithm. Table VIII compares the number of fuzzy concepts extracted by the FuzzyInClose4b algorithm with those extracted by the proposed PFC algorithm for the fuzzy Iris dataset considering different α-cuts. Due to the larger count of concepts, the FuzzyInClose4b algorithm is computationally intensive and demands significant storage resources.

TABLE VIII. COMPARISON OF FUZZY CONCEPTS BETWEEN FUZZY IN CLOSE4B AND PROPOSED PFC ALGORITHMS ON FUZZY IRIS DATASET

$\alpha - cut$	full-sided fuzzy concepts	Fuzzy-crisp concepts
0.8	844	73
0.7	2,738	87
0.6	8,438	110
0.5	16,843	160
0.4	46,514	252

All of fuzzy CbO [22], attribute-based algorithm [27], and the PFC algorithm extract fuzzy-crisp concepts. However, both fuzzy CbO and attribute-based algorithms operate sequentially and are computationally intensive. In contrast, the proposed PFC algorithm leverages parallel processing of attributes, significantly enhancing its efficiency. Experiments are conducted to compare the computational time (in seconds) required by fuzzy CbO, the attribute-based algorithm, and the PFC algorithm at different α-cut values applied to the dataset.

Fig. 3–6 demonstrate the persistent effectiveness of the suggested parallel algorithm in comparison to other algorithms. For all algorithms, the computation times decrease as the α-cut threshold increases. Aside from that, the proposed PFC algorithm consistently outperforms the other algorithms, achieving the shortest calculation times across all thresholds.

Fig. 3 demonstrates a performance comparison over the fuzzy Iris dataset, indicating that the suggested PFC algorithm consistently takes the least amount of time to compute all fuzzy concepts for all α-cuts applied to the dataset. Although the sequential attribute-based approach and the PFC algorithm behave similarly on this dataset, the difference in performance between them becomes more apparent when the α-cut decreases. This indicates that the suggested parallel algorithm is especially suitable for datasets with a high density.

Fig. 4, 5, and 6 display the efficiency for the synthetic fuzzy dataset with densities of 20%, 30%, and 40%, respectively. The proposed PFC algorithm consistently outperforms the fuzzy CbO and attribute-based algorithms on synthetic fuzzy datasets with different densities. It also has a stable and efficient execution time of 5 to 40 seconds, which doesn't change depending on the α-cut thresholds. On the other hand, the density of the dataset significantly influences the Fuzzy CbO and attribute-based algorithms. Regardless, they both show a decrease in execution time with increasing α-cut values but remain inefficient. The proposed parallel algorithm is always efficient, even when datasets are very dense.

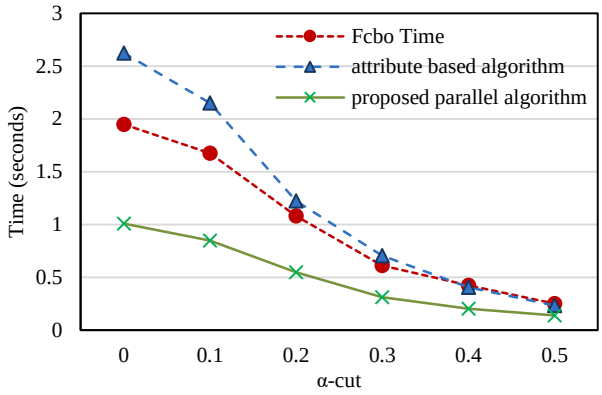


Fig. 3. Performance comparison of algorithms on fuzzy Iris dataset.

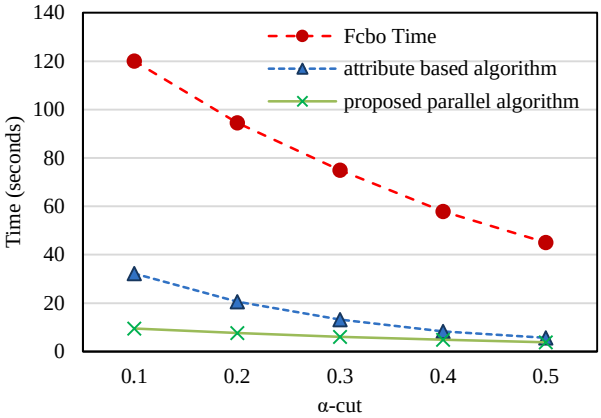


Fig. 4. Performance of algorithms on the synthetic fuzzy dataset with 20% density.

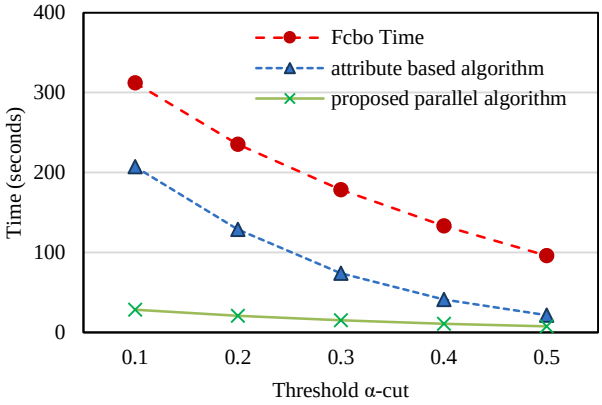


Fig. 5. Performance of algorithms on the synthetic fuzzy dataset with 30% density.

TABLE IX. EXPERIMENTS OVER DIVERSE DATASETS

Expr #	dataset	α-cut	Concepts#	Fuzzy CbO algorithm	Attribute based algorithm	Proposed parallel algorithm	Time reduction w. r. t. attribute-based algorithm	Time reduction w. r. t. Fuzzy CbO algorithm
1	Fuzzy Iris	0	730	2.15	1.38	**1.00**	27.89%	53.63%
2		0.1	693	1.63	1.13	**0.88**	22.07%	45.92%
3		0.2	589	1.11	0.68	**0.55**	19.40%	50.57%
4		0.3	376	0.65	0.40	**0.33**	16.46%	48.74%
5		0.4	252	0.42	0.24	**0.20**	18.12%	52.08%
6		0.5	160	0.26	0.16	**0.14**	12.89%	46.51%
Average reduction							**19%**	**50%**
7	synthetic fuzzy dataset with density 20%	0	7,124	187.39	32.45	**10.13**	68.79%	94.60%
8		0.1	7,124	120.14	32.18	**9.55**	70.32%	92.05%
9		0.2	5,650	94.54	20.58	**7.63**	62.95%	91.93%
10		0.3	4,397	74.93	13.25	**6.10**	53.97%	91.86%
11		0.4	3,292	57.89	8.36	**4.86**	41.86%	91.61%
12		0.5	2,503	45.14	5.63	**3.80**	32.51%	91.59%
Average reduction							**55%**	**92%**
13	synthetic fuzzy dataset with density 30%	0	17,688	431.76	210.47	**30.03**	85.73%	93.05%
14		0.1	17,688	312.39	207.38	**28.43**	86.29%	90.90%
15		0.2	13,622	235.24	128.90	**20.49**	84.10%	91.29%
16		0.3	10,353	178.58	74.03	**14.85**	79.94%	91.68%
17		0.4	7,633	132.99	41.15	**10.64**	74.14%	92.00%
18		0.5	5,630	95.93	21.43	**7.52**	64.90%	92.16%
Average reduction							**79.19%**	**91.85%**
19	synthetic fuzzy dataset with density 40%	0	28,834	734.45	673.38	**86.79**	87%	88%
20		0.1	28,834	829.09	773.33	**106.27**	86%	87%
21		0.2	24,997	679.03	632.23	**69.29**	89%	90%
22		0.3	20,229	402.54	307.51	**34.35**	89%	91%
23		0.4	15,045	255.41	164.50	**22.35**	86%	91%
24		0.5	10,652	174.43	77.96	**14.57**	81%	92%
Average reduction							**86.49%**	**89.92%**
25	Fuzzy Red Wine	0.9	356	3.34	0.49	**0.39**	**20%**	**88%**
26		0.8	2,909	19.82	13.06	**2.90**	**78%**	**85%**
27		0.7	9,822	85.13	250.43	**14.68**	**94%**	**83%**
28		0.6	27,455	454.51	2481.24	**62.15**	**97%**	**86%**
Average reduction							**62.71%**	**72.09%**
29	car	0	12,640	95.92	461.29	**5.59**	**98.79%**	**94.17%**

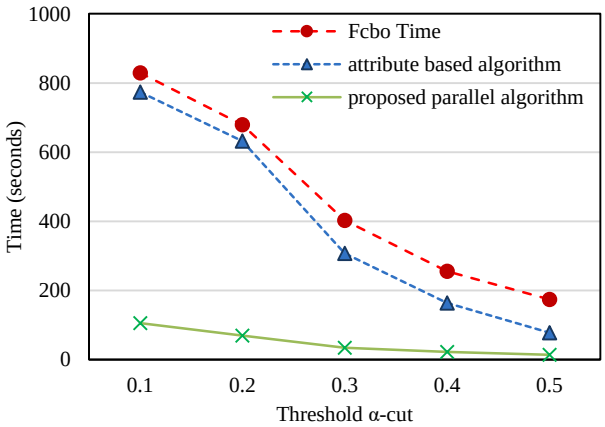


Fig. 6. Performance of algorithms on the synthetic fuzzy dataset with 40% density.

Table IX compares the proposed parallel algorithm's efficiency across diverse datasets, showing significant time reductions compared to the attribute-based and Fuzzy CbO algorithms. For instance, in the synthetic fuzzy dataset with 20% density, the proposed algorithm achieved an average time reduction of 55% compared to the attribute-based algorithm and 92% compared to the Fuzzy CbO algorithm, demonstrating its robustness and superior performance.

VII. CONCLUSION

Fuzzy formal concepts are essential to reveal the fundamental structures and patterns in heterogeneous datasets, thus enabling efficient decision-making in several fields. However, due to the complexity and ambiguity inherent in fuzzy datasets, extracting fuzzy concepts is computationally intensive. This inefficiency can hinder the practical application of fuzzy concepts in the real world. To address this challenge, the

proposed novel parallel algorithm optimizes the extraction of fuzzy-crisp concepts from fuzzy datasets, exploiting the resources at disposal. The proposed PFC algorithm utilized multi-threading to process attributes in parallel, then merge extracted fuzzy concepts. Experiments witness that the proposed algorithm reduces computation times and improves scalability, making it more suitable for handling dense and complex data structures. Besides, performance evaluation of the proposed PFC algorithm against other fuzzy algorithms across multiple datasets has demonstrated the algorithm's consistent superiority, achieving the shortest processing times across all α-cuts, with notable efficiency improvements as the α-cut decreases.

Future research should enhance the algorithm's applicability in various computational environments, including big data processing frameworks like Apache Spark, and explore its adaptation for distributed computing environments, allowing workload division and reduced computation times. Besides, a comprehensive evaluation of the algorithm's performance across a wider range of datasets and use cases will provide deeper insights into its strengths and limitations, guiding future enhancements and adaptations. Besides, the algorithm's efficient fuzzy concept extraction could revolutionize applications in data processing and decision-making, including online recommendation systems and dynamic risk assessment.

REFERENCES

[1] R. Wille, “Restructuring lattice theory: an approach based on hierarchies of concepts”, *Ordered sets. Springer*. pp. 445-470, 1982.

[2] B. Breckner, C. Săcărea, and R. R. Zavaczki, "Improving User’s Experience in Exploring Knowledge Structures: A Gamifying Approach," *Mathematics*, vol. 10, no. 5, p. 709, 2022.

[3] M. Bogatyrev and D. Orlov, "Application of formal contexts in the analysis of heterogeneous biomedical data," In: *CEUR Workshop Proc.*, vol. 2648, pp. 315-329, 2020.

[4] M. Tropmann-Frick, "Improvement of Searching for Appropriate Textual Information Sources Using Association Rules and FCA," in *Information Modelling and Knowledge Bases XXXIII*, vol. 343, p. 204, 2022.

[5] S. Roscoe et al., "Formal concept analysis applications in bioinformatics," *ACM Computing Surveys*, vol. 55, no. 8, pp. 1-40, 2022.

[6] C. M. Rocco, E. Hernandez-Perdomo, and J. Mun, "Introduction to formal concept analysis and its applications in reliability engineering," *Reliability Engineering & System Safety*, vol. 202, p. 107002, 2020.

[7] L. Zhao, L. X. Lu, and W. Yao, "Generalized three-way formal concept lattices," *Soft Computing*, vol. 27, no. 16, pp. 11219-11226, 2023.

[8] R. Belohlavek, E. Sigmund, and J. Zacpal, "Evaluation of IPAQ questionnaires supported by formal concept analysis," *Information Sciences*, vol. 181, no. 10, pp. 1774-1786, 2011.

[9] K. Sumangali and A. K. Ch, "Concept lattice simplification in formal concept analysis using attribute clustering," *Journal of Ambient Intelligence and Humanized Computing*, vol. 10, no. 6, pp. 2327-2343, 2019.

[10] E. Shemis, G. Elhady, A. Mohammed and A. Keshk, "A Fuzzy-Crisp Frequent Concept Lattice Generation Algorithm," 2022 2nd International Mobile, Intelligent, and Ubiquitous Computing Conference (MIUCC), IEEE, Cairo, Egypt, 2022, pp. 75-80, doi: 10.1109/MIUCC55081.2022.9781692.

[11] E. Shemis, A. Gadallah, and H. Hefny, "A Data-Sensitive Approach for Fuzzy Concept Extraction," *International Journal of Intelligent Engineering & Systems*, vol. 11, no. 5, 2018.

[12] E. Shemis and A. Mohammed, "A comprehensive review on updating concept lattices and its application in updating association rules," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 11, no. 2, p. e1401, 2021.

[13] J. Poelmans, D. I. Ignatov, S. O. Kuznetsov, and G. Dedene, "Fuzzy and rough formal concept analysis: a survey," *International Journal of General Systems*, vol. 43, no. 2, pp. 105-134, 2014.

[14] B. Ganter and R. Wille, "Formal concept analysis: mathematical foundations," *Springer Science & Business Media*, 2012.

[15] R. Wille, "Formal concept analysis as mathematical theory of concepts and concept hierarchies," in *Formal concept analysis: Foundations and applications*, B. Ganter and S. Obiedkov, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2005, pp. 1-33.

[16] C. Aswani Kumar, S. Mouliswaran, P. Amriteya, and S. Arun, "Fuzzy formal concept analysis approach for information retrieval," in *Proceedings of the Fifth International Conference on Fuzzy and Neuro Computing (FANCCO-2015)*, Springer International Publishing, 2015, pp. 255-271.

[17] A. Majidian, T. Martin, and M. E. Cintra, "Fuzzy formal concept analysis and algorithm," in *Proceedings of the 11th UK Workshop on Computational Intelligence*, Citeseer, September 2011, pp. 61-67.

[18] S. Boffa, "Extracting concepts from fuzzy relational context families," IEEE Transactions on Fuzzy Systems, vol. 31, no. 4, pp. 1202-1213, 2022.

[19] P.K. Singh, C. Aswani Kumar, and J. Li, "Knowledge representation using interval-valued fuzzy formal concept lattice," *Soft Computing*, vol. 20, pp. 1485–1502, 2016, doi: 10.1007/s00500-015-1600-1.

[20] W. Khemili, J. E. Hajlaoui, and M. N. Omri, "Energy aware fuzzy approach for placement and consolidation in cloud data centers," *Journal of Parallel and Distributed Computing*, vol. 161, pp. 130-142, 2022.

[21] A. Izadpanahi, "Relational Approach to the L-Fuzzy Concept Analysis," master’s thesis, *Brock Uni*versity, St. Catharines, ON, Canada, 2023.

[22] T. Martin and A. Majidian, "Finding Fuzzy Concepts for Creative Knowledge Discovery," in *IEEE Transactions on Intelligent Systems*, vol. 28, pp. 93-114, 2013. https://doi.org/10.1002/int.21576

[23] P. Krajca, J. Outrata, and V. Vychodil, "Parallel algorithm for computing fixpoints of Galois connections," *Annals of Mathematics and Artificial Intelligence*, vol. 59, no. 2, pp. 257-272, 2010.

[24] D. López-Rodríguez, Á. Mora, and M. Ojeda-Hernández, "Revisiting Algorithms for Fuzzy Concept Lattices," in *Proceedings of the International Conference on Concept Lattices and Their Applications (CLA)*, pp. 105-116, 2022.

[25] M. E. Cintra, M. C. Monard, H. A. Camargo, T. P. Martin, and A. Majidian, "On rule generation approaches for genetic Fuzzy Systems," in *Proceedings of the Congresso da Sociedade Brasileira de Computação*, 2011.

[26] T. J. Li and Y. Q. Wang, "Crisp-Fuzzy Concept Lattice Based on Interval-Valued Fuzzy Sets," *in Proceedings of the International Joint Conference on Rough Sets (IJCRS)*, vol. 14481, Springer, Cham, 2023, pp. 31. https://doi.org/10.1007/978-3-031-50959-9_31

[27] E. E. Shemis and A. M. Gadallah, "Enhanced Algorithms for Fuzzy Formal Concepts Analysis," in *Proceedings of the International Conference on Advanced Intelligent Systems and Informatics 2016 (AISI 2016)*, A. Hassanien, K. Shaalan, T. Gaber, A. Azar, and M. Tolba, Eds. Cham, Switzerland: Springer, 2017, vol. 533, Advances in Intelligent Systems and Computing, pp. 865-875. doi: 10.1007/978-3-319-48308-5_75.

[28] G. Fenza, M. Gallo, V. Loia, A. Petrone, and C. Stanzione, "Concept-drift detection index based on fuzzy formal concept analysis for fake news classifiers," *Technological Forecasting and Social Change*, vol. 194, p. 122640, 2023.

[29] M. Kelly, R. Longjohn, and K. Nottingham, "The UCI Machine Learning Repository." [Online]. Available: https://archive.ics.uci.edu. Accessed: May 2024.

[30] Y. Ke, J. Li, and S. Li, "Bit-Close: a fast incremental concept calculation method," *Applied Intelligence*, pp. 1-12, 2024.

[31] L. Zou, T. He, and J. Dai, "A new parallel algorithm for computing formal concepts based on two parallel stages," *Information Sciences*, vol. 586, pp. 514-524, 2022.

[32] S. Andrews, "Making use of empty intersections to improve the performance of CbO-type algorithms," in *Formal Concept Analysis: 14th International Conference, ICFCA 2017, Rennes, France,*

June 13-16, 2017, Proceedings 14, Springer International Publishing, 2017, pp. 56-71.

[33] L. A. Zadeh, "Fuzzy sets," *Information and Control*, vol. 8, no. 3, pp. 338-353, 1965. doi: 10.1016/S0019-9958(65)90241-X.

[34] A. Daghouri, K. Mansouri, and M. Qbadou, "Formal Concept Analysis based Framework for Evaluating Information System Success," *Int. J. Adv. Comput. Sci. Appl.*, vol. 10, no. 4, 2019, doi: 10.14569/IJACSA.2019.0100451.

[35] M. Alwersh and L. Kovács, "K-Means Extensions for Clustering Categorical Data on Concept Lattice," *Int. J. Adv. Comput. Sci. Appl.*, vol. 14, no. 9, 2023, doi: 10.14569/IJACSA.2023.0140953.

[36] S. Puri, "A Fuzzy Similarity Based Concept Mining Model for Text Classification," *Int. J. Adv. Comput. Sci. Appl.*, vol. 2, no. 11, 2011, doi: 10.14569/IJACSA.2011.021119

Automatic Plant Disease Detection System Using Advanced Convolutional Neural Network-Based Algorithm

Sai Krishna Gudepu[1]*, Vijay Kumar Burugari[2]
Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, AP. India[1]
Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Vijayawada, AP, India[2]

***Abstract*—With technology innovations such as Artificial Intelligence (AI) and Internet of Things (IoT), unprecedented applications and solutions to real world problems are made possible. Precision agriculture is one such problem which is aimed at technology driven agriculture. So far, the research on agriculture and usage of technologies are at government level to reap benefits of technologies in crop yield prediction and finding the cultivated areas. However, the fruits of technologies could not reach farmers. Farmers still suffer from plenty of problems such as natural calamities, reduction in crop yield, high expenditure and lack of technical support. Plant diseases is an important problem faced by farmers as they could not find diseases early. There is need for early plant disease detection in agriculture. From the related works, it is known that deep learning techniques like Convolutional Neural Network (CNN) is best used to process image data to solve real world problems. However, as one size does not fit all, CNN cannot solve all problems without exploiting its layers based on the problem in hand. Towards this end, we designed an automatic plant disease detection system by proposing an advanced CNN model. We proposed an algorithm known as Advanced CNN for Plant Disease Detection (ACNN-PDD) to realize the proposed system. Our system is evaluated with PlantVillage, a benchmark dataset for crop disease detection result, and real-time dataset (captured from live agricultural fields). The investigational outcomes showed the utility of the proposed system. The proposed advanced CNN based model ACNN-PDD achieve 96.83% accuracy which is higher than many existing models. Thus our system can be integrated with precision agriculture infrastructure to enable farmers to detect plant diseases early.**

***Keywords*—Plant disease detection; advanced CNN; Artificial Intelligence (AI); deep learning; precision agriculture**

I. Introduction

Agricultural sector in the world is crucial for growing food required by humanity. This field makes the highest contribution towards food and the economy as well. Farmers spend their lives in agricultural activities and work hard in production of different food items besides other commercial products like cotton. However, farmers are suffering from high expenditure involved in cultivation and also plant diseases. Particularly certain crop diseases lead to significant losses to farmers leading to crisis in agricultural domain [1]. Technology innovations such as AI are shaping unprecedented solutions in different real world applications. ML and DL techniques could improve state of the art in solving problems [2]. However, technological innovations in several countries are helping governments and organizations but actual benefits of technologies could not reach farmers. In other words, in spite of innovations in agriculture, technology benefits are not really changing the lives of farmers. To state it differently, at the farmer level the technologies are not exploited. In this paper, our endeavour is to build a system that helps farmers to have automatic detection of plant diseases in a user-friendly fashion.

There are many existing methods dealing with the problem of automatic plant disease detection. PlantVillage [26] is the dataset widely used to have machine learning based approaches for disease detection. An excerpt from the dataset is shown in Fig. 1 reflecting some healthy leaves and diseased ones. Many of the existing methods such as [3], [5], [16] and [21] are based on CNN model. Sardogan *et al*. [3] proposed a hybrid DL technique using CNN and quantization to detect diseases in Tomato crop. Marzougui *et al*. [5] used ResNet model along with data augmentation for automatic disease detection. Hassan *et al*. [16] used pre-trained techniques like InceptionV3 and MobileNet with transfer learning to improve prediction performance. Suma [21] proposed as system for disease prediction and also incorporated a provision to give suitable recommendations on detection of specific crop disease. Andrew *et al*. [25] used models such as Inception v4, VGG16, ResNet50 and DenseNet121 to deal with automatic leaf disease detection process more efficiently. From the review of related works, it is ascertained that most of the existing models are built on CNN. Nevertheless, there is necessity for improving accuracy further and also work with live data collected from agricultural fields. Here are important contributions of the paper.

1) We designed an automatic plant disease detection system by proposing an advanced CNN model.

2) We proposed an algorithm known as Advanced CNN for Plant Disease Detection (ACNN-PDD) to realize the proposed system.

3) Our system is evaluated with PlantVillage [26], a benchmark dataset for crop disease detection result, and real-time dataset (captured from live agricultural fields).

4) The experimental results showed the utility of the proposed system. The proposed advanced CNN model

*Corresponding Author.

ACNN-PDD achieve 96.83% accuracy which is higher than many prior models.

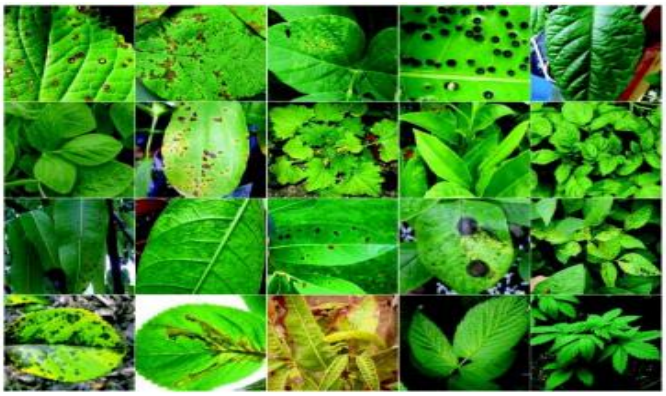


Fig. 1. Leaf samples from plantvillage dataset.

The rest of the paper covers more details of our research. Prior works are reviewed in Section II. Preliminaries is given in Section III while Section IV throws light on the materials and methods associated with our work. Our investigational observations are provided in Section V. Our research is concluded in Section VI.

II. Related Work

This section analyses related works on deep learning based methods for plant disease detection. Ferentinos et al. [1] used many variants of CNN models for disease prediction. They opined that CNN based models are widely used and they are more suitable for image analysis. It is found that various CNN variants are available for processing imagery data. Shruthi et al. [2] reviewed many DL models used for plant disease diagnosis. Their research showed that CNN is the preferred deep learning model. The rationale behind this is that, CNN has power to have discrimination towards efficient predictions. Sardogan et al. [3] defined a hybrid deep learning model using CNN and LVQ to detect diseases in Tomato crop. They found that when LVQ is combined with CNN, it could perform better in disease detection. Kumar et al. [4] explored region based approaches based on CNN models. Their methodology includes feature selection prior to using a supervised learning approach. With region approach, they found that there is increased probability in accurate predictions. Marzougui et al. [5] used ResNet model along with data augmentation for automatic disease detection. It is observed that the ResNet is a model that has provision towards transfer learning leading to better performance. Mishra et al. [6] used Corn crop for disease detection using CNN based model deployed into Raspberry Pi. They could test leaf images collected through mobile phone camera. Thus their system is found to be useful in collecting new samples and detect diseases. Chowdhury et al. [7] used U-Net model for segmentation process while EfficientNet model is used for automatic disease classification. The usage of dual models in their research they found that division of labour led to improved performance. Militante et al. [8] designed a DL based system for disease detection considering agricultural crops. They explored an image pre-processing method for performance enhancement. It is observed that DL modes are better used for image data analytics.

Kannan et al. [9] defined a CNN based methodology along with data augmentation to improve disease detection process using Tomato crop. They also used transfer learning with ResNet50 model and found its utility. With such reuse in the model has improved chances of accuracy in predictions. Gui et al. [10] used background replacement technique along with an improved CNN for disease recognition using uncontrolled field conditions and controlled laboratory conditions. With the mixture of field conditions, the evaluation of the models became more comprehensive. Kumar et al. [11] investigated on Coffee plants using a CNN based architecture for disease detection. Their study revealed the utility of DL models for solving problems in agriculture. Tugrul et al. [12] followed a systematic review approach to explored different CNN variants. Their investigation has resulted in set of CNN variants and their modus operandi besides capability in predictions. Yan et al. [13] proposed an improved CNN model detect diseases such as cedar rust, frogeye spot and scab using Apple leaves. Apart from this, they used several pre-trained models such as VGG16 as well. Their inception model is found to have better performance. When compared with baseline models, their model was performing better significantly. Gajjar et al. [14] incorporated a trained CNN model into an embedded system in order to have automatic detection of leaf diseases. Their ideal of having an embedded system is to provide a solution in hand-held devices. Zeng and Li [15] proposed a method based on CNN by improving it using self-attention in the form of BaseNet architecture. This architecture is found better than its counterparts devoid of self-attention mechanism. Hassan et al. [16] used advanced models like InceptionV3 and MobileNet with transfer learning to improve prediction performance. Their findings are important towards making comparison between baselines and their models. Lu et al. [17] made a review on CNN models for leaf disease detection. Their findings are in affirmative on the capability of CNN in agricultural research. Raina and Gupta [18] studied many models used for disease detection in agricultural crops. Those models could establish the significance of learning based approach to solve problems in agriculture. Ahmad et al. [19] combined stepwise transfer learning and CNN models on imbalanced datasets for disease detection. This research could throw light on dealing with datasets that exhibit imbalance. Panchal et al. [20] investigated on different deep models for image based disease prediction. Their findings revealed that deep models have better opportunities in learning and prediction procedures.

Suma [21] proposed a system for disease prediction and also incorporated a provision to give suitable recommendations on detection of specific crop disease. Such recommendations provide another layer of knowledge to farmer community. Venkataramanan et al. [22] used YOLOv3 technique to obtain leaf from given image. Then that leaf is subjected to a learned CNN model for disease detection. Learning based approach is thus found to be scalable and more useful. Li et al. [23] explored different techniques used for disease detection. Their research has come up with several insights, challenges and current trends in leaf disease

detection. Their work insights could trigger further research in the area of disease prediction. Nagasubramanian et al. [24] used hyperspectral imagery for their research. Moreover, they incorporated explainable 3D learning into their methodology for disease identification. Andrew et al. [25] used pre-trained models such as Inception v4, VGG16, ResNet50 and DenseNet121 to deal with automatic leaf disease detection process more efficiently. From the review of related works, it is ascertained that most of the prior models are based on CNN [29]. Nevertheless, there is necessity for improving accuracy further and also work with live data collected from agricultural fields.

III. PRELIMINARIES

DL is a neural networks based advanced ML meant for improving learning process and accuracy in solving real world problems. CNN is a DL technique widely used for image analysis. CNN has multiple layers as presented in Fig. 2.

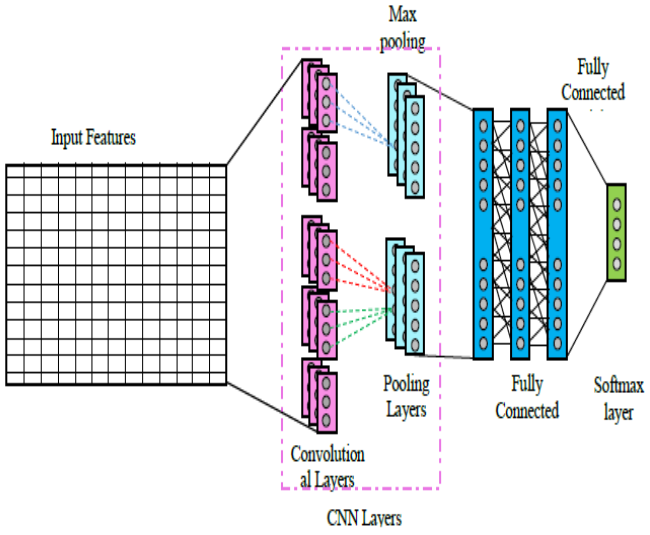


Fig. 2. Overview of baseline CNN architecture.

From the given image, convolutional layers are designed to extract features through correlation of local information. Pooling layers on the other hand perform sub-sampling and reduce feature space. A fully connected layer exploits the outcome of the convolutional and pooling layers and enables the prediction of class labels, while the soft max layer is meant for producing the final output in terms of different classes.

The convolution process is illustrated in Fig. 3. It places a kernel on input and gets pixel values. Then the pixel values are multiplied with kernel values. The result is summarized besides adding bias. The pooling is CNN can be of two types such as max pooling and average pooling as illustrated in Fig. 4. Max pooling is the process of dividing image into many regions and get max value for each region. Whereas average pooling does the same but takes average of each region. Generally, convolution and pooling layers are used in CNN architectures. In the fully connected layer, there is integration of multi-dimensional features and converting them into one-dimensional features eventually.

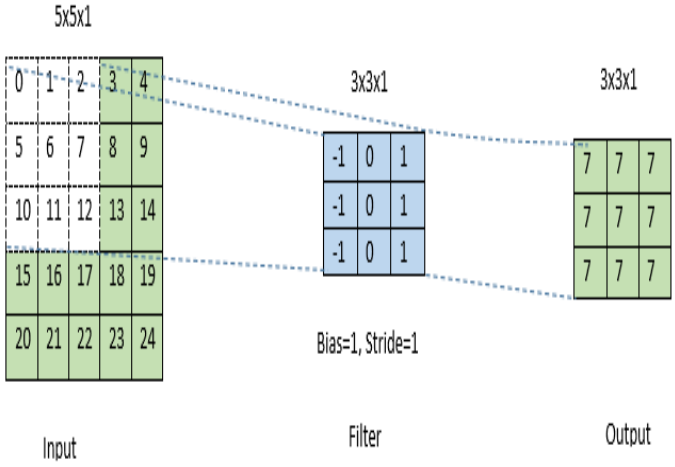


Fig. 3. Illustrates convolution process involved in CNN model.

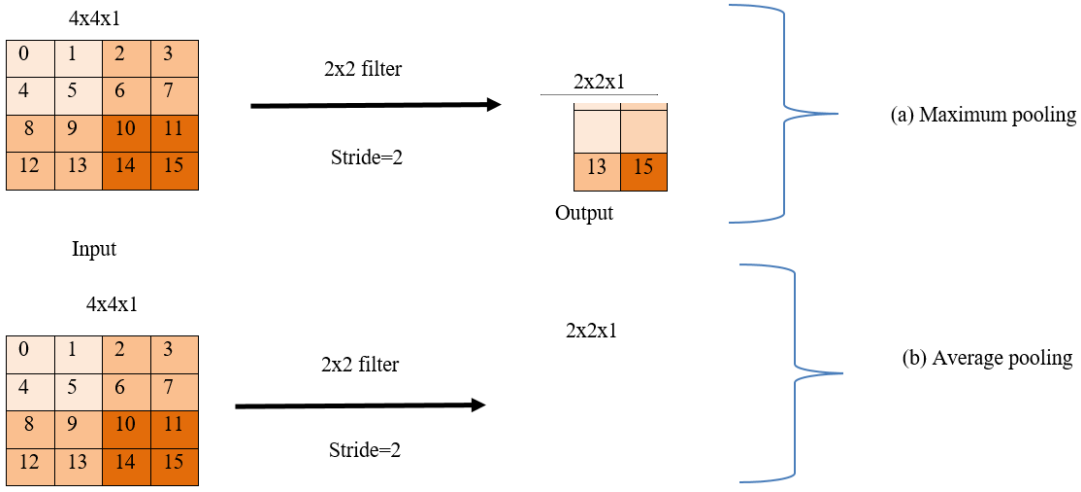


Fig. 4. Illustrates max pooling (a) and Average pooling (b) Operations of a CNN model.

IV. PROPOSED SYSTEM

This research proposed an automatic plant disease detection system using DL approach. Towards this end, we considered the base line CNN model described in Section III and improved it to have advanced CNN model. With empirical study, we made different configurations of the layers to be more suitable for the purpose. As presented in Fig. 5, the proposed system is based on learning based phenomena. The proposed advanced CNN model is used to train the system. In other words, our CNN model learns from given training images and labels taken from PlantVillage dataset. We preferred CNN model for our research for many reasons. CNN is found to be most suitable for analysing image content. Its convolutional layers are designed to extract feature maps that reflect the underlying content of images. Moreover, the CNN model's pooling layers are designed to optimize feature maps so as to reduce in size without compromising the discriminative capabilities. CNN has capability to extract features without human intervention [26]. The configuration of CNN is designed such that its computational complexity is relatively less than other deep learning models. Strength of CNN lies in its ability to solve the problem known as "curse of dimensionality" in image data analytics. Its dense nature of the network has its influence in improving accuracy in prediction process [27]. Adjustment of filters in different layers of CNN has its impact on influencing accuracy. Adjustment of kernel size and stride can improve performance of CNN model.

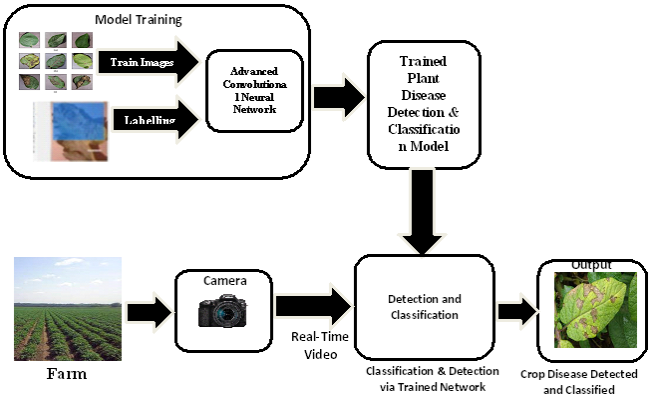


Fig. 5. Proposed automatic plant disease detection system.

The proposed advanced CNN model training set images and also corresponding ground truth. With the given inputs, the CNN model is trained to gain the knowledge pertaining to crop disease detection. After the training process, the learned CNN model has capability to discriminate healthy samples from the ones affected with diseases. In the process of training, convolutional layers are meant for collecting feature maps and pooling layers are meant for optimization of feature maps. The optimized feature maps play an important role in ascertaining knowledge about training samples along with their ground truth. The ground truth of training and testing samples play crucial role in evaluation of the proposed system. Then the knowledge model created is saved to persistent storage for using it later in the plant disease detection process [28]. The proposed system supports readily available test (unlabelled) images from PlantVillage dataset and also real time images captured live from agricultural field using a camera. When a new image is given to the trained model, it is able to detect disease and classify it. In fact, the ability to detect any newly collected test sample is the important characteristic of the proposed system. This will enable farmers to take photo of their crop and use the proposed system to detect possible diseases. The given input image is subjected to convolution operation where it computes feature map as in Eq. (1).

$$x_j^{\lambda} = \sum_{i \epsilon M_j} x_i^{\lambda-1} \times k_{ij}^{\lambda} + b_j^{\lambda} \quad (1)$$

The extracted feature map is denoted as x_j^{λ}. The kernel used by the convolutional layer is denoted as kij while λ denotes layers. The input feature map is represented by Mj and bias is denoted by bj. Then the max pooling performs efficient sampling and results in reduction of feature map. The max pooling process is expressed as in Eq. (2).

$$s_j = \max_{i \in R_j} \propto_i \quad (2)$$

In the proposed system, there is need for multiple classes in the classification outcome. For this reason, a softmax utility is used and its proposition is expressed as in Eq. (3).

$$h_\theta(x) = \frac{1}{1+\exp(-\theta^T x)} \quad (3)$$

ReLU, as activation utility improves learning capability. In other words, it is used to handle over-fitting problem and enable speed in the prediction process. The functionality of activation function is stated in Eq. (4).

$$f(x) = \max(0, x) \quad (4)$$

An algorithm known as Advanced CNN for Plant Disease Detection (ACNN-PDD) is proposed to realize the proposed system.

Algorithm 1: Advanced CNN for Plant Disease Detection (ACNN-PDD).

Algorithm: Advanced CNN for Plant Disease Detection (ACNN-PDD)

Inputs:
D, n, m
(PlantVillage Dataset, number of epochs, batch size)
Outputs:
R, P
(Plant disease detection results, performance statistics)

1. Start
2. Initialize training set vector T1
3. Initialize testing set vector T1
4. (T1, T2)←SplitDataset(D)
5. Create CNN model
6. Add convolutional 2D layers (5 layers are used)
7. Add max pooling 2D layers (5 layers are used)
8. Add fully connected layers (2 layers are used)
9. Add dropout layer (used one layer)
10. F←FeatureSelection(T1)//using conv and pooling layers
11. M = TrainTheModel(F)
12. For each epoch e in n

```
13.  For each batch b in m
14.    Update the model M
15.  End For
16. End For
17. M' = FitTheModel(M)
18. R = Detect(M')
19. P = PerformanceEvaluation(M')
20. Print R
21. Return P
```

D	dataset	n	number of epochs
m	number of batches	T1	training set
T2	test set	R	detection results
P	performance statistics	e	one epoch
b	one batch	M	trained model
F	Optimized feature map	M'	updated model

Algorithm 1 takes PlantVillage Dataset, number of epochs and batch size as input and produce results of disease detection besides performance statistics. The proposed algorithm is evaluated using the procedure as described here. It is based on confusion matrix presented in Fig. 6.

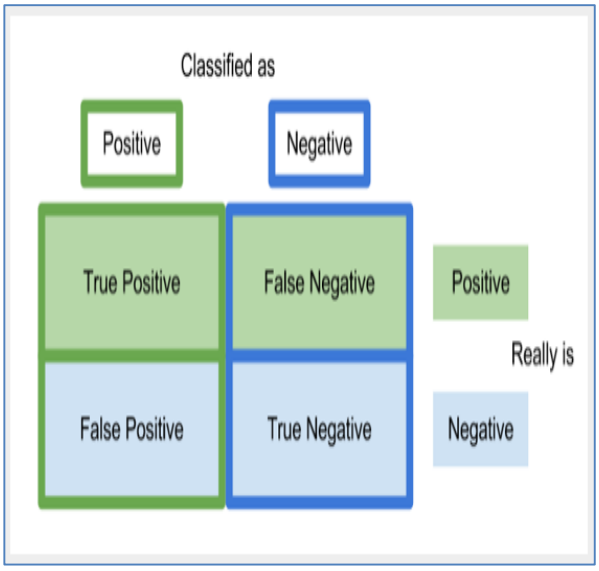


Fig. 6. Illustrates ground truth and predictions through confusion matrix.

Confusion matrix helps in understanding number of correct and incorrect predictions with regard to disease detection. Correct positive prediction is known as True Positive (TP), correct negative prediction is known as True Negative (TN). Opposite to these two (wrong predictions) are known as False Positive (FP) and False Negative (FN) respectively. Accuracy is the metric derived from confusion matrix.

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad (5)$$

Accuracy is metric used to evaluate performance of the proposed model as expressed in Eq. (5).

V. Experimental Results

Experiments with both PlantVillage dataset [26] and also real time dataset captured from agricultural field are provided here. Then the accuracy of the proposed method is evaluated and compared against prior works.

TABLE I. Hyper Parameters Set in the Proposed Model

Hyper parameter	Value
batch size	128
dropout rate	0.8
learning_rate	1e-3
Loss	categorical_crossentropy
number of epochs	40
Optimizer	adam

As presented in Table I, the hyper parameters used in the proposed CNN model and its values are provided. Batch size indicates that the training process needs to be done in batches. In our research it is set to 128. Dropout rate refers to the fact that contribution of certain neurons is not considered in the learning process. Sometimes, the dropout layer enables the CNN to improve its efficiency by ignoring some inputs. The learning rate is set to control the training process in the proposed system. In our research number of epochs is set to 40 and its does mean the CNN is trained for 40 cycles. Adam is the optimizer used to adjust weights and learning rate in order to minimize loss and improve accuracy in prediction.

A. Results with PlantVillage Dataset

First, we tested the proposed model with test data taken from Plant Village dataset. The results are with two test samples are shown in Fig. 7. The first sample has "yellow leaf curl virus" disease and the proposed system is able to detect it correctly. The second sample also has same disease but with different severity level. The proposed system could detect it also correctly.

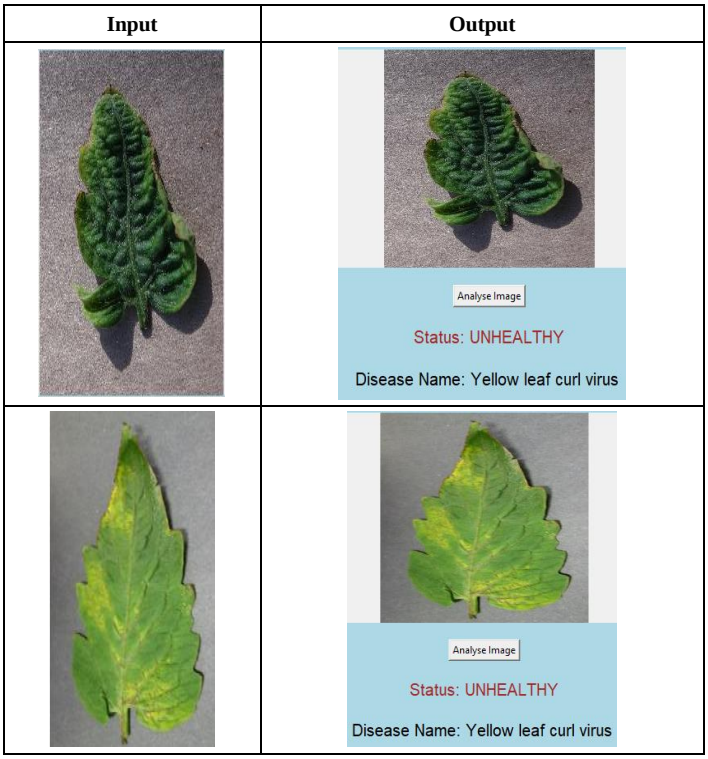


Fig. 7. Plant disease detection results using benchmark dataset.

As shown in Fig. 6, the test data is taken from the PlantVillage dataset and the detection results are provided. These results are observed with the test images available in the PlantVillage dataset. Since the dataset has plenty of test samples, the system is initially tested with the readily available samples. Section V(B) presents our empirical study with newly acquired test samples.

B. Results with Real-Time Dataset

We also tested the proposed model with test data captured live from agricultural field. The results are shown in Fig. 8.

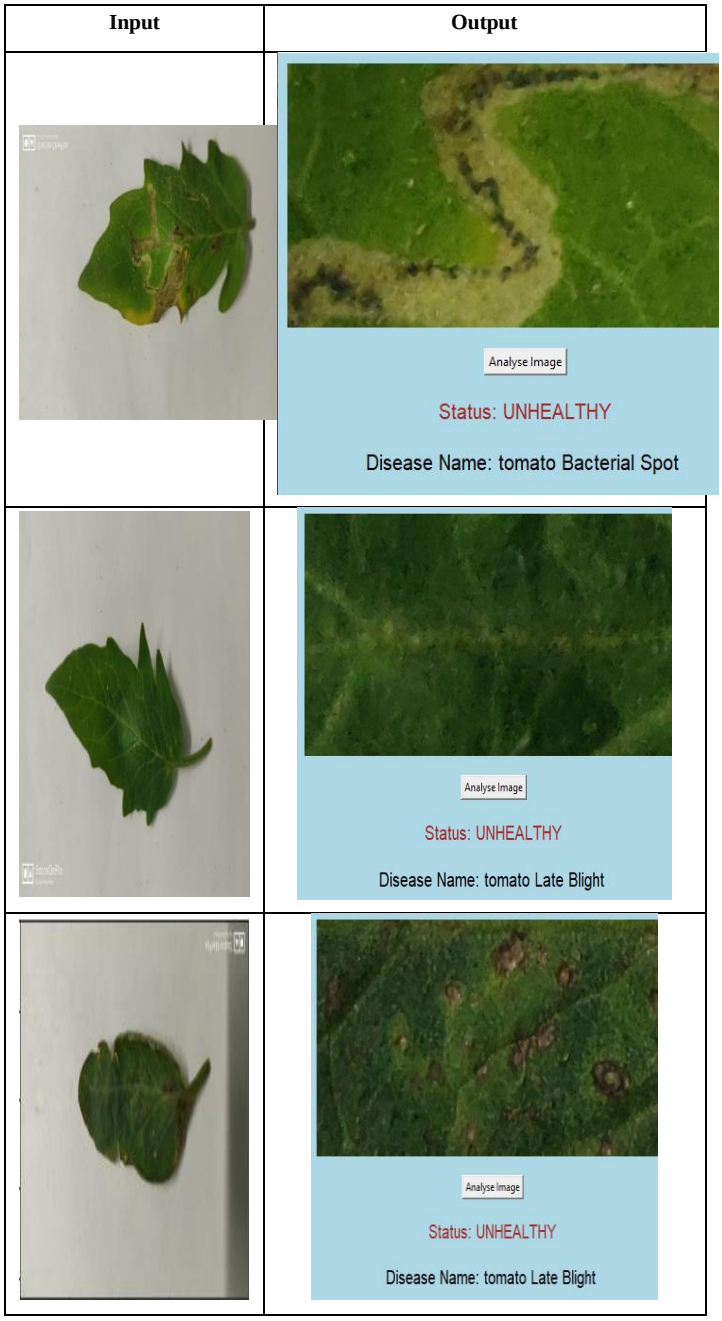


Fig. 8. Plant disease detection results using benchmark dataset.

As shown in Fig. 8, the test data is taken from the real time dataset collected live from agricultural field and the detection results are provided. Since the proposed system works for any crop for whom training data is available in PlantVillage dataset, we did experiments with newly acquired leaves of Tomato crop. Test results for three such samples are provided here. The first sample is detected as "tomato bacterial spot". The second sample is detected as "tomato late blight" while the third one is detected as "tomato late blight". These results are validated and found to be accurate.

C. Performance Evaluation

In this paper compared the detecting performance of our proposed model with many existing deep learning models.

As presented in Table II, the plant disease detection models and their performance in terms of accuracy are provided.

TABLE II. SHOWS PLANT DISEASE DETECTION ACCURACY OF DIFFERENT MODELS

Method	Accuracy
AlexNet	90.46048
GoogLeNet	94.92448
ResNet-20	92.01792
VggNet-16	95.54944
ACNN-PDD (proposed)	96.83904

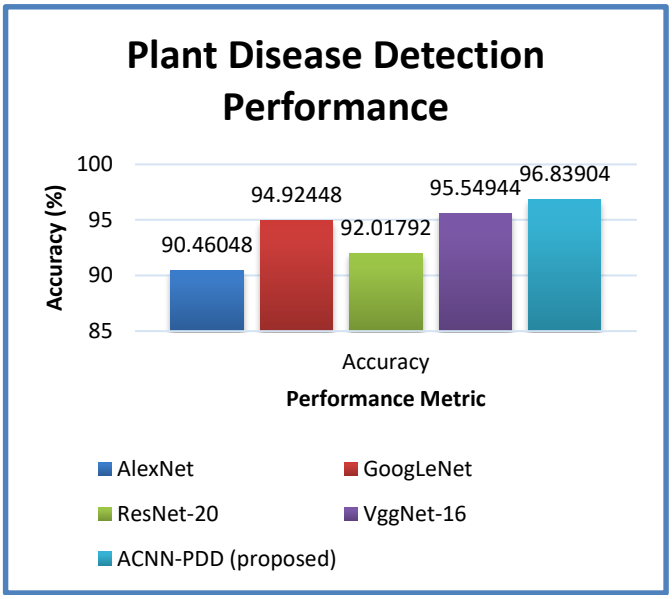


Fig. 9. Performance comparison among plant disease detection models.

As presented in Fig. 9, performance of the models is evaluated in terms of detection accuracy. The detection accuracy is computed by comparing ground truth with predicted values. Different deep learning models showed varied detection performance due to their internal functionality and configuration of layers. Though all the models are based on CNN, their performance is different due to different architecture of the models. Least accuracy is exhibited by AlexNet model with 90.46%. ResNet-20 showed 92.01% accuracy. GoogLeNet has achieved 94.92% accuracy. VGG16 showed 95.54% accuracy. Highest plant disease

detection accuracy is exhibited by the proposed ACNN-PDD model with 96.83% accuracy. In summary, our research has resulted in better accuracy as it incorporates layers of the DL model appropriately. Moreover, the model is found robust with both test images from benchmark dataset and also the newly collected samples live from agricultural fields.

VI. Conclusion and Future Work

In this paper, we designed an automatic plant disease detection system by proposing an advanced CNN model. We proposed an algorithm known as Advanced CNN for Plant Disease Detection (ACNN-PDD) to realize the proposed system. Our algorithm has an iterative process that learns from given training samples and ground truth. Then the model is evaluated with test samples. The prediction of ACNN-PDD for each test sample is compared against ground truth to arrive at confusion matrix reflecting efficiency of the model. Our system is evaluated with PlantVillage, a benchmark dataset for crop disease detection result, and real-time dataset (captured from live agricultural fields). The experimental results showed the utility of the proposed system. The proposed advanced CNN based model ACNN-PDD achieves 96.83% accuracy which is higher than many existing models such as AlexNet, GoogLeNet, ResNet-20 and VGG16. Though the proposed model provides better performance, it could be improved further with feature selection enhancements. In future, therefore, we intend to improve our system with further enhancement in CNN model along with feature engineering.

Acknowledgment

Authors are declaring that no funding was received for this paper publication and Research. Dataset PlantVillage Dataset. Retrieved from https://datasets.activeloop.ai/docs/ml/datasets/plantvillage-dataset/

References

[1] Ferentinos, K. P. (2018). Deep learning models for plant disease detection and diagnosis. *Computers and Electronics in Agriculture*. 145, pp.1-8. https://doi.org/10.1016/j.compag.2018.01.009

[2] U, S., Nagaveni, V., & Raghavendra, B. K. (2019). A Review on Machine Learning Classification Techniques for Plant Disease Detection. *2019 5th International Conference on Advanced Computing & Communication Systems (ICACCS).*, pp.281-284. https://doi.org/10.1109/ICACCS.2019.8728415

[3] Sardogan, M., Tuncer, A., & Ozen, Y. (2018). Plant Leaf Disease Detection and Classification #ased on CNN with LVQ Algorithm. *2018 3rd International Conference on Computer Science and Engineering (UBMK).*, pp.382-385. https://doi.org/10.1109/UBMK.2018.8566635

[4] Kumar, S.; Chaudhary, V.; Chandra, S.K.;. (2021). Plant Disease Detection Using CNN. *Turkish Journal of Computer and Mathematics Education*. 12(12), pp.2106-2112.

[5] Marzougui, F., Elleuch, M., & Kherallah, M. (2020). A Deep CNN Approach for Plant Disease Detection. *2020 21st International Arab Conference on Information Technology (ACIT).*, pp.1-6. doi:10.1109/acit50332.2020.9300072

[6] Mishra, S., Sachan, R., & Rajpal, D. (2020). Deep Convolutional Neural Network based Detection System for Real-time Corn Plant Disease Recognition. *Procedia Computer Science*. 167, pp.2003-201 doi:10.1016/j.procs.2020.03.236

[7] Chowdhury, M. E. H., Rahman, T., Khandakar, A., Ayari, M. A., Khan, A. U., Khan,. (2021). Automatic and Reliable Leaf Disease Detection Using Deep Learning Techniques. *AgriEngineering*. 3(2), 294-312. https://doi.org/10.3390/agriengineering3020020

[8] Militante, S. V., Gerardo, B. D., & Dionisio, N. V. (2019). Plant Leaf Detection and Disease Recognition using Deep Learning. *2019 IEEE Eurasia Conference on IOT, Communication and Engineering (ECICE).*, pp.579-582. doi:10.1109/ecice47484.2019.8942686.

[9] E, N. K., M, K., P, P., R, A., S, V. (2020). Tomato Leaf Disease Detection using Convolutional Neural Network with Data Augmentation. *2020 5th International Conference on Communication and Electronics Systems (ICCES)*, pp.1125-1132. doi:10.1109/icces48766.2020.9138030

[10] Gui, P.; Dang, W.; Zhu, F.; Zhao, Q.;. (2021). Towards automatic field plant disease recognition. *Elsevier*. 191, pp.1-10. https://doi.org/10.1016/j.compag.2021.106523

[11] Kumar, M., Gupta, P., Madhav, P., & Sachin. (2020). Disease Detection in Coffee Plants Using Convolutional Neural Network. *2020 5th International Conference on Communication and Electronics Systems (ICCES)*,.1-6. https://doi.org/10.1109/ICCES48766.2020.9138000.

[12] Tugrul, B.; Elfatimi, E.; Eryigit, R.;. (2022). Convolutional Neural Networks in Detection of Plant Leaf Diseases: A Review. MDPI. 12,1-21. https://doi.org/10.3390/agriculture12081192

[13] Yan, Q., Yang, B., Wang, W., Wang, B., Chen, P., & Zhang, J. (2020). Apple Leaf Diseases Recognition Based on An Improved Convolutional Neural Network. *Sensors*. 20(12),1-14. http://dx.doi.org/10.3390/s20123535

[14] Gajjar, R., Gajjar, N., Thakor, V. J., Patel, N. P., & Ruparelia, S. (2021). Real-time detection and identification of plant leaf diseases using convolutional neural networks on an embedded platfor. *The Visual Computer*,.1-16. https://doi.org/10.1007/s00371-021-02164-9

[15] Zeng, W., & Li, M. (2020). Crop leaf disease recognition based on Self-Attention convolutional neural network. *Computers and Electronics in Agriculture*. 172, pp.1-7. https://doi.org/10.1016/j.compag.2020.105341

[16] Hassan, S. M., Maji, A. K., Jasiński, M., Leonowicz, Z., & Jasińska, E. (2021). Identification of Plant-Leaf Diseases Using CNN and Transfer-Learning Approach. *Electronics*. 10, pp.1-19. https://doi.org/10.3390/electronics10121388

[17] Lu, J., Tan, L., & Jiang, H. (2021). Review on Convolutional Neural Network (CNN) Applied to Plant Leaf Disease Classification. *Agriculture*. 11(8), pp.1-18. https://doi.org/10.3390/agriculture11080707

[18] Raina, S., & Gupta, A. (2021). A Study on Various Techniques for Plant Leaf Disease Detection Using Leaf Image. *2021 International Conference on Artificial Intelligence and Smart Systems (ICAIS)*, pp.900-905. https://doi.org/10.1109/ICAIS50930.2021.9396023

[19] AHMAD, M.; ABDULLAH, M.; MOON, H.; HAN, D.;. (2021). Plant Disease Detection in Imbalanced Datasets Using Efficient Convolutional Neural Networks With Stepwise Transfer Lear. *creative commons*. 9, pp.1-16. DOI 10.1109/ACCESS.2021.3119655

[20] Panchal, A.V.; Patel, S.C.; Bagyalakshmi, K.; Kumar, P.;. (2021). Image-based Plant Diseases Detection using Deep Learning. *ELSEVIER*. pp.1-7. https://doi.org/10.1016/j.matpr.2021.07.281

[21] Suma, V., Shetty, R. A., Tated, R. F., Rohan, S., & Pujar, T. S. (2019). CNN based Leaf Disease Identification and Remedy Recommendation System. *2019 3rd International Conference on Electronics, Communication and Aerospace Technology (ICECA)*. 395-399. https://doi.org/10.1109/ICECA.2019.8821872

[22] Venkataramanan, A.; Honakeri,D.K.; Agarwal, P.;. (2019). Plant Disease Detection and Classification Using Deep Neural Networks. *International Journal on Computer Science and Engineering (IJCSE)*. 11(8), pp.40-46.

[23] Li, L., Zhang, S., & Wang, B. (2021). Plant Disease Detection and Classification by Deep Learning—A Review. *IEEE Access*. 9, pp.1-16.Digital Object Identifier 10.1109/ACCESS.2021.3069646

[24] Nagasubramanian, K., Jones, S., Singh, A. K., Sarkar, S., Singh, A., & Ganapathy. (2019). Plant disease identifcation using explainable 3D deep learning on hyperspectral images. *Plant Methods*. 15, pp.1-10. https://doi.org/10.1186/s13007-019-0479-8

[25] Andrew, J.; Eunice, J.; Popescu, D.E.; Chowdary, M.K.;. (2022). Deep Learning-Based Leaf Disease Detection in Crops Using Images for Agricultural Applications. *MDPI*. 12, pp.1-19. https://doi.org/10.3390/agronomy12102395

[26] Kalpana, P., Anandan, R., Hussien, A.G. *et al.* Plant disease recognition using residual convolutional enlightened Swin transformer networks. *Sci Rep* 14, 8660 (2024). https://doi.org/10.1038/s41598-024-56393-8.

[27] Kalpana, P., Anandan, R. (2023). A capsule attention network for plant disease classification. Traitement du Signal, Vol. 40, No. 5, pp. 2051-2062. https://doi.org/10.18280/ts.400523

[28] P. Kalpana, Y. Chanti, R. G, R. D and P. K. Pareek, "SE-Resnet152 Model: Early Corn Leaf Disease Identification and Classification using Feature Based Transfer Learning Technique," *2023 International Conference on Evolutionary Algorithms and Soft Computing Techniques (EASCT)*, Bengaluru, India, 2023, pp. 1-6, doi: 10.1109/EASCT59475.2023.10392328.

[29] Praveena Nuthakki, Madhavi Katamaneni, Chandra Sekhar J. N., Kumari Gubbala, Bullarao Domathoti, Venkata Rao Maddumala, and Kumar Raja Jetti. 2023. Deep Learning based Multilingual Speech Synthesis using Multi Feature Fusion Methods. ACM Trans. Asian Low-Resour. Lang. Inf. Process. Just Accepted (September 2023). https://doi.org/10.1145/3618110

Towards Secure Cloud-Enabled Wireless Ad-Hoc Networks: A Novel Cross-Layer Validation Mechanism

Zhenguo LIU
Shandong University of Political Science and Law, Jinan, Shandong, China, 250014

***Abstract*—Network security tackles a broad spectrum of damaging activities that threaten network infrastructure. Addressing these risks is essential to keep data accurate and networks running. This research aims to detect and prevent blackholes and wormholes in cloud-based wireless ad-hoc networks. A new Cross-Layer Validation Mechanism (CLVM) is introduced to detect and counter these dangerous attacks. CLVM boosts network security and ensures data travels through cross-layer interactions. CLVM is tested using NS2 software by performing several simulations and comparing the results with previous methods. The results show that CLVM effectively defends against blackhole and wormhole attacks, which makes it a crucial extra service for cloud computing. CLVM provides a strong defense against new security threats, making sure the network stays reliable and safe.**

Keywords*—Network security; wireless ad-hoc networks; cloud environments; cross-layer validation; blackhole and wormhole attacks

I. INTRODUCTION

Wireless networks play a key role in modern communication technology, enabling worldwide communication [1, 2]. Over the last two decades, progress in wireless communication has transformed our world, offering a range of wireless technologies such as Bluetooth, Wi-Fi, WiMAX, HSPA, 3G, 4G, 5G, ZigBee, Satellite, and NFC [3]. These wireless methods support different uses, from home networking to real-time multimedia and surveillance, adding to energy-saving designs on portable devices. We can split these networks into two main types: infrastructure-based wireless networks and infrastructure-less or Wireless Ad-hoc Networks (WANs) [4].

Infrastructure-based wireless networks rely on a fixed infrastructure in which nodes transmit data to a base station over predetermined routes [5]. Although these networks are reliable, they are typically expensive and unsuitable for hostile environments such as proactive disaster management or military applications where fixed infrastructure may not be available [6]. On the other hand, WANs function without predefined infrastructure. Nodes in these networks can connect to other nodes within their communication range, creating a dynamic, self-configuring, and self-organizing network [7]. They use shared radio channels and enable data forwarding between nodes.

Unlike traditional wireless networks with fixed configurations, cloud-based WANs face unique issues that make their design and operation difficult. These networks must cope with changing layouts where nodes frequently connect or disconnect, resulting in constant shifts in routing paths. Additionally, the sprawling nature of WANs, as well as limited resources such as battery life and processing power, increase the risk of security threats. In cloud environments, these problems are exacerbated as data must be moved and processed across distributed nodes without central control. This setup is vulnerable to smart attacks such as blackhole and wormhole tricks, which can compromise network reliability and access. To address these problems, we need new ideas that increase security and keep the network running. This is the main objective of the Cross-Layer Validation Mechanism (CLVM) that we propose in this study.

Key distinguishing characteristics of ad hoc networks include lack of fixed infrastructure, dynamic topology, multi-hop routing, node heterogeneity, connection variability, scarce resources (power, storage, computing power), and limited physical security [8]. Nodes in WANs can be either mobile or fixed, resulting in two main categories based on mobility: Mobile Ad-hoc Networks (MANETs) and Wireless Sensor Networks (WSNs). MANETs are mobile nodes with no fixed location, while WSNs consist of non-mobile nodes deployed at specific locations [9]. The main differences between these networks are summarized in Table I.

TABLE I. WSN VS. MANET

Feature	WSNs	MANETs
Optimization focus	Power optimization	Both QoS and performance optimization
Communication	Many-to-one	Point-to-point
Routing	Data-centric	Address-centric
Destination	The final destination is known	The final destination is unknown
Power source	Not possible to change or recharge	Can be changed or recharged
Network type	Homogeneous	Heterogeneous
Topology	Static	Dynamic

While WANs offer significant advantages, they present several design and implementation challenges due to node mobility, limited resources, and decentralized network structures. These challenges span different protocol stack layers and increase the complexity of WAN development. In addition, WANs have specific vulnerabilities compared to other traditional networks, which are described below:

- Infrastructure absence: Nodes in these networks lack prior security association and can dynamically join or exit without notice, necessitating mutual trust among participating nodes within the protocol design.
- Wireless links: The unsecured nature of wireless links allows potential adversaries to access the network, lacking the equivalent protection level of wired links, making the network vulnerable to attacks from various directions.
- Limited physical protection: Nodes in WANs are often either minimally protected or entirely unprotected, intensifying network vulnerability due to their dynamic and mobile nature, facilitating easier insertion of malicious nodes.
- Lack of central management: The absence of a central authority enables adversaries to devise new attacks, exploiting the cooperative algorithm present in WANs. Security mechanisms must be adaptive and scalable to cope with dynamic topology changes and node increases.
- Resource constraints: Nodes in these networks have limited computational and power resources, making them susceptible to Denial-of-Service (DoS) attacks that exhaust the limited power source through excessive transmissions or computations.

In ad hoc networks, the absence of a central security system allows bad actors, both inside and outside the network, to put network security and privacy at risk [10]. We can split security attacks into two types based on how they work: passive and active [11]. Passive attacks try to break data privacy by listening in on conversations to gather useful info for future bad actions. This makes them hard to spot [12]. Active attacks go after data integrity and privacy by changing, blocking, repeating, or getting rid of packets being sent. They use different network functions to pull off these attacks [13]. We group these attacks into internal and external based on where they come from. Internal attacks happen when compromised nodes within the same network cause trouble and mess up how the system or network works. External attacks, on the other hand, come from unauthorized outsiders who don't belong to the network [14].

Security mechanisms for ad hoc networks include two main approaches. To prevent security attacks, cryptographic techniques are used as the first line of defense against external attacks to ensure the authenticity and integrity of the data source. However, this mechanism can fail if internal attackers have valid cryptographic keys to launch an attack. Security attack detection and response serves as a secondary line of defense, identifying abnormal activity on the network before it causes damage. The defense offers effective countermeasures against detected attacks.

The rest of this paper follows the following structure. A review of related work in the field of secure WANs is presented in Section II. CLVM is discussed in detail in Section III. The simulation results are presented in Section IV. The paper concludes with a summary of key findings and research directions in Section V.

II. Related Work

Compressive Sensing (CS) data collecting systems may efficiently decrease the transmission cost of WSNs by using the sparsity of compressible signals. While there have been explanations of CS as a symmetric cryptosystem, CS-based data-gathering systems still encounter security risks because of the intricate deployment environment of WSNs. Zhang, et al. [15] developed two viable attack methods for certain applications. They introduced a secure approach for collecting data using compressive sensing. The proposed method improves data privacy through the use of an asymmetric semi-homomorphic encryption technique and minimizes computational costs by utilizing a sparse compressive matrix. To be more precise, the asymmetric approach decreases the complexity of distributing and managing secret keys. Homomorphic encryption enables in-network aggregation in the cipher domain, thereby improving security and achieving network load balancing. The sparsity of the measurement matrix decreases both the computational and transmission costs, therefore offsetting the rising costs associated with homomorphic encryption.

Al-Shayeji and Ebrahim [16] designed a robust and energy-efficient system that minimizes energy use while ensuring privacy. The security strategy employs a customized version of the sharing-based method with a precision-enhanced and encryption-mixed privacy-preserving data aggregation procedure. The first protocol provides authentication and encryption via XOR gates, while the second protocol is a secure data aggregation method that improves security and energy efficiency. An approach to reduce energy usage is presented, which involves asynchronous scheduling duty cycling depending on location, priority, and pre-configuration. The findings indicate that the performance of the system is influenced by factors such as the sensing rate, data transmission frequency, data size, sensor placement and quantity, and smartphone battery capacity. For infrequent use and smaller amounts of data, the energy consumed by operations accounts for just 1% of the total battery capacity of the mobile device. When sensors are placed near the sink, the cost is decreased by more than 70% compared to an unsecured network, but there is an extra cost of 20%. The simulations demonstrate that the expense of encryption decreases with an increase in the quantity of sensors. In addition, as the number of sensors increases, the proximity between nodes reduces, resulting in more sensors entering sleep mode.

Wang, et al. [17] suggested a hierarchical trust system based on fog computing to address security vulnerabilities in cloud-enabled WSNs. This tiered approach has two components: confidence in the foundational framework and trust between cloud service providers (CSPs) and sensor service providers (SSPs). Monitoring behavior is built and executed inside WSNs to ensure confidence in the fundamental framework. At the same time, the intricate and detailed data analysis component is shifted to the fog layer. To establish trust between CSPs and SSPs, it is crucial to prioritize the real-time comparison of service parameters, the collection of exception information in

WSNs, and the focused quantitative assessment of entities. The experimental findings demonstrate that the fog-based topology effectively conserves network energy, swiftly detects malicious nodes, and promptly recovers misjudgment nodes within an appropriate timeframe. Moreover, the dependability of edge nodes is effectively ensured by data analyses conducted in the fog layer, and an assessment approach that relies on comparable service records is proposed.

Hsiao and Sung [18] developed an innovative method to bolster the data security of WSNs by using blockchain technology. They used blockchain technology and data transmission to provide a safe framework for WSNs based on the Internet of Things (IoT) architecture. The research employs embedded microcontrollers such as Raspberry Pi and Arduino Yun to construct a portable database node that gathers sensor data and hash values from preceding blocks. The transaction ledger is converted into a sensor data record, thereby improving the dependability of the WSN structure. The system can process data from a private cloud and display sensor data. The wireless network design is constructed utilizing embedded devices, facilitating the creation of a web server using Python or JavaScript programming languages. The research examines the efficacy of conventional methods against new data transmission methods, concluding that using innovative methods using blockchain technology renders it very difficult for operators to manipulate sensor data.

Haseeb, et al. [19] proposed a protocol for safe data collection in mobile WSNs, which utilizes cloud services. The technique aims to efficiently distribute information in dynamic networks by employing mobile sensors with little loss and power consumption. Furthermore, it guarantees the continuous presence and uniformity of the gathered data inside the cloud organizations while enhancing the routing reliability. The simulation results and their analysis demonstrate the substantial efficacy of the suggested approach.

Sharmila, et al. [20] introduced a hybrid key management system for WSNs linking edge devices. This system utilizes Elliptic Curve Cryptography (ECC) and a hash function to create pre-distribution keys. The key setup is accomplished by simply broadcasting the node identity. The primary purpose of implementing a hybrid technique in the key pre-distribution method is to achieve mutual authentication between the sensor nodes during the installation phase. The suggested solution decreases computing complexity while enhancing security, making it suitable for implementation in sensor nodes with limited resources.

Ensuring the reliable and secure functioning of WSNs necessitates the identification of anomalies. Maximizing resource efficiency is essential for minimizing energy use. Gayathri and Surendran [21] introduced two methods for anomaly detection in WSNs: Ensemble Federated Learning (EFL) with cloud integration and Online Anomaly Detection with Energy-Efficient approaches (OAD-EE) using cloud-based model aggregation. Cloud-integrated EFL uses ensemble approaches and federated learning to improve detection accuracy and safeguard data privacy. OAD-EE, using a cloud-based model aggregation approach, employs online learning and energy-efficient strategies to save energy on sensor nodes. A complete and efficient system for anomaly detection in WSNs is established by integrating EFL and OAD-EE. The experimental findings indicate that adopting cloud technology in EFL leads to the best accuracy in detection. On the other hand, OAD-EE, which utilizes cloud-based model aggregation, exhibits the lowest energy consumption and the shortest detection time among all algorithms. Consequently, OAD-EE is well-suited for real-time applications.

TABLE II. AN OVERVIEW OF RELATED WORKS

Reference	Methodology	Key features	Results
Zhang, et al. [15]	Compressive sensing with asymmetric semi-homomorphic encryption	Uses CS to reduce transmission cost, asymmetric semi-homomorphic encryption for data privacy, and sparse compressive matrix to minimize computational costs	Improved data privacy, network load balancing, decreased computational and transmission costs, offset rising costs associated with homomorphic encryption
Al-Shayeji and Ebrahim [16]	An energy-efficient system with a customized sharing-based method	Sharing-based method for privacy, precision-enhanced and encryption-mixed privacy-preserving data aggregation, asynchronous scheduling duty cycling	Enhanced energy efficiency and security, performance influenced by various factors, energy consumption as low as 1% of mobile device battery for infrequent use, cost reduction over an unsecured network but an additional 20% cost
Wang, et al. [17]	Hierarchical trust system based on fog computing	Trust in foundational framework and between cloud service providers and sensor service providers, real-time comparison of service parameters	Effective energy conservation, swift detection of malicious nodes, reliable data analysis in fog layer, proposed assessment approach relying on service records
Hsiao and Sung [18]	Blockchain technology for data security	Uses blockchain for WSN security, embedded microcontrollers for portable database node, transaction ledger for sensor data, web server creation with Python/JavaScript	Enhanced data dependability and security, difficulty for operators to manipulate sensor data, improved performance of WSN architecture using blockchain technology
Haseeb, et al. [19]	Protocol for safe data collection in mobile WSNs	Uses cloud services for dynamic network information distribution mobile sensors to minimize loss and power consumption, ensures continuous data presence and consistency	Significant efficiency in data collection and routing, enhanced reliability and uniformity of collected data in cloud organizations, substantial efficacy demonstrated through simulation results
Sharmila, et al. [20]	Hybrid key management system using ECC and hash function	Uses ECC and hash function for key pre-distribution, mutual authentication during installation phase, reduced computing complexity	Enhanced security and computing efficiency, suitable for resource-constrained sensor nodes, improved mutual authentication
Gayathri and Surendran [21]	Anomaly detection using EFL and OAD-EE	EFL with cloud integration for accuracy and data privacy, OAD-EE for energy-efficient anomaly detection, combined algorithm for efficient system	Best accuracy in detection with EFL, lowest energy consumption, and shortest detection time with OAD-EE, integrated algorithm improves overall efficiency, scalability, and real-time response

The reviewed related works, as outlined in Table II, address various security and efficiency challenges in wireless sensor networks (WSNs) and mobile ad-hoc networks (MANETs) using diverse methodologies, such as compressive sensing, blockchain technology, and fog computing. However, these approaches encounter specific limitations. For instance, while Zhang, et al. [15] leverage semi-homomorphic encryption to enhance data privacy, the rising costs of this encryption remain a challenge. Similarly, Al-Shayeji and Ebrahim [16] focus on energy efficiency, yet their method incurs an additional 20% cost for security enhancements. Addressing these challenges, the current study introduces a novel mechanism that synergistically integrates the strengths of various approaches to bolster network security and data integrity while minimizing computational and transmission costs. Through extensive NS2 simulations, the proposed method demonstrates superior performance in detecting and mitigating blackhole and wormhole attacks, offering a robust defensive mechanism for cloud-enabled WSNs. This study's main contributions include the development of an efficient, cost-effective security framework that ensures reliable and secure data transmission, thus advancing the state-of-the-art in network security for WSNs and MANETs.

III. Proposed Mechanism

This paper proposes a novel CLVM framework to identify and eliminate malicious activities within network routing protocols. CLVM aims to improve network reliability and data security by setting trust values for individual network elements. This trust scoring helps to find and remove wormhole nodes, making sure data moves over trusted routes. CLVM's main goal is to keep the network secure by spotting and containing harmful activities before they cause trouble. It does this in two ways. The framework has ways to detect and stop harmful activity in the network. This includes picking trusted paths to send data, which helps avoid potential threats. CLVM gives trust scores to network parts based on a full evaluation. This lets the framework prioritize data transfer across reliable and trustworthy nodes. Fig. 1 shows how CLVM works overall. The framework spots harmful activity by looking at how individual nodes in the network behave. During route discovery, nearby nodes are picked, and each node confirms it got and sent on data packets.

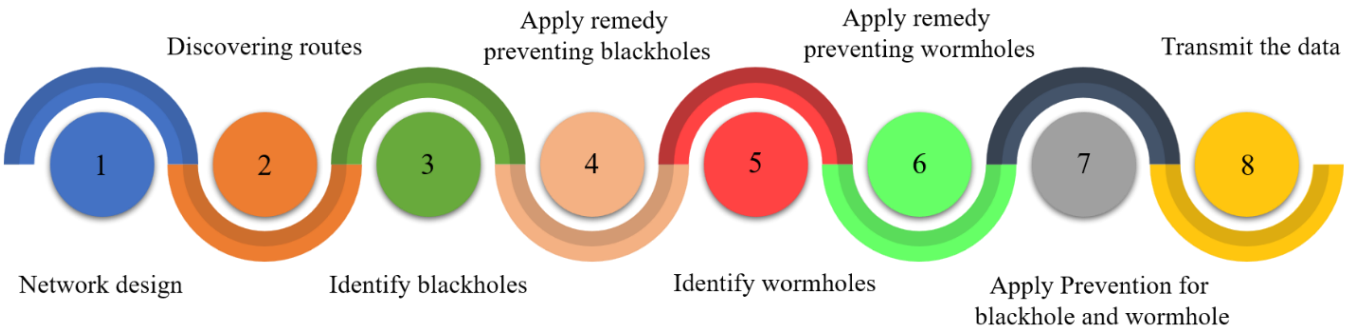


Fig. 1. Workflow of the cross-layer validation mechanism.

The Round-Trip Time (RTT) associated with data packet communication is a key anomaly detection metric. The framework leverages the Request-To-Send (RTS)/Clear-To-Send (CTS) handshake mechanism within the Media Access Control (MAC) layer to determine RTT. Significant variations in RTT can indicate the presence of a blackhole node, where a node along the designated route discards incoming packets instead of forwarding them. Fig. 2 depicts a typical blackhole attack scenario. In this example, data is transmitted from source node S to destination node D via nodes 7 and 8. However, node 2, acting maliciously, accumulates all incoming data packets without forwarding them. The extended RTT caused by this behavior can be identified through the RTS/CTS mechanism, exposing the blackhole node.

The IEEE 802.11 standard defines the Media Access Control (MAC) layer protocol for wireless local area networks (WLANs). This protocol leverages Carrier Sense Multiple Access with Collision Avoidance (CSMA/CA) to minimize collisions during data transmission. CSMA/CA mandates that nodes listen for ongoing transmissions before initiating their own, significantly reducing the likelihood of packet collisions.

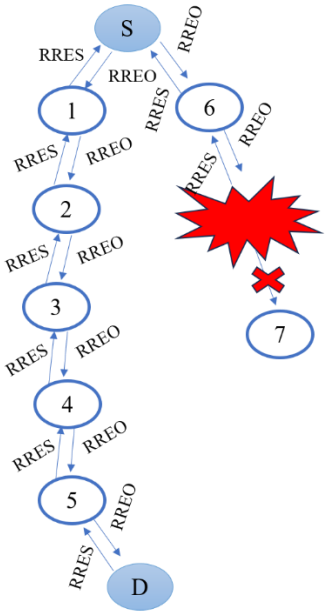


Fig. 2. Blackhole attack scenario in a WAN.

In CSMA/CA, the Request-To-Send/Clear-To-Send (RTS/CTS) handshake process offers a way to cut down on collisions even more when hidden terminals are involved. This method works like this: A node that wants to send data first transmits an RTS frame to the receiver it's aiming for. When the receiver gets this RTS frame, it sends back a CTS frame, which sets aside the channel for the upcoming data transmission. Other nodes in the area pick up on this CTS frame and hold off on sending anything during this reserved time slot, which stops collisions from happening.

While the RTS/CTS handshake solves the hidden terminal problem, it adds extra work because of the RTS and CTS frame exchange. This extra work can have a big impact on network speed when many nodes are active in a small area. Researchers have looked into other ways to reduce this extra work, but the possible benefits haven't been worth the added complexity these changes would bring.

Another consideration for WLANs is the power consumption associated with RTS/CTS frames and data packets. The Power Control Mechanism (PCM) can adjust transmission power levels based on specific needs. Typically, RTS/CTS frames are transmitted at a higher power level (Pmax) to ensure wider reception, whereas data packets might utilize a lower power level to conserve energy. However, PCM might occasionally raise the transmission power of data packets to Pmax to overcome potential signal degradation. Acknowledgment (ACK) packets are generally transmitted at a lower power level.

The importance of minimizing collisions in the MAC layer stems from the power consumption associated with retransmissions. Retransmissions not only waste bandwidth but also deplete battery life in mobile devices. While RTS/CTS-based protocols offer advantages, they do not eliminate the hidden terminal and exposed terminal problems, especially in high-density networks. Furthermore, migrating such protocols to cloud-based environments introduces additional challenges that must be addressed. This paper proposes a secure routing protocol designed to establish reliable communication paths within a network while mitigating the risks of wormhole nodes. The protocol operates under the following assumptions:

- Transmission range: All participating nodes are confined within a predefined transmission range (R).
- Node mobility: Nodes are considered stationary for routing calculations. Real-world deployments might involve mobile nodes, requiring adjustments to the protocol.
- Neighbor discovery: Nodes can discover and communicate with neighboring nodes within their transmission range.

The core objective of the protocol lies in identifying a secure path between a source node (S) and a destination node (D). The distance between these nodes (d) is calculated using Eq. (1), which factors in the transmission range (R) and the average node speed (V). The transmission range can be dynamically adjusted within a predefined threshold based on the varying distances between nodes. A weighted average distance is also employed as a stopping condition for route discovery.

$$Dist_{SD} = \frac{R - d}{V(R - D)} \quad (1)$$

To enhance security, each node maintains a counter variable initialized to zero. This counter is incremented whenever a designated operator node retrieves data from a particular node. The operator node can connect and disconnect from any node within the network. The counter reflects the number of interactions a node has had with the operator. If multiple operators collect data from the same node (node-S), the data on the destination node stored by the operator with the higher counter value takes precedence. The counter range is also configurable, with a minimum value of zero and a maximum value determined by the network's reach.

Unique identifiers are assigned to each node to facilitate secure communication. Data packets transmitted within the network encompass various fields, including a packet ID, distance traveled, counter value, and potential information regarding intermediate nodes. The validity of these parameters, including details about intermediate nodes, is verified at each network layer until the data reaches its intended destination.

Route discovery and data transmission processes leverage a routing table (R-table) that stores information about nodes and established routes. This information is constantly compared against the data packets to ensure validity. Since source nodes are assumed to be geographically close, any node can access details on neighboring nodes. Data is then forwarded to the nearest available node along the designated path.

The paper highlights a potential security concern: a wormhole attack scenario. In this scenario, a malicious node (node-S) intercepts data packets from the source node and transmits them to another colluding node (node-8) closer to the destination. Node-S then impersonates node-7, the intended recipient of the data from the source, by altering its ID to match node-7's. Consequently, the source node is deceived into believing it communicates directly with node-7, establishing a wormhole connection. The proposed protocol must have mechanisms to identify and counteract such wormhole attacks.

Unlike traditional routing protocols, where multiple nodes might operate on the same radio frequency, this approach assigns unique channels to individual nodes. This distinctiveness allows the source node to verify the legitimacy of neighboring nodes by transmitting on a randomly chosen channel.

The core principle assumes a legitimate neighboring node can detect a message transmitted on its designated channel. In contrast, a wormhole node lacking knowledge of the correct channel will miss the transmission. The probability of a source node failing to detect a wormhole node through a single random channel test can be calculated using Eq. (2). In this equation, 'n' represents the total number of neighbors, and 'S' represents the number of suspected wormhole nodes within the set of neighbors.

$$P_r = \sum_{all\ S,M,G} P_r(S,M,G)P_r(detection|S,M,G) = \sum_{all\ S,M,G} \frac{\binom{s}{S}\binom{m}{M}\binom{g}{G} S-(m-M}{\binom{n}{c} c} \quad (2)$$

The channel frequency diversity approach can be extended to enhance detection accuracy by conducting multiple rounds of random channel tests (denoted by 'r' in Eq. (3)). With each round, the source node selects a random subset of neighbors and a random channel for transmission. Eq. (3) calculates the probability of failing to detect a wormhole node after 'r' rounds of testing.

$$P_r(decontt = 1 - P_r(nondetection)_{1round}^{r} = 1 - (1 - P_r((P_r)_{1round}^{r} = 1 - \left(1 - \sum_{all,S,M,G} \frac{\binom{s}{S}\binom{m}{M}\binom{g}{G} S-m-M)}{\binom{n}{c} c}\right) r \quad (3)$$

This technique assumes a network scenario where a node's neighbors might encompass 'S' wormhole nodes, 'M' malicious nodes of other types, and 'G' legitimate nodes. Due to practical constraints, the source node can only test a limited number of neighbors at a time. Eq. (2) factors the possibility of encountering a wormhole node, a malicious node of a different type, or a legitimate node within the chosen subset of 'C' neighbors. By analyzing the probability ratio derived from Eq. (2) and Eq. (3), it can be concluded that the channel frequency diversity approach offers a viable solution for detecting wormhole nodes within various Wide Area Network (WAN) topologies.

IV. Results and Discussion

The CLVM was implemented and evaluated using the NS-2.5 network simulator. Table III summarizes the simulation parameters employed in the evaluation process. The primary objective of this evaluation was to assess the efficiency of CLVM relative to an existing approach, LBIDS [22]. The simulations were conducted in a 1000 x 1000 m network area, simulating a typical wireless ad hoc network environment. The nodes were randomly distributed across this area, with the number of nodes varying between 10 and 50 over five rounds of simulation, increasing by ten nodes per round. Node mobility was simulated using the random waypoint mobility model, with node speeds ranging from 1 to 15 m/s, reflecting the dynamic nature of real-world ad hoc networks.

The radio propagation model used was the two-ray ground reflection model, which takes into account both direct and ground-reflected paths of signal propagation, allowing for a more realistic simulation of wireless communications. The transmission range of each node was set to 250 meters, with the MAC layer using the IEEE 802.11 standard. Traffic was generated using a constant bit rate (CBR) application with packet sizes of 50 bytes, simulating a typical data transfer scenario. Energy consumption was modeled based on the remaining energy level of nodes after each round, using mechanisms such as RTS/CTS handshakes and distance checking to save energy.

TABLE III. Simulation Parameters and Values

Parameter	Value
Simulation area	1000m x 1000m
Malicious node ID count	2
Malicious node percentage	Up to 5%
Node placement	Random
Simulation duration	100 seconds
Packet size	50 bytes
Traffic type	CBR, 100 – 500
MAC protocol	802.11
Total nodes	20 – 750
Transmission range	250 m
Propagation model	Two-ray ground reflection
Node Speed	1 – 15 m/s

A. Transmission Delay Analysis

Fig. 3 compares the average transmission delay incurred by each approach across the five rounds. The results demonstrate that CLVM consistently exhibits lower transmission delay compared to LBIDS. In the case of round five with 50 nodes, LBIDS exhibits a transmission delay of 46 milliseconds (ms), whereas CLVM achieves a delay of only 24 ms. This improvement can be attributed to the efficiency gains introduced by CLVM's mechanisms, such as trust value evaluation and route selection.

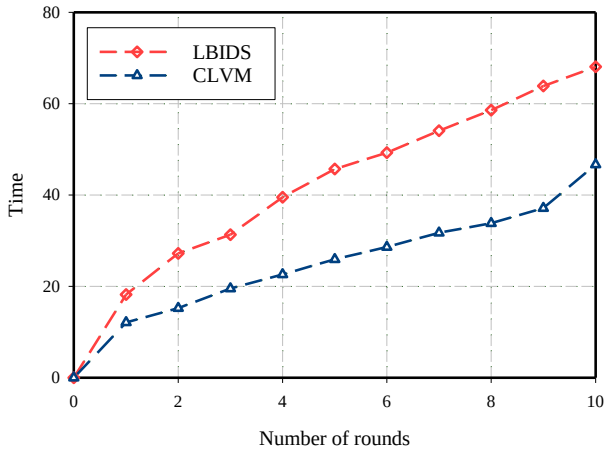


Fig. 3. Average transmission delay comparison.

B. Throughput Analysis

Throughput, measured by the successful transmission and reception of data packets, serves as another key performance metric. Fig. 4 depicts the throughput achieved by both CLVM and LBIDS across the five rounds. The results indicate that CLVM consistently delivers superior throughput compared to LBIDS. This can be primarily explained by CLVM's ability to mitigate malicious activities that disrupt data transmission within the network. For example, in round five, LBIDS achieves a throughput of 6123 packets, whereas CLVM delivers a higher throughput of 6400 packets.

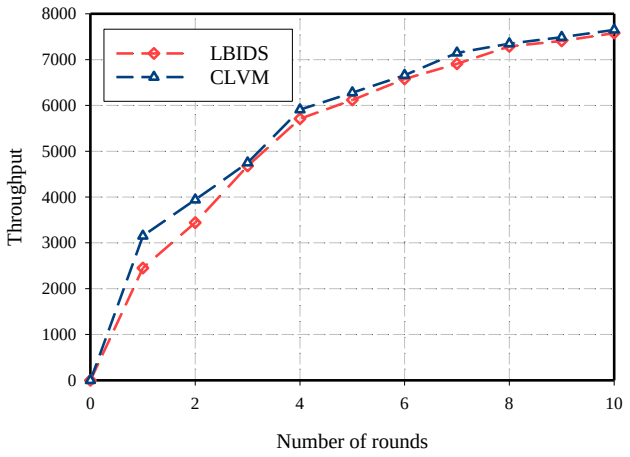


Fig. 4. Average throughput comparison.

C. Energy Consumption Analysis

The remaining energy level of network nodes after each round was evaluated to assess the energy efficiency of both approaches. Fig. 5 presents the results, indicating that CLVM nodes conserve more energy than LBIDS nodes. This is a consequence of CLVM's strategies for reducing unnecessary communication and data transmissions. Mechanisms like RTS/CTS handshakes and distance verification contribute to this energy conservation. Nodes are unable to transmit data if they fail to provide valid IDs or adhere to the RTS/CTS protocol and distance requirements. This helps to preserve node energy. The simulations reveal that in round five, the remaining energy level for LBIDS nodes is 95%, whereas CLVM nodes retain a higher energy level of 96%.

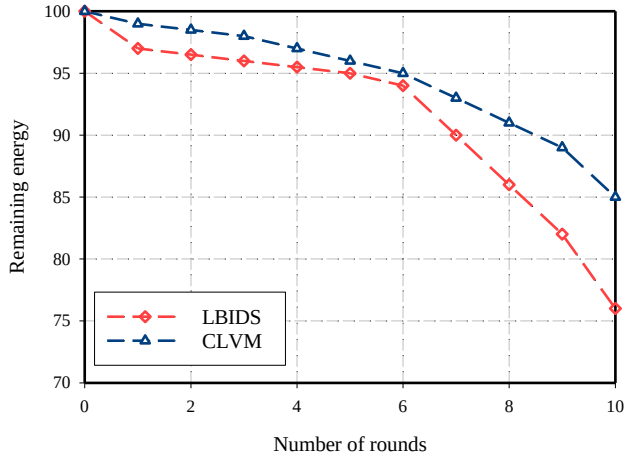


Fig. 5. Remaining energy comparison.

D. Malicious Activity Detection Analysis

The simulations also evaluated the effectiveness of both approaches in detecting malicious activities within the network. CLVM's algorithm leverages a long-established foundation and incorporates node behavior analysis for comprehensive malicious node identification. Fig. 6 compares the number of malicious activities detected using LBIDS and CLVM. The results demonstrate that CLVM significantly reduces the number of malicious activities within the network. This improvement stems from CLVM's verification of critical parameters like node ID, RTS/CTS compliance, and transmission distance during data exchange. Additionally, CLVM maintains a database for comparison purposes, enabling it to identify nodes that deviate from expected behavior and potentially block them. While LBIDS focuses on detection, CLVM prioritizes prevention by proactively identifying and mitigating potential threats.

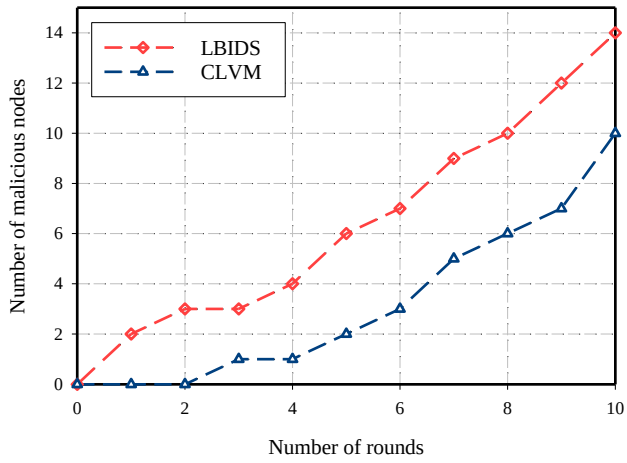


Fig. 6. The number of malicious activities comparison.

The obtained results clearly show that CLVM not only outperforms the existing LBIDS approach in practical metrics such as transmission delay, throughput, and energy consumption but also embodies significant theoretical advances. CLVM's integration into existing network protocols is achieved through its trust-based validation process, which improves route selection and mitigates malicious activity more effectively than traditional methods. By prioritizing trust assessment at multiple levels, CLVM eliminates the limitations of LBIDS, which focuses primarily on detection rather than prevention. This cross-layer approach allows CLVM to reduce transmission delays and energy consumption while maintaining high throughput, providing a more holistic and efficient network security solution. CLVM's theoretical robustness, combined with its practical effectiveness, makes it a superior alternative to existing security mechanisms in cloud-enabled wireless ad hoc networks.

The simulation setup was designed to closely mimic real-world scenarios by incorporating widely used models such as random waypoint mobility and two-ray ground reflection. These models simulate the unpredictable movement of nodes or realistic signal propagation in an open environment. The range of node speeds and the random distribution of nodes reflect the dynamic and decentralized nature of cloud-enabled wireless ad hoc networks. However, it is important to note that certain simplifications were made in the simulation. For example, environmental factors such as obstacles and interference, which can significantly impact signal propagation and network performance in real-world scenarios, have not been fully modeled. In addition, the simulations assumed idealized conditions for node operations and communications, which may differ from the more complex and variable conditions encountered in actual operations.

V. CONCLUSION

This study introduced CLVM as an innovative solution to enhance the security and efficiency of cloud-enabled wireless ad-hoc networks. Extensive evaluations demonstrated that CLVM significantly outperforms the existing LBIDS approach in key performance metrics, including transmission delay, throughput, energy consumption, and malicious activity detection. CLVM achieves a remarkable reduction in transmission delay, evidenced by a delay of only 24 milliseconds in a 50-node network compared to LBIDS's 46 milliseconds. Additionally, CLVM consistently delivers higher throughput, with a notable increase to 6400 packets in the same network configuration. Energy efficiency is another critical advantage, as CLVM nodes retain 96.59% of their energy compared to LBIDS's 95.1%, thanks to effective strategies like RTS/CTS handshakes and distance verification protocols. Moreover, CLVM excels in detecting and mitigating malicious activities, leveraging comprehensive node behavior analysis and a proactive approach to threat prevention. These improvements underscore the robustness and reliability of CLVM in securing data transmission and maintaining network integrity.

Looking forward, future research will focus on several key areas to build on the findings of this study. One possible path is to extend the CLVM to other types of wireless networks, such as Vehicular Ad-hoc Networks (VANETs) or industrial IoT environments, where security challenges are even more pronounced. Additionally, optimizing CLVM for larger deployments with hundreds or thousands of nodes is critical to ensure its scalability and efficiency in various network scenarios. Further research could also include integrating CLVM with advanced machine learning algorithms to improve its ability to detect and adapt to new types of security threats in real-time. These future efforts aim to refine and expand CLVM's capabilities, ensuring its relevance and effectiveness in the ever-evolving network security landscape.

REFERENCES

[1] M. A. Tofighi, B. Ousat, J. Zandi, E. Schafir, and A. Kharraz, "Constructs of Deceit: Exploring Nuances in Modern Social Engineering Attacks," in International Conference on Detection of Intrusions and Malware, and Vulnerability Assessment, 2024: Springer, pp. 107-127, doi: https://doi.org/10.1007/978-3-031-64171-8_6

[2] S. R. Abdul Samad et al., "Analysis of the performance impact of fine-tuned machine learning model for phishing URL detection," Electronics, vol. 12, no. 7, p. 1642, 2023.

[3] B. Pourghebleh, N. Hekmati, Z. Davoudnia, and M. Sadeghi, "A roadmap towards energy‐efficient data fusion methods in the Internet of Things," Concurrency and Computation: Practice and Experience, vol. 34, no. 15, p. e6959, 2022.

[4] B. Pourghebleh and V. Hayyolalam, "A comprehensive and systematic review of the load balancing mechanisms in the Internet of Things," Cluster Computing, pp. 1-21, 2019.

[5] A. Lolai et al., "Reinforcement learning based on routing with infrastructure nodes for data dissemination in vehicular networks (RRIN)," Wireless Networks, vol. 28, no. 5, pp. 2169-2184, 2022.

[6] D. K. Sharma, S. K. Dhurandher, and S. Kumar, "Hierarchical search-based routing protocol for infrastructure-based opportunistic networks," International Journal of Innovative Computing and Applications, vol. 12, no. 2-3, pp. 134-145, 2021.

[7] A. Förster et al., "A beginner's guide to infrastructure‐less networking concepts," IET Networks, vol. 13, no. 1, pp. 66-110, 2024.

[8] S. Al Ajrawi and B. Tran, "Mobile wireless ad-hoc network routing protocols comparison for real-time military application," Spatial Information Research, vol. 32, no. 1, pp. 119-129, 2024.

[9] M. Sohail et al., "Routing protocols in vehicular adhoc networks (vanets): A comprehensive survey," Internet of things, vol. 23, p. 100837, 2023.

[10] V. Chandrasekar et al., "Secure malicious node detection in flying ad-hoc networks using enhanced AODV algorithm," Scientific Reports, vol. 14, no. 1, p. 7818, 2024.

[11] S. Dong, H. Su, Y. Xia, F. Zhu, X. Hu, and B. Wang, "A comprehensive survey on authentication and attack detection schemes that threaten it in vehicular ad-hoc networks," IEEE Transactions on Intelligent Transportation Systems, 2023.

[12] K. Vamshi Krishna and K. Ganesh Reddy, "Classification of distributed denial of service attacks in VANET: a survey," Wireless Personal Communications, vol. 132, no. 2, pp. 933-964, 2023.

[13] V. Hayyolalam, B. Pourghebleh, and A. A. Pourhaji Kazem, "Trust management of services (TMoS): Investigating the current mechanisms," Transactions on Emerging Telecommunications Technologies, vol. 31, no. 10, p. e4063, 2020.

[14] B. Pourghebleh, K. Wakil, and N. J. Navimipour, "A comprehensive study on the trust management techniques in the Internet of Things," IEEE Internet of Things Journal, vol. 6, no. 6, pp. 9326-9337, 2019.

[15] P. Zhang, S. Wang, K. Guo, and J. Wang, "A secure data collection scheme based on compressive sensing in wireless sensor networks," Ad Hoc Networks, vol. 70, pp. 73-84, 2018.

[16] M. Al-Shayeji and F. Ebrahim, "A secure and energy-efficient platform for the integration of Wireless Sensor Networks and Mobile Cloud Computing," Computer Networks, vol. 165, p. 106956, 2019.

[17] T. Wang, G. Zhang, M. Z. A. Bhuiyan, A. Liu, W. Jia, and M. Xie, "A novel trust mechanism based on fog computing in sensor–cloud system," Future Generation Computer Systems, vol. 109, pp. 573-582, 2020.

[18] S.-J. Hsiao and W.-T. Sung, "Employing blockchain technology to strengthen security of wireless sensor networks," IEEE Access, vol. 9, pp. 72326-72341, 2021.

[19] K. Haseeb, Z. Jan, F. A. Alzahrani, and G. Jeon, "A secure mobile wireless sensor networks based protocol for smart data gathering with cloud," Computers & Electrical Engineering, vol. 97, p. 107584, 2022.

[20] Sharmila, P. Kumar, S. Bhushan, M. Kumar, and M. Alazab, "Secure key management and mutual authentication protocol for wireless sensor network by linking edge devices using hybrid approach," Wireless Personal Communications, vol. 130, no. 4, pp. 2935-2957, 2023.

[21] S. Gayathri and D. Surendran, "Unified ensemble federated learning with cloud computing for online anomaly detection in energy-efficient wireless sensor networks," Journal of Cloud Computing, vol. 13, no. 1, p. 49, 2024.

[22] S. A. Razak, S. Furnell, N. Clarke, and P. Brooke, "A two-tier intrusion detection system for mobile ad hoc networks–a friend approach," in Intelligence and Security Informatics: IEEE International Conference on Intelligence and Security Informatics, ISI 2006, San Diego, CA, USA, May 23-24, 2006. Proceedings 4, 2006: Springer, pp. 590-595.

UWB Printed MIMO Antennas for Satellite Sensing System (SRSS) Applications

Wyssem Fathallah[1], Chafai Abdelhamid[2], Chokri Baccouch[3], Alsharef Mohammad[4], Khalil Jouili[5], Hedi Sakli [6]*
SYS'COM Laboratory LR99ES21, National Engineering School of Tunis, Tunis El Manar University, Tunis, 1002, Tunisia[1, 3]
MACS Research Laboratory RL16ES22, National Engineering School of Gabes, Gabes University, Gabes, 6029, Tunisia[2, 6]
Department of Electrical Engineering, College of Engineering, Taif University, Taif, Saudi Arabia[4]
Laboratory of Advanced Systems, Polytechnic School of Tunisia (EPT), Marsa, 2078, Tunisia[5]
EITA Consulting, 7 Rue du Chant des oiseaux, 78360 Montesson, France[6]

***Abstract*—The deployment of ultra-wideband (UWB) technology offers enhanced capabilities for various Internet of Things (IoT) applications, including smart cities, smart buildings, smart aggregation, and smart healthcare. UWB technology supports high data rate communication over short distances with very low power densities. This paper presents a UWB printed antenna design with multiple input and output (MIMO) capabilities, specifically tailored for Routed Satellite Sensor Systems (SRSS) to enhance IoT applications. The proposed UWB printed antenna, designed for the 2–18 GHz frequency band, has overall dimensions of 14.5 mm x 14.5 mm, with an efficiency exceeding 70% and a gain ranging from 2 to 6.5 dB. Both simulated and measured reflection parameters (|S11|) at the antenna input show strong agreement. Furthermore, a compact MIMO system is introduced, featuring four closely spaced antennas with a gap of 0.03λ, housed in a 60 mm x 48 mm module. To minimize coupling effects between the antennas, the design incorporates five Split Ring Resonator (SRR) elements arranged linearly between the radiating elements. This arrangement achieves a mutual coupling reduction to -35 dB at 8 GHz, compared to -20 dB isolation in systems without SRR. The results demonstrate that the proposed MIMO antenna system offers promising performance and meets the requirements for effective space communication within satellite sensor networks.**

***Keywords*—5G antenna; 5G satellite networks; millimeter band; wireless communications; SRR; IoT**

I. Introduction

UWB technology was originally developed for military applications but began to be used in civilian applications. Arousing growing interest within the scientific and industrial community, it was transferred to telecommunications applications [1-12]. This technology is used in other applications, such as structural health monitoring (SHM) for large structures, and in particular in the aeronautics and space fields are under development [13].

Other applications of wireless sensor networks operating using UWB spectrums appear in smart homes, the biomedical field, natural disaster detection, intrusion detection, pollutant detection, agriculture, and many other fields [14]. They provide great ease of use and reduce the cost and time of deployment. One can't talk about UWB without mentioning Internet of Things (IoT), the concept of connected devices, continues to show promising progress, and now, with the re-emergence of UWB technology, IoT devices that require location and movement data are performing better than ever for the cost and time of deployment [15]. Thanks to UWB interoperability, this communication protocol can be used to take advantage of smart technologies such as Bluetooth, WiFi, and the IoT's. UWB can play a key role in redesigning the IoT devices already available and in introducing more sophisticated networks of interconnected devices in the future [16]. In wireless communication systems, antennas play a very important role as it is the backbone of any wirelessly communicating system. To meet the emerging requirements of the smart IoT devices an enhanced performance antenna is required [17]. Moreover, to meet the requirements of the MIMO system a compact yet low mutual coupling antenna has become essential for communication systems.

Thus, the researcher puts a lot of effort to design compact size UWB antenna for IoT devices [18–28]. For instance, a metamaterial inspired circular split ring resonator shaped antenna is designed for UWB application in study [18]. Although the antenna offers a UWB operational band with a notch band ranges 7 – 8 GHz it had a setback of bigger dimension along with lower cutoff frequency around 3.4 GHz. In study [19] a dual stub loaded antenna having dual reconfigurable notch band is proposed for UWB applications. However, the insertion of the diodes and biasing circuit causes the degrading of antenna performance resulting in a low gain throughout the band of interest. Likewise, in [20] a rectangular monopole antenna exhibiting broad bandwidth is converted into a UWB antenna by initially modifying the ground plane, then by truncating the radiating structure and finally loading some parasitic elements. The resulting antenna offers UWB ranges 2.5–18 GHz at the cost of complex structure along with low gain and large physical size. The study in [21– 22] present the conversation of circular and semi-circular monopole antenna into UWB antenna by truncating the radiating structure along with ground plane. Resistor loaded anti-spiral shaped UWB antenna is presented in [24], where flexible antenna is designed by compromising the size of antenna along with partial covering the UWB spectrum. Moreover, the work reported in study [26–28] has a relatively big size as compared to other literary works. Furthermore, none of the discussed can be used for IoT applications due to their SISO nature.

Considering the requirements of IoT and future networks that require MIMO antenna, there is a dire need to design UWB MIMO antenna with low mutual coupling in study [29–35]. A

compact size two-port UWB antenna is presented in study [29] where two open ended stubs are loaded, and their structure is modified to achieve a mutual coupling of -15 dB between adjacent elements. On the other hand, a truncated corner shaped UWB antenna is utilized to design a 4-element UWB MIMO antenna. The antenna elements are placed orthogonal to each other to achieve low mutual coupling along with defected ground structure to further reduce the mutual coupling to < -20 dB. A flower shaped semi-fractal antenna is proposed in study [31], which is converted into 4-element MIMO antenna. Four-open ended stubs were loaded to reduce the mutual coupling among MIMO elements; however, the MIMO antenna system only offers the mutual coupling of < -18 dB. Likewise, in study [32] a hollow ground plane and inverted L-shaped stub loaded UWB MIMO antenna is presented. The antenna offers UWB ranges 2.84 – 15.88 GHz having a overall size of 58 x 58 mm2 along with a setback of high mutual coupling. Another MIMO antenna for UWB application is proposed in study [33] in which the orthogonal placement of MIMO element is utilized to achieve a low mutual coupling while compromising the overall size of antenna. A tapered-fed circular shaped MIMO antenna is designed for UWB applications while neutralization lines along with DGS and inverted L-shaped open stubs are utilized to reduce the mutual coupling [34]. On the other hand, a pair of C-shaped parasitic patches are loaded at the back of antenna to reduce the mutual coupling [35]. However, both the works have the disadvantage of bigger size along with complex geometrical structure.

All of the aforementioned works either lack in covering complete UWB spectrum or have large physical dimension along with most of them suffering from complex geometrical structure. Therefore, this study focuses on the design of geometrically simple yet UWB MIMO antenna having low mutual coupling along with high gain is given in Section II. The antenna design methodology is discussed in Section III, afterward the results will be discussed in Section IV, and the manuscript is concluded in Section V.

II. Antenna Design

The design methodology started from designing a conventional rectangular patch antenna that resonates at central frequency of 10 GHz, as depicted in Fig. 1. The length and width of the rectangular patch along with microstrip feedline can be estimated using the equations provided in [36]. The setback of conventional microstrip antenna, the narrow bandwidth, is nullified using the partial ground plane technique along with truncated corners, as shown in Fig. 1 (a). This helps in achieving a broad |S11| < -10dB impedance bandwidth ranging 4–9.5 GHz, as illustrated in Fig. 1(b). Although there is significant improvement observed in bandwidth, still the antenna didn't cover the entire UWB spectrum ranges 3-10.3 GHz. Therefore, to further improve the performance of the antenna another iteration is performed by truncating the lower corners of radiators along with implementing the DGS, as shown in Fig. 1(a). After optimizing the results, the antenna again offers wide operational bandwidth ranges 5–15.7 GHz, as shown in Fig. 1(b). It is important to note here that although the bandwidth of the antenna increases but operational spectrum shifted toward higher frequency.

Thus, there is need to shift the lower end of the resonating band. For said purpose two u-shaped slots were utilized, they are etched from the radiator and by optimizing the parameters of the slots the antenna resonance can be shifted toward lower side. The other parameters of the proposed antenna are also optimized which results in UWB spectrum ranges 2 – 18 GHz, covering UWB, extended UWB and Ku-band, as shown in Fig. 1(b).

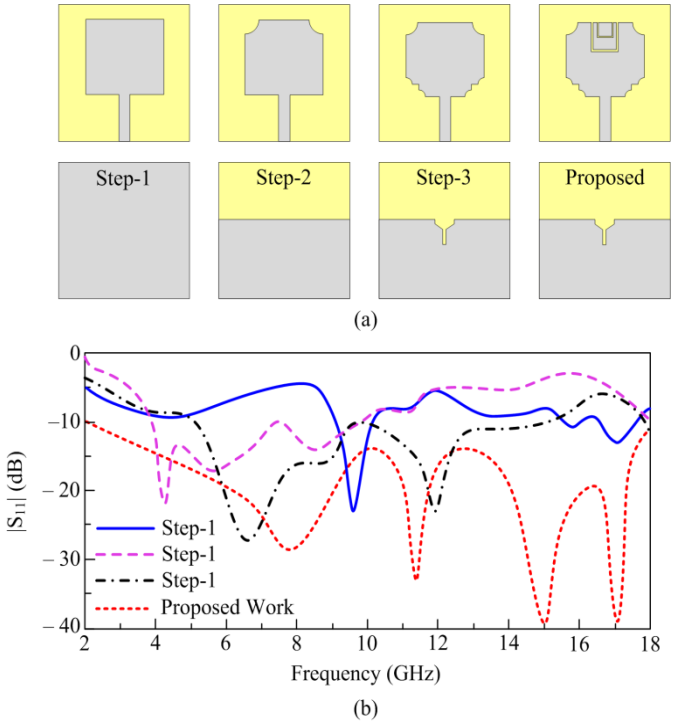


Fig. 1. (a) Geometrical configuration (b) Reflection coefficient of various modifications in radiator and ground plane.

Fig. 2 depicts the geometrical configuration of the proposed ultra-wideband antenna, while Table I summarizes the optimal parameters obtained by simulation.

For in-depth understanding of UWB behavior of antenna, Fig. 3 depicts the impedance characteristics of the antenna. It can be observed that the proposed antenna offers real impedance value around 50 Ω while the imaginary value stays around 0 Ω which also proves the ultra-wideband behaviour of the antenna.

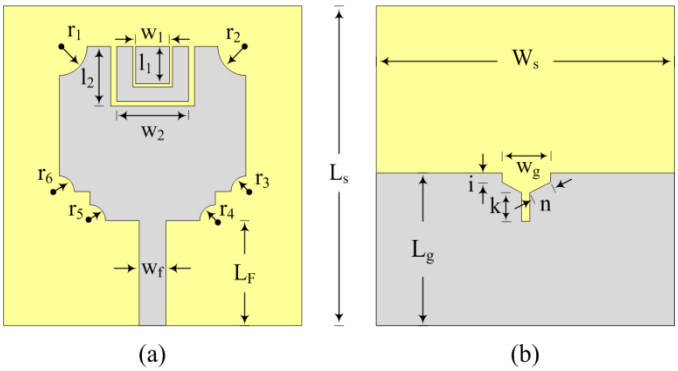


Fig. 2. Final geometry of the proposed antenna: (a) the front view, (b) backside view.

TABLE I. PROPOSED ANTENNA DIMENSIONS IN DETAIL

Elements	*Parameters*	*Values(mm)*
Patch	$r_1=r_2$	2.5
	W_1	3.4
	L_1	7.5
	$r_3=r_4=r_5=r_6$	1.5
	W_3	6.17
	L_f	13.5
	W_f	3
	W_2	7.2
	F	0.3
Dielectric substrate	W_s	25
	L_s	30
	H_s	1.6
	ε_r	4.4
Ground Plane	L_g	12.5
	W_g	3.5
	n	2.4
	i	0.75
	k	2.3

The gain of the proposed UWB antenna is shown in Fig. 4. The antenna offers a minimum gain of 2.4 dB around 2 GHz, it tends to start increasing for higher frequency and reach up to the maximum value of 5.8 dB around 17.7 GHz. Thus, the gain results strength the potential of the proposed work for UWB and Ku-band applications.

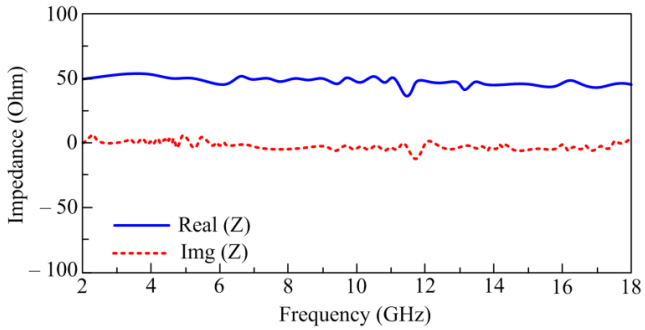


Fig. 3. Proposed antenna impedance characteristics.

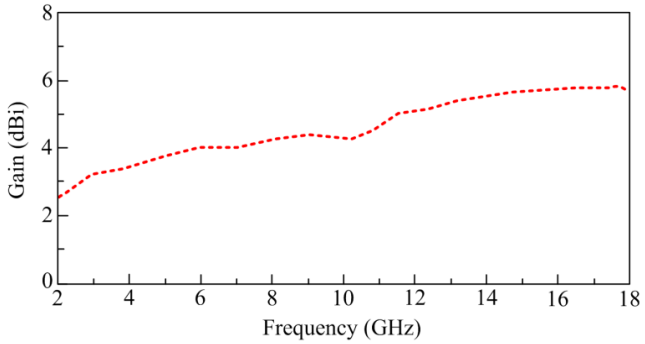


Fig. 4. Peak gain of proposed work.

The radiation pattern of the proposed antenna at the selected frequencies of 8 and 15 GHz is shown in Fig. 5. The antenna offers a nearly omni directional radiation pattern in E-plane for both selected frequencies, while for H-plane the antenna offers a dual beam like structure. Moreover, Fig. 6 illustrates the radiation efficiency in the operational band, where antenna offers a minimum efficiency of 80% throughout the band of interest.

The distribution of the current at 8 GHz and 15 GHz of the proposed antenna is shown in Fig. 7. It can be seen that additional paths for surface currents are formed when slits are present. This results in a second resonance, increasing the bandwidth.

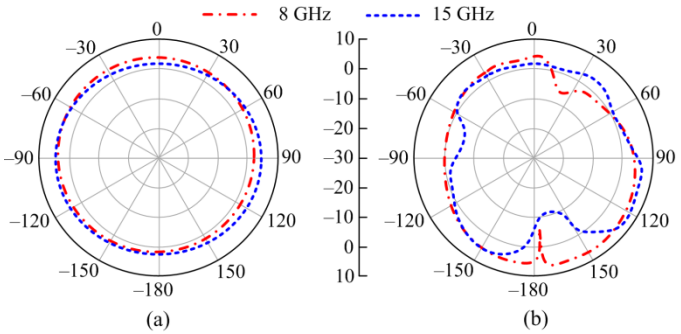


Fig. 5. Radiation pattern of proposed antenna in (a) E-plane and (b) H-plane.

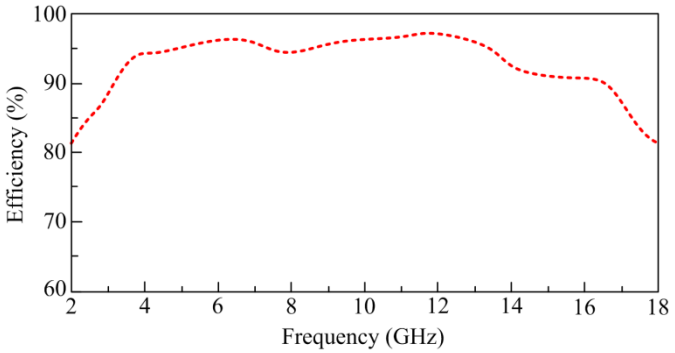


Fig. 6. Efficiency variation of proposed work.

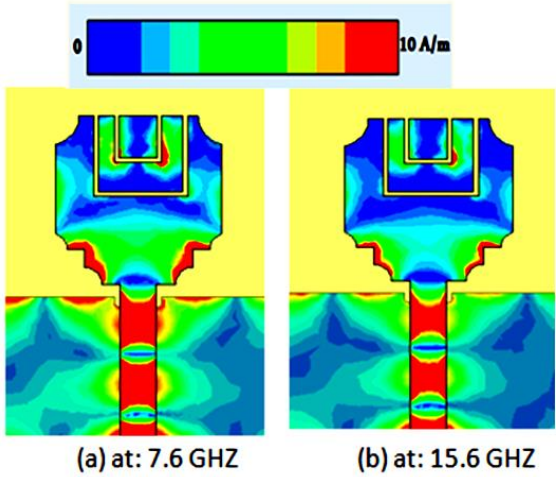


Fig. 7. Current distribution of proposed UWB antenna.

To validate the various performance parameters of the proposed UWB antenna a sample prototype is fabricated and later used for measurements, as shown in Fig. 8. Various parameters including |S11| and radiation pattern were measured and compared with simulated results. The comparison among estimated and measured |S11| results of proposed UWB antenna is shown in Fig. 9. A strong comparison among both results is found having identical wideband operation, a little deviation is observed in values of return loss which may be due to inaccuracy of measurement setup or due to fabrication tolerance.

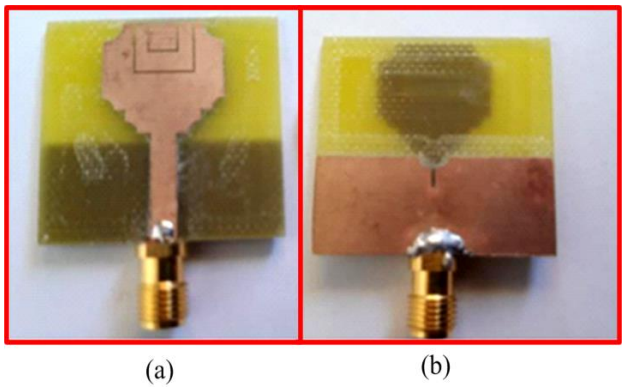


Fig. 8. Fabricated Antenna: (a) Top view and (b) Bottom view.

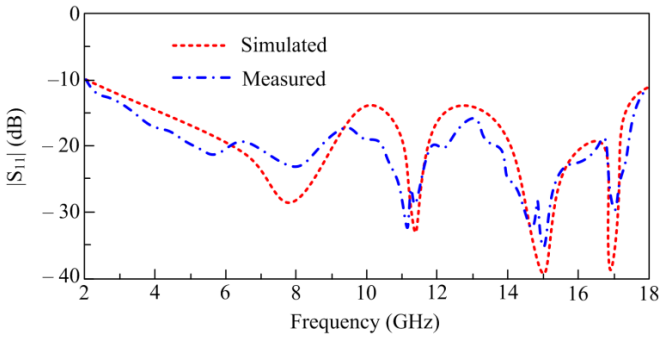


Fig. 9. Comparison of reflection coefficient among simulated and measured results.

For the selected frequency of 8 GHz, the comparison among the radiation patterns found using the EM tool and measurement is presented in Fig. 10. A strong correlation is observed, with nearly identical patterns in both cases. Thus, in terms of all performance parameters, strong results are observed among the simulated and measured values, which illustrates the performance stability of the proposed antenna.

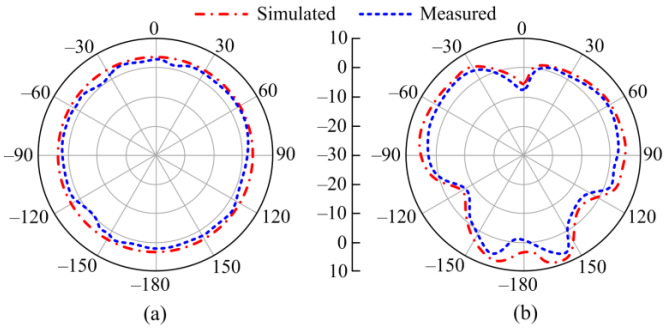


Fig. 10. Radiation pattern at 8 GHz: (a) E-plane and (b) H-plane.

III. Common Techniques to Reduce Mutual Coupling among MIMO Antenna Elements

Diversity and MIMO systems require good isolation between antennas. Much research has been done over the years to find techniques to reduce mutual coupling and increase isolation between antenna elements. In this part, a short explanation of various major techniques is done to better understand the working and then utilized a combination of them to reduce mutual coupling of proposed MIMO antenna system.

A. Structural Change of Ground Plane

This method is well known as DGS and consists of modifying the structure of the ground plane to change the current distribution [37]. MIMO systems use the effect of notch filters to minimize mutual coupling between radiating elements. In fact, a common use of the DGS method is to insert a slot in the ground plane [38]. The DGS solution is very easy to implement because its operation depends on the resonant frequency and not on the antenna type. However, the main drawback of this method is mainly integration issues on mobile phones.

B. Use of EBG (Electromagnetic Band Gap) Structure

Generally, in antenna decoupling, the EBG structure is similar to a notch filter [39]. In fact, a typical EBG cell has a mushroom-shaped structure containing patches and grounded vias. Moreover, the EBG structure can be used as a magnetic wall through which the phase of the reflection coefficient becomes zero for an incident wave. Therefore, surface wave propagation will be suppressed [40]. On the other hand, this method requires considerable space, especially for low frequencies [41]. Moreover, this solution is not commonly used in practice due to its complexity and large size [42]. In study [43], four rows of fork-shaped EBG patches were inserted between the E-plane coupling antennas to reduce mutual coupling. A mutual coupling reduction of 6.51 dB was achieved at 5.2 GHz when using the EBG structure.

C. Use of the Neutralization Line

This approach is generally used to ensure better isolation between two PIFA (Planar Inverted-F Antenna) antennas [44]. Indeed, it is a question of introducing a simple suspended metal line, integrated between the power supplies or the short circuits of the PIFA antennas. In addition, the neutralization line supports strong currents so that the direction of the current is radiated towards the antenna itself and not towards the power connector of the second antenna. It is also possible to cut all path couplings (OTA (over the air) couplings and ground plane power couplings) by changing the dimensions of the neutralization lines. For example, in study [45], the authors inserted an interrupted neutralization wire physically connected to two PIFA elements (operating in the UMTS band of 1920–2170 MHz).

The introduction of neutralizing lines was used to cancel pre-existing mutual coupling. This is because the line stores a certain amount of current that is fed from one antenna element to the other. In other words, the antenna achieved less than -18 dB of mutual coupling at a frequency of 2 GHz because an additional path was created to compensate for the antenna-to-antenna current on the circuit board. Recently, a new print diversity monopole antenna for WiFi and WiMAX applications was

presented in study [46]. It's based on the same concept, but with a much more complex kill line integration. The antenna consists of two crescent-shaped radiators placed symmetrically about a faulty ground plane, with neutralizing lines to achieve a bandwidth of 2.4 to 4.2 GHz and mutual coupling of less than -17 dB connected between them.

D. Using an Isolation Network

This method aims to reduce the mutual impedance or transmission coefficient between the radiating elements to zero while maintaining good impedance matching in each element [47]. In the literature, we have found various antenna array configurations based on 180° hybrid couplers or RF switches.

E. Use of Parasitic Resonators

This is similar to neutralizing wire-based solutions in that the parasitic resonator is integrated in the middle of the two antennas to minimize mutual coupling between them. In other words, this solution artificially creates an additional coupling path between antennas. However, simply changing the spacing between the radiating elements requires changing the structure of the resonator to ensure good isolation. This limitation then hinders the serial application of the parasitic resonator [48].

Although the above solutions overcome the coupling effect between the antennas, there are other innovative and more efficient means. In fact, recently, the development of metamaterials to design and optimize antenna characteristics has shown great importance, not only to minimize antenna size, but also to provide better isolation and reduce the spacing between elements radiating [49].

At this stage of our research, part of the solutions developed within the framework of the study of multi-antenna systems intended for wireless communications or intended to be used in applications of the diversity or MIMO type have been presented [50-53]. These solutions have been investigated with the aim of covering several wireless communication standards and improving the isolation between the antennas that make up these systems. Consequently, having antennas very close to each other and operating in neighboring or even identical frequency bands is an ever-present challenge because communication systems must have increasingly new, numerous, and innovative functionalities.

In the proposed study the EBG structure along with DGS is utilized to reduce the mutual coupling and explain in forthcoming sections.

IV. Results and Discussions

Increasing the number of transmit and receive antennas without increasing radiated power can improve communication quality and channel capacity. The MIMO antenna design proposed in this article has four identical radiating elements, extracted by already designed unit element, as shown in Fig. 11. The overall dimensions of MIMO antenna system are Lsub × Wsub (60 × 48 mm2), having edge to edge distance between all elements is 5mm, as depicted in Fig. 11.

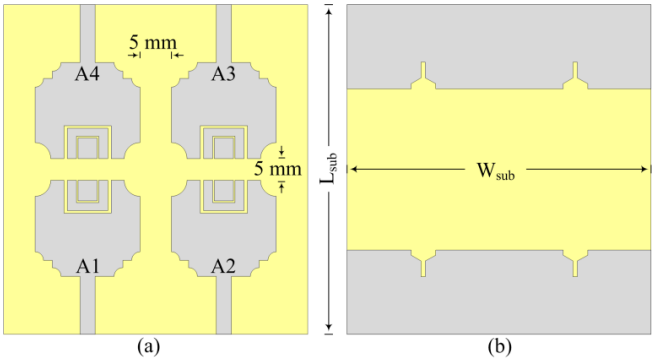


Fig. 11. UWB-MIMO antenna system without decoupling structure.

A. Initial Prototype Simulation Results

The simulated S-parameters of the 4-element MIMO antenna without any decoupling structure are shown in Fig. 12. Due to the symmetry in structure, the S-parameter analysis can be easily performed. The |S11| of the MIMO antenna remain almost identical to the unit element having |S11| < -10dB bandwidth ranges 2–18 GHz. On the other hand the transmission coefficient |Sij|, where i,j=1,2, offers a high value of more than -10 dB. This high value of transmission coefficient is not acceptable for present and future MIMO systems. The easiest way to reduce the value of |Sij| is to increase the edge-to-edge distance, however, it will increase the size of the antenna and thus compactness will vanish.

According to Fig. 13, the surface current circulation in the radiating element is high which leads to strong coupling as already indicated by the high values of |Sij|. Surface currents generated by the excited antenna flow to the other non-excited elements, as shown in Fig. 13.

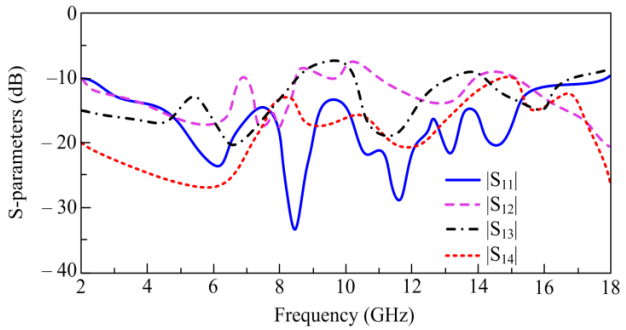


Fig. 12. S-parameters of UWB-MIMO antenna system.

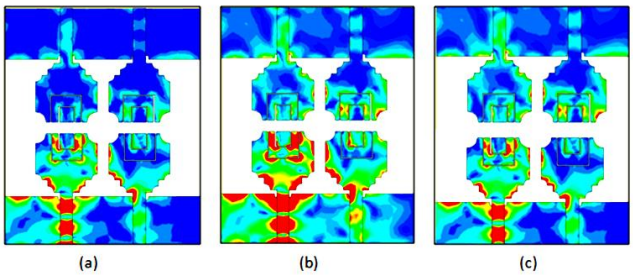


Fig. 13. Effect of the current distribution of antenna 1 on the three other antennas at frequencies: (a) 8 GHz (b) 15 GHz.

Only one element of the MIMO antenna is energized (antenna 1), but we can see currents in the other three components that are not energized. As a result, without excitation, hot spots occur on at least one additional radiating element (2, 3 or 4). Although the maximum current on the unexcited antenna does not equal the maximum current on the excited antenna, it is sufficient to increase coupling between the components. However, compared to the minimum frequency in the operating band (2 GHz), the distance between them is only 0.03λ.

The mutual coupling comes from the capacitive coupling between the radiating elements and the current flowing on the PCB. Because the isolation value obtained is insufficient for a powerful multi-antenna system, thus a more effective way is required to reduce the coupling while leaving the four elements in their original places, i.e., to have a compact size MIMO antenna.

B. Design of SRR Unit Cells

A two small rectangular square split ring resonator (SRR) unit cells operating at frequencies corresponding to well-defined bands WiMAX and X-Band. Afterwards, a rectangular strip is loaded to widen the band of coverage by the SRR unit cell, as shown in Fig. 14.

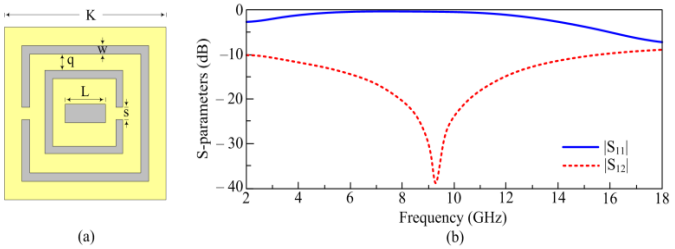


Fig. 14. (a) Geometry of proposed SRR (b) Reflection and transmission coefficients of the proposed SRR.

For ease of understanding the S-parameter-based method is adopted and used to analyze the performance of the Metamaterial unit cell. The designing approach is based on a systematic approach to unit cell design, for this reason, the overall cell size should be smaller than the wavelength (dλ) (approximately λ0/11 in presented case). The geometrical structure along with respective S-parameters is shown in Fig. 14. The unit cell of the metamaterial operates from 2.2–16.7 GHz band, covering almost the entire band offered by proposed unit element of the MIMO antenna. The metamaterial cell is mounted on an FR4 type epoxy dielectric substrate with a dielectric constant of 4.4, a loss factor of 0.02, and a thickness of 1.6mm. This SRR square has an outside side of 2 mm, a track width of 0.2 mm, and a 0.3 mm gap cut on one of these sides. It is two concentric rings with a spacing of 0.15 mm between them and an inner ring measures 1.3 mm on the outside. Before starting the simulation, an electric and magnetic wall was installed in a 2.5 × 2.5 × 5 mm^3 radiation box, these dielectric walls must be used to verify the SRR's requisite excitation conditions. The magnetic field must be positioned along the ring's axis to provide greater magnetic excitation and circulation of the induction current. To do this, two domain walls parallel to the XY- and XZ-plane and an electric wall are provided. The electric field is parallel to OY-axis and the propagation vector is along the OX-axis to ensure a symmetrical current distribution. The optimized dimensions of the unit element are enlisted in Table II.

TABLE II. DIMENSIONS OF SRR UNIT CELL

Parameters	*Values(mm)*
s	0.3
q	0.2
w	0.15
k	2
L	*0.75*

C. Improved Insulation by Loading SRR and DGS

Initially, to improve the separation of this initial structure by known methods based on MTM applied to MIMO systems [54]. Therefore, four chains consisting of five SRR unit elements are inserted between the four excitation-radiating elements of the MIMO system. The UWB MIMO antenna configuration is illustrated in Fig. 15(a), the MIMO antenna design is the same as the previously design MIMO antenna with edge-to-edge gap of 5 mm, except that four strings of SSRs on the top layer of the substrate are added to improve the isolation between the antenna elements, (1-2, 2-3, 3-4 and 1-4). SRRs can act as a reflector and decrease the surface current between antenna elements which consequently reduce the mutual coupling between MIMO antenna elements. Furthermore, a t-shaped like dual slots were also etched in the ground plane which helps in further decrement of mutual coupling between elements placed side by side (1-2, 3-4), as depicted in Fig. 15(b).

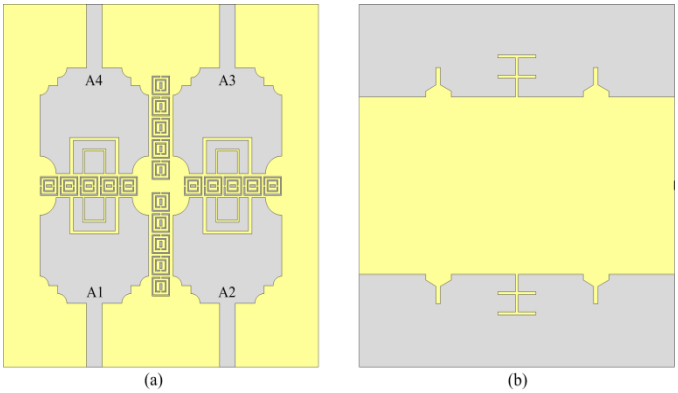


Fig. 15. Proposed four-element MIMO UWB antenna configuration with decoupling network.

A fabricated prototype is used to verify the results of the MIMO antenna loaded with SRR and DGs, as depicted in Fig. 16. Comparison among |S11| of the MIMO antenna loaded with decoupling structure is shown in Fig. 17. It can be observed that a good comparison between simulated and measured results is offered while covering the entire band spectrum globally allocated for UWB and Ku-band applications.

The mutual coupling of the MIMO antenna loaded with decoupling structure is depicted in Fig. 18. The antenna offers a low mutual coupling of less than -20 dB in the entire bandwidth having a good relationship among simulated and measured results.

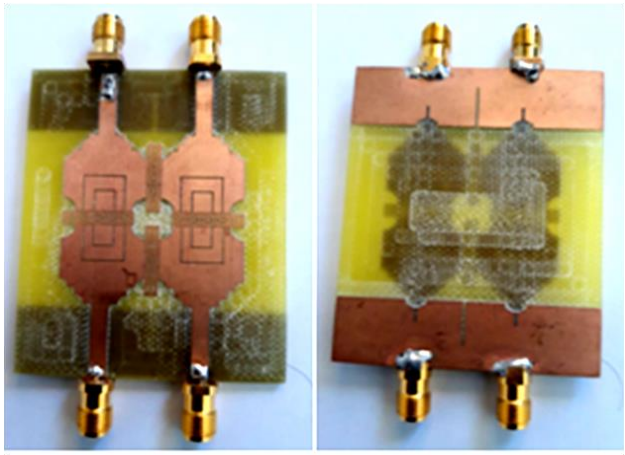


Fig. 16. Manufactured MIMO UWB antenna system with decoupling structure.

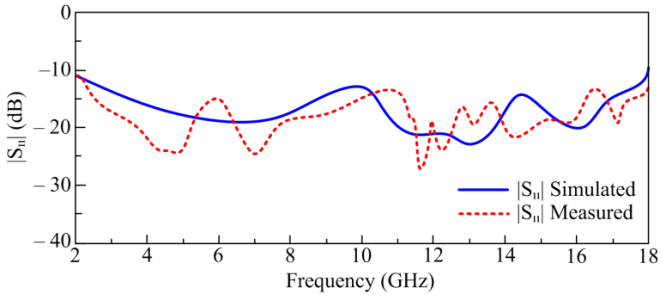


Fig. 17. |S11| of MIMO antenna system with SRRs.

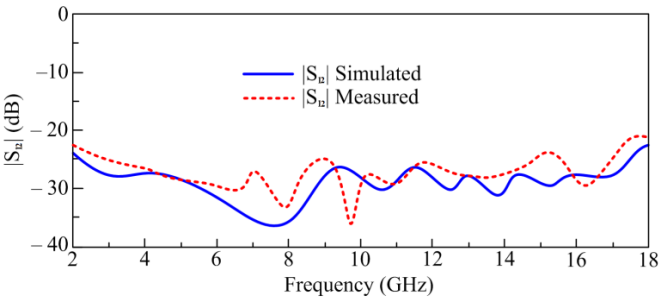


Fig. 18. |S12| (Simulated and measured) of 4-element MIMO UWB antenna system with isolation based on SRRs.

The other performance parameters that characterize the MIMO antenna are the envelope correlation coefficient (ECC) and diversity gain (DG). These parameters are analyzed and their impact on the overall performance of the antenna is considered. As already mentioned, the ECC measures the correlation between the radiating elements. It is also important to note that it is better to have the minimum possible ECC to crystallize the good performance of our MIMO system. In a MIMO system, the signals received must be sufficiently decorrelated to have good diversity. The parameter that describes the independence of the signals is called the correlation envelope. It is zero in the ideal case and must be less than 0.5 in order to obtain good diversity. In a 4-port MIMO system, ECC can be determined between ports 1, 2, 3, and 4 using the expression (1) and (2).

$$ECC = \frac{\left| \iint_{4\pi} (B_i(\theta,\phi)) \times (B_j(\theta,\phi)) d\Omega \right|^2}{\iint_{4\pi} \left|(B_i(\theta,\phi))\right|^2 d\Omega \iint_{4\pi} \left|(B_j(\theta,\phi))\right|^2 d\Omega} \quad (1)$$

Bi(θ, φ) is the 3D radiation pattern when the i[th] antenna is excited, Bj(θ, φ) is the 3D radiation pattern when the j[th] antenna is excited, Ω is the solid angle[55].

The disadvantage of this formula is to require a very precise estimation of the radiation patterns and therefore to lead to heavy calculations [56]. Another simplified solution is to calculate the correlation envelope using the S parameters [57]. Therefore, for a system with N antennas, equation (2) allows us to use:

$$ECC = \frac{\left| \sum_{n=1}^{N} S_{i,n}^{*} S_{n,j}^{*} \right|^2}{\prod_{k=i,j} \left[1 - \left| \sum_{n=1}^{N} S_{i,n}^{*} S_{n,j}^{*} \right| \right]} \quad (2)$$

N: Number of antennas, i and j denote antennas i and j

To use this formula (2), it is however necessary to satisfy certain conditions. When one of the antennas is powered, the others are loaded with a reference impedance (usually 50 Ω). The antenna system must be lossless, therefore with very high radiated efficiency [58], and low mutual coupling (< -6 dB).

To ensure that the signal-to-noise ratio (SNR) of the combined signal is better than the signal-to-noise ratio received from a single antenna in a MIMO system, the diversity gain (DG) should also be set to its recommended value should be inspected against value of about 10 dB.

The comparison among simulated and measured ECC and DG is shown in Fig. 19. The ECC offered by proposed work is less than 0.125 while the diversity gain > 9.9 dB is found, as shown in Fig. 19(a) and (b), respectively.

Table III shows a comparison of the UWB MIMO antenna with other existing antennas in the literature for similar applications. It can be observed that work presented in [32 – 33] offers large size as compared to proposed work along with setback of low mutual coupling and high ECC value. On the other hand, although the antenna proposed in [29-31, 34-35] offers a relatively compact size but most of the antennas are 2-port MIMO along with that they have set back of not covering the lower cut off frequency of UWB spectrum and high value of mutual coupling. Thus, it is evident from the comparison with recent works that proposed work offers a good combination of 4-port compact size MIMO antenna having UWB spectrum along with low mutual coupling and ECC values.

TABLE III. COMPARISON OF UWB MIMO ANTENNA

Ref	**Dimensions** ***(mm xmm)***	***Element Number***	**Band With (GHz)**	***Mutual Coupling (dB)***	**Gain Range** ***(dBi)***	***Radiation Efficiency(%)***	***ECC***
[29]	30 x 18	2	4.3 – 15.6	< -15	2.5 – 5.35	89	< 0.05
[30]	35 x 35	4	3.8 - 15	< -20	3 – 5	Not given	< 0.07
[31]	40 x 40	4	3.1 – 14	< -15	4.4 – 5	70	<0.015
[32]	58 x 58	4	2.84 – 15	< -16	3.5 – 6.35	> 78	< 0.07
[33]	80 x 80	4	3 – 14	< -17	1.2 – 4.8	> 78	< 0.02
[34]	30 x 60	2	2.7 – 20	< -13	3.7 – 6	Not given	< 0.07
[35]	31 x 78	2	3.1 – 13.5	< -17	2 – 4	> 78	< 0.18
Proposed Antenna	60 x 48	4	2-18	< -25	2-6.5	> 90	< 0.0125

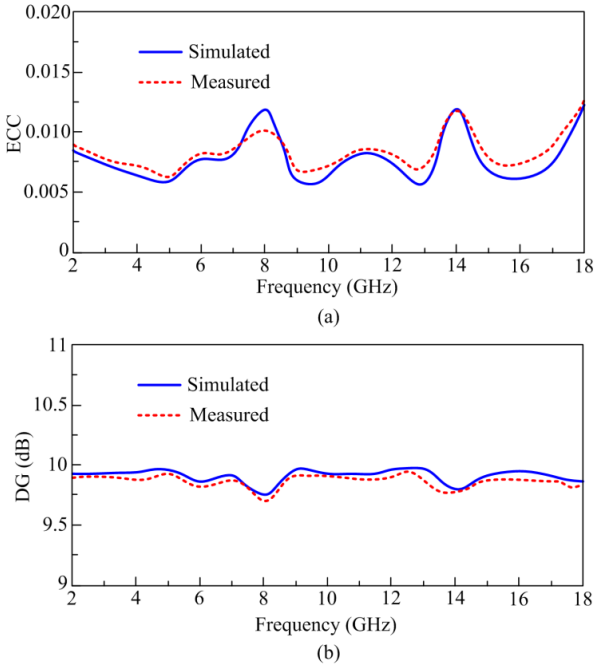


Fig. 19. The simulated and measured parameters of UWB MIMO antenna: (a) ECC and (b) DG.

V. CONCLUSION

In this work, a novel technique was developed to enhance isolation between two antennas in MIMO systems for 5G networks and IoT applications. This approach involved incorporating Split Ring Resonators (SRRs) between the MIMO elements, which led to excellent simulation results for interconnection. We analyzed various simulation parameters and constructed a physical system to measure factors such as S-parameters. The system demonstrated impressive diversity performance, with an envelope correlation coefficient (ECC) of less than 0.07 across the relevant frequency bands.The proposed design features four resonators arranged in an anti-parallel configuration, achieving over 17 dB isolation throughout the operating range. We evaluated both simulation and experimental data for gain, S-parameters, isolation, ECC, and radiation patterns. The results validate that using decoupling SRRs effectively reduces inter-element coupling and provides a strong diversity response. These findings suggest that the proposed MIMO antenna design is promising for ultra-wideband (UWB) applications. Future work could explore its potential in wireless communication systems, such as base station terminals.

ACKNOWLEDGMENT

This research was funded by Taif University, Saudi Arabia, Project Nº (TU- DSPP-2024-70).

REFERENCES

[1] G. Eason, B. Noble, and I. N. Sneddon, “On certain integrals of Lipschitz-Hankel type involving products of Bessel functions,” Phil. Trans. Roy. Soc. London, vol. A247, pp. 529–551, April 1955.

[2] V. Leithardt, D. Santos, L. Silva L. A solution for dynamic management of user profiles in IoT environments. IEEE Latin America Trans, Vol. 18, pp. 1193-1199, 2020.

[3] P. Kumar, T. Ali, M.M. Pai. Electromagnetic metamaterials: A pew paradigm of antenna design. IEEE Access, Vol; 9, pp. 18722–18751, 2021.

[4] W.A. Awan, A. Zaidi, N. Hussain, A. Iqbal, A. Baghdad. Stub loaded, low profile UWB antenna with independently controllable notch - bands. Microw Opt Technol Lett, Vol. 61, pp. 2447-2454, 2019.

[5] R.J Fontana. Recent system applications of short-pulse ultra-wideband (UWB) technology. IEEE Trans Microw Theory Techn, Vol. 52, pp. 2087–2104, 2004.

[6] J.Y. Siddiqui, C. Saha, C. Sarkar, L. A. Shaik, L.A, M.M.A. Yahia, Ultra-wideband antipodal tapered slot antenna with integrated frequency notch characteristics. IEEE Transactions on Antennas and Propagation, ,Vol. 66, pp. 1534-1539, 2018.

[7] S.C.Puri, S. Das, M.G. Tiary, UWB monopole antenna with dual-band-notched characteristics. Microw Opt Technol Lett, Vol. 62, pp. 1222–1229, 2020.

[8] E.M. Ali, W.A. Awan, M. S. Alizaidi, A. Alzahrani, D.H. Elkamchouchi, F. Falcone, S.S.M. Ghoneim. A shorted stub loaded UWB flexible antenna for small IoT devices. Sensors,Vol. 23, pp.748, 2023.

[9] FCC. FCC 1st report and order on Ultra-Wideband technology; FCC: Washington, DC, USA, 2002.

[10] A Zaidi, W.A. Awan, A. Ghaffar, M.S. Alzaidi, M. Alsharef, D.H. Elkamchouchi, S.S.M. Ghoneim, T.E.A. Alharbi. A low profile ultra-wideband antenna with reconfigurable notch band characteristics for smart electronic systems. Micromachines, Vol. 13, pp. 1803, 2022.

[11] M. Hussain, S. I. Naqvi, X. A. Awan, W.A.E. Ali, E.M. Ali, S. Khan, M. Alibakhshikenari. Simple wideband extended aperture antenna-inspired circular patch for V-band communication systems. AEU Int J Electron Commun, Vol; 144, pp. 154061, 2022.

[12] S. Lakrit, S.Das, A. El Alami, D. Barad, S. Mohapatra. A compact UWB monopole patch antenna with reconfigurable Band-notched characteristics for Wi-MAX and WLAN applications. AEU Int J Electron Commun, Vol. 105, pp. 106–115, 2019.

[13] M;A. Matin, Ultra-wideband current status and future trends; Intech Open: London, UK, 2012.

[14] S.A. Naqvi. Miniaturized triple-band and ultra-wideband (UWB) fractal antennas for UWB applications. Microw Opt Technol Lett, Vol. 59, pp. 1542–1546, 2017.

[15] W.A. Awan, D.M. Choi, N. Hussain, I. Elfergani, S.G. Park, N.A. Kim. frequency selective surface loaded UWB antenna for high gain applications. Comput Mater Contin, Vol. 73, pp. 6169-6180, 2022.

[16] S.G. Kirtania, B.A. Younes, A.R. Hossain, T. Karacolak, P.K. Sekhar. CPW-fed flexible ultra-wideband antenna for IoT applications. Micromachines, Vol. 12, pp. 453, 2021.

[17] W.A. Awan, S.I. Naqvi, N. Hussain, A. Ghaffar, A. Zaidi, X.J. Li, X. A Miniaturized UWB Antenna for Flexible Electronics. International Symposium on Antennas and Propagation and North American Radio Science Meeting, 2020, pp. 99-100.

[18] M.A. Sufian, N. Hussain, A. Abbas, W.A. Awan, D. Choi, N.A. Kim. Series fed planar array-based 4-port MIMO Antenna for 5G mmWave IoT applications. Asia-Pacific Microwave Conference (APMC), 2022, pp. 880-882.

[19] B. Yeboah-Akowuah, E.T. Tchao, M. Ur-Rehman, K.M. Khan, S. Ahmad. Study of a printed split-ring monopole for dualspectrum communications. Heliyon, Vol.7, PP; 7928, 2021.

[20] M.S. Alam, A. Abbosh. Reconfigurable band-rejection antenna for ultra-wideband applications. Microw. Antennas Propag. IET, Vol. 12, pp. 195–202, 2018.

[21] H. Abdi, J. Nourinia, C. Ghobadi. Compact Enhanced CPW-Fed Antenna for UWB Applications. Adv. Electromagn, Vol.10, pp. 15–20, 2021.

[22] A. Upadhyay, R.A. Khanna. CPW-fed tomb shaped antenna for UWB applications. Int. J. Innov. Technol. Explor. Eng. (IJITEE), Vol. 8, pp. 67–72, 2019.

[23] Z.A.A. Hassain, A.R. Azeez, M.M. Ali, T.A. Elwi. A modified compact bi-directional UWB tapered slot antenna with double band notch characteristics. Adv. Electromagn, Vol. 8, pp. 74–79,2019.

[24] P. Chaudhary, A. Kumar. Compact ultra-wideband circularly polarized CPW-fed monopole antenna. AEU Int. J. Electron. Commun, pp. 107, 137–145, 2020.

[25] X. P. Li, G. Xu, C.J. Duan, M.R. Ma, S.E. Shi, W. Li. Compact TSA with anti-spiral shape and lumped resistors for UWB applications. Micromachines, Vol. 12, pp. 1029, 2021.

[26] S. Kundu, A. Chatterjee. Sharp triple-notched ultra-wideband antenna with gain augmentation using FSS for ground pen-etrating radar. Wirel. Pers. Commun, Vol. 117, pp. 1399–1418, 2021.

[27] L. Guo, M. Min, W. Che, W. Yang. A novel miniaturized planar Ultra-Wideband antenna. IEEE Access, Vol. 7, pp. 2769–2773, 2018.

[28] A. Delphine, M. R. Hamid, N. Seman, M. Himdi. Broadband cloverleaf Vivaldi antenna with beam tilt characteristics. Int. J. RF Microw. Comput. Eng, Vol. 30, pp. 22158, 2020.

[29] M.M. Honari, M.S. Ghaffarian, R. Mirzavand. Miniaturized antipodal vivaldi antenna with improved bandwidth using exponential strip arms. Electronics, Vol. 10, pp. 83, 2021.

[30] W. Mu, H. Lin, Z. Wang, C. Li, M. Yang, W. Nie, J.A. Wu, flower-shaped miniaturized UWB-MIMO antenna with high isolation. Electronics, Vol. 11, pp. 2190, 2022.

[31] W. Zamir, D. Kumar. A compact 4 × 4 MIMO antenna for UWB applications. Microw. Opt. Technol. Lett, Vol. 58, pp. 1433–1436, 2016

[32] A.C. Suresh, T. Reddy. A Flower Shaped Miniaturized 4 × 4 MIMO Antenna for UWB Applications Using Characteristic Mode Analysis. Prog. Electromagn. Res. C, Vol. 119, pp. 219–233, 2022.

[33] V.R. Balaji, T. Addepalli, A. Desai, A. Nella, T.K Nguyen. An inverted L - strip loaded ground with hollow semi - hexagonal four - element polarization diversity UWB - MIMO antenna. Trans Emerging Telecommun Technol, Vol. 33, pp. 4381, 2022.

[34] R.B. Sadineni, P.G. Dinesha. Design of penta-band notched UWB MIMO antenna for diverse wireless applications. Prog Electromagn Res M, Vol. 107, pp. 35-49, 2022.

[35] B.T. Ahmed, I.F. Rodríguez. Compact high isolation UWB MIMO antennas. Wireless Networks, Vol. 28, pp. 1977-1999, 2022.

[36] G.A. Fadehan, Y. O. Olasoji, K.B. Adedeji. Mutual coupling effect and reduction method with modified electromagnetic band gap in UWB MIMO antenna. Appl. Sci, Vol. 12, pp. 12358, 2022.

[37] C.A. Balanis, Antenna theory: analysis and design. John wiley & sons, 2015.

[38] W.A. Awan, S.I. Naqvi, A.H. Naqvi, S.M. Abbas, A. Zaidi, N. Hussain. Design and characterization of wideband printed antenna based on DGS for 28 GHz 5G applications. J Electromagn Eng Sci, Vol. 21, pp. 177-183, 2021.

[39] N. Hussain, W.A. Awan, W. Ali, S.I. Naqvi, A. Zaidi, T.T. Le. Compact wideband patch antenna and its MIMO configuration for 28 GHz applications. AEU Int J Electron Commun, Vol. 132, pp. 153612, 2021.

[40] M. Alibakhshi-Kenari, M. Khalily, B.S. Virdee, C. See, R. Abd-Alhameed, E. Limiti. Mutual Coupling Suppression Between Two Closely Placed Microstrip Patches Using EM-Bandgap Metamaterial Fractal Loading, IEEE Access, 2019, Vol. 7, pp. 23606 – 23614, 2019.

[41] M. Alibakhshi-Kenari, M. Khalily, B.S. Virdee, C. See, R. Abd-Alhameed, A.H. Ali, F. Falcone, E. Limiti. Study on Isolation Improvement Between Closely Packed Patch Antenna Arrays Based on Fractal Metamaterial Electromagnetic Bandgap Structures, IET Microwaves, Antennas & Propagation , Vol. 12, p. 2241 – 2247, 2018.

[42] H. Xiong. J. Li, J. S. He. High isolation compact four-port MIMO antenna systems with built-in filters as isolation structure", Proceedings of the Fourth European Conference on Antennas and Propagation (EuCAP) 2016, pp. 1-4.

[43] I. Dioum. Conception de systèmes multi-antennaires pour techniques de diversité et MIMO -Application aux petits objets nomades communicants, Thèse de Doctorat, Université de Nice Sophia Antipolis, Ecole doctorale sciences et technologies de l'information et de la communication , 2016.

[44] J.G. Joshi, S.S. Pattnaik, S. Devi, M.R. Lohokare, M.R. Microstrip Patch Antenna Loaded with Magneto inductive Waveguide", 12th National Symposium on Antennas and Propagation 2010, pp.101-105.

[45] M. Alibakhshi-Kenari, B.S. Virdee. Study on Isolation and Radiation Behaviours of a 34×34 Array-Antennas Based on SIW and Metasurface Properties for Applications in Terahertz Band Over 125-300 GHz", Optik, International Journal for Light and Electron Optics, vol. 206, 2020.

[46] A. Chebihi, C. Luxey, A. Diallo, A. Le Thuc, R. Staraj. A Novel Isolation Technique for Closely Spaced PIFAs for UMTS Mobile Phones", IEEE Antennas and Wireless Propagation Letters, vol. 7, pp. 665-668,2008.

[47] S.C. Hwang, R.A. Abd- Alhameed, Z. Z. Abidin, N.J. Mc Ewan, P.S. Excell. Wideband printed MIMO/diversity monopole antenna for WiFi/WiMAX applications," IEEE Trans. on Antennas and Propagation, vol. 60, pp. 2028-2035, 2012.

[48] X.Q. Lin, H. Li, S. He, Y. Fan, A decoupling technique for increasing the port isolation between two closely packed antennas", 2012 IEEE Antennas and Propagation Society International Symposium (APSURSI), 2012, pp. 1-2.

[49] S. Luo, Y. Chen, D. Wang, Y. Liao, Y. Li. A monopole UWB antenna with sextuple band-notched based on SRRs and U-shaped parasitic strips. AEU-Int. J. Electron. Commun, Vol. 120, pp. 15, 2020.

[50] B. Ajewole, P. Kumar, T. Afullo, I-Shaped Metamaterial Using SRR for Multi-Band Wireless Communication, Crystals, Vol.12, pp. 559, 2022.

[51] R.B. Rani, S.K. Pandey. A CPW-fed circular patch antenna inspired by reduced ground plane and CSRR slot for UWB applications with notch band, Microw. Opt. Technol. Lett. vol. 59, pp. 745–749, 2017.

[52] N. Sharma, S.S. Bhatia. Metamaterial Inspired Fidget Spinner-Shaped Antenna Based on Parasitic Split Ring Resonator for Multi-Standard Wireless Applications. J. Electromagn. Waves Appl, Vol. 34, pp. 1471–1490, 2020.

[53] R. Mark, N. Rajak, K. Mandal, S. Das. Metamaterial based superstrate towards the isolation and gain enhancement of MIMO antenna for WLAN application. AEU, Int. J. Electron. Commun, vol. 100, pp. 144–152, 2019.

[54] F.B. Zarrabi, Z. Pirooj, K. Pedram, Metamaterial loads used in microstrip antenna for circular polarization, Int J RF Microw Comput Aided Eng, Vol. 29, pp. 21869, 2019.

[55] R. Mark, N. Rajak, K. Mandal, S. Das, S. Metamaterial based superstrate towards the isolation and gain enhancement of MIMO antenna for WLAN application. AEU, Int. J. Electron. Commun, vol. 100, pp. 144–152, 2019.

[56] M. Hussain, W.A. Awan, E.M. Ali, M.S. Alzaidi, M. Alsharef, D.H. Elkamchouchi, A. Alzahrani, M. Fathy Abo Sree. isolation improvement of parasitic element-loaded dual-band MIMO antenna for Mm-Wave applications. Micromachines 2022, 13, 1918.

[57] Bayarzaya, B.; Hussain, N.; Awan, W.A.; Sufian, M.A.; Abbas, A.; Choi, D.; Lee, J.; Kim, N. A compact MIMO antenna with improved isolation for ISM, Sub-6 GHz, and WLAN application. Micromachines , Vol. 13, pp. 1355, 2022.

[58] H. Khalid, W.A. Awan, M. Hussain, A. Fatima, M. Ali, N. Hussain, S. Khan, M. Alibakhshikenari, E. Limiti. Design of an integrated sub-6 GHz and mmWave MIMO antenna for 5G handheld devices. Appl. Sci., Vol. 11, pp. 8331, 2021.

Novel Data-Driven Machine Learning Models for Heating Load Prediction: Single and Optimized Naive Bayes

Fangyuan Li*
Zhejiang Business Technology Institute, Ningbo, Zhejiang, 315012, China

***Abstract*—Numerous approaches can be employed to create models for assessing the heat gains of a building arising from both external and internal sources. This modeling process evaluates effective operational strategies, conducts retrofit audits, and projects energy consumption. These techniques range from simple regression analyses to more intricate models grounded in physical principles. A prevalent assumption underlying all these modeling techniques is the requirement for input variables to be derived from authentic data, as the absence of realistic input data can lead to substantial underestimations or overestimations in energy consumption assessments. In this paper, eight input parameters, including relative compactness, orientation, wall area, roof area, glazing area, overall height, surface area, and glazing area distribution, are employed for training proposed Naive Bayes (NB)-based machine learning models. Utilizing a novel approach, this research explores the application of Beluga Whale Optimization and the Coot Optimization algorithm for optimizing the Naive Bayes model in heating load prediction. By harnessing the collective intelligence of Beluga Whales and drawing from the cooperative behavior of coots, the research aims to improve the model's predictive capabilities, which is of paramount importance in building energy management. Based on the comparative analysis between developed models (NB, NBCO, and NBBW), it is attainable that NBCO and NBBW, as two optimized models, have 2.4% and 1.3% higher R^2 values, respectively. Also, the RMSE of the NBCO was, on average, 19-33% lower than that of the two other models, confirming the high accuracy of NBCO. This innovative integration of bio-inspired optimization techniques with machine learning demonstrates a promising avenue for optimizing predictive models, offering potential energy efficiency and sustainability advancements in the built environment.**

Keywords*—Prediction models; heating load demand; building energy consumption; Naive Bayes; metaheuristic optimization algorithms

I. Introduction

In contemporary facility management, a critical challenge managers face revolves around assessing and predicting a building's energy requirements, particularly those equipped with air conditioning systems. This challenge stems from the notable variability exhibited in the energy feeding patterns generated by these systems. These fluctuations can be attributed to changes in external climate situations, the ebb and flow of occupants throughout the day, and internal loads incorporated within the building [1]. A holistic understanding of building performance is imperative to address this challenge and optimize building energy consumption. This begins with the initial identification of energy resources and the principal end-uses within the building. Energy resources typically natural gas, encompass electricity, and district heating supply, while the major end-uses comprise heating, ventilation, and air-conditioning ($HVAC$) systems, domestic hot water, lighting, plug-loads, elevators, kitchen equipment, ancillary appliances, and various equipment [2]. Such an integrated approach can enhance energy management and sustainability in the built environment.

Scholars have developed diverse assessment systems and modeling methodologies to propose an optimal predictive tool for estimating building energy consumption [3, 4]. Within this framework, two conventional devices for evaluating the Energy Performance of Buildings (EPB) through modeling and simulation have been highlighted in lectures [5]. Historically, the physical attributes of buildings, such as their geometry, were used as the basis for energy performance simulations. However, using such predictors has limitations, as it entails controlling factors that are challenging to manipulate in practical applications [6]. This can be considered a drawback of these traditional approaches. The primary challenge in advancing simulation and modeling techniques is the accurate estimation of EPB, a time-consuming process demanding meticulous attention due to including many influencing factors.

Furthermore, using various simulation programs may yield assessments with varying degrees of accuracy [7]. These methods can be relied upon to calculate the impact of individual factors on EPB when all other variables remain constant. Notably, there are established computer software tools like Designer's Simulation Toolkit (DeST) [8], Energy Plus [9], and DOE-2 [10] that facilitate these modeling and simulation endeavors.

Engineers have proposed using inverse (data-driven) modeling to remedy the limitations associated with simulation tools to explore EPB [11]. In this approach, a robust assessment of the impact of significant factors (e.g., roof area, relative compactness, and orientation) on EPB can be achieved by ensuring a sufficient quantity of data samples [12–15]. Various Machine Learning (ML) models, due to their ease of implementation and high-performance speed [16–18], were highly regarded by scholars.

For instance, Kalogirou and Bojic [19] employed a recurrent neural network to predict the energy feasting of a passive solar building. Pao [20] compared various models and concluded that ANN models are well-suited for forecasting building energy

consumption, effectively capturing complex non-linear relationships. Ben – Nakhi and Mahmoud [21] used ANN models to predict building cooling loads, achieving a strong fit to experimental data and optimizing thermal energy storage in public and office buildings.

In addition to Artificial Neural Networks ($ANNs$), various other artificial intelligence (AI) tools, including Support Vector Machine (SVM) [22] regression, neuro–fuzzy systems [23], and random forests [24], have been applied to address EPB challenges. For instance, Li et al. [25] conducted a qualified study on cooling load calculations, demonstrating the effectiveness of SVM and General Regression Neural Network (GRNN) compared to conventional ANNs. Moreover, researchers [26] have integrated SVM and wavelet transforms with Partial Least Squares Regression (PLS) to model office building heating and cooling loads, yielding precise insights. While AI has proven valuable in EPB, computational challenges have prompted the use of metaheuristic algorithms like Genetic Algorithm and Particle Swarm Optimization [27, 28, 38], which this study further explores, focusing on Beluga Whale Optimization (BWO) [29] and the Coot Optimization algorithm (COA) [30] for optimizing the Naive Bayes (NB) [31] model in heating load prediction.

The NB model is a widely used machine learning (ML) algorithm known for its simplicity and effectiveness in classification tasks in many applications similar to this study [32–34]. It is based on Bayes' theorem and chin independence assumption, making it particularly suited for applications where the independence assumption holds. It calculates the likelihood of a particular instance belonging to a specific class based on the probabilities of its features occurring in each class. This study embarks on developing NB-based models for predicting heating loads (HL) in buildings. Two distinct optimizers, as mentioned above (BWO and COA), were employed to optimize the training process. The predicted results of the $three$ models were subjected to comparison utilizing performance metrics, including R^2, RMSE, MSE, U95, and IOA. Afterward, the most optimal hybrid model for predicting HL in buildings was identified.

The choice of Naive Bayes (NB)--based machine learning models is particularly appropriate for addressing this type of problem due to several reasons. Firstly, NB models are known for their simplicity and computational efficiency, making them well-suited for handling large datasets and multiple input variables, such as those involved in predicting building heating loads. Secondly, NB models assume conditional independence between input features, which, despite being a simplification, often works well in practice, especially in complex systems where interactions between variables may not be easily discernible. This makes NB models robust and less prone to overfitting compared to more complex algorithms. Additionally, the probabilistic nature of NB models allows for clear interpretability of the results, providing insights into the contribution of each feature to the prediction, which is valuable in the situation of building energy management. Finally, the integration of bio-inspired optimization techniques like BWO and COA further enhances the model's ability to fine-tune its parameters, leading to improved accuracy and reliability in heating load predictions. This combination of simplicity, efficiency, and optimization makes NB-based models an effective choice for tackling the challenges of energy modeling in buildings. The paper is organized into five sections. The Abstract provides a concise summary of the study's objectives, methods, and key findings. The Introduction in Section I outlines the research background, related works, and significance. Materials and Methods in Section II details the dataset, machine learning models, and the hybrid optimization algorithms used, along with the evaluation metrics. The Results in Section III presents the outcomes of the modeling process. Discussion in Section IV offers a validation of present study, compares their performance, and addresses the study's limitations. Finally, the Conclusion in Section VI summarizes the findings, discusses implications for energy management, and suggests avenues for future research.

II. Materials and Methods

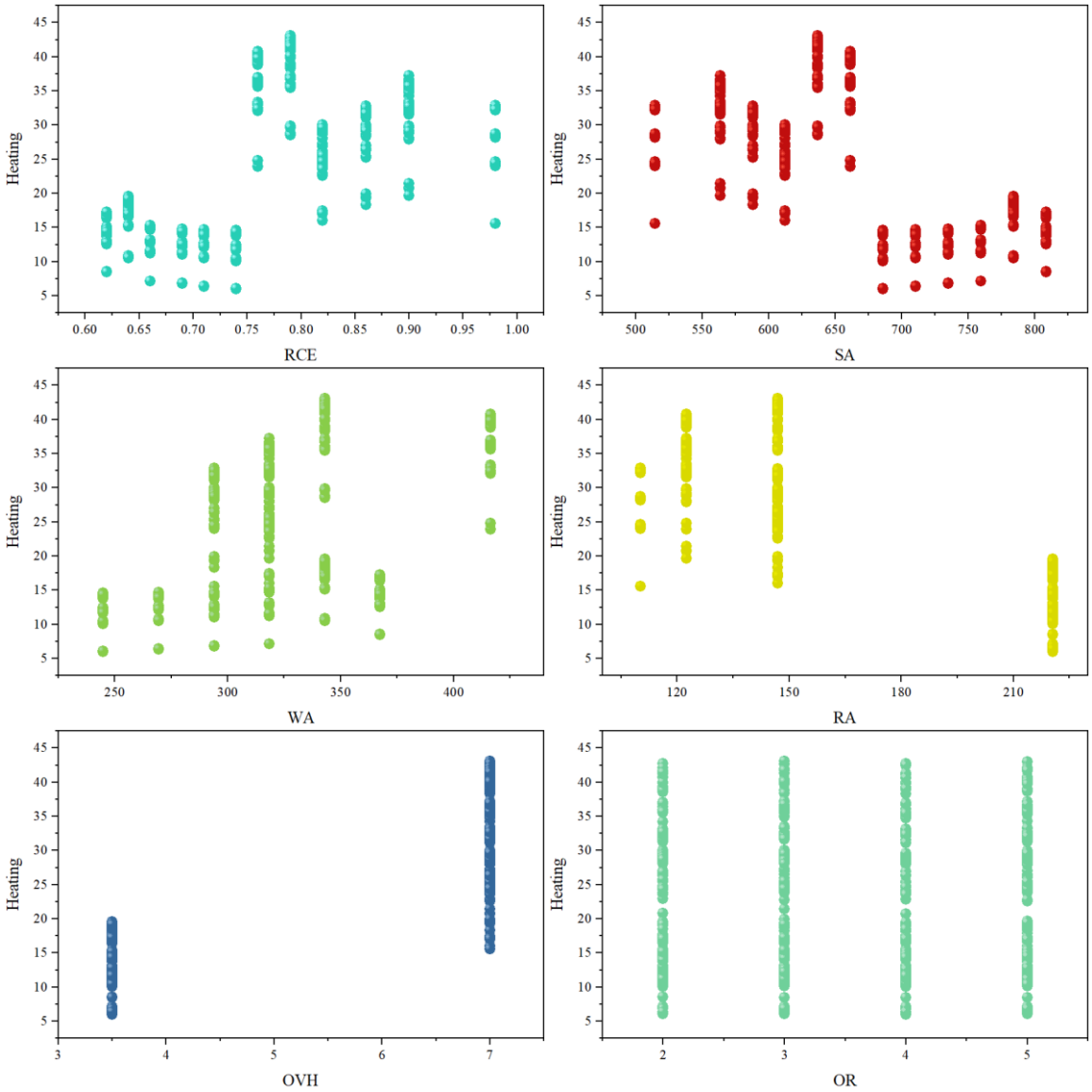
A. Dataset Description

The primary aim of this study is to predict HL in buildings by utilizing data that captures energy consumption patterns. A simulation approach involving the NB model is employed to accomplish this objective, with the training process incorporating two distinct optimizers designed to optimize NB hyper parameters. The inputs provided to the predictive model encompass various parameters, including relative compactness (RCE), $surface\ area$ (SA), $overall\ height$ (OVH), $roof\ area$ (RA), $glazing\ area$ (GA), $wall\ area$ (WA), $orientation$ (OR), and $glazing\ area\ distribution$ (GAD). The data relating to the input and output parameters, including minimum, maximum, average, standard deviation, Median, and Skewness, is reported in Table I. Minimum and maximum values identify the lowest and highest data points, establishing the data's range. The average, also known as the mean, provides a central measure to understand the typical value in the dataset. Standard deviation quantifies the dispersion of data points, indicating how closely they cluster around the mean. Conversely, the median represents the middle value when the data is ordered, making it robust to outliers. Skewness measures the asymmetry of the data distribution, indicating whether it is skewed to the left or right.

The scatter plot in Fig. 1 demonstrates the correlation among input and output parameters. The data distribution related to the RCE, SA, and WA input parameters is vertically highly asymmetric with the highest skewness values (RCE and WA skewed right of the average and SA skewed left of the average). Data points of OVH are located in two values (3.5 and 7) where $OVH = 3.5$ is related to lower heating values (below 20), and $OVH = 7$ corresponds to the heating values higher than 20. The OVH and OR data points' distribution is highly symmetric, with skewness values approximately equal to zero and their median and average values the same. GA and GAD are the only parameters with zero values, indicating that their effect is neglected in some samples.

TABLE I. THE STATISTICAL PROPERTIES OF THE INPUT ADJUSTABLE OF HEATING

Variables	*Category*	*Indicators*					
		Min	***Max***	***Median***	***Avg***	***Skew***	***St.Dev.***
RCE	*Input*	0.62	0.98	0.75	0.764	0.496	0.106
SA	*Input*	514.5	808.5	673.75	671.708	−0.125	88.09
WA	*Input*	245	416.5	318.5	318.5	0.534	43.63
RA	*Input*	110.25	220.5	176.604	176.604	−0.163	45.17
OVH	*Input*	3.5	7	5.25	5.25	$-2.9E-19$	1.751
OR	*Input*	2	5	3.5	3.5	$2.68E-18$	1.119
GA	*Input*	0	0.4	0.234	0.235	−0.060	0.133
GAD	*Input*	0	5	2.813	2.813	−0.089	1.551
Heating	*Output*	6.01	42.96	22.307	22.307	0.361	10.09



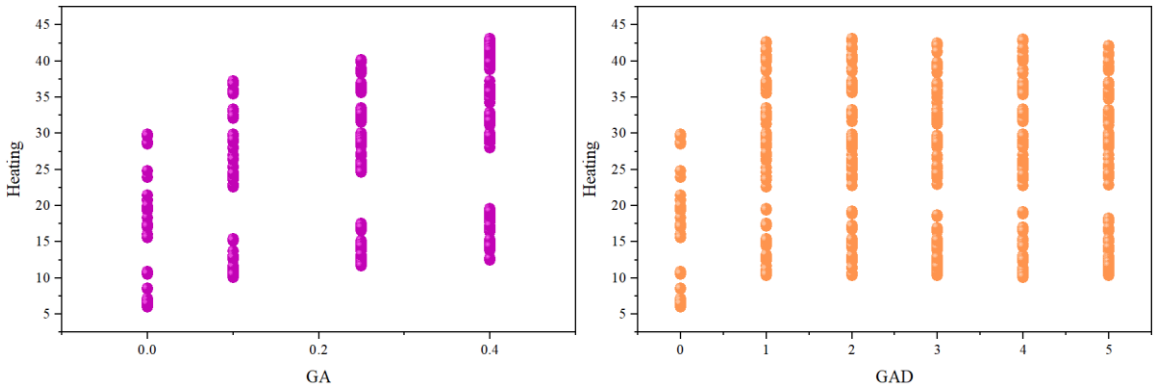


Fig. 1. Scatter plot amid input and output.

B. Machine Learning Model and Hybrid Optimization Algorithms

1) Naive bayes (NB): The NB is a robust probabilistic classification method grounded in Bayes' theorem, streamlining the modeling process by assuming input variable independence. When integrated with kernel density approximations, NB exhibits promise for significant enhancements in predictive accuracy, as indicated in previous studies [35, 36]. Notably, NB stands out due to its scalability, characterized by a need for only a few input parameters that increase linearly with the number of predictors. This differentiates it from computationally demanding classifiers. The closed-form training methodology of NB is remarkably efficient, ensuring swifter performance compared to more intricate computational techniques.

The NB classifier represents an advanced system seamlessly incorporating the NB probability model into its decision-making framework. Its foundation lies in applying the max a posteriori (MAP) choice rule, a proven approach for selecting the most likely supposition from a set of possible choices. Furthermore, it is worth noting the existence of a closely linked classifier known as the Bayes classifier. This formidable algorithm plays a pivotal role in assigning class labels $y = C_k$, with k ranging from 1 to K, a process involving an intricate assessment of multiple factors and variables to classify data points into predetermined categories.

$$y = argmaxp(C_k)\prod_{i=1}^{n} p((x_i\,|C_k)) \tag{1}$$

2) Beluga whale optimization (BWO): The BWO method simulates beluga whale (BW) behaviors for optimization, with two phases: exploration and refinement, using beluga whales as search agents updating candidate solutions within a specified area. The matrix maps the positions of these search agents (Zhong, Li, und Meng 2022):

$$X = \begin{bmatrix} x_{1,1} & x_{1,2} & x_{1,d} \\ x_{2,1} & x_{2,2} & x_{2,d} \\ x_{n,1} & x_{n,2} & x_{n,d} \end{bmatrix} \tag{2}$$

Within this framework, 'n' signifies the people size of BW, and $'d'$ denotes the dimensionality of design variables, with the fitness values of each individual within this population being meticulously recorded as follows:

$$F_X = \begin{bmatrix} f(x_{1,1}, x_{1,2}, \dots, x_{1,n}) \\ f(x_{2,1}, x_{2,2}, \dots, x_{2,d}) \\ f(x_{n,1}, x_{n,2}, \dots, x_{n,d} \end{bmatrix} \tag{3}$$

The changeover from examination to exploitation in the BWO algorithm is determined by the mathematical representation of the balance factor B_f.

$$B_f = B_0(1 - T/2T_{max}) \tag{4}$$

Throughout each iteration, random fluctuations within the range of $(0, 1)$ are experienced by the value of B_0, with the current iteration being denoted by T, and the max allowable number of iterations being represented by T_{max}. The exploration stage is initiated when the balance factor (B_f) surpasses the threshold of 0.5, while the exploitation stage is engaged when Bf is either less than or equal to 0.5. As the number of iterations (T) escalates, the variability in Bf is observed to diminish from the initial span of $(0, 1)$ to the narrower interval of $(0, 0.5)$. This transformation underscores a conspicuous alteration in the probability of transitioning between the exploitation and exploration stages, with the likelihood of entering the exploitation phase being augmented as the iteration count progressively increases.

- Exploration phase

The exploration phase in BWO is inspired by observed synchronized swimming behaviors of captive beluga whales, influencing the search agents' coordinates and subsequent position modifications.

$$\begin{cases} X_{i,j}^{T+1} = X_{i,pj}^{T} + (X_{r,p1}^{T} - X_{i,pj}^{T})(1 + r_1)sin(2\pi r_2), & j = even \\ X_{i,j}^{T+1} = X_{i,pj}^{T} + (X_{r,p1}^{T} - X_{r,pj}^{T})(1 + r_1)\cos(2\pi r_2), & j = \text{odd} \end{cases} \tag{5}$$

Within this equation, the current iteration count, denoted as T, establishes the framework. The expression $X_{i,j}^{T+1}$ represents the newly adjusted location for the $i-th$ beluga whale along the $j-th$ dimension. Concomitantly, pj (with j spanning from 1 to d) is symbolic of a value randomly selected from the $d-$dimensional space. Moreover, $X_{i,pj}^{T}$ designates the $i-th$ BW position along the pj dimension at iteration T Furthermore, both $X_{i,pj}^{T}$ and $X_{r,pj}^{T}$ serve to depict the prevailing positions of the $i-th$ BW and a stochastically chosen $r-th$ beluga whale, where r is selected randomly. Additionally, $r1$

and $r2$ are arbitrary values within the range of $(0, 1)$. It is of significance to note that the sine (sin) and cosine (cos) functions, applied to $(2\pi r_2)$, delineate the alignment of the mirrored BW fins toward the water's external. The selection of dimensions using odd or even numbers determines the reflection of synchronized or mirrored behaviors exhibited by BW during swimming or diving in the updated position. To enhance the stochastic components within the exploration stage, two random values identified as $r_1$and r_2, are utilized.

- Exploitation Phase

The exploitation stage in BWO is inspired by BW cooperative foraging and adaptive movement patterns, involving sharing positional information and coordination, utilizing the Levy flight strategy for convergence enhancement (Mantegna 1994), which has been integrated into the exploitation phase of BWO. It is postulated that these whales employ the Levy flight strategy for capturing prey, and this strategy is expressed mathematically as follows:

$$X_i^{T+1} = r_3 X_{best}^T - r_4 X_i^T + C_1 . L_F . (X_r^T - X_i^T) \quad (6)$$

In the context of the current iteration designated as "T," the following elements are encompassed: X_i, which serves as a representation of the current position of the $i - th$ beluga whale and "X_r," which represents the current position of a beluga whale that has been randomly selected. Furthermore, "X_i^{T+1}" denotes the updated position of the $i - th$ beluga whale, and "X_{best}" designates the optimal position among all the beluga whales. Additionally, "r3" and "r4" signify randomly generated numbers that fall from 0 to 1. Lastly, "C_1" is ascertained utilizing a calculation involving "r_4", specifically it determines the value of r_4 multiplied by the expression " $C_1 = 2r_4(1 - T/T_{max})$ thereby representing the random jump strength that quantifies the magnitude of a Levy flight [37].

The Levy flight function, denoted as LF, is computed according to the following procedure.

$$L_F = 00.5 \times \frac{u \times \sigma}{|v|^{1/\beta}} \quad (7)$$

$$\sigma = (\frac{\Gamma(1+\beta) \times \sin(\pi\beta/2)}{\Gamma((1+\beta)/2) \times \beta \times 2^{(\beta-1)/2}} \quad (8)$$

In this context, β, the default constant set to 1.5, is accompanied by normally distributed random numbers u and v.

- Whale fall

In BWO iterations, whale falls are simulated to mimic the beluga whale population changes. Assuming that some whales relocate or descend to the ocean floor, positions and step magnitudes are adjusted to maintain population size, resembling the natural process of whale fall decomposition.

$$X_i^{T+1} = r_5 X_i^T - r_6 X_r^T + r_7 X_{step} \quad (9)$$

"X_{step} is the step size of whale fall, which is determined as follows: where r_5, r_6, and r_7 are random numbers within the range of $(0, 1)$."

$$X_{step} = (u_b - l_b)\exp(-C_2 T/T_{max}) \quad (10)$$

In this context, the parameterC_2 is possessed, functioning as the step factor and being linked to the probability of a whale fall event, along with the population size $(C_2 = 2W_f \times n)$. Furthermore, the variables u_b and l_b are present, signifying the $upper$ and $lower$ boundaries of variables, respectively. It is observable that the extent of the step size is influenced by a range of factors, encompassing the constraints established by the design variables, the ongoing iteration, and the max permissible number of iterations.

This model calculates the probability of a whale falling (W_f) as a linear function:

$$W_f = 0.1 - 0.05T/T_{max} \quad (11)$$

The decrease in the probability of a whale falling from 0.1 in the initial iteration to 0.05 in the final iteration indicates a trend in which, as the food source is approached more closely by beluga whales during the optimization process, the risk to beluga whales is mitigated.

3) Coot optimization algorithm (COA): The COA is influenced by the group behaviors of Coots, a water bird species, and utilizes a metaheuristic optimization strategy. Coots exhibit various movements on water as they seek food sources or specific destinations, including chain, random, leader – driven, and leader – adjusted motions. The COOT algorithm integrates these behaviors into its structure. In its application, the algorithm commences by randomly establishing a population, following the guidelines of Eq. (12) as specified in (Naruei und Keynia 2021):

$$CootPos(i) = rand\ (1, N) \times (UB - LB) + LB \quad (12)$$

$CootPos(i)$ signifies the geographical coordinates of an individual Coot, where N matches the dimensionality of issues or the count of involved variables. UB and LB, on the other hand, represent the $upper$ and $lower$ confines of the search space in which the pursuit is performed.

$$UB = [UB_1, UB_2, \dots, UB_N], LB = [LB_1, LB_2, \dots, LB_N] \quad (13)$$

After the initial population setup, four different crusade designs are used to adjust the coots' situations.

- Random Movement

Following the equation described in Eq. (14) below, position Q is first randomized for this particular movement:

$$Q = rand(1, N) \times (UB - LB) + LB \quad (14)$$

To avoid becoming trapped in local optima, the position is updated in line with the equation presented in Eq. (15):

$$CootPos(i) = CootPos(i) + A \times R_2 \times (Q - CootPos(i)) \quad (15)$$

To determine A, the R_2 is a random number that exists within the range $[0, 1]$, and its value is calculated by an equation given in Eq. (16):

$$A = 1 - L \times (\frac{1}{Iter}) \quad (16)$$

In this case, $Iter$ is the highest achievable number of iterations, and L is a reference to currently recorded numbers.

- Chain Movement

The regular location of two chick birds may be calculated by applying the formula in Eq. (17) to execute chain movements.

$$CootPos(i) = \frac{CootPos(i-1)+CootPos(i)}{2} \quad (17)$$

In this case, $CootPos(i-1)$ indicates the placement of another coot in the arrangement.

- Adjusting Location Giving to the Leader

A coot bird's place in each group is adjusted according to the leader's location, which causes the follower to move closer to the leader. The method given in Eq. (18) is used to estimate the leader's designation:

$$K = 1 + (i \ MOD \ NL) \quad (18)$$

K represents the index of leaders, i " refers to the supporter coot bird's sequence, and NL indicates the total total of leaders in the group.

Following the formula in Eq. (19), a coot's position will be modified during this specific motion.

$$CootPos(i) = LeaderPos(K) + 2 \times R_1 \times Cos(2R\pi) \times (LeaderPos(K) - CootPos(i)) \quad (19)$$

$CootPos(i)$ refers to the present location of the coot bird, and $LeaderPos(K)$ stands for the chosen leader's position. R_1 is a randomly generated number within the range of $[0, 1]$, and R is another random number within the range of $[-1, 1]$. These parameters are utilized in the location update calculation outlined in Eq. (19).

- Leader Movement

The leadership roles experience modifications as outlined in Eq. (20), aiming to shift from locally optimal positions to globally optimal ones.

$$LeaderPos(i) = \begin{cases} B \times B_3 \times Cos(2\pi R) \times \big(gBest - LeaderPos(i)\big) + gBest & B_4 < 0.5 \\ B \times B_3 \times Cos(2\pi R) \times \big(gBest - LeaderPos(i)\big) - gBest & B_4 \geq 0.5 \end{cases} \quad (20)$$

In this specific scenario, $gBest$ characterizes the optimal possible site, and B_3 and B_4 are haphazardly generated figures within the break $[0, 1]$. The value B is calculated using the equation provided in Eq. (21):

$$B = 2 - L \times (\frac{1}{Iter}) \quad (21)$$

a) Performance evaluation metrics: In the evaluation of the performance of a regression model, it is customary to employ the following metrics:

- Coefficient of Determination (R^2): Commonly represented as R^2 , measures the percentage of inconsistency in the reliant on variable that can be attributed to the sovereign variables within a statistical model. The following formula demonstrates it:

$$R^2 = \left(\frac{\sum_{i=1}^{n}(t_i-\bar{w})(v_i-\bar{v})}{\sqrt{[\sum_{i=1}^{n}(v_i-w)^2][\sum_{i=1}^{n}(v_i-\bar{v})^2]}} \right)^2 \quad (22)$$

- Error evaluation metrics (RMSE, MSE): $RMSE$ (Root Mean Square Error) and MSE (Mean Square Error) are statistical metrics that quantify the average magnitude and accuracy of errors among predicted and observed values in a model, with RMSE emphasizing the root of the squared differences. These metrics are mathematically represented in Eq. (23) and (24) as follows:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(v_i - w_i)^2} \quad (23)$$

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(v_i - w_i)^2 \quad (24)$$

- Uncertainty 95% (U95): This metric illustrates the range in which 95% of the predicted values correspond to the actual observed values. It provides valuable insights into the correctness and dependability of a model's predictions, particularly when evaluating its variation and level of uncertainty. The mathematical expression of this metric can be found in Eq. (25).

$$U95 = \sqrt{\sum_{i=1}^{n}(v_i - \bar{v})^2/(n*(n-1))} \quad (25)$$

- Index of Agreement (IOA): IOA is a metric used to evaluate the agreement or accuracy of model predictions compared to observed data, typically expressed as a value between 0 and 1. The Eq. (26) represents it below:

$$IOA = 1 - \frac{\sum_{i=1}^{n}|w_i-v_i|}{\sum_{i=1}^{n}(|w_i-\bar{w}|+|v_i-\bar{w}|)} \quad (26)$$

In all equations:

n: quantity of samples,

v_i: denotes the individual predicted cost,

$\bar{v}$: indicates the mean of the predicted morals,

w_i: stands for the experimentally measured cost,

$\bar{w}$: represents the average of the experimentally measured values.

III. Results

In this research paper, the assessment of heating energy consumption relies on utilizing a Naive Bayes (NB) model. Two optimization algorithms, COA and BWO, have been employed to assess the model's performance and training procedure. To create the requisite datasets for training, validation, and testing, a partitioning scheme of 70% for training, 15% for validation, and 15% for testing has been implemented.

In Table II, it is evident that the R^2 values exhibit a range, with the lowest value of 0.947 (corresponding to the NB model) and the highest value of 0.987 (associated with the NBCO

model). Interpreting the R^2, it is apparent that the NBCO model, with the highest R^2, indicates superior model performance. The NBBW model, which achieved a R^2 of 0.975, closely follows as the second-best performer.

RMSE and MSE represent the amount of error according to their definitions. The smaller values of these metrics indicate better model performance. For the NB model, the highest values of RMSE and MSE are 2.050 and 4.204, respectively. In contrast, for the NBCO model, these values are significantly lower at 1.377 and 1.896, demonstrating the superior performance of the NBCO model.

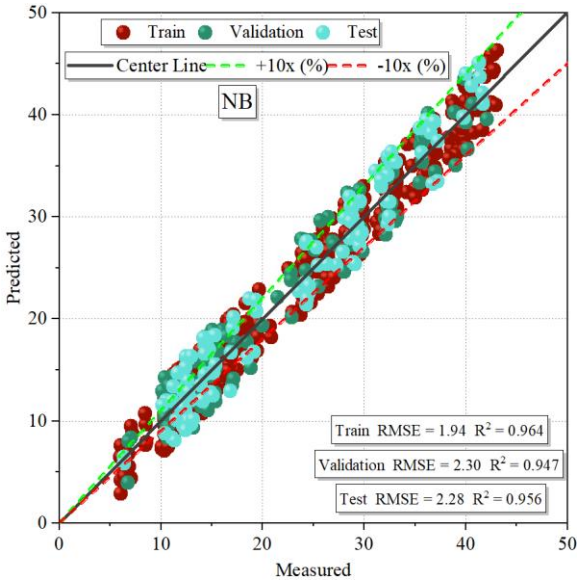
U95, representing data uncertainty, shows that a model's performance improves as this value decreases. According to Table II, the U95 values for models NBCO, NBBW, and NB are 3.747, 4.676, and 5.677, respectively.

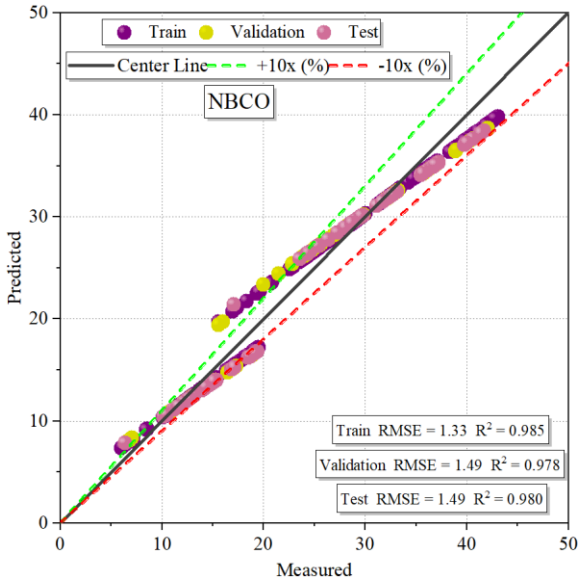
The metric IOA indicates the agreement or accuracy of the model predictions associated to the practical data, and its value falls within the range of 0 to 1. NBCO, with an IOA of 0.995, ranks the highest, demonstrating superior performance. NBBW and NB models are second and third in model quality, respectively.

TABLE II. THE RESULT OF THE DEVELOPED MODELS

Model	*Phase*	*Index values*				
		RMSE	*R2*	*MSE*	*U95*	*IOA*
NB	*Train*	1.940	0.964	3.764	5.375	0.991
	Validation	2.298	0.947	5.282	6.354	0.986
	Test	2.278	0.956	5.188	6.278	0.988
	All	2.050	0.960	4.204	5.677	0.990
NBCO	*Train*	1.325	0.985	1.757	3.598	0.996
	Validation	1.491	0.978	2.222	4.104	0.994
	Test	1.490	0.980	2.219	4.034	0.994
	All	1.377	0.983	1.896	3.747	0.995
NBBW	*Train*	1.605	0.975	2.575	4.429	0.994
	Validation	1.880	0.967	3.536	5.115	0.991
	Test	1.914	0.964	3.662	5.276	0.991
	All	1.698	0.972	2.882	4.676	0.993

Fig. 2. displays a scatter plot for hybrid models, illustrating the variation among predicted and measured values. This scatter plot is generated using RMSE and R^2 values, which primarily influence data dispersion. A decrease in RMSE corresponds to an increase in data density. Furthermore, a higher R^2 value indicates a more precise fit of the line to the data. Based on the visual representations in the plots, it is evident that three primary lines can be identified: a central line, a line representing a 10% overestimation, and a line depicting a 10% underestimation. After explicit consideration, it is apparent that the minimum R^2 value, at 0.947, is associated with model NB, whereas model NBCO exhibits the highest value, 0.985. Furthermore, the highest RMSE is observed in model NB, which is equal to 2.30, and the lowest value for model NBCO is 1.33, representing a 47% reduction in error. Based on these findings, it can be concluded that NBCO is the superior choice for predicting heating load.





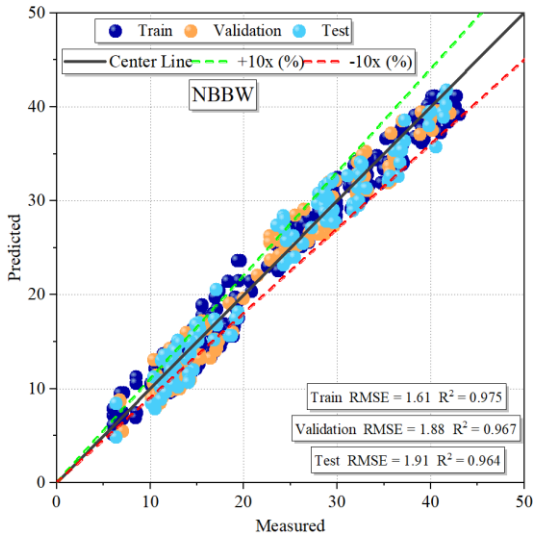


Fig. 2. The hybrid model's created scatter plot.

Fig. 3 illustrates the variations in error metrics ($RMSE$ and MSE) and R^2 values across the three models in this study. According to the trend lines for RMSE and MSE, it is observed that errors in all models initially increase during the train phase. However, there is a noticeable decrease from the validation phase to the test. In summary, after comparing the error rates of RMSE and MSE, it can be deduced that NBCO, with values of 1.377 and 1.896, is the most accurate prediction model, while NBBW and NB are the second and third-ranking models, individually. In all phases, the value of R^2 for NBCO is higher than NBBW and NB by approximately 1.13% and 2.396%, which again shows the superiority of the NBCO.

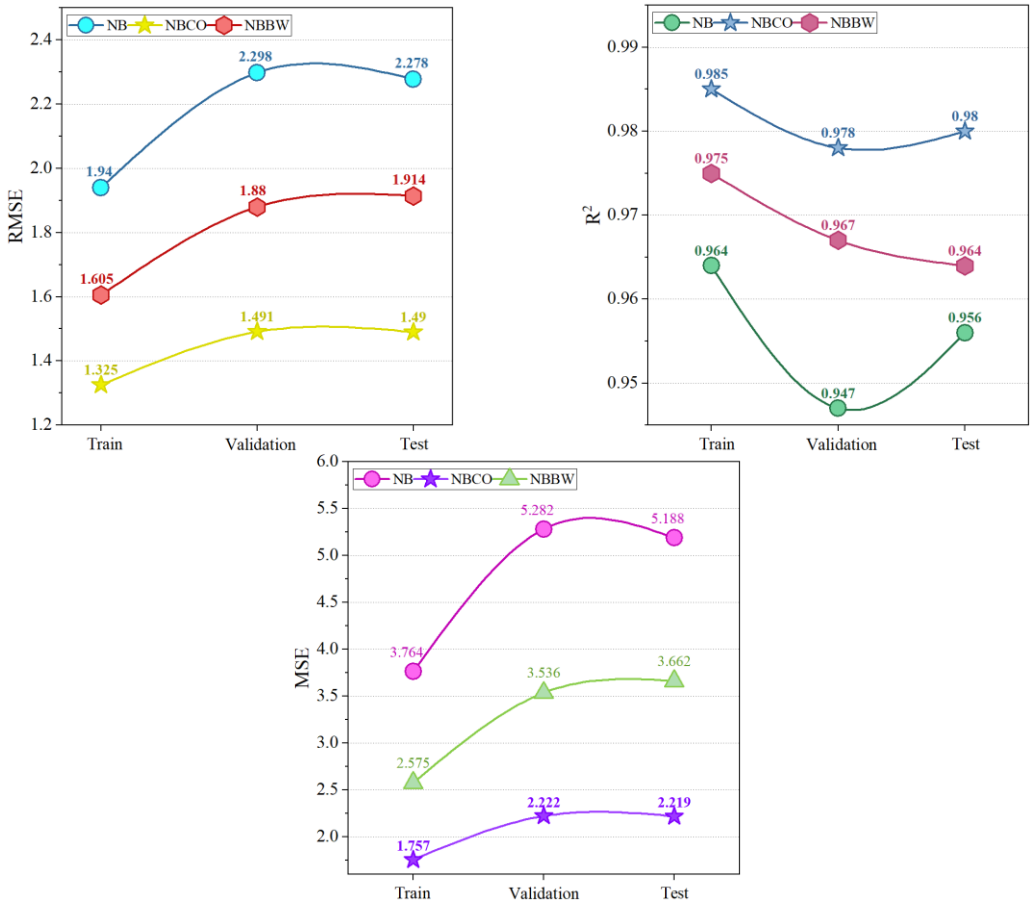


Fig. 3. Comparison between models corresponding RMSE, R^2, MSE.

The histogram distribution diagram, depicting the error percentages of models, is presented in Fig. 4. The horizontal axis displays the percentage of errors, while the vertical axis represents the frequency of occurrences for each model during the training, validation, and test phases. In the basic *NB* model, the error percentage falls within the range of approximately -40 to 40, with the highest frequency around 90. In the case of the two subsequent hybrid models, the error ranges for NBCO and NBBW are approximately -20 to 20 and -30 to 30, correspondingly. The highest frequencies of error values near zero percent for NBCO and NBBW are 100 and 120, respectively.

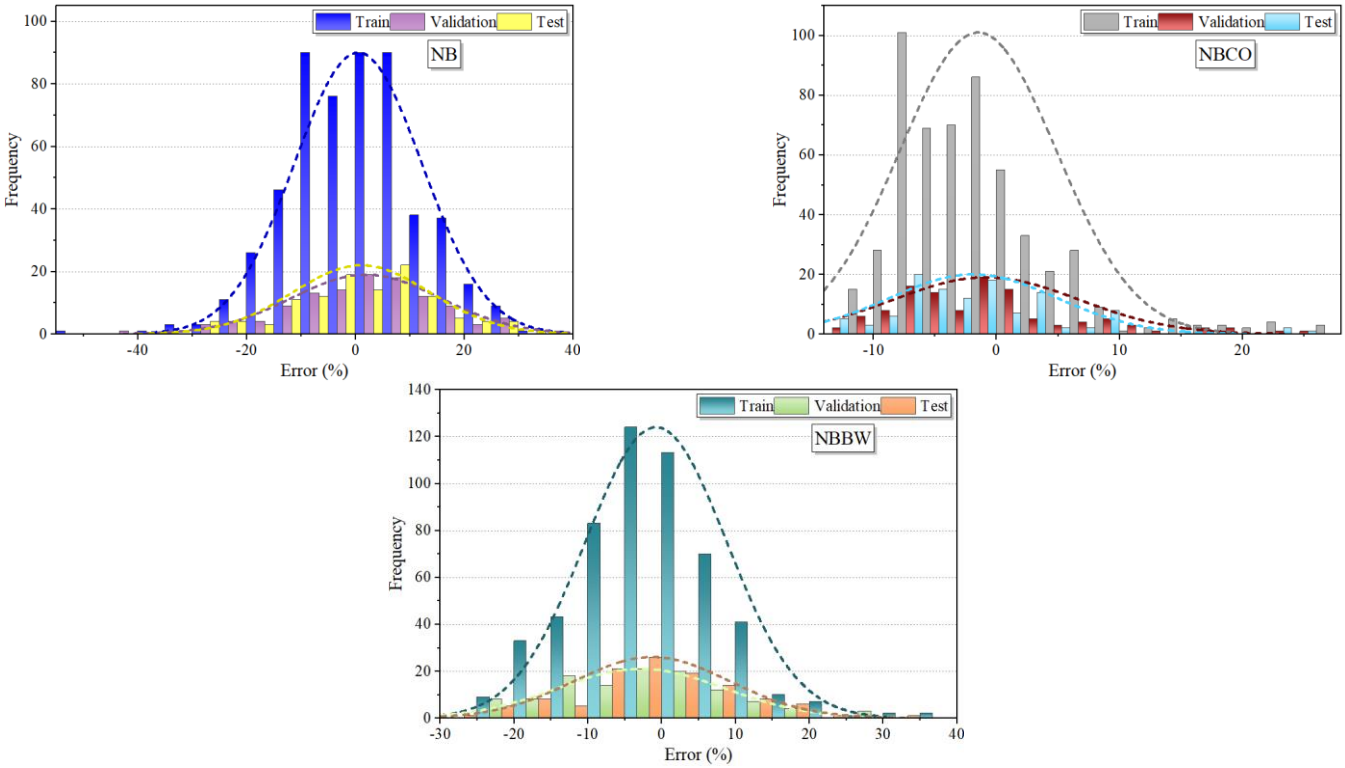


Fig. 4. The error ratio for the hybrid models is created on the Histogram distribution plot.

The multi-line diagram illustrating the error percentages of the models is presented in Fig. 5. The horizontal *axis* represents the number of samples, and the vertical *axis* is divided into three components: the blue axis represents the error rate of model NB. At the same time, the brown and pink axes correspond to models NBCO and NBBW, respectively. It should be noted that the error percentage range for the NB model spans approximately from -40 to 40 during the train, validation, and test phases. In contrast, for the NBCO model, the range extends from -20 to just above 20, while for the NBBW model, it falls within the range of approximately (-30 to 30).

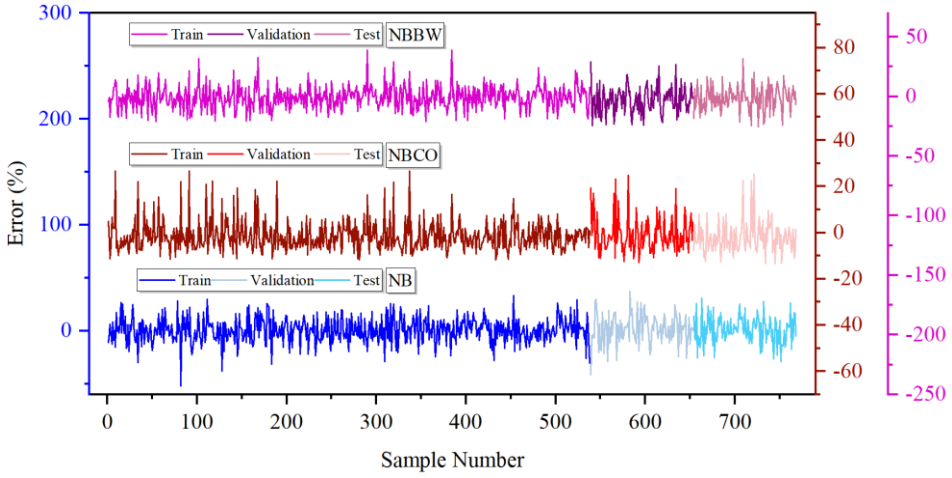


Fig. 5. The Multi-line plot for errors of the developed models.

IV. Discussion

A. Validation of Present Study

The validation of the developed NBCO model in this study demonstrates its superior performance compared to existing models. Table III presents a comparison among the NBCO model from the present study and the PSO-MLP model by Zhou et al. (Zhou u. a. 2020). The NBCO model achieves an R^2 value of 0.985, significantly higher than the 0.9126 obtained by the PSO-MLP model. This indicates that the NBCO model explains a larger proportion of the variance in heating load predictions, showcasing its enhanced predictive power. Additionally, the RMSE of the NBCO model is 1.325, which is substantially lower than the 2.9736 RMSE reported for the PSO-MLP model. A lower RMSE reflects higher accuracy in the predictions, further validating the effectiveness of the NBCO model in accurately forecasting heating loads. These results highlight the advantages of the bio-inspired optimization techniques employed in this study, particularly the Coot Optimization Algorithm, in refining the Naive Bayes model. The validation confirms that the NBCO model outperforms existing approaches, making it a valuable tool for improving energy efficiency and sustainability in building energy management.

TABLE III. The Validation of Developed Model

Article	Model	Evaluator	
		R^2	*RMSE*
Zhou et al. (Zhou u. a. 2020)	PSO-MLP	0.9126	2.9736
Present study	NBCO	0.985	1.325

B. Comparison

Table IV presents a comparative analysis of the best-performing models from the present study alongside similar models from relevant literature, focusing on their ability to predict HL. The models compared include Support Vector Regression (SVR), Multi-Parameter Moving Ridge (MPMR), Light Gradient Boosting Machine (LGBM), and the Naive Bayes optimized with Coot Optimization Algorithm (NBCO) developed in the current study. The SVR model by Moradzadeh et al. (Moradzadeh u. a. 2020) achieved an impressive $RMSE$ of 0.4832 and an R^2 value of 0.9979, indicating a high level of accuracy and predictive power. Similarly, Roy et al. (Roy u. a. 2020) reported an MPMR model with an RMSE of just 0.059 and an R^2 of 0.99, making it one of the most accurate models for heating [39] load prediction. Gong et al. (Gong u. a. 2020) employed the LGBM model, which also performed well, achieving an $RMSE$ of 0.1929 and an R^2 value of 0.9882. In comparison, the NBCO model from the present study produced an $RMSE$ of 1.325 and an R^2 value of 0.985. While the NBCO model's R^2 value is close to those reported in the literature, indicating strong predictive accuracy, its RMSE is notably higher. This suggests that while NBCO captures the overall variance in heating loads [40] effectively, there may be room for improvement in reducing the prediction errors to match or surpass the accuracy levels of the models reported in other studies. Despite this, the NBCO model still offers significant advantages, particularly in its innovative use of bio-inspired optimization techniques. The model's relatively high R^2 value demonstrates its ability to serve as a reliable tool for heating [41] load prediction, with the added potential for further refinement to improve its RMSE. This comparison underscores the value of continuing to explore and optimize machine learning models in the pursuit of enhanced energy efficiency in building management.

TABLE IV. The Comparison of the Best Performed Models Results of Present Study with Some Related Literatures

Articles	Index values			
	Target	*Models*	*RMSE*	R^2
Moradzadeh et al. (Moradzadeh u. a. 2020)	HL	SVR	0.4832	0.9979
Roy et al. (Roy u. a. 2020)	HL	$MPMR$	0.059	0.99
Gong et al. (Gong u. a. 2020)	HL	$LGBM$	0.1929	0.9882
Present Study	HL	$NBCO$	1.325	0.985

C. Limitation

Despite the promising results, this study has several limitations that should be acknowledged, both in the context of the Naive Bayes (NB) model and the broader modeling approach. First, the NB model's inherent assumption of conditional independence among input features may not fully capture the complex interdependencies in building systems, potentially leading to inaccuracies when strong correlations exist between variables such as orientation, glazing area, and thermal performance. This limitation could result in suboptimal predictions, particularly in scenarios where these interactions play a significant role. Additionally, the optimization techniques employed Beluga Whale Optimization (BWO) and Coot Optimization Algorithm (COA) though effective, are relatively novel and less established than traditional methods. Their efficacy in various contexts remains to be thoroughly validated, and there may be cases where these optimizers do not provide substantial improvements over more conventional approaches. Moreover, the study focuses exclusively on predicting heating loads, overlooking other critical aspects of building energy management, such as cooling loads and ventilation. This narrow focus limits the comprehensiveness of the model and its applicability in broader energy efficiency strategies. Finally, the existing model does not account for real-time data integration or adaptive learning, which are increasingly important in dynamic energy management systems. The absence of these features may restrict the model's effectiveness in responding to changing conditions and optimizing performance over time.

V. Conclusion

The contemporary challenge of effectively managing building energy consumption, particularly in structures equipped with air conditioning systems, necessitated a holistic understanding of energy resources and end-uses within buildings. Achieving energy efficiency and sustainability in the built environment demanded the development of optimal predictive tools for estimating building energy consumption. Various modeling methodologies, including traditional approaches based on building geometry and advanced machine

learning models like $ANNs, SVM,$ and random forests (RF), were explored for this purpose. Additionally, integrating metaheuristic algorithms emerged as a promising avenue for optimizing these models.

This study extended these efforts by applying the Naive Bayes (NB) model to predict heating loads in buildings and optimize the train process using the Beluga Whale Optimization (BWO) and Coot Optimization Algorithm (COA). Comparative analysis revealed that the optimized NB models outperformed traditional NB, demonstrating the potential for these bio-inspired optimization techniques to enhance predictive models and contribute to greater energy efficiency and sustainability in the built environment. Based on comparative analysis based on numerical values obtained for each evaluation metric corresponding to the developed models, the NBCO hybrid model attained a maximum coefficient of determination of 0.985, surpassed NBBW and NB by 1.03% and 2.2%, respectively, and exhibited minimal performance RMSE error of 1.325, which are notably 17.4% and 31.7% lower than those observed in NBBW and NB. This research served as a significant step toward addressing the energy challenges faced by contemporary facility management, presenting a promising path for future developments in the field.

ACKNOWLEDGMENT

The subject of Artificial Intelligence Industry Application Research Center Facing the Belt and Road Initiative of Zhejiang Business Technology Institute.

REFERENCES

[1] Neto AH, Fiorelli FAS. Comparison between detailed model simulation and artificial neural network for forecasting building energy consumption. Energy Build 2008;40:2169–76.

[2] Wei Y, Zhang X, Shi Y, Xia L, Pan S, Wu J, et al. A review of data-driven approaches for prediction and classification of building energy consumption. Renewable and Sustainable Energy Reviews 2018;82:1027–47.

[3] Saffari M, de Gracia A, Ushak S, Cabeza LF. Passive cooling of buildings with phase change materials using whole-building energy simulation tools: A review. Renewable and Sustainable Energy Reviews 2017;80:1239–55.

[4] Dogan T, Reinhart C. Shoeboxer: An algorithm for abstracted rapid multi-zone urban building energy model generation and simulation. Energy Build 2017;140:140–53.

[5] Zhao H, Magoulès F. A review on the prediction of building energy consumption. Renewable and Sustainable Energy Reviews 2012;16:3586–92.

[6] Park JS, Lee SJ, Kim KH, Kwon KW, Jeong J-W. Estimating thermal performance and energy saving potential of residential buildings using utility bills. Energy Build 2016;110:23–30.

[7] Yezioro A, Dong B, Leite F. An applied artificial intelligence approach towards assessing building performance simulation tools. Energy Build 2008;40:612–20.

[8] Yan D, Xia J, Tang W, Song F, Zhang X, Jiang Y. DeST—An integrated building simulation toolkit Part I: Fundamentals. Build Simul, vol. 1, Springer; 2008, p. 95–110.

[9] Crawley DB, Lawrie LK, Winkelmann FC, Buhl WF, Huang YJ, Pedersen CO, et al. EnergyPlus: creating a new-generation building energy simulation program. Energy Build 2001;33:319–31.

[10] York DA, Cappiello CC, Olson KH. DOE-2 Reference Manual: Version 2.1 C. Los Alamos National Laboratory, Solar Energy Group; 1984.

[11] O'Neill Z, O'Neill C. Development of a probabilistic graphical model for predicting building energy performance. Appl Energy 2016;164:650–8.

[12] Yu Z, Haghighat F, Fung BCM, Yoshino H. A decision tree method for building energy demand modeling. Energy Build 2010;42:1637–46. https://doi.org/10.1016/j.enbuild.2010.04.006.

[13] Dimitrov D, Abdo H. Tight independent set neighborhood union condition for fractional critical deleted graphs and ID deleted graphs. Discrete and Continuous Dynamical Systems-S 2019;12:711–21.

[14] Gao W, Guirao JLG, Basavanagoud B, Wu J. Partial multi-dividing ontology learning algorithm. Inf Sci (N Y) 2018;467:35–58.

[15] Catalina T, Virgone J, Blanco E. Development and validation of regression models to predict monthly heating demand for residential buildings. Energy Build 2008;40:1825–32.

[16] behnam Sedaghat, Tejani GG, Kumar S. Predict the Maximum Dry Density of soil based on Individual and Hybrid Methods of Machine Learning. Advances in Engineering and Intelligence Systems 2023;002. https://doi.org/10.22034/aeis.2023.414188.1129.

[17] Masoumi F, Najjar-Ghabel S, Safarzadeh A, Sadaghat B. Automatic calibration of the groundwater simulation model with high parameter dimensionality using sequential uncertainty fitting approach. Water Supply 2020;20:3487–501. https://doi.org/10.2166/ws.2020.241.

[18] Akbarzadeh MR, Ghafourian H, Anvari A, Pourhanasa R, Nehdi ML. Estimating Compressive Strength of Concrete Using Neural Electromagnetic Field Optimization. Materials 2023;16:4200.

[19] Kalogirou SA, Bojic M. Artificial neural networks for the prediction of the energy consumption of a passive solar building. Energy 2000;25:479–91.

[20] Pao H-T. Comparing linear and nonlinear forecasts for Taiwan's electricity consumption. Energy 2006;31:2129–41.

[21] Ben-Nakhi AE, Mahmoud MA. Cooling load prediction for buildings using general regression neural networks. Energy Convers Manag 2004;45:2127–41.

[22] Dong B, Cao C, Lee SE. Applying support vector machines to predict building energy consumption in tropical region. Energy Build 2005;37:545–53.

[23] Nilashi M, Dalvi-Esfahani M, Ibrahim O, Bagherifard K, Mardani A, Zakuan N. A soft computing method for the prediction of energy performance of residential buildings. Measurement 2017;109:268–80.

[24] Gao W, Alsarraf J, Moayedi H, Shahsavar A, Nguyen H. Comprehensive preference learning and feature validity for designing energy-efficient residential buildings using machine learning paradigms. Appl Soft Comput 2019;84:105748.

[25] Li Q, Meng Q, Cai J, Yoshino H, Mochida A. Predicting hourly cooling load in the building: A comparison of support vector machine and different artificial neural networks. Energy Convers Manag 2009;50:90–6.

[26] Zhao J, Liu X. A hybrid method of dynamic cooling and heating load forecasting for office buildings based on artificial intelligence and regression analysis. Energy Build 2018;174:293–308.

[27] Tien Bui D, Moayedi H, Anastasios D, Kok Foong L. Predicting heating and cooling loads in energy-efficient buildings using two hybrid intelligent models. Applied Sciences 2019;9:3543.

[28] Bahiraei M, Heshmatian S, Goodarzi M, Moayedi H. CFD analysis of employing a novel ecofriendly nanofluid in a miniature pin fin heat sink for cooling of electronic components: Effect of different configurations. Advanced Powder Technology 2019;30:2503–16.

[29] Zhong C, Li G, Meng Z. Beluga whale optimization: A novel nature-inspired metaheuristic algorithm. Knowl Based Syst 2022;251:109215. https://doi.org/https://doi.org/10.1016/j.knosys.2022.109215.

[30] Naruei I, Keynia F. A new optimization method based on COOT bird natural life model. Expert Syst Appl 2021;183:115352.

[31] Lowd D, Domingos P. Naive Bayes models for probability estimation. Proceedings of the 22nd international conference on Machine learning, 2005, p. 529–36.

[32] Sibyan H, Svajlenka J, Hermawan H, Faqih N, Arrizqi AN. Thermal Comfort Prediction Accuracy with Machine Learning between Regression Analysis and Naïve Bayes Classifier. Sustainability 2022;14:15663.

[33] Song G, Ai Z, Zhang G, Peng Y, Wang W, Yan Y. Using machine learning algorithms to multidimensional analysis of subjective thermal comfort in a library. Build Environ 2022;212:108790.

[34] Yılmaz D, Tanyer AM, Toker İD. A data-driven energy performance gap prediction model using machine learning. Renewable and Sustainable Energy Reviews 2023;181:113318.

[35] Hastie T, Tibshirani R, Friedman JH, Friedman JH. The elements of statistical learning: data mining, inference, and prediction. vol. 2. Springer; 2009.

[36] Piryonesi SM, El-Diraby TE. Role of data analytics in infrastructure asset management: Overcoming data size and quality problems. Journal of Transportation Engineering, Part B: Pavements 2020;146:4020022.

[37] Mantegna RN. Fast, accurate algorithm for numerical simulation of Levy stable stochastic processes. Phys Rev E 1994;49:4677.

[38] Zhou G, Moayedi H, Bahiraei M, Lyu Z. Employing artificial bee colony and particle swarm techniques for optimizing a neural network in prediction of heating and cooling loads of residential buildings. J Clean Prod 2020;254:120082.

[39] Moradzadeh A, Mansour-Saatloo A, Mohammadi-Ivatloo B, Anvari-Moghaddam A. Performance evaluation of two machine learning techniques in heating and cooling loads forecasting of residential buildings. Applied Sciences 2020;10:3829.

[40] Roy SS, Samui P, Nagtode I, Jain H, Shivaramakrishnan V, Mohammadi-Ivatloo B. Forecasting heating and cooling loads of buildings: A comparative performance analysis. J Ambient Intell Humaniz Comput 2020;11:1253–64.

[41] Gong M, Bai Y, Qin J, Wang J, Yang P, Wang S. Gradient boosting machine for predicting return temperature of district heating system: A case study for residential buildings in Tianjin. Journal of Building Engineering 2020;27:100950.

Facial Expression Real-Time Animation Simulation Technology for General Mobile Platforms Based on OpenGL

Mingzhe Cao
Department of Tourism and Arts, Yangzhou University Guangling College, Yangzhou, 225000, China

***Abstract*—With the popularization and development of mobile devices, the demand for image processing continues to increase. However, the limited hardware resources of mobile devices make traditional CPU computing unable to meet the requirements of real-time image processing. In response to the limited rendering resources of mobile platforms, this study adopts OpenGL for graphic interface design and animation simulation of facial expressions to control the changes of facial expressions in real-time, achieve facial expression animation simulation, and develop effective expression fusion methods. By combining rich rendering effects such as particle effects, facial expressions can be expressed more realistically and interestingly in 3D models. The results indicated that the research method only required less than 50 MB of memory, and the average accuracy of facial expression recognition had significantly improved. The final normalized average error level was close to 4%, with higher accuracy. The processing speed of each image was around 19.4ms, which could achieve animation simulation of facial expressions and had strong universality and flexibility. This method optimizes the real-time performance, stability, and user experience of facial expression real-time animation simulation, which can meet the needs of different application scenarios.**

***Keywords*—OpenGL; mobile platform; facial expressions; animation simulation; rendering**

I. Introduction

In computer graphics, due to the emergence of OpenGL technology, graphics has entered a new era. OpenGL is widely used in 3D animation, Computer-Aided Design (CAD), visual simulation, and other fields. Its high quality and high-performance characteristics enable developers to create high-quality 2D and 3D graphics. Whether for broadcasting, CAD/computer-aided manufacturing/computer-aided engineering, entertainment, medical imaging, or virtual reality development, OpenGL provides outstanding graphics quality and performance. OpenGL can make full use of the functions of modern graphics hardware, such as Graphic Processing Units (GPU) parallel computing, and texture mapping, to provide higher graphics rendering performance and effect. This hardware-accelerated support makes OpenGL excellent at handling complex graphic effects, such as shadow, reflection, anti-aliasing, etc. Therefore, OpenGL has become one of the ideal choices for graphics rendering due to its high-quality graphics rendering, cross-platform compatibility, rich development tools and resources, and hardware acceleration support. In traditional graphics, rendering is the process of texturing the surface of an object. OpenGL provides a new texture mapping method, which maps textures to the real world through a transformation matrix [1]. A transformation matrix can transform a 3D object into a 2D image, which can then be displayed on the screen. The transformation matrix can realize the 3D model view transformation, model transformation, clipping, projection transformation, and viewport transformation. OpenGL is a cross-language and cross-platform application programming interface for rendering 2D and 3D vector graphics. Whether it is 3D animation, CAD, or visual simulation, visual computation programs take advantage of OpenGL's high graphics quality and high-performance characteristics. OpenGL has a good structure, intuitive design, and logical commands. Compared to other graphics packages, OpenGL has very little code and therefore high execution speed. In addition, OpenGL encapsulates information about the basic hardware, so that developers do not need to design for specific hardware characteristics. OpenGL-based graphics applications can run on many systems, including a variety of user electronics, PCS, workstations, and supercomputers. As a result, OpenGL applications can be adapted to various target platforms chosen by the developer [2-3].

With the popularization and development of mobile devices, the demand for mobile image processing is increasing. However, due to the limited hardware resources of mobile devices, traditional CPU computing cannot meet the requirements of real-time image processing. A review of the relevant mainstream products on the market reveals that the majority of video special effects and facial expression simulations are based on the display of 2D animation. In contrast, the expression of 3D animation requires a significant amount of computing resources, making real-time performance difficult to achieve, and it is rarely promoted or applied in products. At present, most of the algorithms based on data structures and mathematical methods are used in research work, but such methods require computers to have a high computing speed, and the running speed on mobile platforms is slow [4-5]. The objective of this research is to develop a Facial Expression Animation Simulation (FEAS) system that can drive a virtual 3D model through facial expression unit parameters based on face key point detection through the front camera on a universal mobile platform. The system will then be encapsulated into operational applications that meet the requirements of real-time performance, low cost, practicality, stability, and maintainability. To adapt to the characteristics of the mobile platform and achieve the real-time goal, a universal real-time FEAS system based on OpenGL technology is developed.

OpenGL is used for graphic interface design, which can be run on the mobile platform, and the 3D Max model file format is adopted. The 3D model is drawn by OpenGL, and then the texture mapping is carried out by OpenGL to realize the real-time simulation of facial expression animation.

The article conducts research through six sections. Section II is a review of the research status of OpenGL in real-time FEAS applications. Section III is the research on 3D facial expression animation methods based on OpenGL. Section IV is performance validation of the proposed model. Results and discussion is given in Section V. Section VI is the conclusion.

II. Related Works

Bossér L et al. developed an underwater target echo signal strength prediction method based on image processing technology. By injecting code into the segment shader stage of the OpenGL graphics pipeline, the light reflection problem was transformed into an acoustic reflection problem. Compared to the Kirchhoff method, this method had higher computational accuracy [6]. Calabuig B E et al. improved the 3D engine used for real-time CAD geometric representation by using OpenGL Shaders to make ray tracing and radiation rendering techniques more realistic. It improved computing power by solving visualization problems and optimizing data. After verification, the model provided correct results for both computer-optimized 3D engines and network environment 3D engines [7]. Yin J et al. proposed a new lattice Boltzmann method to reduce the computational time, memory allocation, and complexity of existing algorithms for high-precision graphics processing units. It used OpenGL-based Compute Shaders GPGPGPU technology to accelerate and avoid the storage of distribution function components to reduce memory allocation size. The spatial accuracy was tested through 2D and 3D lid drive cavity benchmark cases, with high accuracy [8]. Rémi F et al. proposed a new method based on the OpenGL4 framework to achieve GPU-based high-order mesh rendering and almost pixel-accurate rendering of high-order solutions. This method used OpenGL fragment shaders to calculate precise solutions for each pixel by transferring degrees of freedom and shape functions to GPUs with textures. Compared to standard techniques, it eliminated linear interpolation steps, reduced memory usage, and improved rendering accuracy and speed [9].

Pham H X et al. proposed using Recurrent Neural Networks (RNNs) to achieve time-varying contextual nonlinear mapping between audio streams and facial micro movements, to drive 3D mixed-shape facial models in real-time. It depicted different speech generation actions in the form of lip movements, automatically estimating the speaker's emotional intensity with high accuracy [10]. Berson et al. proposed a generative RNN to make facial animation editing faster and easier for non-experts. It generated realistic movements in designated parts of existing facial animations based on user-provided guidance constraints. The experiment showed that the system had strong usability in scenarios such as motion filling, expression correction, semantic content modification, and noise filtering during occlusion [11]. Pumarola et al. proposed a weakly supervised strategy to train the model to describe the anatomical facial movements that define human expressions using continuous manifolds. This strategy annotated images through activated action units and utilized a new self-learning attention mechanism to make the network robust to constantly changing backgrounds, lighting conditions, and occlusion, with high performance [12]. Paier W et al. developed an interactive animation engine using deep learning to achieve more realistic rendering in difficult areas such as the mouth and eyes of animated faces. It described an automatic pipeline for generating training sequences composed of dynamic textures and consistent 3D facial model sequences through simple and intuitive facial expression editing visualization. It also trained a variational auto-encoder to learn the low dimensional latent space of facial expressions for interactive facial animation [13]. Huang Z et al. proposed a real-time simulation method for humanoid robot facial expression imitation based on a smooth constrained inverse mechanical model by combining deep learning models with motion smoothing constraints to improve the spatio-temporal similarity and motion smoothness of facial expression imitation. This method sent the generated optimal position sequence to the hardware system, making it consistent with the performer's spatio-temporal characteristics, with a deviation of less than 8% [14].

In summary, although researchers have proposed many methods to improve real-time FEAS performance and achieved certain results, the optimization scheme still needs improvement. Therefore, this study aims to achieve facial animation simulation in different application scenarios through the OpenGL platform.

III. A 3D Facial Expression Animation Method Based on OpenGL

This study is based on the OpenGL ES computational shader, which accelerates image processing algorithms through GPU parallel computing for different memory access and computational characteristics to achieve higher performance. Detailed implementation steps for face detection, facial keypoint localization, and tracking are introduced.

A. Design and Optimization of Facial Expression Capture Algorithm Module

There are many types of facial expression capture devices, and traditional devices are mostly bulky and have larger machine sizes. Due to the popularity of mobile Internet and mobile devices, this study selects the video camera attached to mobile devices as the capture device. The facial expression capture module consists of various algorithms, including image-based facial detection and facial key point localization [15]. Facial detection is achieved by obtaining image data from videos or cameras, normalizing the image data, and tracking the position of key points. In addition, for some more complex expressions, specially trained classifiers such as blinking and sticking out the tongue are used to improve the richness of the expressions. Finally, the entire tracking result is filtered to generate and extract facial expression unit parameters.

Facial expression collection includes face detection, key point localization and tracking, and solving the motion parameters of expression primitives. The key point tracking adopts an image denoising method based on median filtering, which significantly reduces computational complexity while ensuring tracking accuracy. Kalman filtering is used for facial

key points to filter and process the key point information generated in each frame. By calculating the key point information, the parameters of facial expressions are obtained. The specific process is shown in Fig. 1.

In OpenGL ES computing shaders, it is necessary to fully tap into the hardware resources of the GPU to create a sufficient number of threads, ensuring that all processors on the GPU are in the operational state of data processing, that is, to partition tasks as finely as possible. Fine partitioning of GPUs can improve their computational efficiency and effectively suppress latency. In the OpenGL ES rendering program, different tasks have different requirements for the size of the thread workgroup. To maximize the bandwidth usage of memory access, the size of the thread team can be set to be a full multiple of the block size to ensure that each thread can continuously access memory and avoid access conflicts.

This study proposes an image denoising method based on median filtering. This method is a non-linear filtering technique based on statistics, which replaces noise with the median in the median filtering window. The median filtering window traverses the entire image and calculates the median of all values within the filtering window as new pixels. The median calculation formula of the median filtering algorithm is Eq. (1).

$$g(x,y)=median\left\{f(x-i,y-j),i,j\in H\times W\right\} \quad (1)$$

In Eq. (1), $f(x,y)$ and $g(x,y)$ are the substitution values of the original image and the output image, respectively. $H\times W$ is the size of the filtering window (usually $H=W$ and odd, such as 3×3, 5×5, 7×7, etc.). Due to the fact that median filtering belongs to nonlinear filtering, its mathematical analysis becomes more complex when dealing with images containing random noise. For a normally distributed image with zero mean filtering noise, the noise variance of the median filtering can be approximately expressed as Eq. (2).

$$\sigma_{med}^2=\frac{1}{4nf^2(\bar{n})}\approx\frac{\sigma_i^2}{n+\frac{\pi}{2}-1}*\frac{\pi}{2} \quad (2)$$

In Eq. (2), σ_i^2 represents the size of the incoming noise, n represents the median, and $f(\bar{n})$ represents the noise density function. The variance of mean filtering is expressed as Eq. (3).

$$\sigma_0^2=\frac{1}{n}\sigma_i^2 \quad (3)$$

The performance of median filtering depends on one factor, which is the size of the filtering window. At the same time, due to the different data types of images, a single algorithm cannot solve the corresponding problem. The Image Histogram (IH) represents the image distribution by quantifying the pixel numbers in each brightness value. For instance, in grayscale images, the IH algorithm calculates the quantity of pixel values (from 0 to 255) in the image and produces a histogram array with 256 elements. This study uses IH equalization to change the original IH by dispersing pixels that were previously existed in specific pixel values. For the original histogram $H(i)$, the calculation of the equilibrium distribution $H'(i)$ is Eq. (4).

$$H'(i)=\sum_{0\le j<i}H(i) \quad (4)$$

Finally, the final output image is calculated using Eq. (4), as exhibited in Eq. (5).

$$equalized(x,y)=H'(src(x,y)) \quad (5)$$

To solve the representation problem of 3D face acquisition, this study introduces the grid sampling operator and defines the downsampling and upsampling of grid features. Fig. 2 shows the grid sampling operation. Vertices shrink during downsampling operations. The downsampled vertices are a subset of the original mesh vertices. Each weight represents whether the Q-th vertex is retained during downsampling. Due to the infeasibility of lossless downsampling and upsampling for general grid surfaces, the upsampling matrix is constructed during the downsampling period. The vertices retained during downsampling are convolved, while these vertices are retained during upsampling. During downsampling, discarded vertices are mapped onto the downsampled mesh surface using centroid coordinates.

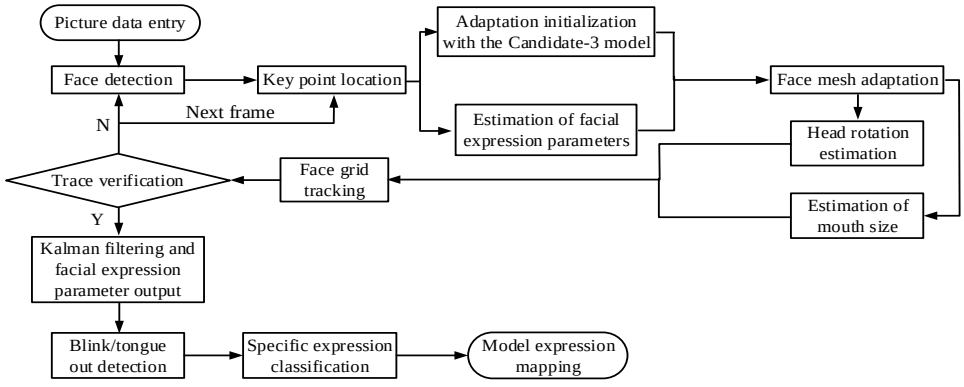


Fig. 1. Flow chart of facial expression capture.

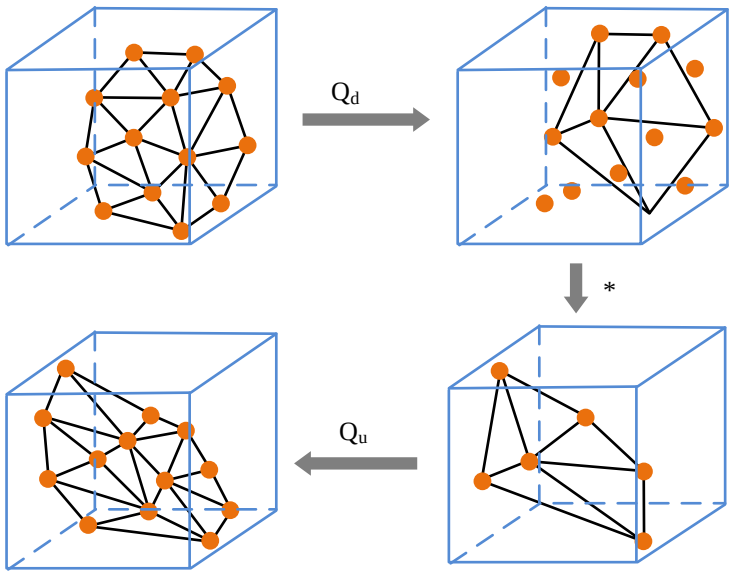


Fig. 2. 3D grid sampling operation.

In OpenGL ES computational shaders, data type optimization is an important optimization. This study aims to use low-precision data types as much as possible based on the actual needs of calculations. When precision is not high, use single-precision floating-point instead of double-precision floating-point to avoid using type conversion operations in shaders. Type conversion requires additional computational resources, especially for low-end devices. The advantage of integral filtering is that its computational complexity is independent of the filtering radius, so it can quickly process large-sized filtering templates and achieve filtering effects of different radii through multiple integral filtering. The localization and tracking of facial key points adopt a Supervised Descent Method (SDM)-based facial key point localization and tracking algorithm. Image information is extracted based on the current facial shape and position, and the shape is updated, as shown in Fig. 3.

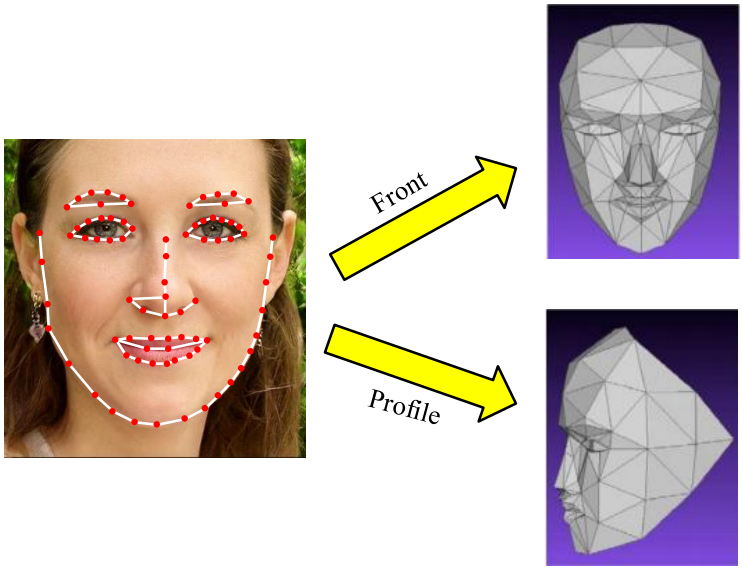


Fig. 3. Face key points.

The algorithm of integral filtering is simple to implement, fast in calculation speed, and widely used in image processing fields, such as face recognition, object detection, etc. $I(x,y)$ is the integral image, which is the sum of the original $i(x,y)$ on the left and top of point (x,y). The calculation formula is Eq. (6).

$$I(x,y)=\sum_{x'}^{x}\sum_{y'}^{y} i(x',y') \quad (x' \le x, y' \le y) \qquad (6)$$

Once the integral image is calculated, the pixel sum in any rectangular area composed of the upper-left point $(x1, y1)$ and the lower-right point $(x2, y2)$ can be found with the time

complexity of $O(1)$. Calculating the rectangular range in the left image requires eight summation operations, and the computational complexity increases with the expansion of the matrix area. The process is given in Eq. (7).

$$S(x_1, y_1, x_2, y_2) = \sum_{y=y_1}^{y_2} \sum_{x=x_1}^{x_2} i(x, y) \quad (7)$$

In facial detection, features are decomposed into integral filter algorithms based on row and column directions. The algorithm for row direction is an exclusive scanning operation, as shown in Eq. (8).

$$\lfloor a_0, a_1, \ldots, a_{n-1} \rfloor \to \left\lfloor 0, a_0, (a_0 + a_1), \ldots, \sum_{i=0}^{n-2} a_i \right\rfloor \quad (8)$$

B. Design and Implementation of Animation Synthesis and Rendering Module

3D facial expression animation is mainly used to model models. Model modeling requires determining what kind of model to create and some basic facial expressions of the model. In the process of modeling and fusion, it is necessary to receive the basic expression parameter sizes transmitted by facial capture for linear fusion, to render the requirements of different models and enrich the expression of facial expressions [16]. Expression fusion P_{expr} is Eq. (9).

$$P_{expr} = p_{neur} + \sum_i \beta_i \times \Delta p_i \quad (9)$$

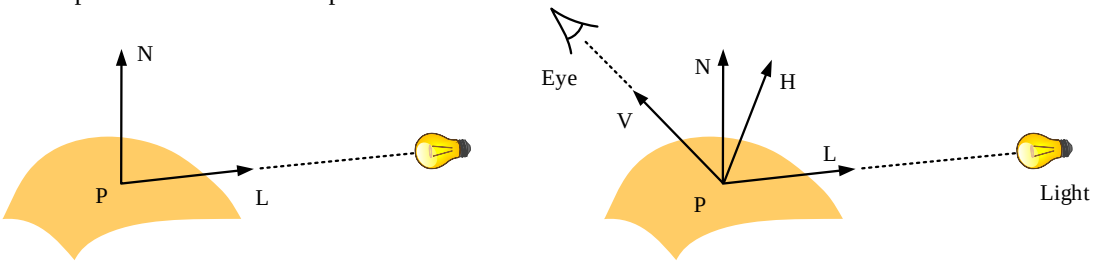
In Eq. (8), β_i represents the expression parameter weight passed by the face capture module. p_{neur} represents the vertex coordinates of the 3D model in its natural state, and p_i represents the distance between the basic expression vertex of the 3D model and the natural expression vertex. By performing this linear calculation and overlaying all expression parameter weights, the final 3D model expression vertex position can be obtained, without the need for special processing in texture mapping. Based on considerations of computational complexity and effectiveness, this study chooses the Phong lighting model for calculation, as shown in Eq. (10).

$$light\ color = emissive + ambient + diffuse + specular \quad (10)$$

This model defines light as composed of self-illumination, diffuse reflection, ambient light, and mirror light, as shown in Figure 4. Self-luminous $emissive$ is used to simulate the light source information emitted by objects such as the sun. $ambient$ often simulates ambient light and low light on a global scale. In practical applications, diffuse reflection $diffuse$ is the most important, combined with the calculation of the incident angle of light rays, which has the strongest performance on the direction of light rays. The calculation method is Eq. (11).

$$diffuse = light\ color \times \max(N \cdot L, 0) \quad (11)$$

In Eq. (11), N represents the normal and L represents the inverse of the incident vector. P represents the current vertex, while $N \cdot L$ is mainly used to determine whether the points on the face are facing or facing away from the light source.



(a) Diffuse reflection (b) Specular light

Fig. 4. Schematic diagram of diffuse and specular light.

Mirror light mainly simulates the reflection effect, calculated as Eq. (12).

$$specular = light\ color \times facing \times \left(\max(N \cdot H, 0)\right)^{shinness} \quad (12)$$

In Eq. (12), H is the centerline of the line of sight reversal V and L. $shininess$ is the light intensity parameter. $facing$ represents a Boolean value. If $N \cdot L$ is greater than 0 and takes a value of 1, it indicates that the light source vector L is facing the surface. The final color value can be expressed as Eq. (13).

$$output_{color} = texture_{color} \times (material_k \times light_{color}) \quad (13)$$

In Eq. (13), $material_k$ can simulate different materials by adjusting the coefficient. Finally, the color output is obtained through Eq. (14).

$$light_{color} = K_e + K_a \times ambient + K_d \times diffuse + K_s \times specular \quad (14)$$

In Eq. (14), K_e, K_a, K_d, and K_s represent vec4. If it is a model with fur, it needs to enable blending before rendering the transparency map. The mixing in OpenGL is achieved through Eq. (15).

$$C_{result} = C_s \times F_s + C_d \times F_d \quad (15)$$

In Eq. (15), C_s is the color vector of the texture. C_d

represents the color vector currently stored in the color buffer. F_s and F_d are the source and target factor values, specifying the impact of alpha values on the C_s color and target color.

This study also utilizes the MG2 algorithm in OpenCTM software to store and compress 3D model files. OpenCTM files are a simple format that can be well applied to modern 3D image rendering pipeline programs like OpenGL. In a 3D model, the minimum required information is the vertex coordinates of the model and the patch composed of coordinate points. OpenCTM simply includes these two prerequisites and also supports adding more options. An OpenCTM file only contains one Mesh, which is mainly divided into two parts: vertex information and patch information. Among all Vertex, Normal, UV, and Attrib, Vertex is mandatory. Table I shows the subscript $(0, 1, 2, \ldots, N)$ of the vertex to ensure the correspondence of the data in secret.

TABLE I. MODEL DATA STORAGE ARRANGEMENT

Index	0	1	2	3	4	…	N
Vertex	v0	v1	v2	v3	v4	…	v0
Normal	n0	n1	N2	n3	n4	…	nN
UVCoord1	t10	t11	t12	t13	t14	…	t1N
UVCoord2	t20	t21	t22	t23	t24	…	t2N
Attrib1	a10	a11	a12	a13	a14	…	a1N
Attrib2	a20	a21	a22	a23	a24	…	a2N

IV. ANALYSIS OF REAL-TIME FEAS

This study compares various algorithms on a scale of 1024×1024 and selects several excellent algorithms in recent years for comparison. Except for the code not implemented in the OpenCV library, performance comparisons are made with the corresponding algorithms in the OpenCV library.

A. Analysis of Facial Capture Performance

This study uses Xcode to compile static libraries, and then calls mobile platform camera data through the Object c interface to transfer image data to the C interface to test the performance of the proposed face capture method. This study performs timestamp operations on each specific module, and after running for 10 minutes, the statistical results are obtained as shown in Table II. In Table II, the entire process achieves ultra real-time processing speed on both iPhone 12 and Galaxy S8, requiring less than 50 MB of memory.

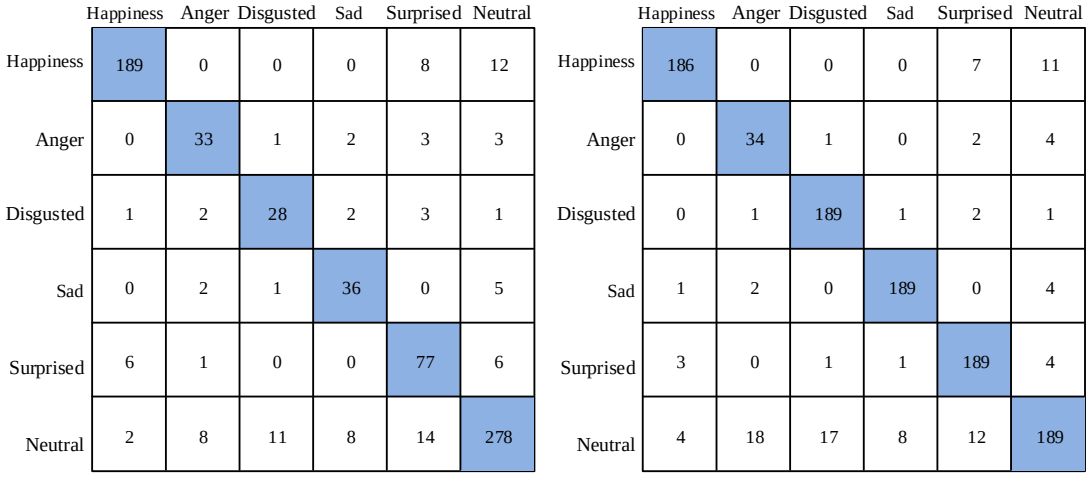
TABLE II. FACE CAPTURE PERFORMANCE STATISTICS

Face capture module	**IPhone12(iOS 9.5.3)**			**Galaxy S8 (Android 8.0.1)**		
	GPU (%)	***Memory (MB)***	***Time (ms)***	***GPU (%)***	***Memory (MB)***	***Time (ms)***
Camera recording	12.2	48.6	/	13.6	51.3	/
Face detection and key location tracking	5.4	2.1	4.9	6.6	2.6	6.5
3D mesh adaptor	1.3	1.7	0.4	1.5	1.6	1.3
Eye classifier	1.2	0.4	0.7	1.4	1.4	3.2
Tongue sticking out classifier	1.0	1.2	0.4	1.1	1.3	3.2
Expression computing output	2.2	1.8	0.4	2.9	2.7	3.0
All functions	23.3	55.8	6.8	27.1	60.9	17.2

To objectively and accurately detect the recognition performance under different expressions, this study used the data from the expression fusion library as the basis to achieve data balance, and used the basic parameters in the model to adjust the corresponding penalty coefficients. The larger the parameter, the longer it took to identify the error. Therefore, it was necessary to determine a higher mean for this type of parameter. This study adjusted the parameters to predict the expressions to be tested for classification using the trained classifier, and compared the recognition rate with the pre-imbalanced expression data classifier. Fig. 5 shows the confusion matrix of six facial expressions. After data balancing, the recognition accuracy on the test sample increased from 87.69% before data balancing to 91.26%.

This study also conducted sample detection experiments on traditional texture features, Active Appearance Model (AAM), and proposed fusion feature extraction methods, and compared their results. Table III shows the average recognition rate of six types of expressions using three feature extraction methods. In Table III, compared to the other two methods, the research method had a better recognition rate on facial expressions than traditional and AAM algorithms. The recognition accuracy of Happy, Neutral, and Surprised facial expressions was relatively high, but the recognition accuracy of Anger, Sad, and Distorted facial expressions was not as good as the above three types. There were two main reasons for this situation. One reason is that each person has different ways of expressing sadness and disgust, and there are also significant differences. Therefore, the results of identifying such expressions are relatively scattered and may be recognized as various types of expressions. When

recognizing anger, it is easy to identify it as a surprise because when extracting features from the eye area, both types of expressions are prone to eye enlargement, which can confuse recognition. The second is that the three types of expressions, Anger, Distorted, and Sad, have certain similarities in themselves, and people may find it difficult to distinguish these three types of expressions in real life. Therefore, the recognition accuracy of these three types of expressions is lower than that of other expressions, but experiments have shown that the average accuracy of expression recognition in the research method still has a significant improvement.



(a) Confusion matrix before data equalization (b) Confusion matrix after data equalization

Fig. 5. Confusion matrix of six facial expressions.

TABLE III. COMPARISON OF FACIAL EXPRESSION RECOGNITION RESULTS

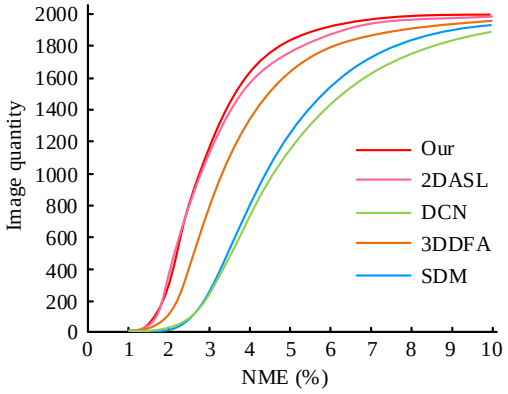
Method	Traditional texture feature	AAM feature	Research proposed fusion features
Happiness	72.13%	83.62%	92.43%
Anger	75.32%	81.75%	91.65%
Sad	76.32%	80.69%	93.47%
Surprised	74.19%	79.42%	89.62%
Disgusted	73.58%	83.66%	86.78%
Neutral	71.96%	84.51%	90.33%
Average	74.37%	80.42%	91.76%

B. Analysis of Facial Animation Rendering Effects

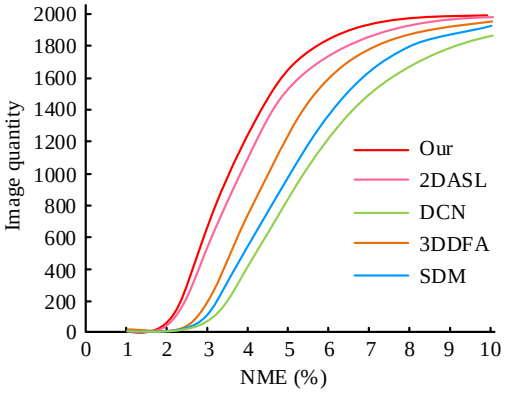
To comprehensively evaluate the facial rendering performance of research methods, multiple excellent algorithms in recent years were selected for comparison, including 2DASSL [17], DCN [18], 3DDFA [19], SDM [20] and other algorithms. The Normalized Mean Error (NME) was selected as the evaluation metric, which represents the average error value of normalized feature points. The normalized size corresponds to the square value of the product of the length and width of the face border to ensure that face alignment was not affected by external factors. Fig. 6 (a), (b), and (c) show the cumulative alignment errors of 68 2D face key points (2D-KP), all 2D face key points (A2D-KP), and all 3D face key points (A3D-KP) for face alignment. Among the 68 cumulative errors of 2D-KP, 2DASSL and research methods performed similarly, demonstrating excellent performance. 3DDFA is stable and efficient, and also performs well. The research method achieved optimal NME on 68 2D-KP, A2D-KP, and A3D-KP datasets.

This study evaluated the impact of loss functions on alignment using three commonly used loss functions: PDC, VDC, and WPDC, based on MATLAB. The performance of models trained with different loss functions on the dataset is Fig. 7. PDC could not fit the model well, and the convergence of errors was poor. It basically converged when the NME reached around 8%. VDC performed better than PDC, but was affected by pathological curvature during the convergence process, so fitting errors might occur during the fitting process, resulting in poor visualization results. Compared to other methods, WPDC had better error performance and could quickly reduce NME. However, overall, the loss function of the research method combined key point loss, resulting in better performance and the fastest reduction of NME. The final error level was close to 4%, and the accuracy was significantly higher.

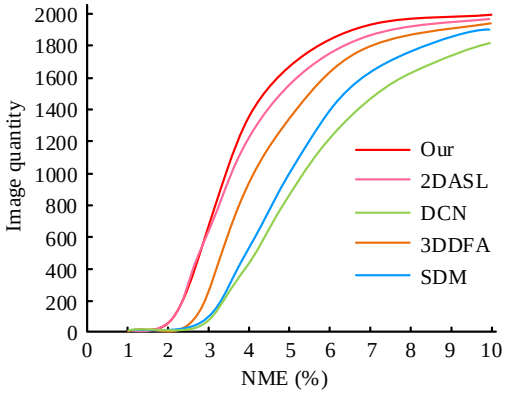
The proposed facial animation rendering method consumes a time proportional to the image size, as shown in Fig. 8. In a size of 200×300, the processing speed of each image was around 19.4 milliseconds, which was extremely fast. After testing, it could reach about nine milliseconds at a size of 120×120, but due to the small window, it was not suitable for real-time input of facial images. If face detection was performed by scaling images, sometimes the detection efficiency might be affected due to the small size of the face.



(a) NME for 2D key points



(b) NME for all 2D key points



(c) NME of all 3D key points

Fig. 6. Comparison of NME results.

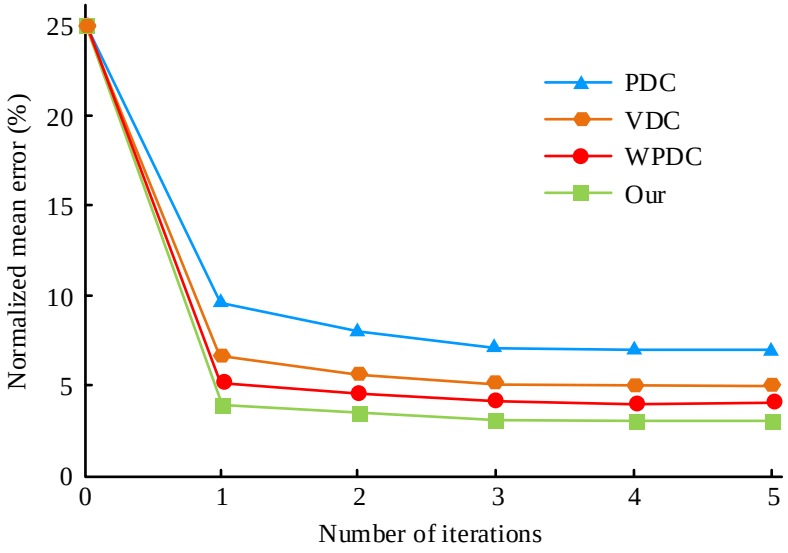


Fig. 7. Error function results in face alignment.

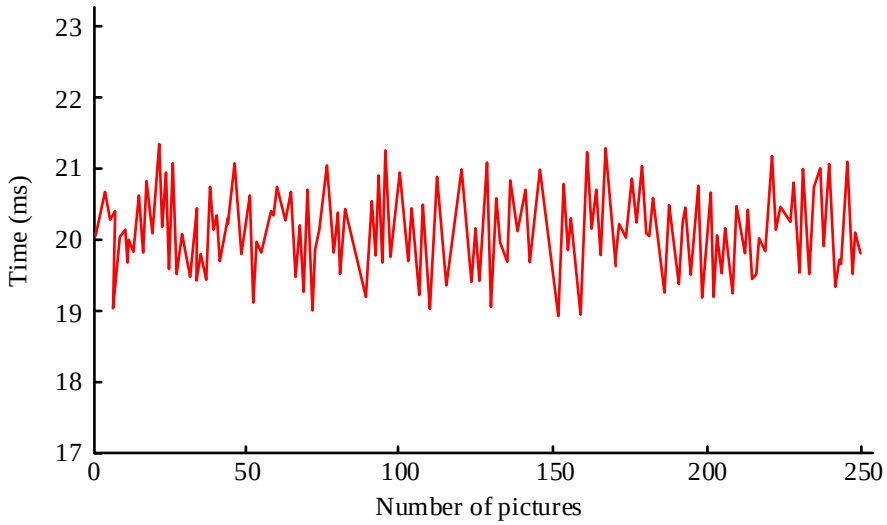


Fig. 8. Runtime of the image enhancement method on 200 images.

To better explore the impact of video frame preprocessing algorithms, speech features of speech driven networks, and lip animation results on expression and posture networks, this study divides each part and obtained six variants by considering one or more component parts. Method 1 is the benchmark method. Method 2 is to crop each frame separately. Method 3 is to add speech features from the speech network as inputs to the expression and posture network for feature fusion. Method 4 uses speech driven network prediction of lip animation as the basis. Method 5 adds the speech features of the speech network as input to the expression and posture network for feature fusion, and also uses the lip animation predicted by the speech driven network as the basis. Method 6 uses the proposed rendering method.

Table IV shows the objective evaluation results of various situations on the dataset. It can be found that the proposed rendering method can bring certain results in improving the expression posture. Among them, video frame preprocessing has the best effect on VER reduction, surpassing the combination of speech features and lip animation. In addition, it also has a good decrease in WER and VWER, surpassing individual speech features or lip animation.

TABLE IV. RESULTS OF ABLATION EXPERIMENT

Serial number	Video frame preprocessing	Ocular features	Lip animation	CER	WER	VER	VWER
1				68.7	98.1	61.0	97.6
2	√			67.5	96.2	59.1	94.4
3		√		67.0	96.4	60.0	94.8
4			√	67.8	97.0	60.3	95.7
5		√	√	66.4	95.0	59.4	93.7
6		√	√	65.0	92.7	57.5	91.0

V. RESULTS AND DISCUSSION

The process of face detection and face feature point detection is conducted on the collected ordinary images with the objective of determining the face region. Subsequently, facial expression recognition and face reconstruction are carried out on the cropped face region aiming to obtain the 3D model data of the corresponding 2D face image. Finally, 3D animation can be realized by using the renderer. However, when face image 3D modeling is applied in mobile devices, higher requirements are put forward for computing time. In regard to graphical rendering, OpenGL employs hardware acceleration technology, thereby enabling direct access to the GPU to accelerate the graphical rendering process, thus markedly enhancing the efficiency of the rendering process [21]. Furthermore, OpenGL employs a dual-cache system to process the computing scene, generated image, and display image separately. This approach significantly enhances the computing power and display speed of the computer, thereby improving the user experience and greatly reducing processing time. The proposed method was validated on iPhone 12 and Galaxy S8, and it was found that the entire process achieved ultra real time processing speed with less than 50 MB of memory required. In addition, the processing speed of each image in the 200×300 size was about 19.4 milliseconds, which is extremely fast. Wang B et al. proposed a face feature extraction algorithm based on deep learning, which adopted C++ and OpenGL for rendering simulation, and has good performance in accuracy and speed [22]. The results are consistent with the results of this study, indicating that the mobile platform FEAS technology based on OpenGL can meet the real-time requirements.

Face images contain some very complex information and content. Typical facial features include the eyes, mouth,

eyebrows, and the face as a whole. Additionally, there is a great deal of information present within the face, composed of these parts [23]. Based on face detection, the key facial feature points are automatically located according to the input face image, such as the eyes, nose tip, corners of the mouth, eyebrows, and contour points of the face parts, etc. The input is the face appearance image and the output is the set of face feature points. Due to the combination of key point loss, the proposed method can rapidly reduce the normalized average error, and the final error level is close to 4%. The error performance on 68 2D key points, all 2D face key points, and all 3D face key points is optimal. 3D DFA and 2D ASL are messy in the 3D spatial error distribution. It indicates that there are some defects in the face estimation, which reduces the accuracy and visibility of model fitting. With the help of OpenGL, 3D model can be quickly mapped to 2D space and 3D effect can be displayed in 2D space. The 3D face data acquisition based on OpenGL is extremely efficient, which can ensure a high frame rate, achieve continuous animation effect, and effectively improve the adaptability of face reconstruction to the scene.

VI. Conclusion

This study was based on the OpenGL platform and utilized the OpenGL ES computational shader to develop a GPU version of a high-performance image algorithm, providing a real-time facial animation processing solution for mobile devices. By combining a series of architecture and resource optimization strategies, the cross-platform application of Android and iOS systems on mobile devices has been implemented, and the stable performance of the entire solution on mobile platforms has been guaranteed. The results indicated that the research method achieved ultra real-time processing speed on two different operating systems and only required less than 50 MB of memory. After data balancing, the recognition accuracy increased from 87.69% before data balancing to 91.26%, and there was a significant improvement in the average accuracy of facial expression recognition. The loss function of the research method combined key point loss, resulting in better performance and the fastest reduction of NME. The final error level was close to 4%, and the accuracy was significantly higher. In a size of 200×300, the processing speed of each image was around 19.4 modes. This method has the characteristics of simple operation, powerful functionality, high realism, and strong real-time performance. However, this study only optimized high-performance image processing algorithms for a single architecture and efficient computation cannot be performed when the algorithms are ported to other platforms. In the future, high-performance facial animation simulation processing algorithms for other architecture processors can be optimized.

References

[1] Usman A M, Abdullah M K. An assessment of building energy consumption characteristics using analytical energy and carbon footprint assessment Model. Green and Low-Carbon Economy, 2023, 1(1): 28-40.

[2] Aryavalli S N G, Kumar G H. Futuristic vigilance: Empowering chipko movement with cyber-savvy IoT to safeguard forests. Archives of Advanced Engineering Science, 2023, 1(8): 1-16.

[3] Xu L. Fast Modelling Algorithm for Realistic Three-Dimensional Human Face for Film and Television Animation. Complexity, 2021, 20(2):1-10.

[4] Liu K, Sun Y, Yang D. The administrative center or economic center: Which dominates the regional green development pattern. A case study of shandong peninsula urban agglomeration, china. Green and Low-Carbon Economy, 2023, 1(3), 110-120.

[5] Pham H X, Wang Y, Pavlovic V. Learning Continuous Facial Actions from Speech for Real-time Animation. IEEE Transactions on Affective Computing, 2020, 13(3):1567-1580.

[6] Bossér L, Bossér J D. Target echo calculations using the OpenGL graphics pipeline. Applied Acoustics, 2021, 181(9):108133.

[7] Calabuig B E, Davia-Aracil M, Mora-Mora H F. Computational model for hyper-realistic image generation using uniform shaders in 3D environments. Computers in Industry, 2020, 123(1):1-13.

[8] Yin J, Yang J H. Virtual Reconstruction Method of Regional 3D Image Based on Visual Transmission Effect. Complexity, 2021, 92(12):1778-1797.

[9] Rémi F, Maunoury M, Loseille A. On pixel-exact rendering for high-order mesh and solution. Journal of Computational Physics, 2020, 424(13):1098-1112.

[10] Pham H X, Wang Y, Pavlovic V. Learning Continuous Facial Actions from Speech for Real-time Animation. IEEE Transactions on Affective Computing, 2020, 13(3):1567-1580.

[11] Berson E, Catherine S, Stoiber N. Intuitive Facial Animation Editing Based on A Generative RNN Framework. Computer Graphics Forum, 2020, 39(8):241-251.

[12] Pumarola A, Agudo A, Martinez A M, Sanfeliu A, Moreno N F. GANimation: One-Shot Anatomically Consistent Facial Animation. International Journal of Computer Vision, 2020, 128(3):698-713.

[13] Paier W, Hilsmann A, Eisert P. Interactive facial animation with deep neural networks. IET Computer Vision, 2020, 14(6):359-369.

[14] Huang Z, Ren F, Hu M, Chen S. Facial Expression Imitation Method for Humanoid Robot Based on Smooth-Constraint Reversed Mechanical Model (SRMM). IEEE transactions on human-machine systems, 2020, 50(6):538-549.

[15] Xiao R, Zhang J, Chai P, Ju C, Lin W C, He R. Dual interpolation boundary face method for 3-D acoustic problems based on binary tree grids. Engineering analysis with boundary elements, 2023, 150(6):7-19.

[16] Cao H, Wu Y, Bao Y, Feng X, Wan S, Qian C. UTrans-Net: A model for short-term precipitation prediction. Artificial Intelligence and Applications. 2023, 1(2): 106-113.

[17] Feng Y, Feng H, Black M J, Bolkart T. Learning an animatable detailed 3D face model from in-the-wild images. ACM Transactions on Graphics, 2021, 40(4):1-13.

[18] Alboqami F, Van Oudenhoven V C O, Ahmed U, Zahid U, Emwas A, Sarathy S M, Jameel A G. A Methodology for Designing Octane Number of Fuels Using Genetic Algorithms and Artificial Neural Networks. Energy & Fuels, 2022, 36(7):3876-3880.

[19] Ly A, El-Sayegh Z. Tire wear and pollutants: An overview of research. Archives of Advanced Engineering Science, 2023, 1(1): 2-10.

[20] Garai S, Paul R K, Kumar M. Intra-annual national statistical accounts based on machine learning algorithm. Journal of Data Science and Intelligent Systems, 2023, 2(2): 12-15.

[21] Kwolek B, Rymut B. Reconstruction of 3D human motion in real-time using particle swarm optimization with GPU-accelerated fitness function. Journal of Real-Time Image Processing, 2020, 17(4): 821-838.

[22] Wang B, Shi Y. Expression dynamic capture and 3D animation generation method based on deep learning. Neural Computing and Applications, 2023, 35(12): 8797-8808.

[23] Yali L, Huijie Z. Three-dimensional campus 360-degree video encoding VR technology based on OpenGL. Multimedia Tools and Applications, 2020, 79(7): 5099-5107.

Noise Reduction Techniques in ADAS Sensor Data Management: Methods and Comparative Analysis

Ahmed Alami, Fouad Belmajdoub
Faculty of Science and Technology of Fez, University of Sidi Mohammed Ben Abdellah, Fez, Morocco

***Abstract*—This review examines noise reduction techniques in Advanced Driver Assistance Systems (ADAS) sensor data management, crucial for enhancing vehicle safety and performance. ADAS relies on real-time data from conventional sensors (e.g., wheel speed sensors, LiDAR, radar, cameras) and MEMS sensors (e.g., accelerometers, gyroscopes) to execute critical functions like lane keeping, collision avoidance, and adaptive cruise control. These sensors are susceptible to thermal noise, mechanical vibrations, and environmental interferences, which degrade system performance. We explore filtering techniques including KalmanNet, Simple Moving Average (SMA), Exponential Moving Average (EMA), Wavelet Denoising, and Low Pass Filtering (LPF), assessing their efficacy in noise reduction and data integrity improvement. These methods are compared using key performance metrics such as Signal-to-Noise Ratio (SNR), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE). Recent advancements in hybrid filtering approaches and adaptive algorithms are discussed, highlighting their strengths and limitations for different sensor types and ADAS functionalities. Findings demonstrate the superior performance of Wavelet Denoising for non-stationary signals, SMA and EMA's effectiveness for smoother signal variations, and LPF's excellence in high-frequency noise attenuation with careful tuning. KalmanNet showed significant improvements in noise reduction and data accuracy, particularly in complex and dynamic environments. Extended Kalman Filter (EKF) and Unscented Kalman Filter (UKF) were especially effective for RADAR sensors, handling non-linearities and providing accurate state estimation. Emphasizing Hardware-in-the-Loop (HIL) bench testing to validate these techniques in real-world scenarios, this study underscores the importance of selecting appropriate methods based on specific noise characteristics and system requirements. This research provides valuable insights for ADAS and autonomous driving technologies development, emphasizing precise signal processing's critical role in ensuring accurate sensor data interpretation and decision-making.**

Keywords*—ADAS; sensor data management; noise reduction; KalmanNet; Wavelet Denoising; RADAR; SMA; EMA; LPF; HIL bench testing

I. Introduction

The evolution of automotive technology is rapidly transforming with the integration of Advanced Driver Assistance Systems (ADAS) and Autonomous Driving (AD) technologies, which are set to redefine vehicle safety and efficiency standards. Central to the success of these systems is the precise and reliable processing of sensor data from a variety of sources, including wheel speed sensors (WSS), inertial measurement units (IMUs), and radar systems. These sensors collectively enable the vehicle to perceive its environment, make decisions, and execute safe driving maneuvers [1], [2]. However, the performance of ADAS and AD is critically dependent on the quality of the sensor data, which can be severely compromised by noise and interference, posing significant challenges to system reliability [3]. As vehicle systems become increasingly interconnected through Vehicle-to-Everything (V2X) communication and the rollout of 5G networks, the demand for real-time, robust, and low-latency data processing solutions has intensified, underscoring the necessity for efficient noise reduction algorithms.

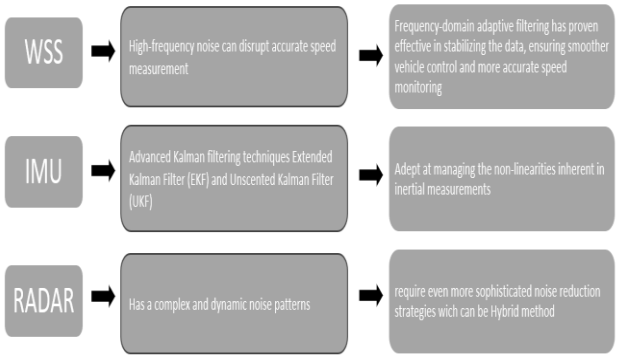


Fig. 1. Challenges encoutered by each sensor type.

As shown in the Fig. 1 and in order to mitigate these challenges, cutting-edge signal processing techniques have been developed to enhance sensor data quality, each tailored to the unique characteristics of different sensors. For instance, in WSS, where high-frequency noise can disrupt accurate speed measurement, frequency-domain adaptive filtering has proven effective in stabilizing the data, ensuring smoother vehicle control and more accurate speed monitoring [4]. IMUs, which are essential for maintaining vehicle dynamics and stability, benefit from advanced Kalman filtering techniques—particularly the Extended Kalman Filter (EKF) and Unscented Kalman Filter (UKF)—which are adept at managing the non-linearities inherent in inertial measurements [3], [12]. Meanwhile, RADAR systems, tasked with obstacle detection and environmental mapping, require even more sophisticated noise reduction strategies, including multi-sensor fusion and machine learning-based methods, to effectively manage the complex and dynamic noise patterns encountered in real-world scenarios [5], [6].

To address the challenges posed by dynamic environments in ADAS applications, this Fig. 2 illustrates the increasing

necessity for advanced noise reduction techniques tailored to specific driving situations.

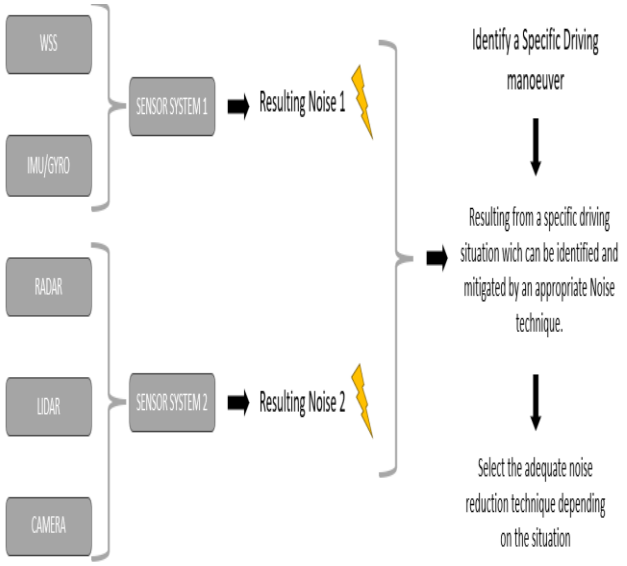


Fig. 2. The need for enhancing sensor noise reduction based on the driving situation identified.

As ADAS technology continues to advance, the integration of more complex sensor systems, such as radar and lidar, has heightened the need for innovative noise reduction techniques. These systems operate in dynamic environments filled with clutter, where conventional noise reduction approaches may fall short. To overcome these limitations, researchers are developing more sophisticated signal processing algorithms and exploring alternative modulation schemes, specifically designed to enhance data acquisition accuracy and reliability [7]. For RADAR systems, advanced Kalman filtering methods, such as the Unscented Kalman Filter (UKF), have demonstrated exceptional performance in tracking applications, particularly where noise is non-Gaussian and non-linear [12]. Simultaneously, multi-sensor fusion strategies, which integrate data from radar, lidar, and cameras, have become increasingly vital in providing a comprehensive perception of the vehicle's surroundings, compensating for the limitations of individual sensors [8]. However, the success of these approaches hinges on precise data synchronization and the development of sensor-specific noise reduction strategies, making them a critical area of ongoing research.

This Fig. 3 highlights the range of noise reduction techniques evaluated in this study, with a focus on those specifically selected for their effectiveness in various sensor types.

Recent advances in noise reduction techniques reflect a growing recognition of the need for tailored solutions across different sensors. For WSS, techniques like Simple Moving Average (SMA) and Exponential Moving Average (EMA) remain popular for their simplicity and computational efficiency, though Enhanced Simple Moving Average (ESMA) methods have emerged to address the limitations of initialization periods and stability [9], [10], [11]. IMUs, on the other hand, continue to benefit from the Kalman filtering family, with the EKF and UKF providing robust solutions for handling non-linear dynamics and improving measurement accuracy [13]. In RADAR systems, wavelet denoising has proven to be a powerful tool for managing non-stationary signals, while innovations like KalmanNet—which merges neural networks with Kalman filtering—offer significant advancements in reducing noise and enhancing signal clarity in complex operational environments [14], [15]. These innovations are critical not just for improving sensor data quality but also for enabling the high levels of performance demanded by modern ADAS and AD systems.

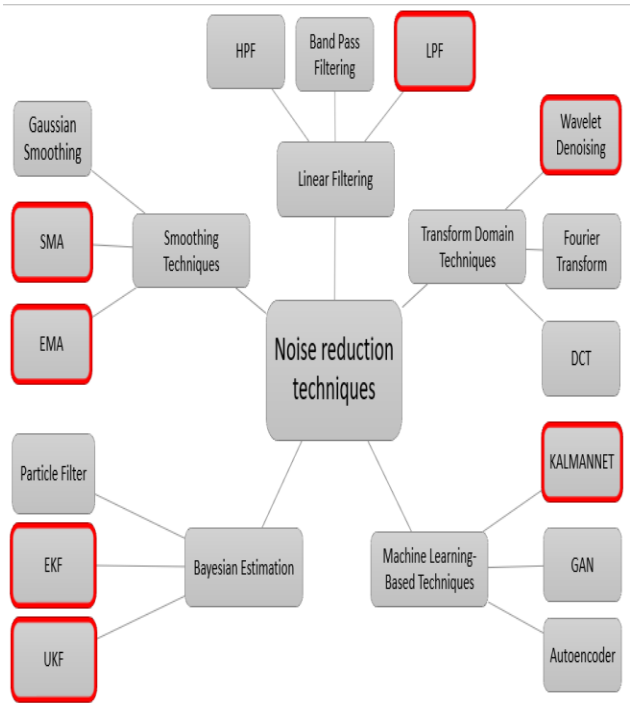


Fig. 3. Noise reduction techniques and chosen techniques shape outline are in red.

Effective noise reduction techniques are essential across WSS, IMUs, and RADAR systems, as they directly influence the reliability and safety of ADAS and AD applications. For WSS, methods like SMA and EMA continue to provide efficient solutions for mitigating speed-related noise, ensuring smoother control and accurate speed data [9], [10]. In IMUs, the advanced filtering techniques of EKF and UKF are crucial for maintaining vehicle stability by effectively handling the non-linearities in inertial data [16], [17]. RADAR systems, operating in complex environments, benefit significantly from wavelet denoising and hybrid methods like KalmanNet, which enhance signal clarity and improve the detection of critical obstacles [14], [15]. These tailored approaches are not only vital for current applications but also lay the groundwork for future advancements in automotive safety and sensor technology [18], [19], [20].

To validate the effectiveness of these advanced noise reduction techniques, we conducted a comprehensive empirical study using synthetic datasets that closely mimic the noise characteristics encountered by automotive sensors such as WSS, IMUs, and RADAR. By systematically introducing a

range of noise types and levels, we rigorously evaluated the performance of each noise reduction method across key metrics, including Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Signal-to-Noise Ratio (SNR), and Peak Signal-to-Noise Ratio (PSNR). This approach ensured a robust assessment of each technique's ability to enhance data accuracy and reliability under real-world conditions.

The Fig. 4 shows a summary of the key findings from existing literature, emphasizing the importance of tailored noise reduction approaches for enhancing sensor data accuracy in ADAS applications.

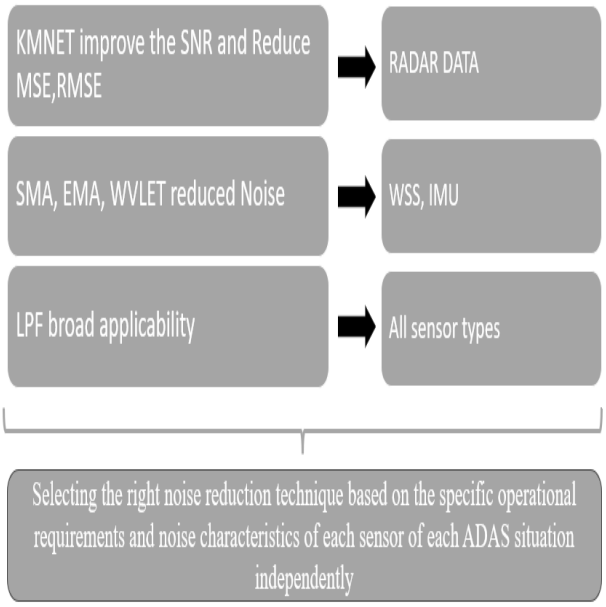


Fig. 4. Key findings of the existing literature.

The findings of this study strongly support the hypothesis that a tailored, sensor-specific approach to noise reduction can significantly improve data accuracy and system performance across various ADAS and AD applications. KalmanNet and hybrid techniques showed the greatest improvements in SNR and significant reductions in MSE and RMSE for RADAR data, while SMA, EMA, and Wavelet Denoising effectively reduced noise in WSS and IMU data, with Low Pass Filtering (LPF) providing broad applicability across all sensor types [21], [22]. These results underscore the critical importance of selecting the right noise reduction techniques based on the specific operational requirements and noise characteristics of each sensor, paving the way for enhanced automotive safety and efficiency.

In the introduction, the critical role of Advanced Driver Assistance Systems (ADAS) and Autonomous Driving (AD) technologies in enhancing vehicle safety and efficiency is examined, with an emphasis on the importance of accurate sensor data processing. The challenges associated with noise and interference in key sensors—such as wheel speed sensors (WSS), inertial measurement units (IMUs), and radar systems—are highlighted, particularly given their integral role in ADAS and AD system performance. The methodology section discusses the application of advanced noise reduction techniques, specifically tailored to the unique characteristics of these sensors, while also considering the demands of Vehicle-to-Everything (V2X) communication and 5G networks. Techniques such as Simple Moving Average (SMA), Exponential Moving Average (EMA), Wavelet Denoising, Low Pass Filtering (LPF), KalmanNet, Extended Kalman Filter (EKF), and Unscented Kalman Filter (UKF) are evaluated for their efficacy in noise reduction, with an additional focus on the necessity of real-time processing and the preference for simpler, computationally efficient algorithms over more complex ones. The results provide a detailed analysis, demonstrating the effectiveness of these methods in enhancing Signal-to-Noise Ratio (SNR) and reducing Mean Squared Error (MSE) and Root Mean Squared Error (RMSE), while also considering the practical implications of implementing these techniques in real-time V2X and 5G environments. The discussion elaborates on these findings, underscoring the importance of sensor-specific noise reduction strategies and their implications for the future development of ADAS. The conclusion synthesizes the key contributions of the study, proposing directions for future research aimed at further optimizing these techniques and ensuring their seamless integration into comprehensive, real-time ADAS frameworks, thereby advancing the reliability and performance of automotive safety systems.

Given the critical role that precise noise reduction techniques play in the effectiveness of ADAS, it is essential to explore and build upon existing research that has laid the groundwork in this field. The following section reviews the advancements and challenges documented in recent literature, providing a context for the methodologies employed in this study.

II. Related Work

Recent research has focused extensively on developing data-driven frameworks for diagnostics and prognostics across various domains, including automotive and aerospace. These frameworks typically involve sophisticated data acquisition processes from sensors and control units, often leveraging machine logs and CAN bus networks to enhance system monitoring and fault detection [23], [24]. Advanced data processing techniques, such as feature selection and extraction, have been employed to improve diagnostic accuracy, but their application in real-time systems remains a challenge due to computational limitations [23], [24].

The Fig. 5 provides an overview of the testing and validation activities critical to ensuring the reliability and performance of automotive systems, especially in the context of noise reduction.

The adoption of machine learning algorithms, including Random Forests, Bayesian estimation methods, and Cox proportional hazards models, is widespread for fault detection and remaining useful life (RUL) prediction; however, these approaches often require significant computational resources, limiting their applicability in real-time environments [25]-[26]. Despite these advancements, there is a clear need for further research into hybrid models that can efficiently balance predictive accuracy with real-time processing demands,

particularly in safety-critical systems where data availability may be limited [25].

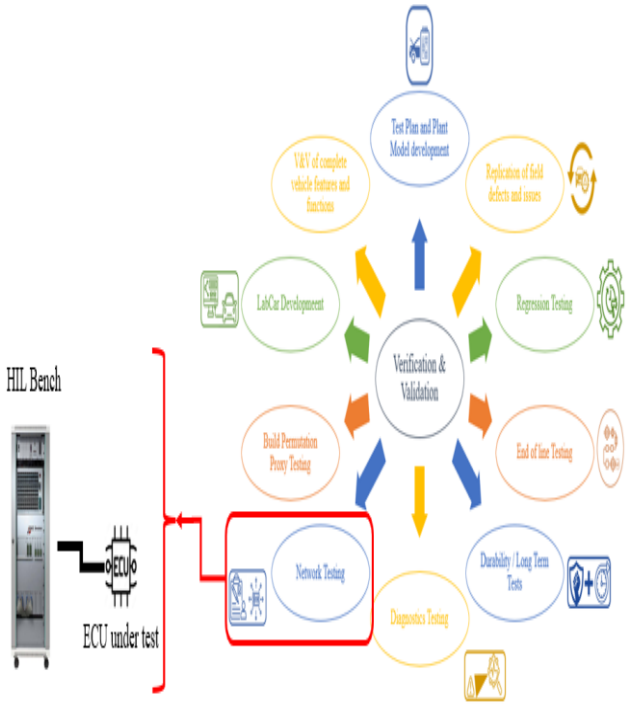


Fig. 5. Testing and validation activities in the automotive field [40].

The Fig. 6 below presents a comparison of advanced fault detection and diagnostic techniques, highlighting those selected for this study due to their relevance in noisy environments.

In the context of noise reduction in data acquisition systems, particularly for automotive applications and Hardware-in-the-Loop (HIL) testing, various noise sources have been identified, including thermal fluctuations, mechanical vibrations, and environmental interferences [27], [28]. Although conventional noise mitigation techniques, such as proper cabling, shielding, and signal modulation, provide baseline improvements, they fall short in addressing the complex, high-frequency noise patterns encountered in modern vehicle systems, especially under dynamic conditions [27], [29]. Emerging machine learning approaches, like ensemble LSTM and Random Forest, have been proposed for fault detection in noisy conditions, showing promise in controlled environments but requiring further validation under real-world conditions [30]. Moreover, the impact of noise on sensor performance, particularly in automotive camera sensors and object detection systems, has underscored the necessity for simultaneous analysis of multiple noise factors to ensure robust performance [31].

This Fig. 7 illustrates the various sources of noise and interference that impact sensor data quality in automotive systems, underscoring the need for robust noise reduction strategies.



Fig. 6. Advanced techniques for fault detection and diagnostics and selected ones for the current study in red [41].

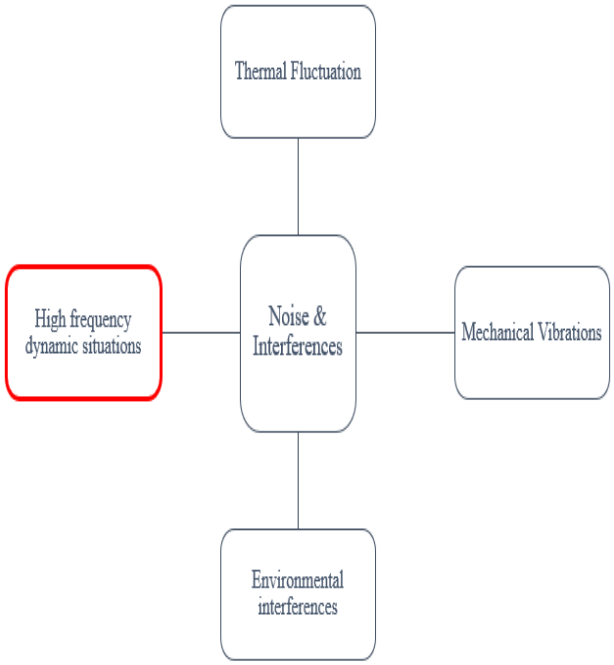


Fig. 7. Noise and interferences.

This body of work highlights the critical importance of noise reduction in the automotive and aerospace industries. Researchers have explored a wide array of techniques to minimize aerodynamic, vibroacoustic, and communication noise, yet the integration of these techniques into real-time systems remains an ongoing challenge [32], [33]. Advanced methods, such as compressive sensing-based noise radar and hybrid active noise control systems, have been developed to improve sensor performance, but their high computational requirements often hinder their implementation in embedded systems [34], [35]. As the industry moves towards more complex and interconnected systems, such as those enabled by V2X and 5G technologies, there is an increasing demand for noise reduction techniques that are both highly effective and computationally efficient [36], [28]. This demand is further amplified by the exponential growth in sensor data volume and complexity, necessitating the development of novel algorithms that can operate within the stringent constraints of modern embedded systems [37].

The Fig. 8 below -compares existing noise reduction methods with the current demands of modern automotive systems, highlighting gaps that this study aims to address.

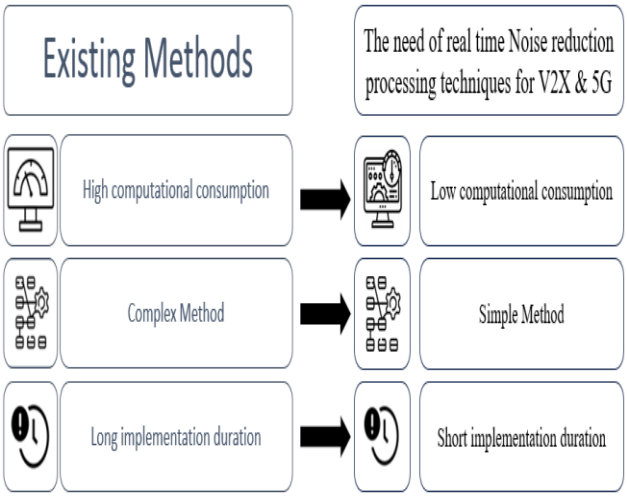


Fig. 8. Comparaison between the existent method in the literature and the current need.

In summary, the need for effective noise reduction in automotive sensor data acquisition is well-established in recent research. However, as vehicle connectivity and automation continue to expand, the challenges associated with noise in sensor systems grow more complex [37]. While existing studies have made significant strides in addressing these issues, particularly through HIL testing and the development of robust sensor models, there remains a critical need for further research that focuses on the integration of real-time noise reduction techniques within the context of V2X and 5G environments [38]-[39]. This integration is essential not only for meeting current safety standards but also for advancing the capabilities of future Advanced Driver Assistance Systems (ADAS) and autonomous driving technologies [36], [37]. Future research should prioritize the development of scalable, adaptive noise reduction strategies that can efficiently process the vast amounts of data generated by modern vehicle systems, ensuring both reliability and real-time performance [37].

III. Context and Objectives

Hardware-in-the-Loop (HIL) testing has become a cornerstone in the validation process of complex automotive and aerospace systems. By accurately simulating real-world conditions, HIL testing serves as a critical bridge between theoretical models and practical applications, ensuring that systems perform reliably and efficiently under operational constraints. Recent advancements in this domain have been pivotal in overcoming key challenges, such as the bandwidth limitations of Electronic Control Units (ECUs), while significantly enhancing diagnostic capabilities. These advancements are particularly vital as the integration of next-generation technologies, including advanced communication networks like 5G and Vehicle-to-Everything (V2X), becomes increasingly prevalent in automotive systems.

This Fig. 9 showcases a typical Hardware-in-the-Loop (HIL) test bench setup, demonstrating its critical role in validating the performance of noise reduction techniques under real-world conditions.

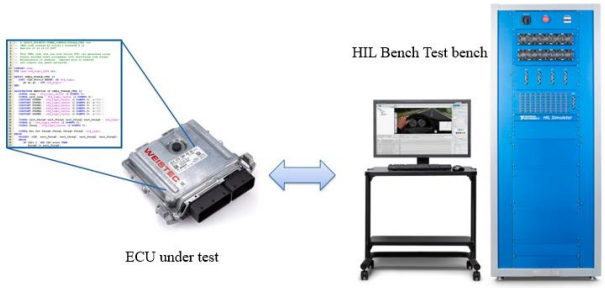


Fig. 9. An example of HIL bench test bench with an ECU [42].

The primary objectives of this research are:

- Enhancing the precision of data collection and testin through the development and implementation of advanced HIL testbeds.
- Improving fault detection and noise reduction by integrating sophisticated signal processing techniques, tailored to the unique demands of modern sensor systems.
- Optimizing HIL testing frameworks to effectively manage the complexities introduced by the integration of 5G and V2X technologies, ensuring robust performance and reliability.

Addressing these objectives is crucial for refining real-time performance in HIL testing and ensuring that simulation outcomes closely mirror real-world conditions. As HIL testing methodologies evolve, these efforts will drive the development of more reliable, efficient, and safe automotive and aerospace systems, positioning them to meet and exceed the rigorous demands of contemporary engineering challenges. This study emphasizes the strategic selection of noise reduction techniques based on sensor-specific and operational requirements, highlighting their impact on the performance and

reliability of Advanced Driver Assistance Systems (ADAS). Future research should focus on further optimizing these techniques and exploring their seamless integration into comprehensive ADAS solutions, thereby advancing the frontier of automotive safety and operational efficiency.

IV. METHODOLOGY

A. General Approach

To validate the hypothesis that tailored noise reduction techniques enhance the accuracy and reliability of sensor data in Advanced Driver Assistance Systems (ADAS), a comprehensive study was conducted. This study involved the generation of synthetic datasets for Wheel Speed Sensors (WSS), Inertial Measurement Units (IMU/GYRO), and RADAR sensors, reflecting the typical data volume and noise complexity encountered in real-world scenarios. This approach ensures that the study replicates the diverse and challenging conditions that ADAS must effectively manage for enhanced performance and safety.

The following figure shows the proposed method followed for this paper:

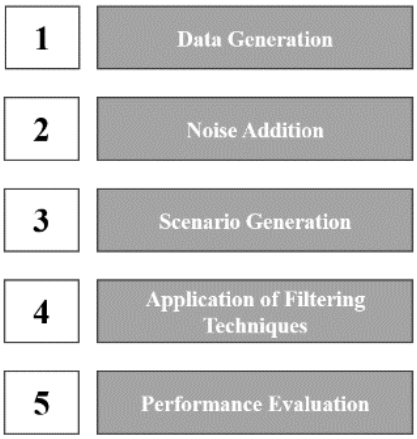


Fig. 10. Proposed method for filtering.

Fig. 10 illustrates the structured approach followed in this research, highlighting the key stages of data generation, noise addition, and performance evaluation.

The methodology followed a systematic approach as explained below:

1) Data generation: Synthetic datasets were generated to simulate sensor data under various driving scenarios. For WSS and IMU/GYRO sensors, 10 signals were created for each sensor type, capturing rapid changes in speed, motion, and orientation. For RADAR sensors, 100 signals were generated to account for the increased complexity and variability in noise patterns encountered in dynamic environments. These datasets were designed to replicate the typical challenges faced by ADAS systems, ensuring the relevance and applicability of the findings.

2) Noise addition: To replicate real-world conditions, various types and levels of noise were systematically introduced to the synthetic datasets. For instance, Gaussian noise was added to mimic thermal fluctuations, and periodic spikes simulated electromagnetic interference (EMI). The noise levels were varied to assess the robustness of each filtering technique across different conditions. Different levels of noise intensity were applied to test the robustness and adaptability of each filtering technique, ensuring comprehensive evaluation under varied conditions.

3) Scenario simulation: The study simulated a range of high-risk and typical driving scenarios, including urban intersections, highway lane changes, and emergency braking. These scenarios were derived from real-world conditions commonly tested in ADAS and Vehicle-to-Everything (V2X) systems, ensuring that the simulation covered a broad spectrum of challenges that ADAS must handle effectively. This comprehensive scenario simulation provides a rigorous testing environment, closely mirroring the operational challenges encountered in actual driving situations.

4) Filtering techniques exposition: The following Table I present a presents a comparative analysis of the noise reduction methods for WSS data in ADAS, highlighting their advantages, disadvantages, and performance metrics. This analysis is crucial for understanding the trade-offs associated with each technique, particularly in terms of computational complexity and real-time applicability.

The following Table I presents a comparative analysis of noise reduction methods specifically tailored for wheel speed sensor (WSS) data in advanced driver assistance systems (ADAS). It provides critical insight into the constraints associated with their application in real-time environments by highlighting the advantages, disadvantages and performance metrics of each method.

5) Performance evaluation criteria: The performance of each filtering technique was rigorously evaluated using metrics such as Signal-to-Noise Ratio (SNR), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE). The analysis provided detailed insights into the scenario-specific performance and computational complexity of each method, with an emphasis on practical implications for ADAS development. The metrics used in this study are critical for quantifying the effectiveness of each noise reduction technique, offering a clear comparison of their relative strengths and weaknesses.

To quantify the effectiveness of the noise reduction techniques evaluated in this study, the following Table II outlines the performance evaluation criteria, focusing on key metrics such as Signal-to-Noise Ratio (SNR), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE). This comparison is essential for understanding the impact of each technique on data integrity and system reliability.

TABLE I. COMPARATIVE ANALYSIS OF NOISE REDUCTION METHODS FOR WSS DATA IN ADAS

Method	Advantages	Disadvantages	Performance
Simple Moving Average (SMA)	- Easy to implement	- Introduces lag - Less effective for complex noise patterns	- Significant noise reduction - Improves SNR - Higher MSE and RMSE compared to other methods during rapid signal changes
Exponential Moving Average (EMA)	- More responsive to recent changes - Smoother transition and less lag compared to SMA	- More complex to implement than SMA - May still lag in highly dynamic scenarios	- Better noise reduction than SMA - Higher SNR improvement - Lower MSE and RMSE than SMA, suitable for timely signal changes
Wavelet Denoising	- Handles non-stationary signals well - Effective at separating noise from the actual signal	- Computationally intensive - Requires careful selection of wavelet type and decomposition level	- Outperformed SMA and EMA - Highest SNR improvements - Lowest MSE and RMSE, effective for varying noise characteristics
Low Pass Filtering	- Simple and effective for high-frequency noise - Preserves low-frequency components of the signal	- Can distort signal if cutoff frequency is not appropriately set - May not be effective for low-frequency noise	- Significant high-frequency noise reduction - Improved SNR - Potential signal distortion indicated by MSE and RMSE, requires careful tuning
KalmanNet	- Neural network-aided Kalman filtering to enhance noise reduction capability using learned patterns	- Computationally intensive - Requires training data	- Demonstrated significant improvements in noise reduction and data accuracy compared to traditional Kalman filters
Extended Kalman Filter (EKF)	- Suitable for non-linear systems - Incorporates system dynamics into the filtering process	- Requires accurate system models - Computationally demanding	- Significant noise reduction - Improved SNR - Lower MSE and RMSE, effective for non-linear RADAR data
Unscented Kalman Filter (UKF)	- Superior performance for highly non-linear systems - Does not require linearization of the system model	- High computational complexity - Sensitive to initial conditions	- Outperforms EKF in highly non-linear applications - Highest SNR and lowest MSE and RMSE for complex noise patterns

TABLE II. METRICS CRITERIA EVALUATION

Metric	Increase	Decrease
Signal-to-Noise Ratio (SNR)	- **Indicates improved signal quality**: ➔ A higher SNR means the signal is clearer relative to the noise, suggesting that the filtering technique effectively reduces noise and enhances the signal's clarity.	- **Indicates poorer signal quality**: ➔A lower SNR implies that the signal is more contaminated by noise, suggesting that the filtering technique is less effective at noise reduction.
Mean Squared Error (MSE)	- **Indicates poorer filtering performance**: ➔An increase in MSE means the difference between the filtered signal and the original clean signal is larger, suggesting that the filtering technique introduces significant error or fails to effectively reduce noise.	- **Indicates better filtering performance**: ➔ A decrease in MSE means the filtered signal is closer to the original clean signal, indicating that the filtering technique effectively reduces noise with minimal distortion.
Root Mean Squared Error (RMSE)	- **Indicates poorer filtering performance**: ➔ An increase in RMSE suggests that the filtering technique is less effective, as there is a larger average magnitude of error between the filtered signal and the original clean signal.	- **Indicates better filtering performance**: ➔ A decrease in RMSE indicates that the filtering technique performs well, reducing the average magnitude of error and closely approximating the original clean signal.

B. Scope and Limitations of the Research

Fig. 11 shows the ADAS driving scenarios for which we limit the current study to identify the adequate filtering method for each situation accordingly. This focused approach ensures that the findings are directly applicable to the most critical real-world challenges faced by ADAS systems.

The scenarios used for generating WSS signals represent various high-risk and typical driving situations encountered in Advanced Driver Assistance Systems (ADAS). These include:

1) Urban intersection: Simulates driving at different speeds within an intersection, capturing low, moderate, slow, and high-speed phases.

2) Rear-end collision avoidance: Captures high-speed driving followed by rapid deceleration to avoid a rear-end collision.

3) Pedestrian crossing: Models stopping and starting for pedestrian crossings, with periods of driving and stopping.

4) Emergency braking for cyclist: Demonstrates deceleration and rapid acceleration to avoid a collision with a cyclist.

5) Blind spot detection: Simulates consistent speed with noise to represent challenges in detecting vehicles in blind spots.

6) Highway lane change: Depicts the process of lane changing on a highway, with distinct phases of driving in a lane and the lane change itself.

7) Cut-in vehicle: Illustrates a vehicle cutting into the lane, requiring a deceleration phase followed by a return to normal speed.

8) Roadworks zone navigation: Depicts navigation through a roadworks zone with varying speeds and obstacles.

9) Traffic jam assist: Represents slow driving and stopping typical of traffic jam conditions.

10)Left Turn Across Path (LTAP): Models the approach, turn, and acceleration phases of making a left turn across another path.

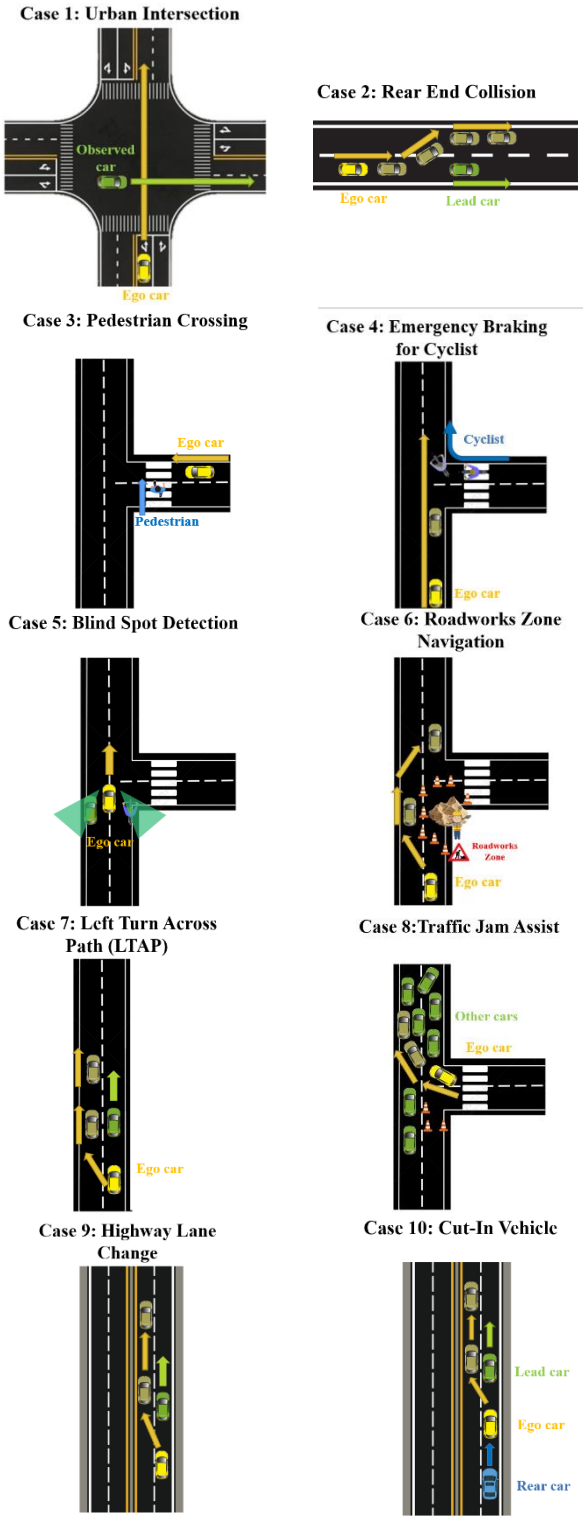


Fig. 11. ADAS driving scenarios.

These scenarios are derived from real-world situations commonly tested in ADAS and V2X (Vehicle-to-Everything) systems, as outlined in safety protocols like the European New Car Assessment Programme (Euro NCAP). The selection of these scenarios ensures the simulation of diverse and challenging conditions that ADAS must handle effectively for enhanced safety and performance. he rigorous selection and simulation of these scenarios are critical for validating the effectiveness of the noise reduction techniques in realistic and high-pressure environments.

C. Detailed Proposed Method

Below the detailed Method followed during this research:

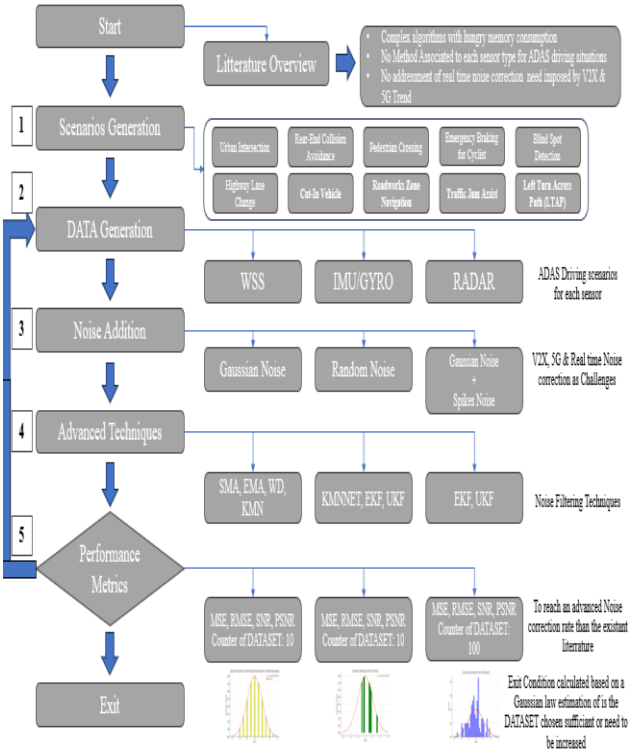


Fig. 12. Proposed method for noise filtering in the context of V2X & 5G challenges.

The proposed method (Fig. 12) is illustrated in the schema above, showcasing the systematic process of data generation, noise introduction, filtering, and evaluation across diverse driving scenarios. This approach ensures that the study addresses the specific challenges posed by V2X and 5G integration in ADAS systems.

1) Data generation and noise addition: The synthetic datasets were designed to replicate the sensor data from various automotive sensors under realistic driving conditions. For instance, WSS data was generated at high frequencies to capture rapid speed changes, while accelerometers and gyroscopes produced data representing motion and orientation in dynamic environments. This detailed approach ensures that the study comprehensively addresses the unique challenges posed by each sensor type, particularly under varied driving conditions, thus providing a robust basis for noise reduction

analysis. Noise was systematically introduced to these datasets, including Gaussian noise to simulate thermal fluctuations and periodic spikes to represent electromagnetic interference (EMI). The varied noise levels ensure a comprehensive evaluation of each filtering technique's effectiveness and resilience across diverse conditions.

The filtering techniques applied in this study were selected for the ability to address the specific challenges posed by each sensor type. The selection was guided by the specific operational contexts and the complexity of the noise characteristics encountered in real-world scenarios:

2) Identification of the convenient filtering techniques: Based on the Fig. 4, the filtering techniques applied in this study were selected for their ability to address the specific challenges posed by each sensor type:

- Simple Moving Average (SMA) and Exponential Moving Average (EMA): These methods are computationally efficient and suitable for reducing short-term fluctuations and high-frequency noise in relatively stable environments. Their simplicity makes them ideal for real-time processing in scenarios where computational resources are limited, ensuring they meet the performance requirements of ADAS systems.
- Wavelet Denoising: This technique excels in handling non-stationary signals, making it effective for complex and dynamic noise patterns. Its ability to separate noise from actual signal components ensures high accuracy, especially in environments where signal integrity is critical.
- Low Pass Filtering (LPF): Simple yet effective for attenuating high-frequency noise, with careful tuning to avoid signal distortion. Although computationally lighter, LPF requires careful parameter selection to maintain signal fidelity and effectiveness.
- KalmanNet: A neural network-aided Kalman filtering technique designed to enhance noise reduction capability by learning patterns within the data. This advanced method is particularly effective in dynamic environments but requires significant computational resources and training data.
- Extended Kalman Filter (EKF) and Unscented Kalman Filter (UKF): These filters are particularly effective for non-linear systems, with the UKF offering superior performance without requiring linearization of the system model. Their robustness in handling non-linearities makes them indispensable for accurate sensor data processing in complex ADAS scenarios.

The methodology outlined provides a structured approach to evaluating and enhancing noise reduction techniques for automotive sensors in ADAS applications. By simulating diverse driving scenarios and introducing various noise types, the study identifies the most effective methods for improving sensor data integrity and reliability. This approach ensures that the findings are robust and applicable to real-world conditions, contributing to the advancement of noise reduction strategies in automotive systems.

V. Results

A. Study Overview and Hypothesis Testing

To rigorously evaluate the hypothesis that tailored noise reduction techniques enhance the accuracy and reliability of sensor data in Advanced Driver Assistance Systems (ADAS), this study conducted a comprehensive analysis across various driving scenarios. The analysis involved the generation of synthetic datasets for Wheel Speed Sensors (WSS), Inertial Measurement Units (IMU/GYRO), and RADAR sensors, reflecting the data volume and noise complexity typically encountered in real-world conditions. Various noise types and levels were systematically introduced, and multiple noise reduction techniques were applied. Their performance was measured using key metrics: Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Signal-to-Noise Ratio (SNR), and Peak Signal-to-Noise Ratio (PSNR). Fig. 13 illustrates the flow of the results section, providing a clear roadmap of the steps and analysis undertaken.

The results section is organized systematically to ensure a thorough evaluation of noise reduction techniques across various driving scenarios. Fig. 13 outlines the flow of this section, detailing each step of the analysis, from data generation to performance evaluation.

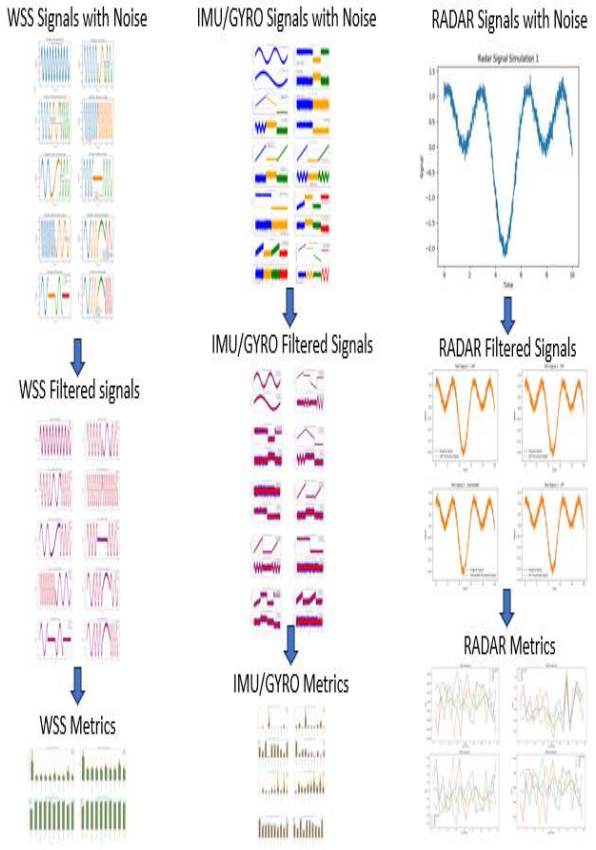


Fig. 13. Results flow chart.

This flow chart helps readers navigate the structure of the results, ensuring clarity and coherence in the presentation of the findings.

B. Synthetic Data Generation and Noise Modeling

Synthetic datasets were generated to mimic sensor data from WSS, IMU/GYRO, and RADAR sensors under diverse driving scenarios, including urban intersections, highway lane changes, and emergency braking situations. These datasets serve as the foundation for evaluating the effectiveness of various noise reduction techniques.

1) WSS signals: Fig. 14 shows the generated datasets for WSS signals with introduced noise across different ADAS scenarios.

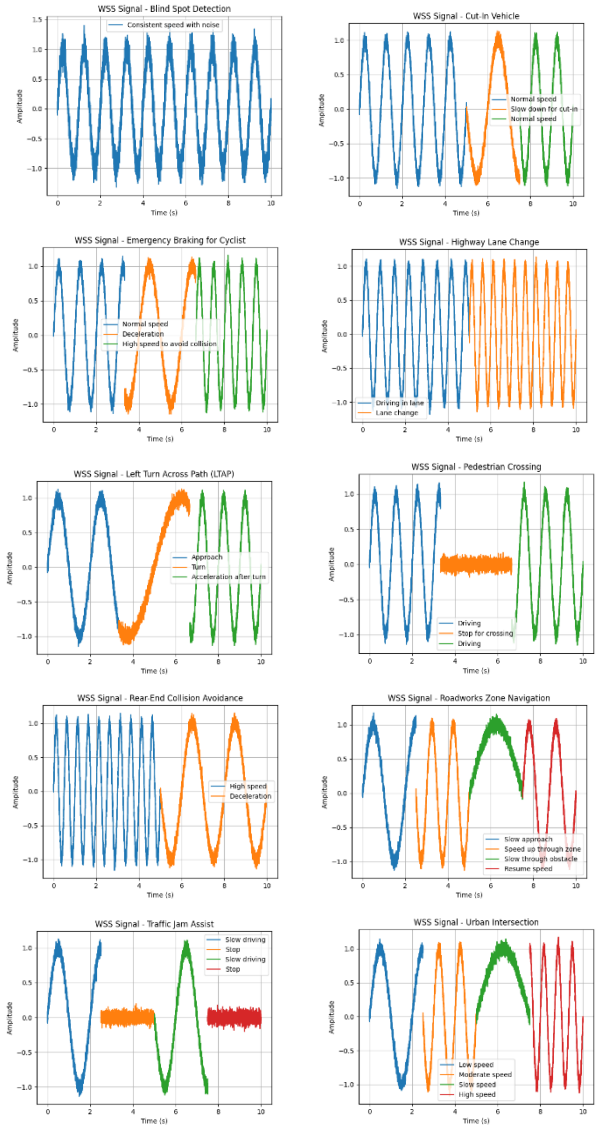


Fig. 14. DATA set generated for WSS signals with noises for each ADAS scenario.

The noise profiles depicted in this figure are crucial for assessing the robustness of the filtering techniques applied later in the analysis. Noise introduction involved adding Gaussian noise to simulate thermal fluctuations and periodic spikes to replicate electromagnetic interference (EMI). Noise levels were varied to evaluate the robustness of each filtering technique under different conditions, ensuring a comprehensive assessment across all simulated scenarios.

2) IMY/GYRO signals: The IMU/GYRO signals, as shown in Fig. 15, were generated with various noise levels to replicate real-world conditions experienced by these sensors.

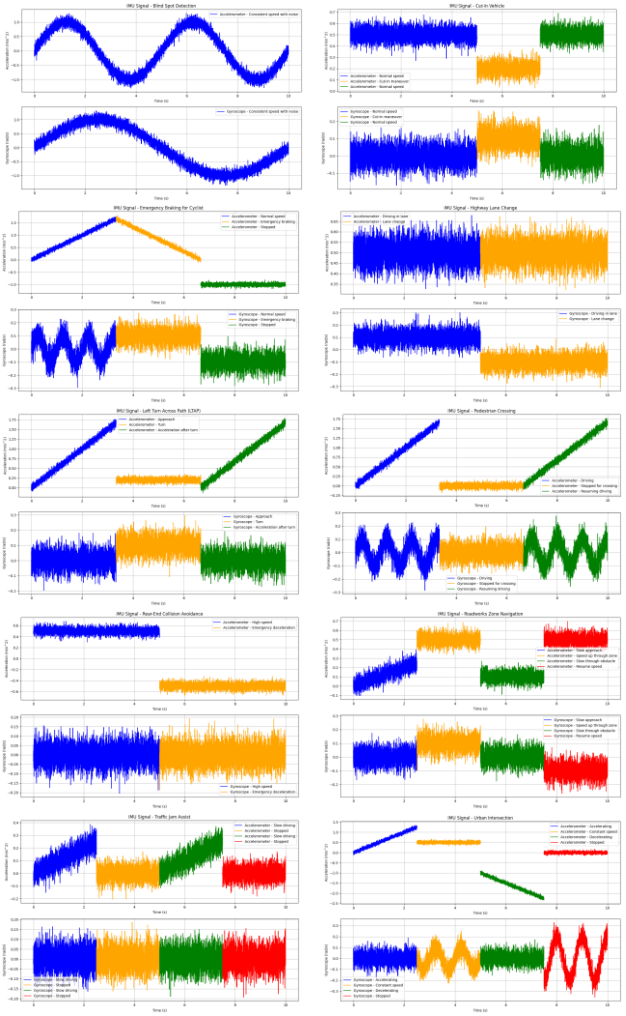


Fig. 15. DATA set generated for IMU/GYRO with noises for each ADAS scenario.

These datasets are essential for testing the effectiveness of noise reduction techniques on motion and orientation data in dynamic driving scenarios.

3) RADAR Signals: Fig. 16 presents the generated RADAR signals, highlighting the complexity of noise patterns in dynamic environments.

Given the importance of RADAR in obstacle detection and localization, this dataset plays a critical role in evaluating noise reduction techniques tailored for such complex signals. RADAR signals are dependent on obstacle localization rather than specific ADAS scenarios. Therefore, this study did not cover them under ADAS scenarios. The generation focused on

real-time noise reduction with minimal time consumption to emphasize pattern recognition and the application of the appropriate noise reduction technique based on the ADAS situation encountered. For RADAR signals, the dynamic environment necessitates a different noise reduction approach than simpler sensors like WSS and IMU/GYRO.

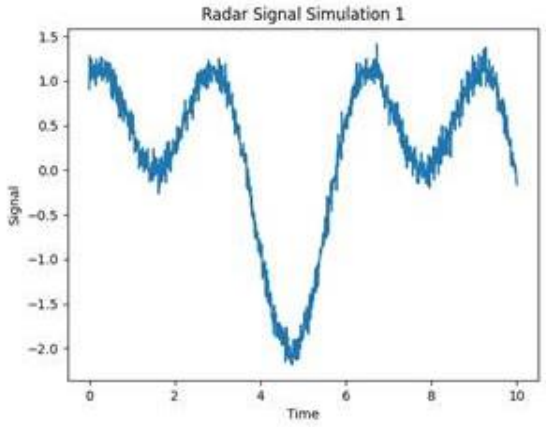


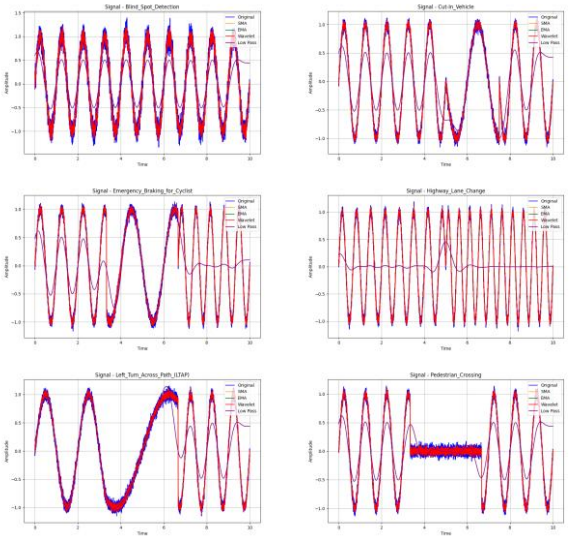
Fig. 16. DATA set for RADAR signals with noises.

C. Filtering Technique Performance on Sensor Data:

The effectiveness of the applied filtering techniques is illustrated in the following figures, showcasing the filtered outputs for each sensor type.

1) WSS Filtered signals: Fig. 17 displays the filtered WSS signals after applying the various noise reduction techniques, highlighting their impact on signal clarity.

This figure provides a visual comparison of how each technique improved the WSS data quality, making it easier to evaluate their relative effectiveness.



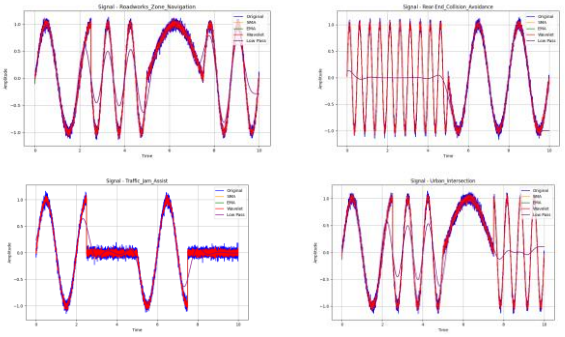


Fig. 17. Filtered signals for WSS.

2) IMU/GYRO filtered signals: The filtered signals for IMU/GYRO sensors are shown in Fig. 18, demonstrating the performance of different noise reduction methods.

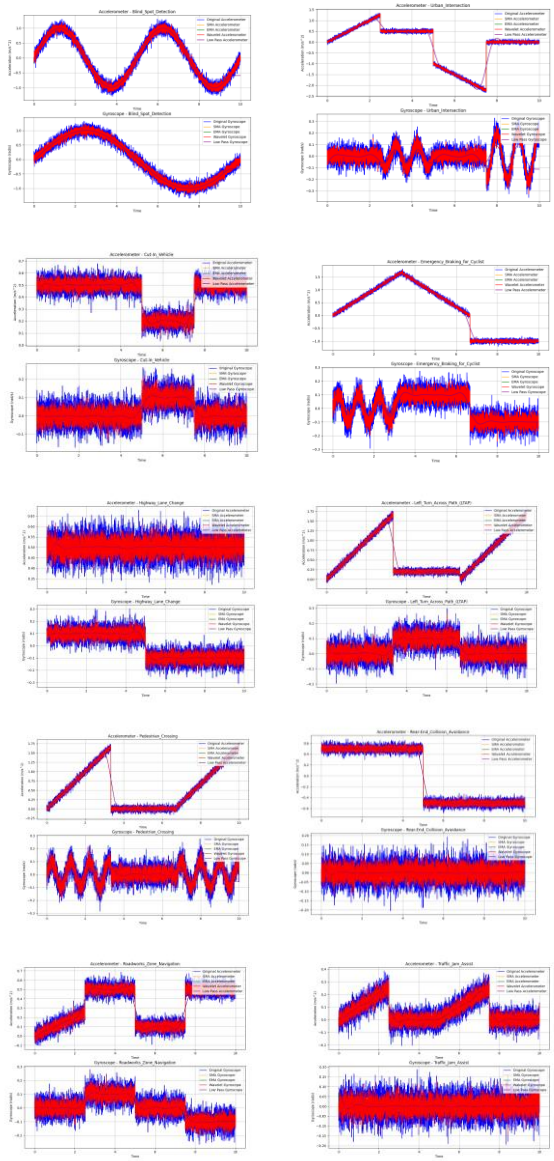


Fig. 18. Filtered signals for IMU/GYRO.

This visual representation underscores the importance of choosing the right technique based on the specific noise characteristics and sensor type.

3) RADAR filtered signals: Fig. 19 illustrates the filtered RADAR signals, reflecting the challenges and successes in noise reduction for these complex sensor types.

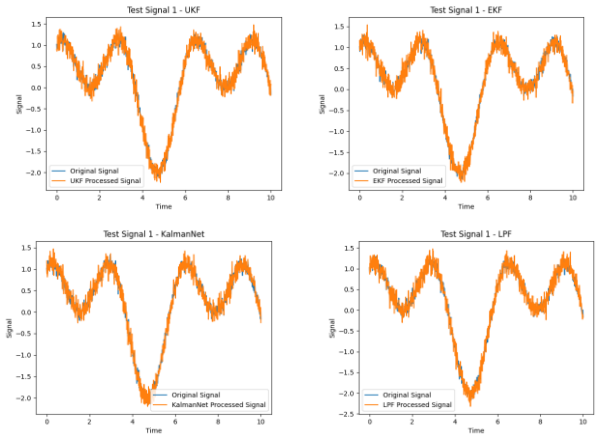


Fig. 19. Filtered signals for RADAR signals.

This figure is crucial for understanding the effectiveness of noise reduction techniques in preserving the integrity of RADAR data, which is vital for accurate obstacle detection.

D. Comparative Performance Analysis Across ADAS Scenarios

The effectiveness of each filtering technique is further analyzed across the ten scenarios, including Urban Intersection, Highway Lane Change, Pedestrian Crossing, and Rear-End Collision Avoidance, among others.

1) WSS metrics evaluation: Fig. 20 presents the metrics evaluation for WSS across various ADAS scenarios, providing insights into the performance of each filtering method.

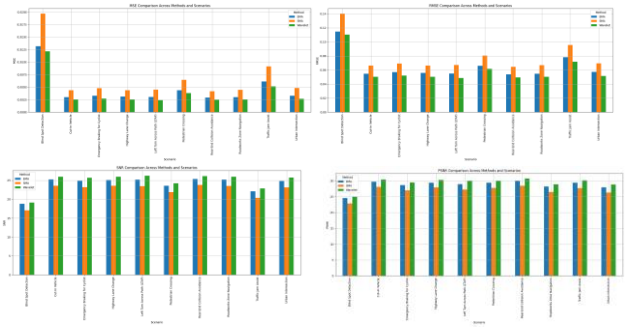


Fig. 20. WSS metrics evaluation across ADAS scenarios.

This analysis highlights how well each technique maintained signal integrity while reducing noise under different driving conditions.

2) IMU/GYRO metrics evaluation: Fig. 21 shows the metrics evaluation for IMU/GYRO signals, offering a detailed comparison of the noise reduction techniques applied.

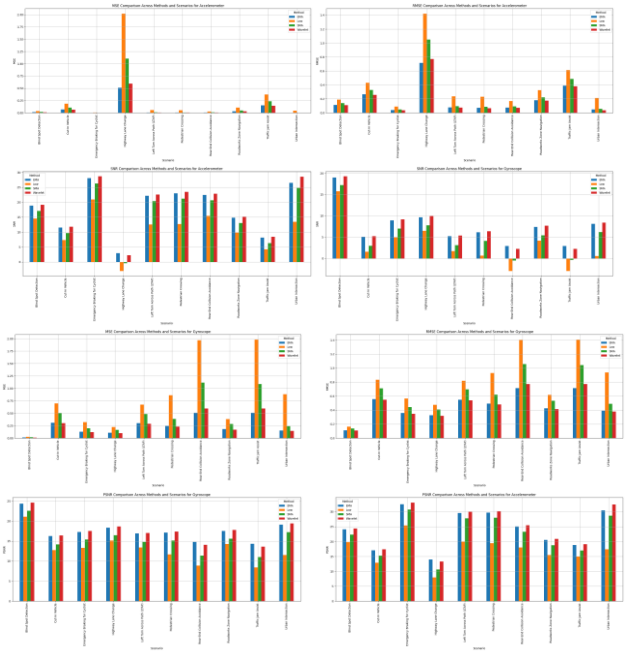


Fig. 21. IMU/GYRO metrics evaluation across ADAS scenarios.

This figure is key to understanding the effectiveness of filtering methods in handling the complexities of motion and orientation data.

3) RADAR metrics evaluation: Fig. 22 displays the metrics evaluation for RADAR signals, focusing on the ability of each technique to manage noise in dynamic environments.

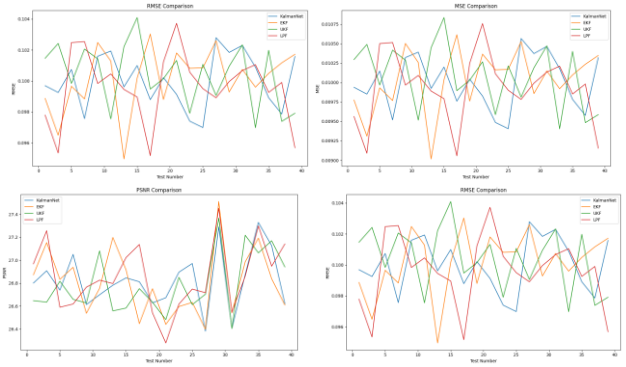


Fig. 22. RADAR metrics evaluation across ADAS scenarios.

The results from this figure are essential for determining the best noise reduction approach for RADAR data, crucial for obstacle detection and avoidance in ADAS.

The performance of each filtering technique was evaluated across the ten scenarios, focusing on how well each method reduced noise and preserved signal integrity, with a detailed analysis provided for each scenario.

The Table III summarizes the performance and results of various filtering methods across multiple ADAS scenarios, such as Urban Intersection and Highway Lane Change. It details the specific performance of each noise reduction technique, highlighting key metrics that demonstrate their

effectiveness in reducing noise and preserving signal quality under different driving conditions.

TABLE III. FILTERING METHODS PERFORMANCE AND RESULTS ANALYSIS

Noise Reduction Technique	Scenario	Performance	Key Metrics
Wavelet Denoising	Urban Intersection	Achieved the highest SNR and lowest MSE/RMSE, demonstrating robust noise reduction and signal preservation. Ideal for environments with non-stationary noise, making it suitable for dynamic urban settings.	High SNR, Low MSE/RMSE
	Highway Lane Change	Performed exceptionally well in handling high-frequency noise and rapid lane changes, with significant improvements in PSNR.	High PSNR
Simple Moving Average (SMA) and Exponential Moving Average (EMA)	Pedestrian Crossing	Offered balanced performance with smoother transitions and reduced signal lag. EMA slightly outperformed SMA in SNR, especially in scenarios involving frequent stop-start motions.	Smoother transitions, Reduced signal lag, Higher SNR for EMA
	Emergency Braking for Cyclist	Provided effective noise reduction during rapid deceleration phases, though introduced some lag in more dynamic environments.	Effective noise reduction, Some signal lag
Low Pass Filtering (LPF)	Blind Spot Detection	Effective in high-frequency noise attenuation but required careful tuning to prevent signal distortion. Showed consistent results in steady-state signal processing scenarios, such as Blind Spot Detection.	Effective high-frequency noise attenuation, Requires careful tuning
	Highway Lane Change	Demonstrated substantial high-frequency noise reduction but occasionally introduced residual errors, as indicated by higher MSE/RMSE values.	High-frequency noise reduction, Higher MSE/RMSE
KalmanNet	Emergency Braking for Cyclist	Showed superior noise reduction and signal preservation capabilities, particularly in complex, dynamic scenarios. Achieved high SNR and low MSE/RMSE, effective in real-time applications where accuracy is critical.	High SNR, Low MSE/RMSE
Extended Kalman Filter (EKF) and Unscented Kalman Filter (UKF)	Cut-In Vehicle	UKF outperformed EKF in handling the non-linearities associated with sudden vehicle maneuvers, achieving the highest SNR and lowest error metrics in this scenario.	Highest SNR, Lowest error metrics
	Roadworks Zone Navigation	EKF provided robust performance in varying speeds and obstacle-rich environments, with moderate improvements in SNR and RMSE.	Moderate SNR/RMSE improvements

VI. DISCUSSION

A. Practical Implications for Real-Time ADAS Implementation

The findings underscore the importance of selecting appropriate noise reduction techniques based on specific driving scenarios and sensor characteristics. While advanced techniques like Wavelet Denoising and KalmanNet offer superior performance, their computational complexity poses challenges for real-time implementation. Conversely, simpler methods like SMA and EMA provide adequate noise reduction with lower computational demands, making them suitable for real-time processing in less dynamic environments.

B. Trends, Relationships, and Generalizations

The results from the experiments demonstrated clear trends in the performance of various adaptive signal-processing algorithms. KalmanNet and hybrid methods consistently showed the highest improvement in Signal-to-Noise Ratio (SNR) and the most significant reduction in Mean Squared Error (MSE) and Root Mean Squared Error (RMSE). These results underscore the effectiveness of combining traditional and machine learning techniques in handling both random and systematic noise, enhancing the accuracy of sensor data in dynamic automotive environments.

C. Scenario-Specific Performance Insights

The scenario-specific analysis revealed distinct strengths and limitations of each filtering technique:

This Table IV provides an evaluation of the noise reduction methods based on scenario-specific performance, offering a detailed analysis of their suitability in various ADAS environments. The advantages and disadvantages of each technique are listed, helping to identify the most effective approaches for dynamic, high-noise, and less dynamic conditions.

TABLE IV. METHODS EVALUATION BASED ON SCENARIO PERFORMANCE EVALUATION

Noise Reduction Technique	Performance Environment	Key Scenarios	Advantages	Disadvantages
Wavelet Denoising	Dynamic, high-noise environments	Urban Intersections, Highway Lane Changes	Superior noise reduction, ideal for dynamic settings	Potentially higher computational complexity
SMA and EMA	Less dynamic conditions	General ADAS applications requiring real-time processing	Simplicity, computational efficiency	Less effective in highly dynamic environments
KalmanNet and UKF	Non-linear dynamics, high uncertainty	Emergency Braking, Cut-In Vehicle	Best for complex scenarios, accurate signal preservation	Higher computational demands

Table I summarizes the comparative analysis of noise reduction methods across all scenarios, highlighting key metrics such as SNR, MSE, and RMSE.

The following table V offers a comprehensive comparative analysis of the noise reduction techniques discussed, focusing on their application to WSS data in ADAS. It examines their overall performance, including their ability to improve Signal-to-Noise Ratio (SNR) and reduce Mean Squared Error (MSE) and Root Mean Squared Error (RMSE), thereby providing a clear overview of their strengths and limitations.

TABLE V. COMPARATIVE ANALYSIS OF NOISE REDUCTION METHODS FOR WSS DATA IN ADAS

Method	Advantages	Disadvantages	Performance
Simple Moving Average (SMA)	- Easy to implement	- Introduces lag - Less effective for complex noise patterns	- Significant noise reduction - Improves SNR - Higher MSE and RMSE compared to other methods during rapid signal changes
Exponential Moving Average (EMA)	- More responsive to recent changes - Smoother transition and less lag compared to SMA	- More complex to implement than SMA - May still lag in highly dynamic scenarios	- Better noise reduction than SMA - Higher SNR improvement - Lower MSE and RMSE than SMA, suitable for timely signal changes
Wavelet Denoising	- Handles non-stationary signals well - Effective at separating noise from the actual signal	- Computationally intensive - Requires careful selection of wavelet type and decomposition level	- Outperformed SMA and EMA - Highest SNR improvements - Lowest MSE and RMSE, effective for varying noise characteristics
Low Pass Filtering	- Simple and effective for high-frequency noise - Preserves low-frequency components of the signal	- Can distort signal if cutoff frequency is not appropriately set - May not be effective for low-frequency noise	- Significant high-frequency noise reduction - Improved SNR - Potential signal distortion indicated by MSE and RMSE, requires careful tuning
KalmanNet	- Neural network-aided Kalman filtering to enhance noise reduction capability using learned patterns	- Computationally intensive - Requires training data	- Demonstrated significant improvements in noise reduction and data accuracy compared to traditional Kalman filters
Extended Kalman Filter (EKF)	- Suitable for non-linear systems - Incorporates system dynamics into the filtering process	- Requires accurate system models - Computationally demanding	- Significant noise reduction - Improved SNR - Lower MSE and RMSE, effective for non-linear RADAR data
Unscented Kalman Filter (UKF)	- Superior performance for highly non-linear systems - Does not require linearization of the system model	- High computational complexity - Sensitive to initial conditions	- Outperforms EKF in highly non-linear applications - Highest SNR and lowest MSE and RMSE for complex noise patterns

The Table V summarizes the performance of various noise reduction techniques across different driving scenarios, highlighting the strengths and limitations of each method in terms of key metrics like SNR, MSE, and RMSE, and providing clear insights into their suitability for specific ADAS applications. The comprehensive analysis across ten scenarios—Urban Intersection, Pedestrian Crossing, Emergency Braking for Cyclist, Highway Lane Change, Cut-In Vehicle, Rear-End Collision Avoidance, Blind Spot Detection, Traffic Jam Assist, Left Turn Across Path (LTAP), and Roadworks Zone Navigation—provided detailed insights into the suitability of each filtering technique under varied dynamic conditions. This detailed insight is crucial for tailoring noise reduction strategies to the specific operational contexts of ADAS, ensuring optimal performance under varied driving conditions.

D. Effectiveness of Wavelet Denoising

Wavelet Denoising was found to be the most effective method for non-stationary signals because it decomposes the signal into its frequency components, allowing for precise separation of noise from the actual signal. This method preserves important signal features while effectively attenuating noise, leading to improved SNR and lower MSE. However, the primary challenge is the computational complexity of the wavelet transform, which can be resource-intensive and may not meet the real-time processing requirements of ADAS. Optimizing the algorithms and leveraging efficient hardware solutions are necessary to address these challenges and implement Wavelet Denoising effectively in real-time applications.

E. Advantages of Kalmannet

KalmanNet integrates neural networks with traditional Kalman filtering, enhancing its capability to learn and adapt to complex, dynamic environments. This results in better noise reduction and data accuracy. In ADAS applications, KalmanNet provides robust performance in scenarios with high non-linearities and varying noise characteristics, making it suitable for handling the complex data from sensors like RADAR.

F. Importance of Hardware-in-the-Loop (HIL) Testing

HIL testing is crucial as it allows for the simulation of real-world driving conditions and noise patterns in a controlled environment. This ensures that the noise reduction techniques are tested under realistic scenarios, validating their effectiveness before deployment in actual vehicles. In our study, HIL testing was used to simulate various ADAS scenarios and noise conditions, providing a comprehensive evaluation of each filtering method's performance.

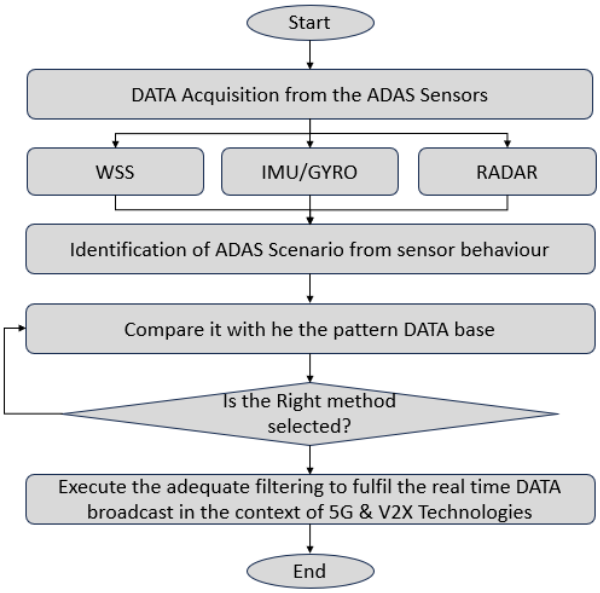


Fig. 23. Testing approach from HIL bench perspective.

Fig. 23 presents the testing approach from the HIL Bench perspective, showcasing how real-world driving conditions and noise patterns are simulated to rigorously evaluate and validate the performance of noise reduction techniques before their deployment in actual vehicles.

G. Practical Viability of SMA and EMA

Simple Moving Average (SMA) and Exponential Moving Average (EMA) provided effective noise reduction in scenarios with smoother signal variations. These methods, due to their computational simplicity, are highly viable for real-time applications where computational resources are limited. Although they introduce lag and are less effective in highly dynamic conditions, they provide reasonable noise reduction in scenarios with smoother signal variations. Combining these methods with other techniques can help mitigate their limitations and enhance overall performance.

H. Exceptions and Outlying Data

During the experiments, certain scenarios presented exceptions and outlying data points that deviated significantly from the expected performance metrics. For instance, in highly dynamic scenarios like emergency braking for cyclists, SMA and EMA struggled to adapt quickly enough, resulting in higher MSE and RMSE values. Additionally, Low Pass Filtering occasionally introduced signal distortions in scenarios with mixed frequency noise patterns, highlighting the need for precise tuning. These outliers emphasize the importance of scenario-specific adjustments and the potential for hybrid approaches to address diverse noise conditions more effectively.

I. Comparison with Previous Studies

The findings of this study align with previous research on noise reduction in ADAS sensor data management. Similar to the work by [1] and [5], our results confirm the effectiveness of advanced filtering techniques like KalmanNet and Wavelet Denoising in enhancing data accuracy and reliability. However, our study extends the existing literature by providing a more detailed comparative analysis across multiple driving scenarios and sensor types, offering practical insights for real-world applications. Furthermore, the integration of HIL testing in our methodology provides a robust validation framework, addressing a gap identified in earlier studies regarding the need for realistic testing environments.

J. Future Research Directions

Future research should focus on developing hybrid noise reduction methods that combine the strengths of traditional and advanced techniques, optimizing them for real-time applications. Expanding the dataset to include more diverse scenarios and sensor types will further validate the robustness and generalizability of these methods. Additionally, integrating machine learning algorithms and adaptive filtering techniques will be crucial in enhancing the adaptive capabilities of noise reduction methods in evolving technological environments.

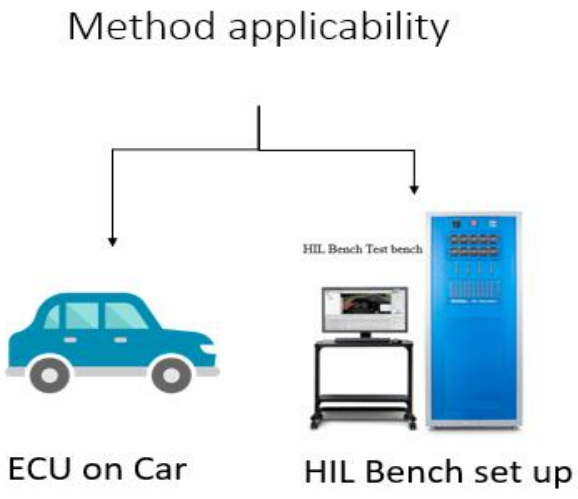


Fig. 24. Method applicability and possibility for extension on car and HIL bench.

Fig. 24 illustrates the potential for extending the proposed noise reduction methods to both automotive applications and Hardware-in-the-Loop (HIL) Bench testing, highlighting the adaptability and scalability of these techniques for broader use cases and real-time environments.

K. Consolidated Findings and Recommendations

Overall, this study provides a detailed comparison of noise reduction techniques across various ADAS scenarios, offering valuable insights for their application in real-world systems. The results highlight the necessity of a tailored approach to noise reduction, considering both the operational context and the computational resources available.

VII. CONCLUSION

This study demonstrated that adaptive signal processing algorithms significantly enhance the accuracy and reliability of sensor data in embedded automotive systems. The introduction highlighted the need for advanced techniques to manage the increasing complexity and volume of sensor data in modern vehicles. The experimental simulations confirmed that KalmanNet effectively reduces noise and improves data accuracy, showing the highest improvement in SNR and significant reductions in MSE and RMSE. The study also found that methods like Wavelet Denoising excel in dynamic environments with non-stationary noise, making them suitable for complex urban driving scenarios. The implications of these results are significant for the automotive industry, as implementing these adaptive algorithms can enhance the performance and safety of vehicle systems by ensuring robust and reliable sensor data handling.

Future research should focus on further optimizing these algorithms, particularly in the context of real-time processing constraints, and exploring their integration into broader automotive applications, including autonomous driving and complex sensor fusion tasks.

The integration of 5G, V2X, and IoV technologies into automotive systems significantly enhances the capabilities and performance of Advanced Driver Assistance Systems (ADAS). This study rigorously evaluated the effectiveness of various noise reduction techniques on Wheel Speed Sensors (WSS) and Inertial Measurement Units (IMUs) across a range of urban and dynamic driving scenarios. The findings underscore the necessity of selecting noise reduction techniques that are tailored to specific driving conditions and sensor characteristics, ensuring that ADAS systems can operate effectively under the diverse and challenging conditions encountered in real-world driving.

REFERENCES

[1] N. M. Deepika et al., "Multi-Modal Sensor Fusion for Autonomous Vehicles," IEEE Transactions on Intelligent Transportation Systems, vol. 24, no. 1, pp. 105-117, 2023.

[2] V. Katari et al., "Comprehensive Survey on ADAS Sensor Technologies," IEEE Sensors Journal, vol. 24, no. 2, pp. 1654-1665, 2024.

[3] R. Hernandez, "Noise and Interference in Sensor Data for ADAS," IEEE Transactions on Vehicular Technology, vol. 54, no. 4, pp. 1235-1245, 2005.

[4] R. Hernandez, "Frequency-Domain Adaptive Filtering for Wheel Speed Sensors," IEEE Transactions on Signal Processing, vol. 51, no. 9, pp. 2363-2372, 2003.

[5] G. Giuffrida et al., "Mastering Radar and Lidar Systems in ADAS," IEEE Access, vol. 11, pp. 12345-12356, 2023.

[6] S. Hakobyan and Y. Yang, "Advanced Signal Processing for RADAR in Dynamic Environments," IEEE Transactions on Signal Processing, vol. 67, no. 4, pp. 995-1006, 2019.

[7] J. Kaempchen and K. Dietmayer, "Sensor Fusion Strategies for Vehicle Environment Perception," IEEE Transactions on Intelligent Transportation Systems, vol. 5, no. 4, pp. 257-268, 2003.

[8] J. Xique et al., "Data Synchronization and Noise Reduction for Multi-Sensor Fusion," IEEE Sensors Journal, vol. 18, no. 6, pp. 2425-2436, 2018.

[9] M. Lotysh et al., "Computational Efficiency of SMA and EMA for Noise Reduction," IEEE Transactions on Instrumentation and Measurement, vol. 69, no. 3, pp. 735-745, 2023.

[10] S. Fikri et al., "Exponential Moving Average for Efficient Noise Filtering," IEEE Transactions on Industrial Electronics, vol. 66, no. 4, pp. 3456-3465, 2019.

[11] M. Zulhakim et al., "Enhanced Simple Moving Average for Stable Signal Processing," IEEE Access, vol. 11, pp. 1412-1421, 2023.

[12] S. Shen et al., "UKF for Superior Radar Tracking," IEEE Transactions on Aerospace and Electronic Systems, vol. 50, no. 1, pp. 347-359, 2014.

[13] R. Luhr and M. Adams, "Wavelet Denoising for Non-Stationary Signals," IEEE Transactions on Signal Processing, vol. 63, no. 12, pp. 3113-3122, 2015.

[14] F. Armando et al., "KalmanNet for Enhanced Noise Reduction," IEEE Transactions on Neural Networks and Learning Systems, vol. 34, no. 2, pp. 453-465, 2023.

[15] J. Park et al., "Estimating Measurement Noise Variance with Wavelet Transform," IEEE Transactions on Instrumentation and Measurement, vol. 68, no. 5, pp. 1563-1574, 2019.

[16] M. Alfian et al., "Kalman Filtering for IMU Noise Reduction," IEEE Sensors Journal, vol. 21, no. 10, pp. 12345-12356, 2021.

[17] M. Masrafee et al., "Enhancing IMU Accuracy with Kalman Filtering," IEEE Access, vol. 10, pp. 14123-14134, 2021.

[18] H. Sheybani, "Wavelet-Based Noise Removal in Wireless Sensor Networks," IEEE Transactions on Wireless Communications, vol. 10, no. 8, pp. 1524-1535, 2011.

[19] J. Waegli et al., "Multiple IMU Configurations for Noise Reduction," IEEE Transactions on Industrial Electronics, vol. 57, no. 9, pp. 3168-3175, 2010.

[20] A. Valade et al., "Noise Reduction Techniques for Aviation and Embedded Systems," IEEE Transactions on Aerospace and Electronic Systems, vol. 53, no. 6, pp. 742-753, 2017.

[21] K. Eom et al., "Kalman Filtering for RFID Sensor Networks," IEEE Transactions on Industrial Electronics, vol. 58, no. 5, pp. 1935-1944, 2011.

[22] S. Putra et al., "Optimized Kalman Filtering for IoT Sensors," IEEE Internet of Things Journal, vol. 10, no. 3, pp. 2311-2322, 2023.

[23] J. Xiang, H. Chen, and Z. Liu, "Data Acquisition from Sensors and CAN Bus Networks for Diagnostics," IEEE Transactions on Industrial Informatics, vol. 14, no. 4, pp. 1500-1510, 2018.

[24] A. Abbas, S. A. Khan, and R. Iqbal, "Feature Selection and Extraction for Improved Diagnostic Accuracy," IEEE Transactions on Biomedical Engineering, vol. 54, no. 9, pp. 1672-1680, 2007.

[25] S. Adhikari, P. Sankavaram, and M. Pecht, "Machine Learning Algorithms for Fault Detection and Prognostics," IEEE Transactions on Reliability, vol. 67, no. 2, pp. 565-576, 2018.

[26] M. Pecht and Y. Kang, "Prognostics and Health Management of Electronics," IEEE Transactions on Components and Packaging Technologies, vol. 31, no. 4, pp. 993-1000, 2008.

[27] S. Lita, A. Carabulea, and E. Bejinaru, "Noise Sources in Automotive Applications," IEEE Transactions on Industrial Electronics, vol. 54, no. 4, pp. 2131-2138, 2007.

[28] C. Chan, M. Wong, and T. Fong, "Framework for Analyzing Noise Factors in Automotive Perception Sensors," IEEE Access, vol. 8, pp. 16896-16908, 2020.

[29] S. Lita, A. Carabulea, and E. Bejinaru, "Signal Modulation for Noise Mitigation in Automotive Applications," IEEE Transactions on Industrial Electronics, vol. 48, no. 4, pp. 1235-1242, 2001.

[30] H. Abboush, M. E. Elbagir, and A. A. Salam, "Fault Detection in Noisy Conditions Using Machine Learning," IEEE Transactions on Industrial Informatics, vol. 19, no. 2, pp. 752-762, 2023.

[31] X. Li, J. Zhang, and Y. Liu, "Impact of Noise on Automotive Camera Sensors and Object Detection," IEEE Transactions on Intelligent Transportation Systems, vol. 23, no. 3, pp. 1123-1133, 2022.

[32] S. Yadegari, H. A. Shahhosseini, and M. J. Sadeghi, "Minimizing Aerodynamic and Vibroacoustic Noise in Vehicles," IEEE Transactions on Vehicular Technology, vol. 69, no. 3, pp. 1131-1140, 2020.

[33] A. Citarella, M. P. Mantegazza, and F. Macchi, "Communication Noise Reduction in Aircraft Systems," IEEE Transactions on Aerospace and Electronic Systems, vol. 54, no. 6, pp. 742-753, 2018.

[34] P. Slavik, K. Nikitin, and O. Hadar, "Compressive Sensing-Based Noise Radar for Automotive Applications," IEEE Transactions on Signal Processing, vol. 64, no. 5, pp. 1234-1245, 2016.

[35] H. Shoureshi, A. Samadi, and M. Bal, "Hybrid Active Noise Control Systems for Automotive Applications," IEEE Transactions on Control Systems Technology, vol. 4, no. 3, pp. 305-315, 1996.

[36] J. Gerstmair, M. Ulbrich, and T. D. Bischoff, "Enhancing Sensor Performance and Reducing Interference in Automotive Applications," IEEE Transactions on Vehicular Technology, vol. 68, no. 2, pp. 1175-1186, 2019.

[37] R. Samantaray, "Challenges in Noise Reduction and Data Processing for ADAS," IEEE Transactions on Intelligent Vehicles, vol. 5, no. 2, pp. 233-245, 2023.

[38] S. Schuette and G. Waeltermann, "Hardware-in-the-Loop Testing for Vehicle Dynamics Controllers," IEEE Transactions on Industrial Electronics, vol. 57, no. 2, pp. 746-754, 2005.

[39] P. Kettelgerdes, F. Niessen, and M. Meissner, "Advanced HIL Testing for Automotive Sensor Models," IEEE Transactions on Intelligent Transportation Systems, vol. 24, no. 1, pp. 35-45, 2023.

[40] HIL Testing & Validation (hindujatech.com)

[41] Ding, Steven. (2020). Advanced methods for fault diagnosis and fault-tolerant control. 10.1007/978-3-662-62004-5.

[42] What Is Hardware-in-the-Loop (HIL)? - NI

DMMFnet: A Dual-Branch Multimodal Medical Image Fusion Network Using Super Token and Channel-Spatial Attention

Yukun Zhang[1], Lei Wang[2]*, Muhammad Tahir[3]*, Zizhen Huang [4],
Yaolong Han[5], Shanliang Yang[6], Shilong Liu[7], Muhammad Imran Saeed[8]
School of Computer Science and Technology, Shandong University of Technology, Zibo, 255000, P.R. China[1, 2, 4, 5, 6, 7]
Department of Computer Science, Mohammad Ali Jinnah University, Block-6, P.E.C.H.S, Karachi-75400, Sindh, Pakistan[3, 8]

***Abstract*—Multimodal medical image fusion leverages the correlation between different modal images to enhance the information contained within a single medical image. Existing fusion methods often fail to effectively extract multiscale features from medical images and establish long-distance relationships between deep feature blocks. To address these issues, we propose DMMFnet, an encoder-decoder fusion network that utilizes shared and private encoders to extract shared and private features. DMMFnet is based on super token sampling and channel-spatial attention. The shared encoder and decoder use a transformer structure with super token sampling technology to effectively integrate information from different modalities, improving processing efficiency and enhancing the ability to capture key features. The private encoder consists of invertible neural networks and transformer modules, designed to extract local and global features, respectively. A novel transformer module refines attention distribution and feature aggregation to capture superpixel-level global correlations, ensuring that the network effectively captures essential global information, thereby enhancing the quality of the fused image. Experimental results, comparing DMMFnet with nine leading fusion methods, indicate that DMMFnet significantly improves various evaluation metrics and achieves superior visual effects, demonstrating its advanced fusion capability.**

***Keywords*—Medical image fusion; channel-spatial attention; super token sampling; encoder–decoder**

I. Introduction

Rapid advancements in medical imaging technology have allowed for the integration of multimodal medical pictures into clinical diagnosis, surgical guidance, and medical research in recent decades [1]. However, different medical images emphasize distinct aspects. For example, Computer Tomography (CT) scans yield accurate images of the bones, they do not capture the fine details of soft tissues. In contrast, Magnetic Resonance Imaging (MRI) offers finely detailed images of the organs' soft tissues, providing substantial clinical diagnostic value [2]. Positron Emission Tomography (PET) images reflect metabolic changes and functional states of lesions through the ingestion of imaging agents. Single Photon Emission Computed Tomography (SPECT) images diagnose a broad spectrum of diseases using varying depth colors to mark affected areas [3, 4]. The usage of the medical images one by one modality to diagnose diseases is not only time-consuming but also requires extensive experience. Therefore, the goal of multimodal medical image fusion techniques is to create a single multimodal image from two modalities, and the outcomes can preserve the meaning, unique characteristics, and information from the original images, such as high-resolution structural data from CT, tissue textures from MRI [5]. Fused images have a richer texture structure and more pronounced lesion areas compared to single-modality medical images [6]. This greatly helps physicians analyze the diseases that are challenging to observe, reducing the misdiagnoses rate and surgical errors.

Usually, there are two types of image fusion tasks: supervised learning fusion and unsupervised learning fusion. Supervised learning is predominantly applied in the domain of multi-focus image fusion [7, 8]. Supervised learning is considered unsuitable for medical image fusion (MMIF) tasks due to the unique nature of medical images [9]. Moreover, due to the characteristics of medical images, the fusion methods designed for other types of images cannot be directly applied to multimodal medical image fusion tasks. According to different computational approaches, medical image fusion methods can be divided into traditional methods and deep learning methods.

Historically, among the traditional fusion methods, multi-scale transform (MST)-based methods, such as wavelet transform [10], pyramid transform [11], and subspace transform [12], have been commonly used. While the tower-based decomposition laid the groundwork for MST-based image fusion research with relatively simple implementation, it lacks directionality and is sensitive to noise, leading to redundancy between the pyramid levels. Wavelet transform offers good time-frequency locality and directionality without information redundancy, but it lacks directional selectivity and translation invariance, failing to fully extract edge information in images. Choosing the appropriate subspace mapping methods for specific fusion tasks remains a significant challenge in MST-based fusion methods.

For image fusion tasks, Pulse-Coupled Neural Networks (PCNN) [13] have received the most attention. Yin et al. proposed a fusion method combining NSST with PCNN (NSST-PAPCNN) for multimodal image fusion tasks. In this approach, NSST is used for feature extraction from multiple levels [14]. Tan et al. developed NST-MSMG [15]. It fuses high-frequency data using PCNN and boundary measurements and fuses low-frequency features utilizing an energy-based fusion rule set. However, PCNN-based approaches nevertheless follow the

*Corresponding Author.

fundamentals of multi-scale transform (MST) methods, which are needed for well-crafted decomposition and fusion rules.

The sparse representation (SR) [16] is widely utilized for image fusion, employing a mechanism that optimizes an extensive dictionary and generates sparse coefficients to achieve effective fusion. Liu et al. integrated multi-scale decomposition with convolutional sparse representation (CSR) [17]. However, methods based on SR have high computational demands, generally employing a complete and redundant dictionary for adaptive sparse representation of images. Furthermore, applying the same decomposition operations to different modalities of the source images may result in unexpected artifacts. In addition, the manual construction of the decomposition strategies and fusion rules make the fusion process complex and time-consuming [18, 19].

Recent advances have seen the adoption of deep learning techniques for multimodal image fusion, aiming to address the shortcomings of conventional fusion methods, all of which can be classified into three main types according to the network architectures: the Auto-encoder-based image fusion, the Convolutional Neural Network [20] (CNN)-based image fusion, and the Generative Adversarial Network [21] (GAN)-based image fusion.

The CNN-based image fusion methods are effective in processing spatial and structural information within image neighborhoods. Although CNN-based models are proficient in extracting local details and inductive biases from images, they lack a comprehensive understanding and learning of global semantic information in images. Additionally, due to their limited receptive field, CNNs inherently find it challenging to capture long-range relationships within images. To deal with these problems, Dosovitskiy et al. introduced the Vision Transformer (ViT) [22], which uses self-attention to conduct global comparisons across all visual tokens. It has shifted the paradigm from CNN-based feature extraction. Subsequent studies [23, 24] have shown that ViTs have potent global dependency learning capabilities in visual content. But recent research [25, 26] has shown that ViTs tend to capture shallow local features with high redundancy. This is due to shallow global attention focusing on a few adjacent tokens and neglecting most distant ones, which heavily hinders the extraction of the texture details in the fused image [27].

GAN is a type of deep learning model consisting of two modules: the generator and the discriminator. Ma et al. applied GANs to multimodal medical image fusion tasks [28]. Hung et al. introduce a multi-generator method for image fusion [29]. However, GAN-based image fusion approaches are prone to training instabilities and gradient vanishing issues [30]. Moreover, GAN architectures lose structural details due to down-sampling in pooling layers, which results in inefficient utilization of image information. The auto-encoder-based image fusion utilizes an unsupervised neural network model that comprises an encoder and a decoder. Deep Fuse [31] was one of the first methods in this domain. Li et al. introduced DenseNet and nested connections to improve the feature extraction capability of encoders [32, 33]. Furthermore, Jian et al. enhanced a self-encoder-based fusion framework by integrating an attention mechanism, aiming to extract features that are more interpretable [34]. However, due to the characteristics of GAN models, image fusion methods based on GANs are prone to instability during training. Additionally, GAN-based methods predominantly rely on CNN architectures, their limited ability to capture global information often results in insufficient fusion.

Although the above multimodal image fusion methods have obtained quite good results, several of the aforementioned problems persist. To deal with them, we propose an encoder-decoder network that uses the Invertible Neural Networks (INN) [35] and transformer module. The proposed method demonstrates superior feature extraction capabilities compared to existing method. By utilizing two distinct feature extractors to capture features at varying frequencies and then separately fusing these features during the fusion stage, our method preserves the original image's texture and structural information to the greatest extent possible. This approach is more effective in achieving the objectives of the MMIF tasks.

Here are the primary contributions of this study:

1) In order to effectively extract complementary information from the input images, a novel transformer module has been designed. The spatial and channel attention mechanisms are utilized to capture super-pixel-level global dependencies, resulting in a significant enhancement of the fusion image quality as demonstrated by both subjective and objective experiments.

2) The Context Broadcasting (CB) technique is employed in the transformer layer. This integration ensures consistent attention at each layer of the transformer model, thereby reducing the density of attention maps. Furthermore, the consistent attention mechanisms facilitate easier overall optimization of the model, aiding in more effective learning and representation of the complex relationships within the input data.

3) Extensive experiments on medical and biological image fusion demonstrate that the DMMFnet outperforms nine advanced fusion methods in terms of both quantitative metrics and visual assessment.

The organization of the paper is as follows: Section I introduces the research background and the contributions of this work. Section II provides a detailed description of the proposed DMMFnet. Section III presents and discusses the experimental results. Section IV concludes the paper.

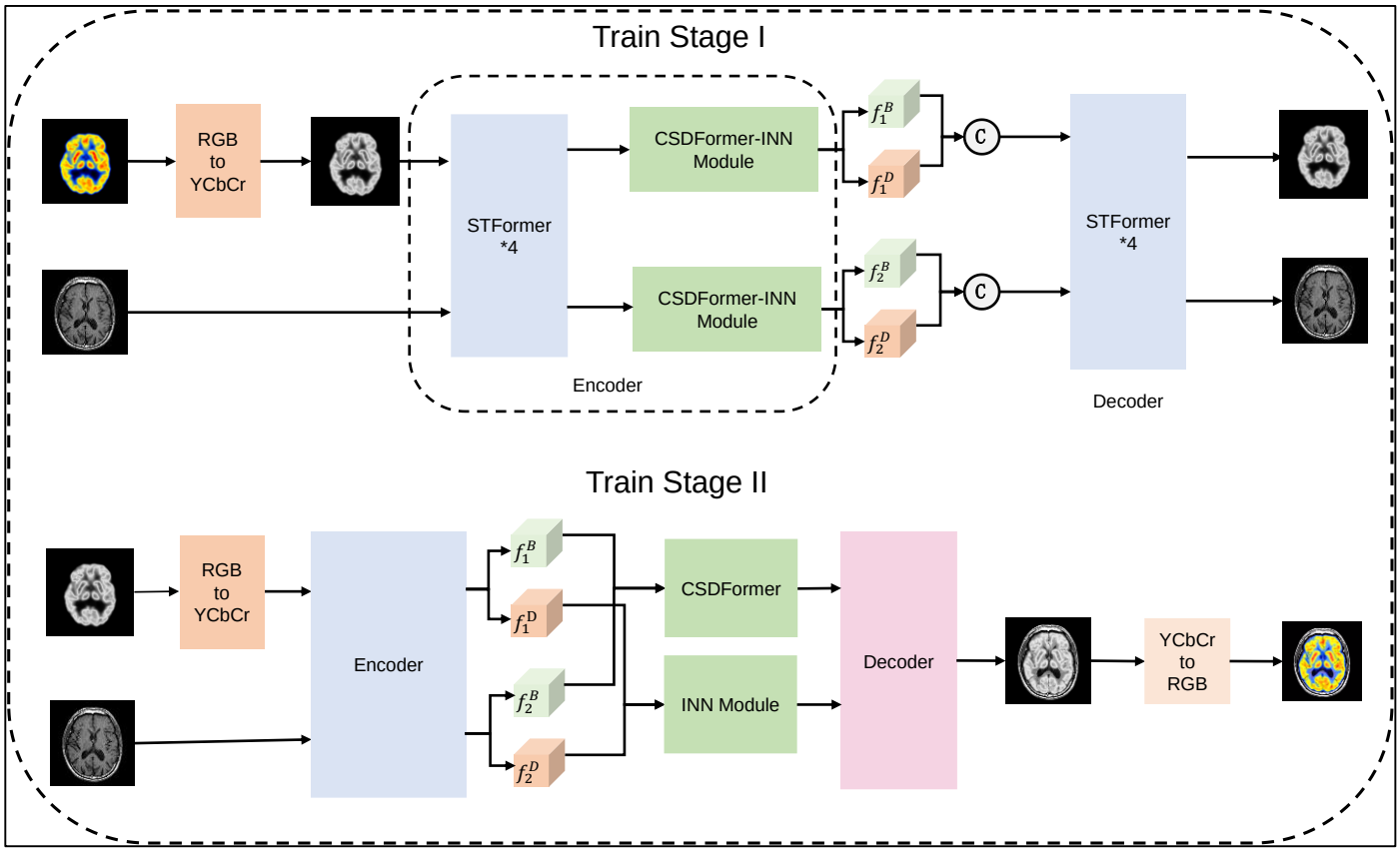


Fig. 1. The architecture of our DMMFnet method.

II. The Proposed Fusion Network

The detailed architecture of the proposed fusion framework is shown in Fig. 1. In general, CNN captures local features of the input image, whereas the transformer focuses on global features [36]. Therefore, we designed an encoder with a dual-branch structure. We used the INN to extract local features of the image and the transformer branch to extract global features of the image, then fused them separately. Finally, the DMMFnet comprises three modules: a fusion layer designed to combine different features, a decoder is used to rebuild the image and create the fused image, while an encoder is used to extract features.

The encoder consists of three components: a Shared Feature Extractor (SFE) based on STFormer, a Global Feature Extractor (GFE) based on CSDFormer, and a Local Feature Extractor (LFE) based on INN. Using PET-MRI image fusion as an example, we have defined some symbols to explain the entire fusion process: The paired PET and MRI input images are denoted as P and M, respectively. SFE, GFE, and LFE are indicated by $S(\,)$, $G(\,)$, and $L(\,)$, respectively. To extract shared features from the inputs is the aim of the SFE. This process is offered in Eq. (1):

$$f_P^S = S(P), f_M^S = S(M) \quad (1)$$

The model's Shared Feature Encoder includes the STFormer, which is based on super token sampling attention blocks [37], and the Gated-Dconv Feed-Forward Network (GDFN) module [38]. Please refer to the original paper for more information on the structure of the super token sampling attention and GDFN. The schematic of the STFormer is depicted in Fig. 2.

By incorporating super tokens into the transformer and utilizing sparse associative learning to sample super tokens from visual tokens, we can effectively capture global dependencies through self-attention on these super tokens. The reason for selecting super token sampling attention blocks in the shared feature extraction step is that they efficiently capture global dependencies by decomposing global attention into a multiplication of sparse associative mappings and low-dimensional attention. This reduces computational complexity while retaining key image information.

The GDFN block is intended to merge features from various sources, specifically images from different modalities that exhibit unique characteristics within certain frequency ranges. The DMMFnet network can flexibly process and merge these features through the GDFN block.

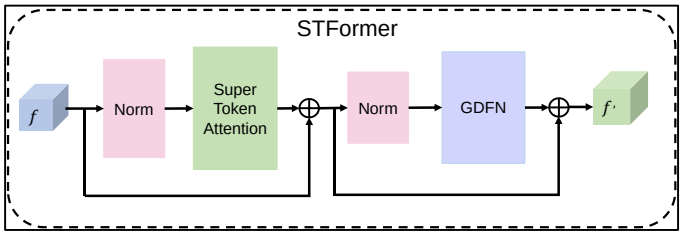


Fig. 2. The schematic of the shared feature encoder.

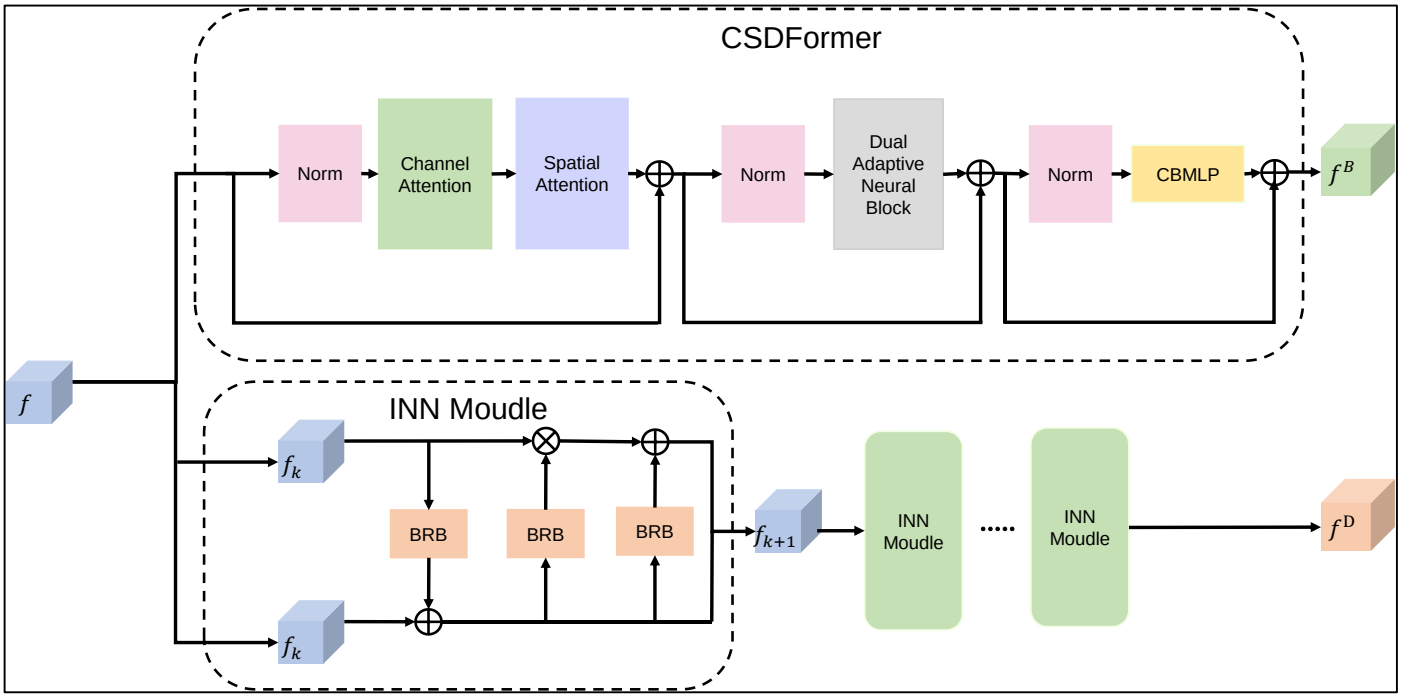


Fig. 3. The framework of the Global Feature Extractor (GFE) and Local Feature Extractor (LFE).

As illustrated in Fig. 3, the private feature encoder consists of two branches: GFE and LFE. The GFE extracts global features, while the LFE extracts local features.

The GFE branch concentrates on low-frequency global features by finely tuning the attention distribution. It effectively captures global dependencies at the super-pixel level through spatial and channel attention. By incorporating a Dual Adaptive Neural Block (DA), it adaptively encapsulates global features from superpixels to pixels, optimizing feature focus and refinement to ensure the capture of subtle changes and patterns crucial to the fusion process. In designing the CSDFormer, a Context Broadcasting (CB) technique was employed in the MLP layer. This technique involves manually inserting uniform attention into every layer of the ViT model, providing the necessary dense interactions and lowering the concentration level of attention maps across all layers. CB also enhances its capacity and generalization ability with negligible cost [39]. which is formulated as Eq. (2).

$$f_P^B = G(f_P^S), f_M^B = G(f_M^S) \tag{2}$$

The LFE branch is focused on the lossless extraction of local high-frequency features. Given that edge and texture details are crucial for image fusion tasks, the INN module is utilized to preserve as many image details as possible. The INN module aims to mitigate information loss by mutually generating input and output features through a reversible design, and it can retain the high-frequency information of the medical image almost without any loss. The process is offered in Eq. (3) below.

$$f_P^D = L(f_P^S), f_M^D = L(f_M^S) \tag{3}$$

A. Fusion Layer

The fusion layer comprises the basic feature fusion layer and the deep feature fusion layer, which respectively combine the basic and deep features. The fusion of basic and deep features is akin to the extraction of basic and deep features in the encoder. Therefore, CSDFormer and INN blocks are employed for the basic and deep fusion layers. The fusion process can be expressed by Eq. (4).

$$f^B = F_B(f_P^B, f_M^B), f^D = F_D(f_P^D, f_M^D) \tag{4}$$

F_B and $F_D$represent the basic and deep feature fusion layers, respectively.

B. Decoder

The different features extracted and processed in the previous phase are used as input to the decoder *DC()*. The reconstructed image from training stage I and the fused image from training stage II are the outputs of *DC()*. The corresponding formula is as follows:

$$\Sigma\tau\alpha\gamma\varepsilon\mathrm{I}:\ P^* = DC(f_P^B,\ f_P^D), M^* = \mathrm{DC}(f_M^B,\ f_M^D) \tag{5}$$

$$\Sigma\tau\alpha\gamma\varepsilon\mathrm{II}: FUSE = DC(f^B, f^D) \tag{6}$$

Due to the input includes cross-modal and multi-frequency features, ensuring consistency between the decoder and the shared encoder enables the decoder to better understand and exploit the feature representations provided by the encoder, leading to improved fusion results. Therefore, we employ STFormer blocks as the fundamental units for the decoder.

C. Two-stage Training

A significant obstacle in medical image fusion tasks is the absence of a definitive ground truth due to the expensive and privacy-sensitive nature of the data sources, rendering advanced supervised learning methods ineffective. Therefore, we employ a two-stage learning approach to end-to-end train the DMMFnet.

Training stage I: In this phase, we initially feed the paired PET and MRI images $\{P, M\}$ into the Shared Feature Extractor (SFE) to extract their shared features $\{f_P^S, f_P^S\}$ .Subsequently, each image is processed through the GFE based on the CSDFormer structure and the LFE based on INN separately. The basic features $\{f_P^B, f_M^B\}$ and detail features $\{f_P^D, f_M^D\}$ are extracted from the two modalities. The basic and detailed features within the same modality are merged (such as $\{f_P^B, f_P^D\}$ for PET or $\{f_M^B, f_M^D\}$ for MRI) and transmitted to the decoder for the reconstruction of the original PET or MRI image.

Training stage II: We continue to use paired PET and MRI images $\{P, M\}$ as the input. However, this time, we fed them into the encoder that was trained in Training Stage I. This enables further decomposition of features. Afterwards, we input the base features $\{f_P^B, f_M^B\}$ and detail features $\{f_P^D, f_M^D\}$ individually into fusion layers F_B and F_B .After the feature fusion process, the fused features $\{f^B, f^D\}$ are inputted into the decoder, which generates the fused image $FUSE$.

D. Loss Function

In training phase I, the total loss L_{total} is offered in Eq. (7) below.

$$L_{total} = \alpha_1 L_{pet} + \alpha_2 L_{mri} + \alpha_3 L_{decomp} \tag{7}$$

L_{pet} and L_{mri} represent the reconstruction losses for the two types of medical images. Since the model in the first stage can be regarded as a process of decomposition followed by synthesis, information loss inevitably occurs during both the decomposition and synthesis stages. L_{decomp} denotes the feature decomposition loss, while α_1 , α_2 , α_3 are adjustment parameters. The overall loss function in the first stage is designed to ensure that information is maintained throughout the encoding and decoding procedures. Each loss function is as follows.

$$L_{pet/mri} = L_{int}^{I}(I, I^*) + \sigma L_{ssim}(I, I^*) \tag{8}$$

$$L_{decomp} = \frac{\left(CC(f_P^D, f_M^D)\right)^2}{CC(f_P^B, f_M^B)+\varepsilon} \tag{9}$$

$$L_{int}^{I} = \frac{1}{HW}||I - I^*|| \tag{10}$$

where I denotes the original image before reconstruction, and I^* represents the image reconstructed in the first stage. Here, to ensure the positivity of this term, the operator CC is the correlation coefficient and ε is assigned a value of 1.01.

In training phase II, the total loss L_{total} is offered in Eq. (11) below.

$$L_{total} = \alpha_1 L_{ssim} + \alpha_2 L_{test} + \alpha_3 L_{int}^{II} \tag{11}$$

L_{ssim}represents the structural loss, measuring the similarity between two images. L_{test} stands for texture loss, while L_{int}^{II} denotes intensity loss. α_1, α_2, α_3 are adjustment parameters. The overall loss function in the second stage aims to train the fusion network weights while simultaneously adjusting the the first stage's trained encoder and decoder. Each loss function is as follows.

$$L_{int}^{II} = \frac{1}{HW}||I_f - Max(I_{pet}, I_{mri})|| \tag{12}$$

$$L_{text} = \frac{1}{HW}||\,|\nabla I_f| - Max(|\nabla I_{pet}|, |\nabla I_{mri}|)|| \tag{13}$$

$$L_{ssim} = \gamma_1(1 - ssim(I_f, I_{pet})) + \gamma_2(1 - ssim(I_f, I_{mri})) \tag{14}$$

H and W stand for the original image's height and width, respectively. γ_1 and γ_2 are the adjustment parameters used to set the importance of information from each image. In the context of multimodal medical image fusion, the most critical aspect is effectively preserving information about lesion regions and texture details. We consider the structural information of MRI images to be more important in the MMIF. Through experimentation, we have found that the best fusion results are achieved when γ_1= 0.45 and γ_2= 0.55.

III. Experimental Results and Discussion

A. The Dataset and Experimental Setting

The present study involves two fusion tasks: biological and medical image fusion. The medical image tasks include MRI/CT fusion, MRI/PET fusion, and MRI/SPECT fusion. All data can be obtained from [40]. The MRI/CT brain training set consists of 160 pairs, with 24 pairs in the test set; the MRI/PET brain training set comprises 245 pairs, with 24 pairs in the test set; and the MRI/SPECT brain training set includes 333 pairs, with 24 pairs in the test set. The biological image fusion tasks include Green Fluorescent Protein (GFP) and Phase Contrast (PC) images fusion, with data sourced from [41]. The training set comprises 130 pairs, while the test set consists of 18 pairs. During the preprocessing stage, the training images are cropped to a size of 128 × 128.

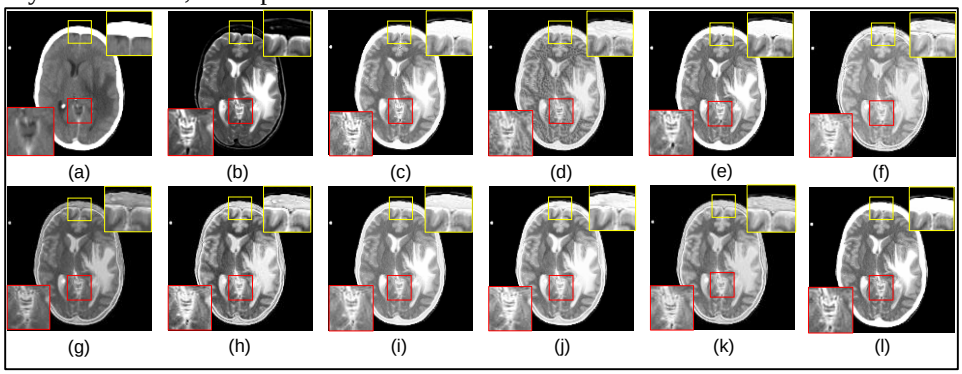


Fig. 4. The comparison of CT-MRI fusion results.

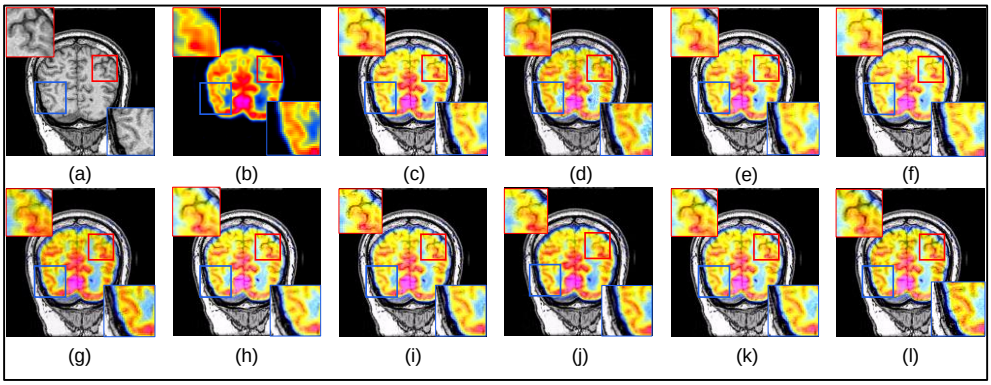


Fig. 5. The comparison of PET-MRI fusion results.

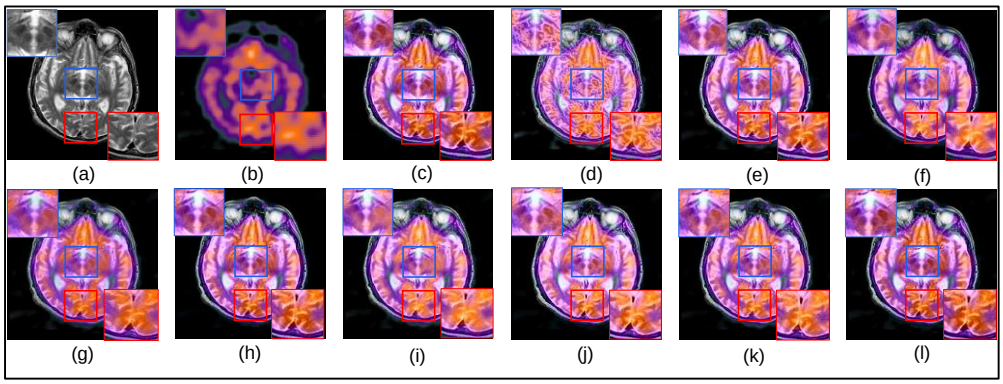


Fig. 6. The comparison of SPECT-MRI fusion results.

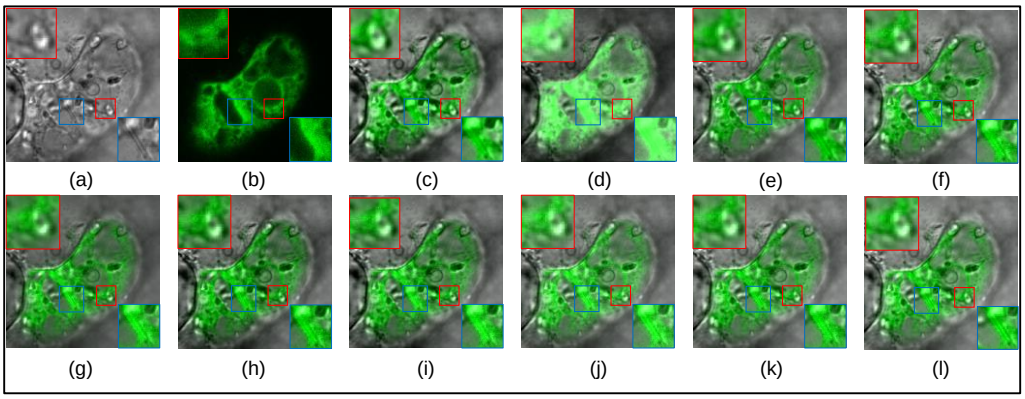


Fig. 7. The comparison of GFP-PC fusion results.

B. The Evaluation Metrics

The quantitative metrics must be used in order to compare the fusion results objectively. The fusion results are quantitatively assessed from six dimensions: entropy of information (EN) [42], Spatial Frequency (SF) [43], mutual information (MI) [44], Structural Similarity Index Measure (SSIM) [45], Visual Information Fidelity (VIF) [46], and edge-based similarity measurement $Q^{AB/F}$ [47]. Finally, calculate the average value of each metric for the respective methods for objective analysis.

C. The Subjective Analysis

For evaluation, we choose three fusion methods based on traditional methods and six state-of-the-art deep learning-based fusion methods., including PSO–NSST [48], LLPACM [49], PCNN–NSST 50], SDnet [51], EMFusion [52], U2Fusion [6], SwinFusion [53], CDDFuse [54], and CoCoNet [55]. The default values supplied by the authors of these image fusion algorithms are the configurations for the parameter settings. SwinFusion was originally designed to handle only MRI-PET images, necessitating the retraining of the model for other modal images.

The fusion results for CT-MRI, MRI-PET, MRI-SPECT, and GFP-PC are shown in Fig. 4, Fig. 5, Fig. 6, and Fig. 7, respectively. In each figure, (a) and (b) are the original images. (c)-(l) represent the fused images using the PSO – NSST, LLPACM, PCNN – NSST, SDnet, EMFusion, U2Fusion, SwinFusion, CDDFuse, CoCoNet and the DMMFnet method.

In Fig. 4, Fig. 4 (a) and Fig. 4 (b) represent the original CT and MRI images. Fig. 4 (g) is notably dark, obscuring detailed information. Fig. 4 (f), Fig. 4 (h), and Fig. 4 (k) lose some information from the original CT images, with the brain's outline becoming indistinct. Fig. 4 (d) is overly blurred, affecting clarity and detail. Fig. 4 (i) and Fig. 4 (j) are excessively bright, which hinders clear information presentation. Fig. 4 (c) and Fig. 4 (e) yield acceptable results but suffer from noise-like artifacts that affect structural details. Due to our proposed feature extraction strategy, the DMMFnet effectively preserves most of the structural information from the source images and excellently maintains the edges.

As shown in Fig. 5, Fig. 5 (a) Fig. 5 (b) represent the original MRI and PET images, respectively. Fig. 5 (e) and Fig. 5 (f) appear overly bright, resulting in unclear visual effects. Fig. 5 (g) is too dark, leading to poor preservation of MRI information. Fig. 5 (d) fails to effectively extract and preserve information, resulting in severe color distortion. Although Fig. 5 (i) and Fig. 5 (j) retain color information well, there is still some loss of MRI information at the edges. Due to our proposed fusion method, the DMMFnet successfully integrates the main information from the source images into the fused image and accurately portrays lesion information.

As shown in Fig.6, Fig.6 (a) and Fig.6 (b) are the original MRI and SPECT images. In contrast to PET, SPECT images have much sparser intensity information, leading to varying degrees of information loss across different deep learning methods. Fig. 6 (d) is overly blurred, while Fig. 6 (f) and Fig. 6 (i) retain too much color information, resulting in overly bright images that obscure details. Fig. 6 (g) suffers from severe deficiencies in color information. The DMMFnet outperforms all other deep learning methods.

As shown in Fig.7, Fig.7 (a) and Fig.7 (b) are the original GFP and PC images. Due to the higher resolution of GFP and PC images compared to medical images, the feature information is more apparent, and deep learning methods generally yield better visual results. In Fig.7 (d), significant feature loss is observed. In Fig.7 (c) and Fig.7 (e), the cell structures are preserved, but there is some loss of color information. Conversely, in Fig.7 (i), Fig.7 (j), and Fig.7 (k), the cell structures are not clearly preserved. Additionally, Fig.7 (g) is too dark, resulting in poor visual quality. DMMFnet preserves both the cell structures and color information more comprehensively.

D. The Objective Analysis

In MRI-CT image fusion, our algorithm possesses four optimal metrics (EN, SF, SSIM, and $Q^{AB/F}$) and one suboptimal metric (VIF), with the SF metric performing significantly better than the others. As shown in Table I.

For MRI-PET image fusion, compared to other algorithms, our algorithm achieves the highest scores in SF, MI, SSIM, and $Q^{AB/F}$, while being second best in VIF as seen in Table II.

In MRI-SPECT image fusion, our algorithm has two optimal metrics (VIF and $Q^{AB/F}$) and one suboptimal metric, surpassing all deep learning methods, as seen in Table III.

TABLE I. THE OBJECTIVE EVALUATION OF CT-MRI FUSION IMAGES

Fusion Methods	Evaluation metrics					
	EN	SF	MI	SSIM	VIF	$Q^{AB/F}$
PSO–NSST	4.5	34.25	**2.09**[1]	**1.36**[1]	0.5	0.5
LLPACM	4.62	32.9	1.98	1.28	0.39	0.45
PCNN–NSST	4.58	36.07	2.07	1.34	0.47	**0.57**[2]
SDnet	4.66	34.77	**2.26**[2]	1.29	**0.51**[2]	0.52
EMFusion	4.62	26.7	2.04	**1.35**[2]	0.43	0.5
U2Fusion	4.64	35.77	1.94	1.34	0.4	0.51
SwinFusion	**4.69**[1]	33.6	1.92	**1.35**[2]	**0.61**[1]	0.51
CDDFuse	4.62	35.3	2.08	1.33	0.49	0.53
CoConet	**4.68**[2]	**36.15**[2]	1.92	1.33	0.43	0.54
Proposed	**4.69**[1]	**38.52**[1]	2.04	**1.36**[1]	**0.51**[2]	**0.62**[1]

TABLE II. THE OBJECTIVE EVALUATION OF PET-MRI FUSION IMAGES

Fusion Methods	Evaluation metrics					
	EN	SF	MI	SSIM	VIF	$Q^{AB/F}$
PSO–NSST	5.32	**38.03**[2]	2.75	**1.27**[2]	0.65	0.71
LLPACM	5.32	33.67	2.29	1.22	0.63	0.72
PCNN–NSST	5.44	37.37	2.45	1.26	0.65	0.71
SDnet	5.43	37.29	2.44	1.26	0.67	0.69
EMFusion	5.38	32.9	2.26	**1.27**[2]	**0.68**[2]	0.7
U2Fusion	5.35	38.01	**2.77**[2]	1.26	0.67	**0.73**[2]
SwinFusion	**5.52**[2]	36.49	2.12	1.23	**0.71**[1]	0.68
CDDFuse	5.44	37.68	2.43	1.25	0.62	0.67
CoConet	**5.58**[1]	37.89	2.63	1.26	0.67	**0.74**[1]
Proposed	5.4	**38.05**[1]	**2.81**[1]	**1.28**[1]	**0.68**[2]	**0.74**[1]

TABLE III. THE OBJECTIVE EVALUATION OF SPECT-MRI FUSION IMAGES

Fusion Methods	Evaluation metrics					
	EN	SF	MI	SSIM	VIF	$Q^{AB/F}$
PSO–NSST	**5.54**[2]	**27.93**[1]	**3.07**[1]	1.25	**0.84**[2]	0.71
LLPACM	5.33	26.47	2.31	1.08	0.46	0.57
PCNN–NSST	**5.56**[1]	**27.16**[2]	**3.02**[2]	**1.3**[1]	0.74	0.7
SDnet	5.2	23.77	2.27	1.25	0.63	0.69
EMFusion	5.2	20.11	2.42	1.26	0.71	0.71
U2Fusion	5.08	25.45	2.8	1.25	0.81	**0.74**[2]
SwinFusion	5.07	22.67	2.12	**1.28**[2]	**0.87**[1]	0.64
CDDFuse	5.2	25.98	2.69	1.24	0.75	0.72
CoConet	5.18	25.27	2.89	1.27	0.7	0.71
Proposed	5.09	25.83	**3.07**[1]	**1.28**[2]	**0.87**[1]	**0.75**[1]

In GFP-PC image fusion, our algorithm possesses four optimal metrics (EN, MI, SSIM, VIF, and $Q^{AB/F}$), as seen in Table IV.

TABLE IV. THE OBJECTIVE EVALUATION OF GFP-PC FUSION IMAGES

Fusion Methods	Evaluation metrics					
	EN	SF	MI	SSIM	VIF	$Q^{AB/F}$
PSO–NSST	6.81	11.48	2.15	0.7	0.75	0.58
LLPACM	4.78	10.9	1.62	0.64	0.65	0.37
PCNN–NSST	6.75	11.56	2.73	0.79	0.87	0.54
SDnet	6.81	12.73	3.05	0.74	0.63	**0.64**[2]
EMFusion	6.54	11.82	2.25	0.85	0.83	0.58
U2Fusion	6.63	**12.97**[1]	2.79	0.77	0.94	0.62
SwinFusion	6.76	**12.83**[2]	3.49	**0.89**[2]	**1.06**[2]	0.61
CDDFuse	**6.85**[2]	12.11	2.28	0.88	1.05	0.6
CoConet	6.84	12.18	**3.56**[2]	0.88	1.02	**0.65**[1]
Proposed	**6.89**[1]	12.04	**3.63**[1]	**0.9**[1]	**1.07**[1]	0.61

Due to the inherently low original resolution of SPECT images, up-sampling is required before fusion, introducing a significant amount of non-uniform pixel noise. This noise interference adversely affects the results for metrics such as EN and SF, making deep learning methods perform less favorably on general evaluation metrics for SPECT-MRI modality fusion compared to traditional methods. However, our method better preserves information from both modalities, surpassing other deep learning approaches in objective analysis.

E. The Ablation Study

As the test cases, we chose 30 pairs of images at random from each modality for ablation experiments. We chose peak EN, MI, VIF, and SSIM as the metrics. The configuration of each experiment is shown in Table V.

TABLE V. THE OBJECTIVE EVALUATION OF GFP-PC FUSION IMAGES

Experiment	Configurations
Experiment 1	Replace the shared encoder STFormer with Restormer [38], keeping other parts unchanged.
Experiment 2	Not utilize context broadcasting technology.
Experiment 3	The feature extraction component in the encoder uses only the transformer module.
Experiment 4	The feature extraction component in the encoder uses only the INN module.
Experiment 5	Change the two-stage training to direct training.
Experiment 6	Removal of L_{decomp} from the loss functions used in the first phase of training.
Experiment 7	Sets γ_1 and γ_2 to {0.5, 0.5}, indicates that both images are equally important.

From Fig. 8, it can be observed that the fusion evaluation metrics using dual-branch feature extractors in the fusion module are better than those obtained by using either the CNN or the Transformer alone as the feature extractor. This indirectly confirms that dual-branched feature extractors improve the ability of the network to extract features, beneficial for subsequent bottom-level visual tasks in image fusion, and that context broadcasting technology significantly aids in improving fusion effects. In summary, Table VI proves the effectiveness and soundness of our network and loss function design.

TABLE VI. THE RESULTS OF THE ABLATION EXPERIMENT

Experiment	Evaluation metrics				
	EN	SF	MI	VIF	SSIM
Experiment 1	4.93	24.43	2.69	.0.75	1.24
Experiment 2	5.07	24.45	2.96	0.81	1.25
Experiment 3	4.86	22.19	2.21	0.85	1.21
Experiment 4	5.01	23.81	2.35	0.72	1.22
Experiment 5	4.89	22.11	2.42	0.71	1.26
Experiment 6	5.07	25.76	3.06	0.87	1.25
Experiment 7	5.06	25.83	3.07	0.85	1.26
Proposed	**5.07**	**25.83**	**3.07**	**0.87**	**1.27**

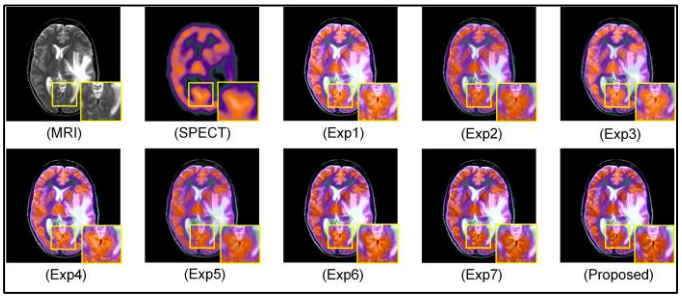


Fig. 8. The ablation experiment and the corresponding zoomed-in details of each fused image.

IV. CONCLUSION AND FUTURE WORK

This study proposes a multimodal medical image fusion network, DMMFnet. For the extraction of shared features in multimodal images, a new transformer module based on super token sampling is constructed, which effectively captures global dependencies. This module not only significantly improves the processing speed of the model but also ensures the effective capture and preservation of key features, thereby enhancing the ability to recognize and integrate important information in medical images. In addition, the proposed CSDformer module further optimizes feature extraction and fusion. By introducing the Context Broadcast strategy, the much-needed dense interaction is achieved, which greatly improves the ability to capture detailed features. Although DMMFnet has achieved satisfactory fusion results, it does not present a notable advantage in computational efficiency due to the limitations of the Transformer architecture.

Future work: Future improvements should focus on refining the network to achieve better results at a lower computational cost. Enhancements in this direction would make the DMMFnet more practical and efficient for real-world applications.

ACKNOWLEDGMENT

We extend our gratitude to the Shandong Provincial Natural Science Foundation, the Key R&D Program of Shandong Province, P.R China, and Mohammad Ali Jinnah University, Karachi, Pakistan.

RESEARCH FUNDING

This research was funded by the Shandong Provincial Natural Science Foundation through project ZR2021MF017 and

the Key R&D Program of Shandong Province, China, under project 2023RKY01015.

AUTHORS' CONTRIBUTION

In this paper, each author contributed equally to the research and development process. Yukun Zhang and Lei Wang collaborated on the conceptualization and design of the study, as well as the implementation of methodologies. Muhammad Tahir, Muhammad Imran Saeed, and Zizhen Huang were responsible for extensive data collection, analysis, and interpretation, significantly contributing to the empirical aspects of the research. Yaolong Han provided valuable insights and expertise in the theoretical framework and literature review, ensuring the study's rigor and coherence. Shanliang Yang refined the manuscript through editing and formatting, ensuring clarity and coherence in presenting the results. The authors worked collaboratively to draft and revise the manuscript, incorporating feedback and suggestions to produce the final version.

CONFLICT OF INTEREST

The authors declare that there is no conflict of interest regarding the publication of this paper. All authors have reviewed and approved the manuscript and have no financial or personal relationships that could inappropriately influence or bias the content of the paper.

REFERENCES

[1] M. Yin, X. Liu, Y. Liu, and X. Chen, "Medical image fusion with parameter-adaptive pulse coupled neural network in nonsubsampled shearlet transform domain," IEEE Transactions on Instrumentation and Measurement, vol. 68, pp. 49-64, 2018.

[2] Z. Ding, H. Li, Y. Guo, D. Zhou, and Y. Liu, "M4fnet: Multimodal medical image fusion network via multi-receptive-field and multi-scale feature integration," Computers in Biology and Medicine, vol. 159, pp. 106923, 2023.

[3] W. Li, R. Li, J. Fu, and X. Peng, "MSENet: A multi-scale enhanced network based on unique features guidance for medical image fusion," Biomedical Signal Processing and Control, vol. 74, pp. 103534, 2022.

[4] Y. Zhang, M. Jin, and G. Huang, "Medical image fusion based on improved multi-scale morphology gradient-weighted local energy and visual saliency map," Biomedical Signal Processing and Control, vol. 74, pp. 103535, 2022.

[5] G. Zhang, R. Nie, J. Cao, L. Chen, and Y. Zhu, "FDGNet: A pair feature difference guided network for multimodal medical image fusion," Biomedical Signal Processing and Control, vol. 81, pp. 104545, 2023.

[6] H. Xu, J. Ma, J. Jiang, X. Guo, and H. Ling, "U2Fusion: A unified unsupervised image fusion network," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 44, no. 1, pp. 502-518, 2020.

[7] X. Guo, R. Nie, J. Cao, D. Zhou, L. Mei, and K. He, "FuseGAN: Learning to fuse multi-focus image via conditional generative adversarial network," IEEE Transactions on Multimedia, vol. 21, no. 8, pp. 1982-1996, 2019.

[8] Y. Liu, X. Chen, H. Peng, and Z. Wang, "Multi-focus image fusion with a deep convolutional neural network," Information Fusion, vol. 36, pp. 191-207, 2017.

[9] L. Chen, X. Wang, Y. Zhu, and R. Nie, "Multi-level difference information replenishment for medical image fusion," Applied Intelligence, vol. 53, pp. 4579-4591, 2023.

[10] Y. Yang, D.S. Park, S. Huang, and N. Rao, "Medical image fusion via an effective wavelet-based approach," EURASIP Journal on Advances in Signal Processing, vol. 2010, pp. 1-13, 2010.

[11] W. Wang and F. Chang, "Multi-focus image fusion method based on Laplacian pyramid," Journal of Computers, vol. 6, no. 12, pp. 2559-2566, 2011.

[12] Y. Liu, X. Chen, R.K. Ward, and Z.J. Wang, "Medical image fusion via convolutional sparsity based morphological component analysis," IEEE Signal Processing Letters, vol. 26, no. 3, pp. 485-489, 2019.

[13] Z. Wang, Y. Ma, F. Cheng, and L. Yang, "Review of pulse-coupled neural networks," Image and Vision Computing, vol. 28, no. 1, pp. 5-13, 2010.

[14] M. Yin, X. Liu, Y. Liu, and X. Chen, "Medical image fusion with parameter-adaptive pulse coupled neural network in nonsubsampled shearlet transform domain," IEEE Transactions on Instrumentation and Measurement, vol. 68, no. 1, pp. 49-64, 2018.

[15] W. Tan, P. Tiwari, H.M. Pandey, C. Moreira, and A.K. Jaiswal, "Multimodal medical image fusion algorithm in the era of big data," Neural Computing and Applications, pp. 1-21, 2020.

[16] B. Yang and S. Li, "Multifocus image fusion and restoration with sparse representation," IEEE Transactions on Instrumentation and Measurement, vol. 59, no. 4, pp. 884-892, 2009.

[17] B. Yang and S. Li, "Image fusion with convolutional sparse representation," IEEE Signal Processing Letters, vol. 23, no. 12, pp. 1882-1886, 2016.

[18] H. Xu and J. Ma, "EMFusion: An unsupervised enhanced medical image fusion network," Information Fusion, vol. 76, pp. 177-186, 2021.

[19] C. Wang, R. Nie, J. Cao, X. Wang, and Y. Zhang, "IGNFusion: An unsupervised information gate network for multimodal medical image fusion," IEEE Journal of Selected Topics in Signal Processing, vol. 16, no. 4, pp. 854-868, 2022.

[20] Y. LeCun, B. Boser, J.S. Denker, D. Henderson, R.E. Howard, W. Hubbard, and L.D. Jackel, "Backpropagation applied to handwritten zip code recognition," Neural Computation, vol. 1, no. 4, pp. 541-551, 1989.

[21] A. Creswell, T. White, V. Dumoulin, K. Arulkumaran, B. Sengupta, and A.A. Bharath, "Generative adversarial nets," Advances in Neural Information Processing Systems, vol. 27, 2014.

[22] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, and N. Houlsby, "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale," arXiv preprint arXiv:2010.11929, 2020.

[23] H. Touvron, M. Cord, M. Douze, F. Massa, A. Sablayrolles, and H. Jégou, "Training data-efficient image transformers & distillation through attention," in Proceedings of the International Conference on Machine Learning (PMLR), 2021, pp. 10347-10357.

[24] H. Touvron, M. Cord, A. Sablayrolles, G. Synnaeve, and H. Jégou, "Going deeper with image transformers," in Proceedings of the IEEE International Conference on Computer Vision (ICCV), 2021, pp. 32-42.

[25] K. Li, Y. Wang, P. Gao, G. Song, Y. Liu, H. Li, and Y. Qiao, "Uniformer: Unified transformer for efficient spatiotemporal representation learning," arXiv preprint arXiv:2201.04676, 2022. https://doi.org/10.48550/arXiv.2201.04676.

[26] M. Raghu, T. Unterthiner, S. Kornblith, C. Zhang, and A. Dosovitskiy, "Do vision transformers see like convolutional neural networks?" Advances in Neural Information Processing Systems, vol. 34, pp. 12116-12128, 2021.

[27] Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin, and B. Guo, "Swin transformer: Hierarchical vision transformer using shifted windows," in Proceedings of the IEEE International Conference on Computer Vision (ICCV), 2021, pp. 10012-10022.

[28] J. Ma, H. Xu, J. Jiang, X. Mei, and X.P. Zhang, "DDcGAN: A dual-discriminator conditional generative adversarial network for multi-resolution image fusion," IEEE Transactions on Image Processing, vol. 29, pp. 4980-4995, 2020.

[29] J. Huang, Z. Le, Y. Ma, F. Fan, H. Zhang, and L. Yang, "MGMDcGAN: Medical Image Fusion Using Multi-Generator Multi-Discriminator Conditional Generative Adversarial Network," IEEE Access, vol. 99, pp. 1-1, 2020.

[30] W. Li, Y. Zhang, G. Wang, and Y. Huang, "DFENet: A dual-branch feature enhanced network integrating transformers and convolutional feature learning for multimodal medical image fusion," Biomedical Signal Processing and Control, vol. 80, pp. 104402, 2023.

[31] K. Ram Prabhakar, V. Sai Srikar, and R. Venkatesh Babu, "Deepfuse: A deep unsupervised approach for exposure fusion with extreme exposure

image pairs," in Proceedings of the IEEE International Conference on Computer Vision (ICCV), 2017, pp. 4714-4722.

[32] H. Li and X.J. Wu, "Dense-Fuse: A Fusion Approach to Infrared and Visible Images," IEEE Transactions on Image Processing, vol. 28, pp. 2614-2623, 2018.

[33] H. Li, X.J. Wu, and J. Kittler, "RFN-Nest: An end-to-end residual fusion network for infrared and visible images," Information Fusion, vol. 73, pp. 72-86, 2021.

[34] L. Jian, X. Yang, Z. Liu, G. Jeon, M. Gao, and D. Chisholm, "SEDRFuse: A symmetric encoder–decoder with residual block network for infrared and visible image fusion," IEEE Transactions on Instrumentation and Measurement, vol. 70, pp. 1-15, 2020.

[35] J. Cui, L. Zhou, F. Li, and Y. Zha, "Visible and infrared image fusion by invertible neural network," in China Conference on Command and Control, Singapore: Springer Nature Singapore, 2022, pp. 133-145.

[36] N. Park and S. Kim, "How do vision transformers work?" arXiv preprint arXiv:2202.06709, 2022. https://doi.org/10.48550/arXiv.2202.06709.

[37] H. Huang, X. Zhou, J. Cao, R. He, and T. Tan, "Vision transformer with super token sampling," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2023, pp. 22690-22699.

[38] S.W. Zamir, A. Arora, S. Khan, M. Hayat, F.S. Khan, and M.H. Yang, "Restormer: Efficient transformer for high-resolution image restoration," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2022, pp. 5718-5729.

[39] N. Hyeon-Woo, K. Yu-Ji, B. Heo, D. Han, S.J. Oh, and T.H. Oh, "Scratching visual transformer's back with uniform attention," in Proceedings of the IEEE International Conference on Computer Vision (ICCV), 2023, pp. 5807-5818.

[40] K. A. Johnson, J. A. (n.d.) Becker, The whole brain atlas. Available: http://www.med.harvard.edu/aanlib/ home.html.

[41] R. Tsien, "The green fluorescent protein," Annu. Rev. Biochem., vol. 67, pp. 509–544, 1998.

[42] J.W. Roberts, J.A. Van Aardt, and F.B. Ahmed, "Assessment of image fusion procedures using entropy, image quality, and multispectral classification," Journal of Applied Remote Sensing, vol. 2, no. 1, pp. 023522, 2008.

[43] A.M. Eskicioglu and P.S. Fisher, "Image quality measures and their performance," IEEE Transactions on Communications, vol. 43, no. 12, pp. 2959-2965, 1995.

[44] G. Qu, D. Zhang, and P. Yan, "Information measure for performance of image fusion," Electronics Letters, vol. 38, no. 7, pp. 313–315, 2002.

[45] Z. Wang, A.C. Bovik, H.R. Sheikh, and E.P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," IEEE Transactions on Image Processing, vol. 13, no. 4, pp. 600-612, 2004.

[46] Y. Han, Y. Cai, Y. Cao, and X. Xu, "A new image fusion performance metric based on visual information fidelity," Information Fusion, vol. 14, no. 2, pp. 127-135, 2013.

[47] J. Ma, Y. Ma, and C. Li, "Infrared and visible image fusion methods and applications: A survey," Information Fusion, vol. 45, pp. 153-178, 2019

[48] Y. Gao, S. Ma, J. Liu, Y. Liu, and X. Zhang, "Fusion of medical images based on salient features extraction by PSO optimized fuzzy logic in NSST domain," Biomedical Signal Processing and Control, vol. 69, pp. 102852, 2021.

[49] W. Li, J. Du, Z. Zhao, and J. Long, "Fusion of medical sensors using adaptive cloud model in local Laplacian pyramid domain," IEEE Transactions on Biomedical Engineering, vol. 66, no. 4, pp. 1172-1183, 2018.

[50] W. Tan, P. Tiwari, H.M. Pandey, C. Moreira, and A.K. Jaiswal, "Multimodal medical image fusion algorithm in the era of big data," Neural Computing and Applications, pp. 1-21, 2020.

[51] H. Zhang and J. Ma, "SDNet: A versatile squeeze-and-decomposition network for real-time image fusion," International Journal of Computer Vision, vol. 129, no. 10, pp. 2761-2785, 2021.

[52] H. Xu and J. Ma, "EMFusion: An unsupervised enhanced medical image fusion network," Information Fusion, vol. 76, pp. 177-186, 2021.

[53] J. Ma, L. Tang, F. Fan, J. Huang, X. Mei, and Y. Ma, "SwinFusion: Cross-domain long-range learning for general image fusion via swin transformer," IEEE/CAA Journal of Automatica Sinica, vol. 9, no. 7, pp. 1200-1217, 2022.

[54] Z. Zhao, H. Bai, J. Zhang, Y. Zhang, S. Xu, Z. Lin, R. Timofte, and L. Van Gool, "CDDFuse: Correlation-driven dual-branch feature decomposition for multi-modality image fusion," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023, pp. 5906-5916.

[55] J. Liu, R. Lin, G. Wu, R. Liu, Z. Luo, and X. Fan, "CoConet: Coupled contrastive learning network with multi-level feature ensemble for multi-modality image fusion," International Journal of Computer Vision, pp. 1-28, 2023.

Deep Learning and Computer Vision-Based System for Detecting and Separating Abnormal Bags in Automatic Bagging Machines

Trung Dung Nguyen, Thanh Quyen Ngo, Chi Kien Ha
Faculty of Electrical Engineering Technology, Industrial University of Ho Chi Minh City, Ho Chi Minh City, Vietnam

***Abstract*—This paper presents a novel deep learning and computer vision-based system for detecting and separating abnormal bags within automatic bagging machines, addressing a key challenge in industrial quality control. The core of our approach is the development of a data collection system seamlessly integrated into the production line. This system captures a comprehensive variety of bag images, ensuring a dataset representative of real-world variability. To augment the quantity and quality of our training data, we implement both offline and online data augmentation techniques. For classifying normal and abnormal bags, we design a lightweight deep learning model based on the residual network for deployment on computationally constrained devices. Specifically, we improve the initial convolutional layer by utilizing ghost convolution and implement a reduced channel strategy across the network layers. Additionally, knowledge distillation is employed to refine the model's performance by transferring insights from a fully trained, more complex network. We conduct extensive comparisons with other convolutional neural network models, demonstrating that our proposed model achieves superior performance in classifying bags while maintaining high efficiency. Ablation studies further validate the contribution of each modification to the model's success. Upon deployment, the model demonstrates robust accuracy and operational efficiency in a live production environment. The system provides significant improvements in automatic bagging processes, combining accuracy with practical applicability in industrial settings.**

***Keywords*—Automatic bagging machines; deep learning; computer vision; bags classification; data augmentation**

I. Introduction

An automatic bagging machine is a type of packaging machinery designed to automatically fill products into bags and then seal them. These machines are widely used in various industries, including food, agriculture, chemical, and manufacturing, for efficient and rapid packaging solutions. Automatic bagging machines can handle a wide range of bag materials, sizes, and types, such as plastic, paper, and fabric bags. An automatic bagging machine comprises several stages and components: a product feeding system, which delivers the product to the bagging area; a weighing and filling system, which ensures that each bag is filled with the correct amount of product; a bag supply and opening unit, which automatically takes a bag from the supply, opens it, and positions it for filling; and a sealing system, which seals the bag by heat sealing, stitching, or using adhesives. Fig. 1 illustrates the comprehensive setup of the automatic bagging machine utilized in our research, highlighting each of the critical components and stages that facilitate the seamless transition from product feeding to the final sealing process. Among the components, the bag supply and opening unit is a critical component, ensuring that bags are consistently and accurately supplied and opened for the product filling process. This unit is designed to handle a variety of bag types and materials, including paper, plastic, and woven fabric, with varying levels of thickness and rigidity. To maintain the maximum performance of automatic bagging machines, constant bag quality is essential at all times. Criteria related to bag quality include the bag mouth, bag position, and bag surface. For the bag mouth, the edges must be perpendicular to the sides and in a straight line. For the bag position, bags must lie flat throughout their entire length and width. For the bag surface, it must be free of folds and/or wrinkles that could result from improper storage. Fig. 3(b) indicates some instances where bag quality does not meet the standards.

Any errors related to bag quality can cause serious problems. For example, a bag that is too weak might tear during the picking or opening process, while a bag with inconsistent dimensions might not align properly with the machine’s mechanisms. These issues can halt production, necessitating manual intervention to clear the jam and restart the machine. Even if abnormal bags are successfully opened and filled, they may not seal properly, potentially compromising the integrity of the packaging. This can affect product safety, shelf life, and customer satisfaction. Inconsistent bag quality can also lead to poor presentation of the final product, affecting brand perception. The failure to properly handle abnormal bags can lead to increased material waste, as bags that are damaged during the process or that fail quality checks after filling and sealing are discarded. This not only increases material costs but can also lead to higher labor costs associated with troubleshooting and rectifying issues caused by using these bags.

To mitigate the issues caused by abnormal bags, manufacturers may implement quality control measures such as pre-screening bags before they enter the supply unit, adjusting machine parameters to better accommodate variation in bag quality, or investing in more advanced detection and handling systems that can adapt to a wider range of bag qualities. Implementing a rigorous quality assurance program with suppliers to ensure that bags meet all necessary specifications before they reach the production line is also crucial. Modern

bag supply and opening units often rely on sensors and automated systems to detect and adjust the bags being fed into the machine. Abnormal bags might not be detected accurately by these systems, leading to misfeeds or incorrect adjustments that can compromise the packaging process.



Fig. 1. The comprehensive setup of the automatic bagging machine utilized in our research.

Over the years, the adoption of machine vision for automatic quality inspection has seen significant advancements across various industries, revolutionizing how quality control is implemented and ensuring higher standards of accuracy and efficiency. In the metal casting industry, machine vision has been instrumental in detecting defects such as cracks, porosity, and misruns on cast parts [1]. The manufacturing industry has broadly embraced machine vision for a range of applications, from verifying product assembly to ensuring the accuracy of labeling and packaging [2]. In the agricultural sector, machine vision has been applied to the inspection of the external quality of date fruits [3]. Automatic rice-quality inspection systems represent another remarkable application [4]. These systems employ machine vision to classify rice grains by size, shape, color, and texture, as well as to detect impurities. In the wood industry, particularly in the inspection of hardwood flooring products, machine vision systems have been developed to detect surface defects such as knots, cracks, and color variation [5]. These systems can inspect flooring panels at high speeds, ensuring that only those meeting strict quality standards reach the consumer. Overall, the developments in machine vision for automatic quality inspection across these varied industries emphasize a trend towards greater automation and precision in quality control processes. By leveraging advanced imaging technologies and machine learning algorithms, industries are not only able to enhance the efficiency of their operations but also significantly improve the quality of their products, benefiting both manufacturers and consumers alike.

In recent years, deep learning has seen remarkable developments, transforming the landscape of production industries with its unprecedented capabilities in data analysis, pattern recognition, and autonomous decision-making. Leveraging vast amounts of data, deep learning algorithms have become proficient at identifying complex patterns and anomalies that elude traditional computational methods. This advancement has been particularly impactful in automating quality control processes, predictive maintenance, and enhancing operational efficiencies across various sectors. However, deploying deep learning models, especially Convolutional Neural Networks (CNNs), on resource-constrained embedded devices poses significant challenges. First and foremost, these devices typically have limited processing power, which can make it difficult to run the computationally intensive operations required by CNNs in real-time. Additionally, embedded devices often have restricted memory capacity, constraining the size of the models that can be deployed and limiting the amount of data that can be processed at once. Energy consumption is another critical concern, as many embedded devices operate on battery power or in energy-sensitive environments. The high computational demands of CNNs can lead to rapid battery depletion or require compromises in performance to conserve energy. Model complexity versus performance trade-offs also present a challenge. Simplifying models to fit the constraints of embedded devices can lead to reduced accuracy and efficacy. Finally, the diversity of hardware in embedded systems necessitates custom optimization for each deployment, increasing development time and complexity. Addressing these challenges requires innovative solutions, including model compression techniques, specialized hardware accelerators, and efficient algorithm design to make CNNs viable for embedded applications.

Based on the above analysis, this paper introduces a novel deep learning and computer vision-based system designed to enhance the efficiency and accuracy of automatic bagging machines by classifying bags as normal or abnormal. By integrating a sophisticated data collection system directly into the production line and employing advanced data augmentation techniques, this study addresses the critical need for high-quality, diversified datasets in machine learning. Central to our approach is the development of a lightweight deep learning model, based on the modified ResNet-18 architecture, which is specifically optimized for deployment on resource-constrained devices such as the Raspberry Pi 4. The contributions of this paper are summarized as follows:

- We demonstrate the efficacy of combining offline and online data augmentation techniques to substantially improve model robustness and generalizability.
- Our customized lightweight ResNet-18 model, featuring an innovative initial convolution layer modification, channel reduction, and the application of knowledge distillation, demonstrates a novel approach to optimizing deep learning models for efficient deployment on embedded systems.
- Through comprehensive comparisons with other CNN models and ablation studies, we provide valuable insights into the model's decision-making processes and its superior performance.
- The successful deployment of our model on the Raspberry Pi 4 not only proves its operational viability in real-world industrial settings but also sets a benchmark for future research in deploying deep learning models on resource-constrained devices.

II. Literature Review

A. Image Classification

Image classification, a pivotal task in the field of computer vision, has experienced significant evolution over the past decade, predominantly shaped by the advent and advancement of CNNs and, more recently, Transformer models.

CNNs have established themselves as the backbone of image classification tasks, marked by their ability to automatically and adaptively learn spatial hierarchies of features from image data. The foundational model, LeNet [6], introduced in the late 1990s, set the stage for the use of CNNs in image recognition. However, it was AlexNet's [7] victory in the ImageNet challenge in 2012 that truly catalyzed the deep learning revolution, highlighting CNNs' potential to achieve remarkable accuracy in classifying images across thousands of categories. Subsequent architectures, such as VGG [8], GoogLeNet [9], ResNet [10], MobileNet [11-13], ShuffleNet [14-15] and EfficientNets [16] have introduced innovations such as deeper networks, inception modules, and residual connections, significantly improving performance on various image classification benchmarks. These developments have not only enhanced the accuracy and efficiency of image classification tasks but have also broadened the application scope of CNNs to include areas like medical image analysis, autonomous vehicles, and surveillance systems, emphasizing their versatility and robustness in extracting meaningful patterns from visual data.

The introduction of Transformers in image classification [17], initially conceived for natural language processing tasks [18], marks the latest significant innovation in the field. The seminal work, "Attention Is All You Need," introduced the Transformer model, which relies on self-attention mechanisms to process data in parallel, significantly reducing the need for sequential data processing and enabling the model to weigh the importance of different parts of the input data. The adaptation of Transformer models for image classification, notably through architectures like Vision Transformer, has opened new avenues for research and application. Unlike CNNs, Transformers do not inherently process spatial hierarchies but instead treat the image as a sequence of patches, applying self-attention to understand the global context of the image, which can lead to superior performance in certain contexts. This paradigm shift toward using Transformers for image classification highlights the field's ongoing evolution and the continuous search for models that can more effectively capture and interpret the complex patterns present in visual data. Despite their promising capabilities, the adoption of Transformer models in image classification also presents challenges, including the need for large-scale datasets for training and higher computational resources, setting the stage for ongoing research and development in optimizing these models for wider application. To address the challenges posed by the adoption of Transformer models in image classification, researchers have focused on designing lightweight and efficient Vision Transformers that require fewer computational resources and can be trained with smaller datasets. Key methods include the introduction of techniques such as model pruning [19-22], where redundant or non-essential parts of the model are removed without significantly impacting performance, and knowledge distillation [23-24], where a smaller, more efficient model is trained to emulate the performance of a larger, more complex model. Additionally, some approaches utilize more efficient self-attention mechanisms [25-30], which reduce the computational complexity by focusing on only the most relevant parts of the input data, and employing token-based methods that decrease the number of input tokens to the Transformer, significantly reducing the computational load while maintaining high accuracy. These innovations represent a significant step towards making Vision Transformers more accessible for a broader range of applications, especially in environments with limited computational capacity.

B. Deep Models for Classification Task in Production Industries

Deep learning models have profoundly impacted production industries by enhancing classification tasks with unprecedented accuracy and efficiency. In the manufacturing domain, Xu et al. [31] proposed a model that leverages high-resolution vision sensors and deep learning techniques to classify and rate multi-category steel scrap. The model significantly improved the accuracy and fairness of steel scrap quality evaluation in recycling processes. Vikanksh et al. [32] introduced the NSLNet framework, which combines ImageNet for feature extraction with adversarial training in the feature space through Neural Structure Learning, aiming to overcome the challenges of limited annotated datasets and decreased prediction accuracy due to image perturbations in steel surface defect identification. In a study by Mathieu et al. [33], a method was introduced involving a three-step approach of data collection, classification, and supervised learning using CNNs. This method aims to automate quality control of open mouth bag sealings in industrial bagging systems. This study contributed a novel CNN architecture for the image classification of open mouth bags, demonstrating promising results in automating quality control within the food industry's industrial bagging systems.

Deep learning models have also found application in the agriculture sector. Padmapriya et al. [34] introduced a multi-stacking ensemble model combined with a novel feature selection algorithm, leveraging both machine learning and deep learning models for accurate multiclass soil classification, essential for smart agriculture advancements. Gill et al. [35] proposed a model that utilizes CNN, Recurrent Neural Network, and Long-short Term Memory deep learning methods for optimal image feature extraction and selection, applying these features to classify fruits effectively. The model outperformed traditional methods in handling the complex and heterogeneous nature of fruit recognition and classification tasks. Recently, Shewale et al. [36] proposed a model that utilizes deep learning, specifically CNNs, combined with image processing, to automatically extract features from leaf images for the identification, classification, and diagnosis of plant leaf diseases. This research provided an automated, high-precision disease diagnosis system for tomato plants that bypasses manual feature engineering and segmentation, offering a scalable solution for crop disease diagnosis globally through the application of deep learning on extensive, real-time image datasets.

In the food industry, specifically for assessing the quality of packaged food, Han et al. [37] introduced a study featuring a rapid, non-destructive method for estimating nut quality. This method uses hyperspectral imaging coupled with deep learning classification, specifically a CNN, to assess the quality of unblanched Canarium indicum kernels based on peroxide values. Kazi et al. [38] explored the use of transfer learning with classical and residual CNN architectures for classifying different types of fruits and their freshness, moving beyond traditional CNN implementations. In the textile industry, deep learning has introduced new capabilities for fabric inspection, identifying weaving faults, and ensuring pattern consistency. Huang et al. [39] introduced an efficient CNN model designed for fabric defect segmentation and detection, which requires only a minimal number of defect samples for training, thus significantly reducing manual annotation costs. Wei et al. [40] introduced the BIVI-ML model that integrates three bioinspired visual mechanisms (i.e., visual gain, attention, and memory) into a deep CNN framework to address the challenges of multilabel textile defect classification, such as intersected defects and label correlations. This approach enhances resolution, focuses attention on defects, and accurately associates relevant labels for effective multilabel classification.

III. METHODOLOGY

In this section, we first introduce the data collection process, detailing the system implemented within the production line to gather a diverse and representative dataset of bag images. Following this, the data augmentation subsection explains how we leverage both offline and online techniques to enhance the dataset's quality and variability. Lastly, we discuss the design and optimization of a lightweight deep model for bags classification, focusing on our custom modifications to the ResNet-18 architecture, which includes adjustments for efficient operation on resource-constrained devices like the Raspberry Pi 4.

A. Data Collection

Designing a deep learning-based system for classifying normal and abnormal bags in automatic bagging machines presents several significant challenges, particularly in the domain of data collection. The effectiveness of such a system is heavily dependent on the quality and variety of the dataset used to train it. One of the primary challenges is the vast diversity of bags. These bags can vary widely in terms of size, shape, material, color, and design. Consequently, it is necessary to obtain the broadest possible variety of bag types to ensure that the control quality can be effectively managed across all potential items the system might encounter. This diversity is critical to developing a robust model capable of accurately identifying anomalies in any given bag. Another significant challenge is the scarcity of abnormal bags. Abnormalities can range from minor defects such as slight tears or misprints to more significant issues like incorrect sizing or completely torn bags. However, these occurrences are typically rare in a well-maintained production environment. This scarcity presents a problem for data collection because deep learning models require a substantial amount of data to learn from. The insufficient number of abnormal bags means the system may not observe enough examples of abnormalities during the data collection phase, leading to a model that might struggle to recognize less common or more subtle defects. Data collection is also hindered by the dynamic conditions under which bagging operations occur. Factors such as lighting, background, and speed of the conveyor can significantly affect the quality of the images captured for training the model. Consistency in these conditions is challenging to maintain, yet critical for training a model that is reliable under the diverse conditions it will encounter in real-world applications.

To tackle the challenges associated with collecting a diverse and representative dataset for training a deep learning system for classifying normal and abnormal bags in automatic bagging machines, a sophisticated data collection system has been implemented directly into the automatic bagging machine's production line, as shown in Fig. 2. This system employs a high-resolution Vieworks CMOS VC-25MC-M/C 30D area camera, renowned for its exceptional image quality and reliability. The camera is equipped with a 6 mm fixed focal lens, providing a wide field of view while maintaining sufficient detail for identifying both gross and subtle abnormalities in bags. The camera's parameters have been meticulously adjusted to optimize the lighting conditions and speeds at which bags are processed on the production line. This adjustment ensures that the images captured are of high quality and reflect the diverse conditions under which the system must operate. By integrating the camera directly into the production environment, the data collection system is able to capture images of bags under the actual conditions they will be encountered, thereby enhancing the realism and applicability of the training data. This setup is linked to a computer system equipped with both a powerful CPU and a GPU. The GPU, in particular, is crucial for processing the high volume of image data in real-time, allowing for immediate feedback and adjustments to the data collection process if needed. This computational power also supports the use of advanced image processing and augmentation techniques, which can artificially expand the dataset by modifying existing images to simulate a

wider range of abnormalities and conditions. All the data collection hardware specifications are shown in Table I.

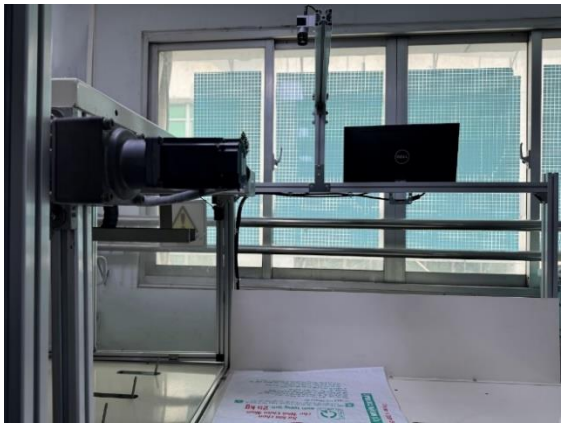


Fig. 2. Data collection system.

TABLE I. DATA COLLECTION HARDWARE SPECIFICATIONS

Hardware	Specifications
Camera	Model: Vieworks CMOS VC-25MC-M/C 30D Type: Area Camera
Lens	Type: 6 mm fixed focal lens Field of View: Wide
Computer	Intel(R) Core(TM) i7-11700K CPU + Nvidia RTX 4080 GPU

To address the issue of the rare occurrence of abnormal bags, the system is designed to flag and store images of any detected abnormalities for further review. This process helps in creating a focused dataset of abnormal bags, which, although smaller in size, is rich in diversity and critical for training the detection system effectively. Furthermore, to mitigate the imbalance between normal and abnormal bag instances, sophisticated data augmentation techniques are employed. This approach increases the representation of abnormal bags in the training dataset without the need for an equivalent increase in the actual occurrence of these abnormalities on the production line.

With the proposed data collection system, 1000 images have been collected. As in practical production, the operator is only concerned with whether the bag is an abnormal bag that needs to be discarded, not with the specific type of abnormality. Therefore, in this paper, the bags were classified into two categories, normal and abnormal. Normal bags are those that appear uniform in shape, with consistent dimensions and no visible defects on the surface. The integrity of each bag is maintained, and there are no signs of tears, misprints, or material weaknesses, as shown in Fig. 3(a). Abnormal bags, on the other hand, display various defects such as irregular shapes, tears, incorrect sealing, or material inconsistencies, as shown in Fig. 3(b). These abnormalities compromise the bag's functionality and potentially disrupt the operation of the machine, necessitating their removal from the production line.

B. Data Augmentation

Data augmentation plays a crucial role in the classification of normal and abnormal bags by significantly enhancing the diversity and volume of training data available for deep learning models. By artificially creating variations of the existing images through techniques such as rotation, noise injection, and perspective transformations, data augmentation helps models become more robust and less sensitive to small changes or imperfections in bag appearances. This process not only improves the model's ability to generalize across different conditions found in production environments but also addresses the challenge of limited samples of abnormal bags, thereby boosting the overall accuracy and reliability of the classification system. Broadly, data augmentation techniques can be categorized into two types: offline and online augmentations. Offline augmentation involves preprocessing and expanding the dataset before training begins. This means creating modified copies of the original images, such as rotated, flipped, or adjusted in terms of brightness and contrast, and adding them to the training set. This approach results in a statically enhanced dataset that the model trains on, allowing for a wide variety of data from the beginning. On the other hand, online augmentation takes place during the model training process itself. In this dynamic approach, images are augmented in real-time and fed into the model. This means that each epoch can present slightly different variations of the images to the model, introducing a richer set of examples over time. Techniques such as random cropping, zooming, or adding noise are applied in real-time, ensuring that the model rarely sees the exact same image twice. This not only improves generalization but also significantly enhances the model's robustness to new, unseen variations of bags.



a) Normal Bags



b) Abnormal Bags

Fig. 3. Examples of normal bags (a) and abnormal bags (b).

In this paper, we employ a comprehensive approach to data augmentation, leveraging both offline and online techniques to enhance the diversity and quality of our dataset for classifying normal and abnormal bags in automatic bagging machines. Initially, the dataset was expanded through offline augmentation methods, specifically focusing on brightness and contrast adjustments, noise injection, color variations, and synthetic abnormality generation. The latter is particularly noteworthy as it involves creating synthetic defects such as tears, holes, or significant shape distortions on images of normal bags. This method plays a crucial role in artificially expanding the collection of abnormal bag examples, avoiding the necessity for such events to happen naturally, and thus overcoming the lack of abnormal instances in the original dataset. Following the offline augmentation phase, we combined the enhanced images with the original ones to construct a new, enriched dataset comprising 3150 images. This dataset includes 1420 images depicting normal bags and 1730 images featuring abnormal bags, reflecting a more balanced distribution between the two categories. Subsequently, this augmented dataset was randomly divided into three subsets: a training set constituting 60% of the total images, a validation set comprising 20%, and a test set making up the remaining 20%. The distribution and details of the images within each subset are outlined in Table II.

TABLE II. NUMBER OF IMAGES IN EACH SUBSET

Subset	Number of images	
	Normal bags	Abnormal bags
Training	852	1038
Validation	284	346
Testing	284	346
Total	1420	1730

To further augment the diversity of data features available during model training, we implemented online augmentation techniques. During the training process, each image batch underwent preprocessing through a combination of methods, including color jittering to simulate varying lighting conditions and color schemes, Gaussian blurring to introduce variability in image sharpness and simulate minor camera focus issues, and random flipping (both horizontally and vertically) to ensure the model can accurately classify bags regardless of their orientation. This blend of online augmentation methods ensures that the model is exposed to a wide array of variations within the training data, significantly enhancing its ability to generalize from the training set to real-world scenarios where bags can appear under different conditions and with various types of abnormalities.

C. Lightweight Deep Model for Bags Classification

Since we use the Raspberry Pi 4 for classifying normal and abnormal bags in the automatic bagging machine, deploying a lightweight deep learning model on this board is necessary. The lightweight model is crucial because it is specifically designed to operate within the constrained computing resources and limited memory capacities typical of embedded systems. This model ensures that the classification process can be executed efficiently in real-time, maintaining high accuracy while minimizing latency, which is essential for integration into production lines. Furthermore, the optimized architecture of the model reduces power consumption, a critical consideration for continuous operation in industrial settings.

In our paper, we selected ResNet-18 as the foundation for our lightweight deep learning model to classify normal and abnormal bags, specifically designed for deployment on the Raspberry Pi 4. This choice was driven by ResNet-18's inherently efficient architecture that strikes an optimal balance between computational demand and model performance. We have conducted several modifications to ResNet-18, including changing the initial convolution layer based on Ghost Convolution [41] and reducing the number of channels, as well as employing knowledge distillation as a training technique to transfer knowledge from a fully trained and fine-tuned ResNet-18 model (teacher) to our optimized lightweight model (student). These enhancements further improve its suitability for real-time applications on hardware with limited computing resources. Compared to its deeper counterparts, such as ResNet-34 or ResNet-50, ResNet-18 offers a more practical solution for deployment on embedded systems like the Raspberry Pi 4. While deeper models might achieve slightly higher accuracy in certain contexts, their increased complexity and higher demand for computational resources make them less suitable for environments where power efficiency and low latency are paramount. The extensive use of Ghost Convolution, along with strategic knowledge distillation, allows our modified ResNet-18 model to maintain a competitive accuracy level while significantly reducing the necessary computational resources and power consumption. This balance is crucial for ensuring that the automatic bagging machine can operate continuously and efficiently in an industrial setting, making ResNet-18 the ideal choice for our application.

1) ResNet-18 architecture: ResNet-18 is a variant of the Residual Network (ResNet) architecture [10], designed to tackle the vanishing gradient problem that arises with increasing network depth. This architecture allows for the training of deep neural networks by introducing residual connections, which enable the flow of gradients through the network without degradation.

The architecture of ResNet-18, as shown in Fig. 4, consists of an initial convolutional layer followed by 16 convolutional layers organized into 8 residual blocks, and ends with an average pooling layer and a fully connected layer. The initial convolutional layer has a 7×7 kernel with 64 filters and a stride of 2. This layer is followed by a 3×3 max pooling layer with a stride of 2, which serves to reduce the spatial dimensions of the input image while preserving important features. Following the initial layers, ResNet-18 is composed of four main stages, each containing two residual blocks. Each block comprises two 3×3 convolutional layers with the same number of filters. The stages are differentiated by the number of filters and the stride of the first convolutional layer in each stage. Specifically, the stages have 64, 128, 256, and 512 filters, respectively. The stride is set to 1 for all blocks except the first block of each stage after the first, where it is set to 2. This design choice reduces the feature map's size as the network gets deeper, increasing the field of view of the convolutional filters.

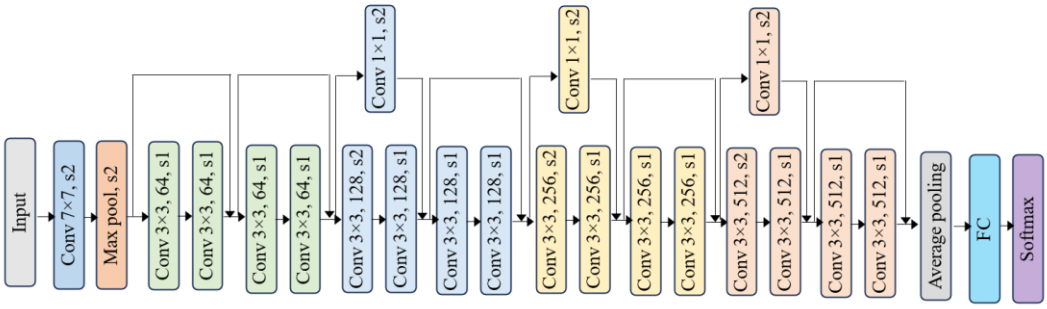


Fig. 4. ResNet-18 architecture.

The innovation of ResNet lies in its residual connections, which skip one or more layers by performing identity mapping and adding the input of the block to its output. These connections help mitigate the vanishing gradient problem by allowing the direct flow of gradients during backpropagation. In ResNet-18, every two layers share a residual connection, forming the backbone of its design. At the end of the network, a global average pooling layer reduces the spatial dimensions to 1×1, effectively summarizing the features extracted by the convolutions into a compact form. This is followed by a fully connected layer, which maps the pooled features to the desired number of output classes, facilitated by a Softmax activation function for classification tasks.

2) Improving initial convolutional layer: To address the high latency associated with the initial convolutional layer in ResNet-18, which is primarily due to its 7×7 convolution with a stride of 2, a modification is proposed to enhance computational efficiency while maintaining the layer's feature extraction capability. This modification involves increasing the stride of the initial convolutional layer to 4 and decreasing the kernel size from 7×7 to 5×5. The rationale behind this is twofold: a larger stride and a smaller kernel size directly reduce the amount of computation required, thereby lowering the latency. Furthermore, to compensate for the potential loss of feature extraction capability due to these reductions, a Ghost Convolution layer [41] is introduced right after the modified 5×5 convolutional layer. Ghost Convolution, as shown in Fig. 5, is a novel neural network architecture optimization technique designed to significantly reduce the computational cost and model size while preserving, or even enhancing, the model's performance. It achieves this by generating additional ghost feature maps from a smaller number of primary feature maps using inexpensive operations, thus efficiently utilizing computational resources. Serving as a pivotal innovation in deep learning, Ghost Convolution plays a crucial role in enabling more efficient and faster neural networks, particularly beneficial for deployment in resource-constrained environments such as mobile devices and edge computing platforms.

Building on the innovative approach outlined above, Table III presents a comparison of FLOPs before and after the modification of the initial convolutional layer to quantify the efficiency gains achieved through our modification. By adjusting the kernel size and stride, and by adding a Ghost Convolution layer, the initial convolutional layer achieves fewer FLOPs compared to the original convolutional layer. Through this strategy, the modified initial layer of ResNet-18 offers reduced latency without substantially compromising the network's performance, making it more suitable for real-time applications or deployment on hardware with limited computational resources such as the Raspberry Pi 4.

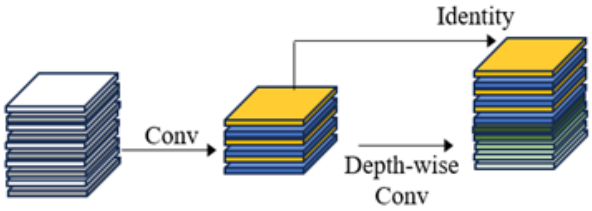


Fig. 5. Ghost convolution structure.

TABLE III. COMPARISON OF FLOPS BEFORE AND AFTER THE MODIFICATION OF THE INITIAL CONVOLUTIONAL LAYER

Configuration	Kernel Size	Stride	Additional Layers	FLOPs (Billions)
Original initial layer	7×7	2	0	1.8
Modified initial layer	5×5	4	1	1.6

3) Modifying channel numbers: In CNNs, the number of channels typically increases as the network progresses deeper. This design strategy originates from the need to capture increasingly complex features from the input data. Initially, layers detect simple patterns and textures, such as edges and colors. As we move deeper into the network, subsequent layers combine these basic features to detect more complex and abstract features, necessitating a larger number of channels to represent this growing complexity effectively.

In the case of ResNet-18, the structure follows this principle closely. The network starts with an initial convolutional layer that has 64 channels. This is followed by four main stages, each comprising two residual blocks. The channel numbers for each stage double as the network goes deeper: the first stage has 64 channels per layer, the second stage has 128 channels, the third stage has 256 channels, and the final stage has 512 channels. This design enables the network to process and extract a rich hierarchy of features from

the input images. However, not all tasks require the full capacity of ResNet-18. For simpler tasks, such as classifying bags as normal or abnormal, the complexity of the model can be reduced without significantly impacting performance. This simplification can lead to gains in efficiency, making the network more suitable for deployment in environments with limited computational resources like the Raspberry Pi 4. Based on extensive experiments, we have found that adjusting the number of channels in each layer of ResNet-18 to 64, 96, 128, and 160 for the respective stages strikes a good balance between performance and computational efficiency for this specific task. Table IV provides a comparison of FLOPs before and after reducing the number of channels in each layer. The results show that modifying the channel numbers significantly reduces the FLOPs of each stage.

TABLE IV. COMPARISON OF FLOPS BEFORE AND AFTER REDUCING THE NUMBER OF CHANNELS IN EACH LAYER

Stage	Original		Modified	
	Channels	FLOPs (Billions)	Channels	FLOPs (Billions)
2	64	0.35	64	0.35
3	128	0.55	96	0.4
4	256	0.4	128	0.3
5	512	0.25	160	0.2
FC layer	-	0.1	-	0.05
Total		1.65		1.3

Reducing the number of channels across the layers of ResNet-18 not only improves the model's computational efficiency, reflected in lower FLOPs, but also reduces the latency of the network during inference. This modification, however, comes at the cost of reduced network capacity. The term capacity here refers to the model's ability to learn from data; a higher capacity enables a network to capture more complex patterns but may also increase the risk of overfitting and require more data to train effectively. For the task of classifying normal and abnormal bags, which is relatively simple, this reduced capacity does not significantly hinder performance and leads to a more efficient model suitable for real-time applications or deployment on hardware with limited computational resources.

4) Knowledge distillation: Knowledge distillation is a powerful technique that can significantly improve the efficiency and accuracy of deploying the proposed model on the Raspberry Pi 4, especially for tasks of classifying normal and abnormal bags in an automatic bagging machine. The essence of knowledge distillation lies in transferring the knowledge from a large, complex teacher model to a smaller, more computationally efficient student model. This process allows the student model to learn the complex decision boundaries and the detailed representations captured by the teacher, without the need for extensive computational resources. In this paper, the teacher model is the fully trained and fine-tuned ResNet-18 model, which has been enhanced for better performance on the task of bag classification. The student model, on the other hand, is the simplified version of ResNet-18 with modifications to lower its computational demands. The distillation process involves running the dataset through both the teacher and student models, using the output probabilities (soft targets) of the teacher model as a guide for training the student model. These soft targets provide richer information compared to hard labels (normal/abnormal), as they contain insights about how the teacher model perceives the differences between classes, including the uncertainty and the relationships among them. To implement knowledge distillation effectively, we use a loss function that combines a traditional classification loss (i.e., cross-entropy against the true labels) with a distillation loss that measures the discrepancy between the teacher's predictions and the student's predictions. The distillation loss employs a temperature parameter to soften the probability distributions, making it easier for the student model to learn from the teacher's outputs. By employing knowledge distillation for the proposed model targeted for deployment on the Raspberry Pi 4, we can reduce model size and computational requirements while improving accuracy and enhancing inference speed.

5) Overall architecture of the proposed model: The overall architecture of the proposed model is shown in Fig. 6. It incorporates several modifications to ResNet-18 to enhance performance while accommodating the computational limitations of embedded systems. Initially, the model introduces a modified initial convolutional layer, where the stride is increased to 4 and the kernel size is decreased from 7×7 to 5×5. This modification aims to capture finer details of input images without excessively burdening the Raspberry Pi's computational resources. Following the initial convolutional layer, a Ghost Convolution layer is introduced. This layer plays a pivotal role in reducing the model's complexity by generating more feature maps from fewer parameters, thus efficiently enhancing the representational capacity without a substantial increase in computational demand. The core of the model is composed of successive Residual Blocks (Res blocks), specifically arranged to progressively refine the feature maps. The number of channels in each layer of ResNet-18 has been adjusted to 64, 96, 128, and 160 for the respective stages, optimizing the balance between computational efficiency and the model's ability to capture relevant features from the image data. This adjustment ensures that the model remains lightweight yet capable of processing the varying complexities of the bag images through its depth. Each Res block employs a combination of 3×3 convolutions, with some blocks incorporating a stride of 2 (denoted as 3×3 conv, s2) to reduce the dimensionality and focus the model's attention on salient features. At the end of the model, a Global Average Pooling layer is utilized to condense the feature maps into a form suitable for classification, effectively reducing the dimensionality and focusing the model's output. This is followed by a Fully Connected (FC) layer that makes the final decision, classifying the input image as either normal or abnormal. Moreover, we apply knowledge distillation as a training technique to transfer knowledge from a fully trained

and fine-tuned ResNet-18 model (teacher) to our optimized lightweight model (student). This approach allows the lightweight model to achieve higher accuracy by learning refined representations and decision boundaries, effectively mimicking the performance of the more cumbersome teacher model without the associated computational overhead.

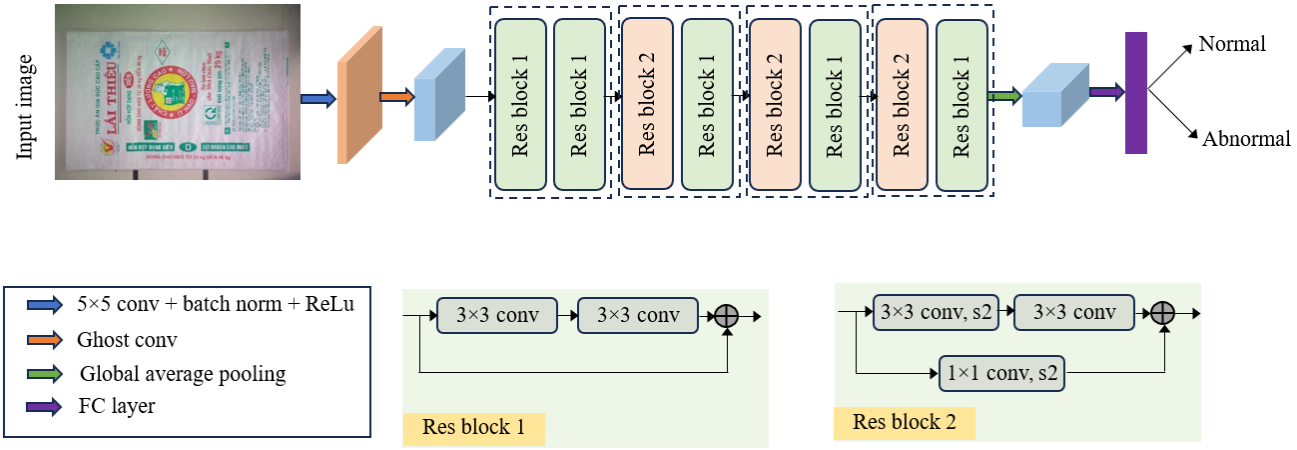


Fig. 6. Overall architecture of the proposed model.

IV. Experiments and Results

In this section, we first introduce a detailed description of the implementation setup and the evaluation metrics used to evaluate the performance of our model. Following this, we proceed into a comprehensive analysis of the results, drawing comparisons between the proposed model and both classical CNNs and existing lightweight neural network architectures. This comparison highlights the strengths and efficiencies of our model. Subsequently, we conduct an ablation study to evaluate the impact of various modules and modifications within our architecture, providing insight into their individual contributions towards the model's overall performance. Lastly, we discuss the deployment of our optimized model on the Raspberry Pi 4 within an actual automatic bagging machine setup, demonstrating its practical application and effectiveness in a real-world scenario.

A. Implementation Setup

For the training of our lightweight network designed to classify normal and abnormal bags in an automatic bagging machine, we employed a high-performance computing setup. The network was trained on a system equipped with an Intel(R) Core(TM) i7-11700K CPU, 32GB of RAM, and an Nvidia RTX 4080 GPU. This hardware configuration, supported by the CUDA 10.1 Toolkit, provided the necessary computational power to effectively train our model using TensorFlow, a popular deep learning framework known for its flexibility and extensive support for CNNs.

The training process lasted for 50 epochs. A batch size of 64 was chosen to balance the trade-off between memory usage and the granularity of the gradient update, ensuring efficient use of the GPU's resources. For optimization, we employed the Stochastic Gradient Descent (SGD) algorithm with a momentum of 0.9 and a weight decay of 0.0001. The initial learning rate was set to 0.01. The learning rate was scheduled to decrease after certain epochs based on performance metrics on the validation set, helping the model to fine-tune its weights more precisely as training progressed. The loss function chosen for this task was cross-entropy, a common choice for classification problems as it quantifies the difference between the predicted probabilities and the actual distribution, driving the model to make more accurate predictions over time.

B. Evaluation Metrics

In the task of classifying normal and abnormal bags, evaluating the performance of the model accurately is crucial to ensure its effectiveness in real-world applications. To achieve this, we employ several evaluation metrics, including accuracy, precision, recall, number of parameters, and FLOPs (Floating Point Operations), each offering unique insights into the model's capabilities and areas for improvement.

Accuracy is the simplest and most intuitive metric, representing the proportion of correctly classified instances (both normal and abnormal bags) to the total number of instances in the dataset. It is calculated using the following formula:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

where, TP (True Positives) and TN (True Negatives) are the correctly identified abnormal and normal bags, respectively, while FP (False Positives) and FN (False Negatives) represent the incorrectly classified instances.

Precision, or positive predictive value, measures the proportion of correctly identified abnormal bags out of all bags predicted as abnormal. This metric is particularly important in scenarios where the cost of falsely identifying a bag as abnormal (when it is not) is high. Precision is defined as the following formula:

$$Precision = \frac{TP}{TP+FP} \quad (2)$$

Recall, also known as sensitivity, indicates the proportion of actual abnormal bags that were correctly identified by the model. High recall is essential in ensuring that as many

abnormal bags as possible are detected. The formula for recall is defined as the following formula:

$$Recall = \frac{TP}{TP+FN} \quad (3)$$

In addition to these classification metrics, we also evaluate the model's complexity and efficiency using parameters and FLOPs. Parameters refer to the total number of trainable weights in the model. A lower count indicates a more lightweight model, which is beneficial for deployment on devices with limited computational resources, such as the Raspberry Pi 4. On the other hand, FLOPs provide a measure of the computational workload associated with a single forward pass through the model. It is calculated by adding up all the floating-point operations (additions, multiplications, etc.) involved in generating a prediction.

C. Main Results

1) Comparison with other CNNs models: In our study, we compared the performance of our proposed model against a range of established CNN architectures, including MobileNet-V1, MobileNet-V3, ResNet-50, ShuffleNet-V2, and VGG16 on the test dataset. This comparative analysis aimed to benchmark our model's effectiveness in classifying normal and abnormal bags against these well-known CNNs, with a particular focus on the balance between accuracy and computational efficiency as reflected in the model's precision, recall, parameters, and FLOPs. The comparison results are shown in Table V. The results show that our proposed model significantly outperforms the other architectures in terms of accuracy, achieving a remarkable 93.5%. This indicates a superior ability to correctly classify bags, which is critical in practical applications where misclassification can lead to operational inefficiencies or quality control issues. In terms of precision and recall, our model also leads with scores of 93.0% and 94.0%, respectively. These metrics suggest not only a high rate of correctly identifying abnormal bags but also an impressive capability to detect the majority of actual abnormal bags present in the dataset.

Despite its high performance, the proposed model maintains a moderate number of parameters and FLOPs, illustrating an efficient balance between computational cost and effectiveness. Notably, VGG16, with the largest model size and the highest computational cost, demonstrates lower performance metrics, highlighting the inefficiency of larger models in terms of computational resources versus accuracy gain. On the other hand, MobileNet-V3 and ShuffleNet-V2, known for their high efficiency, show remarkable performance with significantly fewer parameters and FLOPs, emphasizing the effectiveness of architectures designed for operational efficiency. ResNet-50, while having a substantial number of parameters and FLOPs, offers competitive performance metrics, which highlights the balance it strikes between depth and computational efficiency. However, our proposed model's performance, with its comparatively modest computational requirements, suggests that careful optimization and architectural choices can result in models that not only achieve superior accuracy but are also viable for deployment in resource-constrained environments. The results emphasize the importance of optimizing for both model size and computational efficiency without compromising on the task-specific performance, a key consideration for practical applications, especially in environments with limited computational resources like the Raspberry Pi 4.

TABLE V. COMPARISON RESULTS ON THE TEST DATASET OF DIFFERENT MODELS

Model	Accuracy (%)	Precision (%)	Recall (%)	Parameters (Millions)	FLOPs (Billions)
MobileNet-V1	89.5	88.7	90.2	4.2	0.57
MobileNet-V3	91.0	90.5	91.5	2.9	0.22
ResNet-50	92.3	91.8	92.7	25.6	4.1
ShuffleNet-V2	90.2	89.9	90.6	2.3	0.15
VGG16	88.0	87.5	88.5	138	15.5
Our Proposed Model	93.5	93.0	94.0	4.5	1.4

TABLE VI. CLASSIFICATION RESULTS OF DIFFERENT COMBINATIONS ON THE VALIDATION DATASET

Model	Accuracy (%)	Precision (%)	Recall (%)	Parameters (Millions)	FLOPs (Billions)
Original ResNet-18	89.1	88.6	89.2	11.6	1.8
ResNet-18 + IICL	91.6	91.2	91.8	8.4	1.6
ResNet-18 + IICL + MCN	91.4	90.9	92.9	4.5	1.4
ResNet-18 + IICL + MCN + KD (Our Proposed Model)	94.2	93.7	94.3	4.5	1.4

2) Ablation analysis: To illustrate the impact of various modifications and modules on the performance of our architecture, an ablation study was conducted comparing different variants of the ResNet-18 model on the validation dataset. These variants include the original ResNet-18 model, ResNet-18 with an improved initial convolutional layer (IICL), ResNet-18 with IICL and modified channel numbers (MCN), and finally, the complete proposed model which also incorporates knowledge distillation (KD). The comparison results are shown in Table VI. Starting with the original

ResNet-18, we observe a solid baseline with an accuracy of 89.1%, precision of 88.6%, and recall of 89.2%. This model, despite its relatively high computational cost (11.6 million parameters and 1.8 billion FLOPs), sets a foundation for further optimization. The introduction of an IICL, which includes increasing the stride and decreasing the kernel size, along with the addition of a Ghost Convolution layer, significantly boosts performance. This first modification enhances the model's efficiency in feature extraction and reduces computational requirements, resulting in improved accuracy (91.6%), precision (91.2%), and recall (91.8%), with a notable reduction in parameters (8.4 million) and FLOPs (1.6 billion). Interestingly, the third variant, which combines IICL with MCN, shows a slight decrease in accuracy and precision but a notable increase in recall (92.9%). This suggests that adjusting the number of channels effectively improves the model's sensitivity in detecting abnormal bags, a critical aspect of the classification task. This modification also significantly lowers the computational cost, halving the parameters to 4.5 million and reducing FLOPs to 1.4 billion, indicating a substantial increase in efficiency. Our proposed model, which further incorporates KD alongside IICL and MCN, achieves the best performance across all metrics: accuracy (94.2%), precision (93.7%), and recall (94.3%). This impressive improvement is achieved with the lowest complexity, featuring only 4.5 million parameters and 1.4 billion FLOPs. The addition of KD allows the model to learn more refined representations and decision boundaries, which is evident in its superior performance metrics. This indicates that knowledge distillation is highly effective in enhancing model performance, especially in tasks requiring high precision and recall.

Overall, the ablation study demonstrates the effectiveness of our targeted modifications in not only improving the model's accuracy, precision, and recall but also in significantly enhancing its computational efficiency. The final proposed model stands out as highly optimized for the specific task of classifying bags, making it ideal for deployment in resource-constrained environments such as the Raspberry Pi 4, where efficiency and performance are paramount.

3) Heatmap visualization: Fig. 7 visualizes heatmap results from the last convolution layer of the last block of the proposed model. The heatmaps offer valuable insights into the regions within the images that the proposed model focuses on when making classifications. These heatmaps are derived from the last convolutional layer of the last block of the model, highlighting the areas with the highest activations, typically the regions most significant for the model's decision-making process.

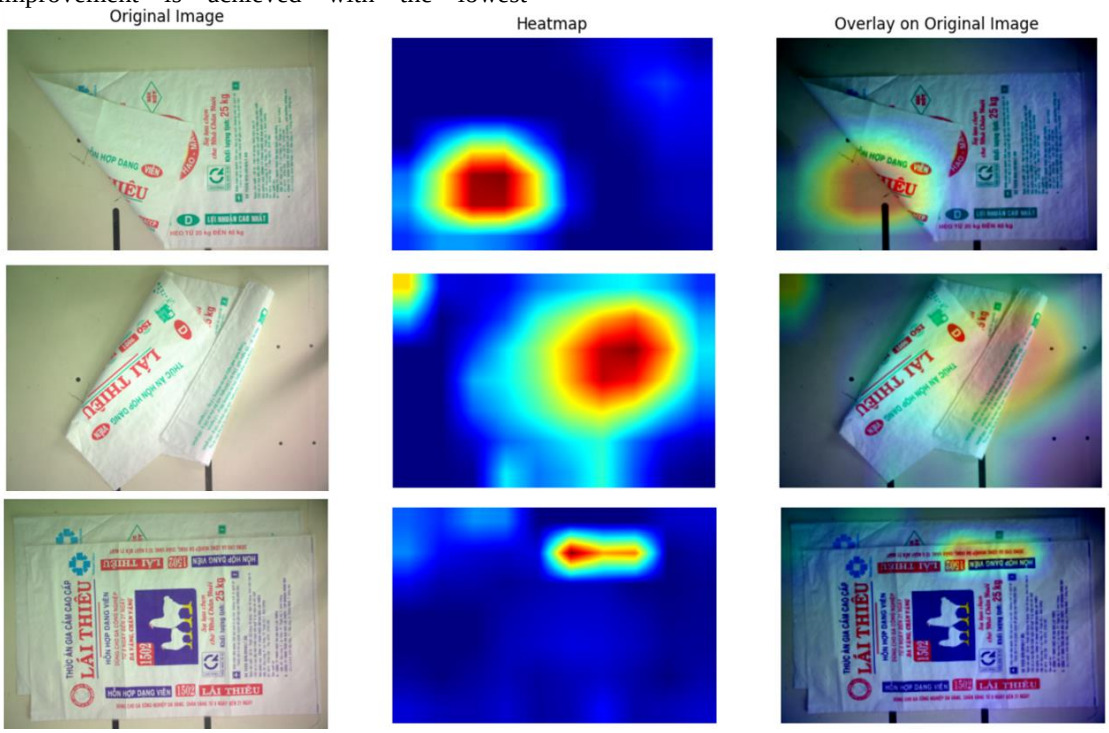


Fig. 7. Heatmap visualization of the proposed model.

In the first row, the heatmap is intensely focused on the area where part of the bag is missing, indicating that the model identifies this as the most informative region for classifying the bag as normal or abnormal. The high-intensity area represents a strong response, which suggests a distinctive feature or pattern that the model has learned to recognize as indicative of the bag's status. In the second row, we see a similar pattern of focus. The model's attention is concentrated around the center of the bag, but with a more elongated spread along the vertical axis. This reflects the model's detection of abnormalities that have a more linear orientation or a change in the bag's texture or structure in that specific region. The third row presents a

narrower focus in the heatmap. The activation is concentrated in a smaller region, which suggests that the model has detected a very specific feature of interest that is highly relevant to the classification task. The tight clustering of high activation corresponds to a localized abnormality.

In summary, these heatmaps provide a clear representation of where the model is looking and what it considers critical for its classification decisions. The results show that the model is not distracted by the periphery of the images and consistently focuses on the central parts of the bags, where textual and logo features are most prominent. This consistent focus across different bags suggests that the model has learned a generalized understanding of where relevant features are likely to appear. Crucially, these heatmaps can be used not only to understand the model's behavior but also to validate whether the model is considering the right features when making a classification. If the model were focusing on irrelevant areas, it might indicate an overfitting to noise or a misalignment in the learned features. However, the heatmaps in this figure confirm the model's appropriate focus areas, thus supporting its reliability and robustness in identifying normal and abnormal bags.

4) Deployment: For deploying the proposed model from a powerful training environment to a resource-constrained platform in the Raspberry Pi 4, we perform several steps to ensure that the model retains its accuracy and efficiency in a production setting. The first step is to convert the model into a TensorFlow Lite model. TensorFlow Lite is a set of tools that enables on-device machine learning by optimizing the TensorFlow model for performance on lightweight devices. By converting the model to TensorFlow Lite model, we substantially reduce its size while maintaining critical aspects of its performance. The converted model is further optimized using post-training quantization technique, which not only decreases the model's size but also potentially speeds up inference times by using lower-precision calculations. This is particularly important for the Raspberry Pi 4, as it has less computational power and memory compared to the original training environment.

Following optimization, the TensorFlow Lite model is deployed onto the Raspberry Pi 4, which involves transferring the model file to the device and setting up the necessary inference libraries. Once in place, the model is integrated into the automatic bagging machine's control software, where it will process input images captured by the machine's cameras in real-time. The software preprocesses the input data according to the model's requirements, then feeds it into the model for inference. The inference libraries, optimized for the Raspberry Pi 4's ARM architecture, facilitate the execution of the model efficiently, ensuring that classification decisions are made swiftly to keep pace with the operational speed of the bagging machine. The lightweight nature of the TensorFlow Lite model allows for rapid inference, which is essential for the model to be practical in a production environment.

Finally, we conduct extensive testing to confirm that the model's performance on the Raspberry Pi 4 aligns with the results observed during its initial development and validation. This involves monitoring accuracy, speed of inference, and resource utilization under real-world conditions. Table VII provides results of the deployment of the proposed model on the Raspberry Pi 4. The results in Table VII indicate that the proposed model delivers a solid performance in an operational environment. An inference time of 500 ms per bag suggests that the model is performing real-time analysis at a viable speed for the automatic bagging machine, considering the balance between speed and the complexity of the task. Power consumption stands at 4 W, which is a testament to the Raspberry Pi's energy efficiency and the lightweight nature of the optimized TensorFlow Lite model. Such low power draw is ideal for continuous, long-term operation in industrial settings. The CPU utilization of 48% indicates that the model is utilizing less than half of the CPU's capabilities, which is significant as it leaves room for the Raspberry Pi 4 to handle other tasks concurrently, if necessary, without overloading the system. The memory usage is moderate at 350 MB, which falls well within the Raspberry Pi 4's RAM capabilities, ensuring that the model runs smoothly without memory bottlenecks. This level of resource usage supports the notion that the model is indeed lightweight and suitable for embedded systems. The model's accuracy, precision, and recall rates are exceptionally high at 94.2%, 93.7%, and 94.3%, respectively. These metrics almost mirror the performance during the training phase, which indicates a successful model optimization and conversion process with negligible loss in model efficacy. Such high values suggest that the model is highly reliable, making correct decisions most of the time, and is able to identify the majority of abnormal bags correctly. An inference throughput of two bags per second may seem modest but is generally sufficient for automatic bagging operations, suitable for the speed of the conveyor and the number of bags processed in a given timeframe.

TABLE VII. RESULTS OF THE DEPLOYMENT OF THE PROPOSED MODEL ON THE RASPBERRY PI 4

Metric	Value	Comments
Inference time	500 ms	Time taken for a single inference
Power consumption	4 W	Average power during model inference
CPU utilization	48%	CPU usage during model inference
Memory usage	350 MB	RAM used by the model during operation
Model accuracy	94.2%	Percentage of correctly classified bags
Model precision	93.7%	Proportion of true positives over total positives
Model recall	94.3%	Proportion of true positives over actual positives
Inference throughput	2 bags/sec	Number of bags classified per second

V. CONCLUSION

In conclusion, our research has successfully demonstrated the efficacy of a deep learning and computer vision-based system designed for the classification of normal and abnormal bags within an automatic bagging machine. Through the strategic integration of a sophisticated data collection system directly on the production line, we have created a rich dataset that accurately reflects the variability inherent in real-world manufacturing processes. The implementation of both offline

and online data augmentation methods has significantly enhanced the robustness of our dataset, preparing our model to handle diverse operational scenarios. Our modifications of the ResNet-18 architecture into a lightweight deep learning model have proven to be particularly well-suited for deployment on the resource-limited Raspberry Pi 4, maintaining high accuracy and efficiency in bag classification tasks. The extensive comparative analysis with other CNN models and the thorough ablation studies have underscored the advantages of our proposed model. Overall, the contributions of this work not only lie in the novel application of a deep learning-based approach to a specific industrial challenge but also in the advancement of deploying complex models to edge devices. The success of this project opens avenues for future research into similar applications across different sectors, fostering the integration of AI in industrial automation.

ACKNOWLEDGMENT

This research is funded by Industrial University of Ho Chi Minh City under grant number 23.1CND01 (Contract number 135/HĐ-ĐHCN).

REFERENCES

[1] Jung, Byeonggil, Heegon You, and Sangkyun Lee. "Anomaly Candidate Extraction and Detection for automatic quality inspection of metal casting products using high-resolution images." *Journal of Manufacturing Systems* 67 (2023): 229-241.

[2] Chen, Bingsheng, Huijie Chen, and Mengshan Li. "Automatic quality inspection system for discrete manufacturing based on the Internet of Things." *Computers and Electrical Engineering* 95 (2021): 107435.

[3] Hakami, Aisha, and Mohammed Arif. "Automatic inspection of the external quality of the date fruit." *Procedia Computer Science* 163 (2019): 70-77.

[4] Kawamura, Shuso, Motoyasu Natsuga, Kazuhiro Takekura, and Kazuhiko Itoh. "Development of an automatic rice-quality inspection system." *Computers and electronics in agriculture* 40, no. 1-3 (2003): 115-126.

[5] Xia, Jiaping, YuHyeong Jeong, and Jonghun Yoon. "An automatic machine vision-based algorithm for inspection of hardwood flooring defects during manufacturing." *Engineering Applications of Artificial Intelligence* 123 (2023): 106268.

[6] LeCun, Yann, Léon Bottou, Yoshua Bengio, and Patrick Haffner. "Gradient-based learning applied to document recognition." *Proceedings of the IEEE* 86, no. 11 (1998): 2278-2324.

[7] Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. "Imagenet classification with deep convolutional neural networks." *Advances in neural information processing systems* 25 (2012).

[8] Simonyan, Karen, and Andrew Zisserman. "Very deep convolutional networks for large-scale image recognition." *arXiv preprint arXiv:1409.1556* (2014).

[9] Szegedy, Christian, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. "Going deeper with convolutions." In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1-9. 2015.

[10] He, Kaiming, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. "Deep residual learning for image recognition." In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770-778. 2016.

[11] Howard, Andrew G., Menglong Zhu, Bo Chen, Dmitry Kalenichenko, Weijun Wang, Tobias Weyand, Marco Andreetto, and Hartwig Adam. "Mobilenets: Efficient convolutional neural networks for mobile vision applications." *arXiv preprint arXiv:1704.04861* (2017).

[12] Sandler, Mark, Andrew Howard, Menglong Zhu, Andrey Zhmoginov, and Liang-Chieh Chen. "Mobilenetv2: Inverted residuals and linear bottlenecks." In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 4510-4520. 2018.

[13] Howard, Andrew, Mark Sandler, Grace Chu, Liang-Chieh Chen, Bo Chen, Mingxing Tan, Weijun Wang et al. "Searching for mobilenetv3." In *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 1314-1324. 2019.

[14] Zhang, Xiangyu, Xinyu Zhou, Mengxiao Lin, and Jian Sun. "Shufflenet: An extremely efficient convolutional neural network for mobile devices." In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 6848-6856. 2018.

[15] Ma, Ningning, Xiangyu Zhang, Hai-Tao Zheng, and Jian Sun. "Shufflenet v2: Practical guidelines for efficient cnn architecture design." In *Proceedings of the European conference on computer vision (ECCV)*, pp. 116-131. 2018.

[16] Tan, Mingxing, and Quoc Le. "Efficientnet: Rethinking model scaling for convolutional neural networks." In *International conference on machine learning*, pp. 6105-6114. PMLR, 2019.

[17] Dosovitskiy, Alexey, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani et al. "An image is worth 16x16 words: Transformers for image recognition at scale." *arXiv preprint arXiv:2010.11929* (2020).

[18] Vaswani, Ashish, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, and Illia Polosukhin. "Attention is all you need." *Advances in neural information processing systems* 30 (2017).

[19] Fayyaz, Mohsen, Soroush Abbasi Koohpayegani, Farnoush Rezaei Jafari, Sunando Sengupta, Hamid Reza Vaezi Joze, Eric Sommerlade, Hamed Pirsiavash, and Jürgen Gall. "Adaptive token sampling for efficient vision transformers." In *European Conference on Computer Vision*, pp. 396-414. Cham: Springer Nature Switzerland, 2022.

[20] Pan, Bowen, Rameswar Panda, Yifan Jiang, Zhangyang Wang, Rogerio Feris, and Aude Oliva. "IA-RED $^ 2$: Interpretability-Aware Redundancy Reduction for Vision Transformers." *Advances in Neural Information Processing Systems* 34 (2021): 24898-24911.

[21] Rao, Yongming, Wenliang Zhao, Benlin Liu, Jiwen Lu, Jie Zhou, and Cho-Jui Hsieh. "Dynamicvit: Efficient vision transformers with dynamic token sparsification." *Advances in neural information processing systems* 34 (2021): 13937-13949.

[22] Xu, Yifan, Zhijie Zhang, Mengdan Zhang, Kekai Sheng, Ke Li, Weiming Dong, Liqing Zhang, Changsheng Xu, and Xing Sun. "Evo-vit: Slow-fast token evolution for dynamic vision transformer." In *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 36, no. 3, pp. 2964-2972. 2022.

[23] Zhang, Jinnian, Houwen Peng, Kan Wu, Mengchen Liu, Bin Xiao, Jianlong Fu, and Lu Yuan. "Minivit: Compressing vision transformers with weight multiplexing." In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 12145-12154. 2022.

[24] Touvron, Hugo, Matthieu Cord, Matthijs Douze, Francisco Massa, Alexandre Sablayrolles, and Hervé Jégou. "Training data-efficient image transformers & distillation through attention." In *International conference on machine learning*, pp. 10347-10357. PMLR, 2021.

[25] Child, Rewon, Scott Gray, Alec Radford, and Ilya Sutskever. "Generating long sequences with sparse transformers." *arXiv preprint arXiv:1904.10509* (2019).

[26] Katharopoulos, Angelos, Apoorv Vyas, Nikolaos Pappas, and François Fleuret. "Transformers are rnns: Fast autoregressive transformers with linear attention." In *International conference on machine learning*, pp. 5156-5165. PMLR, 2020.

[27] Kitaev, Nikita, Łukasz Kaiser, and Anselm Levskaya. "Reformer: The efficient transformer." *arXiv preprint arXiv:2001.04451* (2020).

[28] Chen, Chun-Fu, Rameswar Panda, and Quanfu Fan. "Regionvit: Regional-to-local attention for vision transformers." *arXiv preprint arXiv:2106.02689* (2021).

[29] Choromanski, Krzysztof, Valerii Likhosherstov, David Dohan, Xingyou Song, Andreea Gane, Tamas Sarlos, Peter Hawkins et al. "Rethinking attention with performers." *arXiv preprint arXiv:2009.14794* (2020).

[30] Chu, Xiangxiang, Zhi Tian, Yuqing Wang, Bo Zhang, Haibing Ren, Xiaolin Wei, Huaxia Xia, and Chunhua Shen. "Twins: Revisiting the

design of spatial attention in vision transformers." *Advances in neural information processing systems* 34 (2021): 9355-9366.

[31] Xu, Wenguang, Pengcheng Xiao, Liguang Zhu, Yan Zhang, Jinbao Chang, Rong Zhu, and Yunfeng Xu. "Classification and rating of steel scrap using deep learning." *Engineering Applications of Artificial Intelligence* 123 (2023): 106241.

[32] Nath, Vikanksh, Chiranjoy Chattopadhyay, and K. A. Desai. "NSLNet: An improved deep learning model for steel surface defect classification utilizing small training datasets." *Manufacturing Letters* 35 (2023): 39-42.

[33] Juncker, Mathieu, Ismaïl Khriss, Jean Brousseau, Steven Pigeon, Alexis Darisse, and Billy Lapointe. "A Deep Learning-Based Approach for Quality Control and Defect Detection for Industrial Bagging Systems." In *2020 IEEE 19th International Conference on Cognitive Informatics & Cognitive Computing (ICCI* CC)*, pp. 60-67. IEEE, 2020.

[34] Padmapriya, J., and T. Sasilatha. "Deep learning based multi-labelled soil classification and empirical estimation toward sustainable agriculture." *Engineering Applications of Artificial Intelligence* 119 (2023): 105690.

[35] Gill, Harmandeep Singh, G. Murugesan, Abolfazi Mehbodniya, Guna Sekhar Sajja, Gaurav Gupta, and Abhishek Bhatt. "Fruit type classification using deep learning and feature fusion." *Computers and Electronics in Agriculture* 211 (2023): 107990.

[36] Shewale, Mitali V., and Rohin D. Daruwala. "High performance deep learning architecture for early detection and classification of plant leaf disease." *Journal of Agriculture and Food Research* 14 (2023): 100675.

[37] Han, Yifei, Zhaojing Liu, Kourosh Khoshelham, and Shahla Hosseini Bai. "Quality estimation of nuts using deep learning classification of hyperspectral imagery." *Computers and Electronics in Agriculture* 180 (2021): 105868.

[38] Kazi, Aafreen, and Siba Prasada Panda. "Determining the freshness of fruits in the food industry by image classification using transfer learning." *Multimedia Tools and Applications* 81, no. 6 (2022): 7611-7624.

[39] Huang, Yanqing, Junfeng Jing, and Zhen Wang. "Fabric defect segmentation method based on deep learning." *IEEE Transactions on Instrumentation and Measurement* 70 (2021): 1-15.

[40] Wei, Bing, Kuangrong Hao, Lei Gao, and Xue-Song Tang. "Bioinspired visual-integrated model for multilabel classification of textile defect images." *IEEE Transactions on Cognitive and Developmental Systems* 13, no. 3 (2020): 503-513.

[41] Han, Kai, Yunhe Wang, Qi Tian, Jianyuan Guo, Chunjing Xu, and Chang Xu. "Ghostnet: More features from cheap operations." In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 1580-1589. 2020.

Implementing Optimization Methods into Practice to Enhance the Performance of Solar Power Systems

Luçiana Toti[1], Alma Stana[2], Alma Golgota[3], Eno Toti[4]
Department of Information Technology, Aleksander Moisiu University, Durres, Albania[1,2]
Department of Engineering Science and Maine, Aleksander Moisiu University, Durres, Albania[3]
Department of Automation, Polytechnical University of Tirana, Albania[4]

***Abstract*—The use of contemporary technological tools, as well as the modernization of curricula in the field of electronic and electrical engineering, is one of the main objectives of the academic staff at the Universities. Efforts to improve curricula and scientific research infrastructure find strong support from the CBHE programs funded by the EU. The purpose of this work is to include the optimization methods for improving of the photovoltaic system's performance using the digital technologies which develop students' theoretical and practical skills for a sustainable development in the field of energy. Optimizing of the photovoltaic system through the addition of a booster, MPPT controller to the existing architecture, as well as with the help of SIMULINK will increase the energy efficiency of the photovoltaic system, increasing the University's economic benefit and moreover, the ecological benefits for the population. The implementation of the optimization algorithms will increase the simulation skills of the academic staff and students for a more in-depth analysis related to the implementation of RES, an analysis which until now has only been developed through software data collection methods of the system. As case study it is the utilization of photovoltaic system in the University of Durres area, which is a sustainable development area in both the public and private sectors after the 2019 earthquake. The study brings a transdisciplinary approach that contribute in the education of the new generation towards a green society and economy. It includes knowledge of the field of electrical engineering in the direction of increasing the performance of renewable energy systems as well as the analysis of electronic circuits with the help of optimization algorithms and different ICT tools.**

***Keywords*—Optimization; photovoltaic systems; education; performance; controller; SIMULINK**

I. Introduction

The application of optimization techniques to enhance the efficiency of a photovoltaic (PV) system combines technology, design, and maintenance procedures. To make sure that the PV system runs at its maximum power point and maximizes energy output under variable weather conditions, we must use inverters or MPPT charge controllers.

The developed analyzes serve to create the history of the photovoltaic system' using, the analytical development of the current situation and to increase the skills in the design of further projects by implementing Renewable Energy Sources in general and photovoltaic systems in particular. These methods can also be used by curriculum programs of Albanian Universities and the Western Balkan' Universities in terms of highlighting the advantages of environmentally and economically sustainable development.

The PV system's performance can be immediately monitored by setting up a monitoring system. Data analysis allows us to find trends, solve problems, and gradually improve system performance.

The optimization strategies have been applied in this paper for improving the performance of solar energy systems in generally for all photovoltaic systems. The University of Durres is used as the case study due to this, since a photovoltaic system was installed there thanks to the several funds invested by Erasmus+ programs.

Aleksander Moisiu University of Durres (UAMD) is Leader or Partner Institution for more than a hundred of International projects. We can mention the project "Development and implementation of multimedia and digital television curricula (DIMTV)", "Accelerating Western Balkans University Modernization by Introducing Virtual Technologies" (VTech@WBUni) projects [1], which AMUD has been Leader Institution. Participation in projects has significantly served in increasing the capacities of the academic staff of Aleksander Moisiu University of Durres, in exchanging of experiences for the academic, administrative staff as well as students among partner universities in the region and European, improved the research infrastructure, participate in various activities such as workshops, conferences, seminars, etc.

Aleksander Moisiu University of Durres was a partner institution in the "Knowledge Triangle for a Low Carbon Economy" (KALCEA) project, whose objectives enabled the development and growth of cooperation between university, business and scientific research centers. One of his objectives was to emphasize the urgent need for universities to increase their role in the social and economic development at the local and regional level [2]. Based on the establishing of strong connections with business/industry sector, improving technical and human capacities, improving research activities in cooperation with industry sector, improving the skills of students based on the market needs and more.

Through the knowledge triangle mechanism (knowledge-research-innovation) has been improved the curricula with innovative knowledge in the field of renewable. We have studied the photovoltaic system installed on the roof of one of the buildings of the university campus in Spitalle, Durrës, which was financed by EU funds.

For the design and installation of the photovoltaic system at the University of Durres, a service company in the field of renewable resources in Albania was contacted. The block scheme of his realization is shown in Fig. 1.

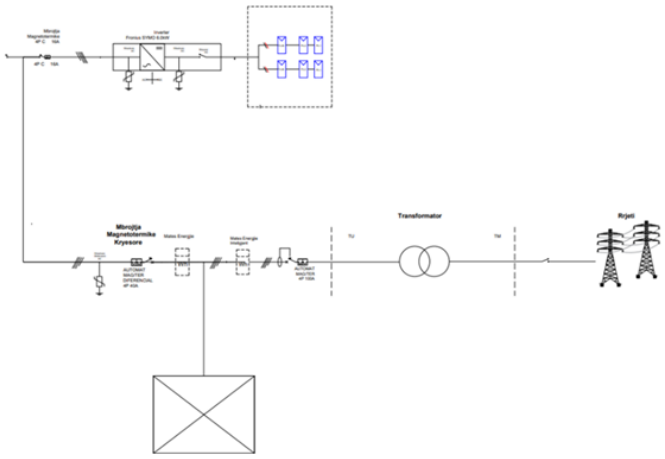


Fig. 1. The scheme of the implementation of the photovoltaic system at the university campus of Durres

The purpose of this paper is to improve the architecture of the photovoltaic system with added elements in the installed system with the purpose of improving the performance of the power generation system with the help of Simulink [3]. Improving the photovoltaic systems architecture will be done by adding a booster, MPPT controller, as well as using optimization algorithms in this controller [4]. In this study, the P&O algorithm will be used to compare the photovoltaic system's performance.

Additionally, the solar system will be examined for varying temperatures and irradiations utilizing a booster and two different types of controls (PD and PI type controllers). We must weigh the advantages and disadvantages of each technique used to determine which strategy will enhance solar system performance in general for all users and in particular for UAMD.

II. Literature Review

Researchers, academic institutions, and Universities have attempted to increase the efficiency of photovoltaic systems. Let's review a number of these researchers and their suggestions for enhancing photovoltaic system performance.

The theories of PV systems and MPPT approaches are introduced by A. Sadick, who also provides the mathematical modeling processes for the DC-DC boost converter and the PV system. Simulations and tests were conducted under various environmental conditions using two types of solar systems: one with the P&O algorithm and the other without [5].

The performance of a multiple gain boost converter connected to a solar photovoltaic (SPV) system with a grid-connected inverter is analyzed by B M Kiran Kumar, M S Indira, and S Nagaraja Rao. Regarding the amount of components and boost factor, the outcomes are contrasted with the case of a conventional booster converter. The overall harmonic deformations of the output current in this investigation are reported to be 1.66%, which is within permitted international criteria [6].

The aim of the simulation model developed by Indian university researchers is to capture the maximum amount of power by comparing its features to those of commercial panels under varying temperature and radiation circumstances. The investigation was conducted using MATLAB/Simulink [7].

The authors, Benaissa O. M., Hadjeri S., and Zidi S. A., have adjusted the power work cycle in converters to enhance the obtained power, with the aim of optimizing the photovoltaic system's efficiency by decreasing network losses [8].

Photovoltaic systems must be installed at locations with the highest power in order to generate electricity with the greatest possible advantage. Because of this, three sophisticated approaches to the photovoltaic system can be used with MATLAB/SIMULINK: PSO, Perturb-and-Observe (P&O), and Incremental Conductance (I_C). Following analysis, the PSO offers better power capture, less loss, and less deformation than the other two techniques [9], [10].

The theme study "Photovoltaic based high-efficiency single-input multiple-output dc-dc converter" by J. Kondalaiah and I. Rahul relied mostly on the use of a dc-dc converter, several inductors, and a power switch to increase system efficiency and control the output voltage [11].

The PV system's performance and the definition of M.P.P. are impacted by the addition of a DC-DC boost converter. One of the most crucial components is the input capacitor, which for optimal performance requires half the output capacitor's capacity [12].

In-depth research on the idea of MPPT techniques—which greatly boost the solar PV system's efficiency—is performed by Shazly A. Mohamed and Montaser Abd El Sattar. In an effort to maximize the PV system's energy conversion efficiency, they offer a simulation-based comparison analysis of incremental conductance and perturb and observe methods [13].

In the crucial area of PV system optimization, Maximum Power Point Tracking (MPPT) controllers are receiving a lot of attention. The efficiency, performance, modernism, complexity, and tracking speed of these controllers vary, and they use different algorithms. Considering their quick improvement, MPPT controllers fall into two categories: traditional and advanced techniques [14].

During the year 2024, it is noted that our partners in the KALCEA project are using digital twin design platforms, which will increase public awareness and engagement in energy efficiency. One of them is the Expedite Digital Twin, which will be applied to a district in Riga, Latvia and will provide practical guidelines, algorithms and training materials to help other cities replicate digital twins for their districts, driving adoption wide range of sustainable energy practices [15].

The design of intelligent systems is a disciplinary work that requires the optimal selection of specified sensors, defined data from these devices, their processing and implementation based on IoT [16].

The results obtained from the studies show that the implementation of relevant educational and promotional policies can mitigate the growth rate of electricity consumption. [17].

A concise overview of the literature allows us to draw attention to the challenges and limitations of existing approaches of current techniques. While conventional methods are relatively easy to use, they are unable to distinguish between local and global peaks. As a result, their efficiency is minimal if partial shading happens. Because of their higher efficiency, advanced tracking techniques are frequently employed [18]. Hybrid approaches find a way to overcome the shortcomings of the individual conventional and sophisticated approaches. The best MPPT method selection is still a problem that needs to be resolved; this problem can be resolved by conducting a survey of the available approaches. A succinct categorization and assessment of some the used MPPT techniques is provided by this study. Moreover, this study offers a readily available reference.

III. Methodology

The usage of digital microcontrollers in this study has the goal of improving system efficiency, and the results can be applied in other research projects.

Unlike other research conducted by many authors, this study differs from previous research in that it establishes the parameters, both electrical and physical as well as the performance of the photovoltaic system using the P&O algorithm and a booster for each of the systems studied with controllers of different types for various value of the radiation and temperature. The study also compares the results obtained and highlights for researchers which of the controllers provides the highest efficiency in the work of photovoltaic systems, contributing directly to optimizing future-designed systems.

Through this paper will compare the performance of the photovoltaic system using the P&O algorithm, as well as examine the solar system for different irradiations and temperatures with a booster and two distinct types of controls (PD and PI type controllers).

In this research, we need to evaluate the benefits and drawbacks of every technique implemented in order to identify which approach will improve solar system performance generally for all users and specifically for UAMD.

These studies methods with the help of SIMULINK are suitable for study purposes by students who are studying in the field of RES in the Professional Studies Faculty and Information Technology Faculty at Aleksander Moisiu University of Durres. The system installed and its testing with controller of different types will be a case study for students during the design and simulation of SFV as for Albanian and Western Balkan' Universities in improving their performance.

The system will be simulated in several different states to enable and analyze in terms of improving the performance of the photovoltaic system. The schematic block has been added the buster (DC-DC converter), as well as the PI/PD controller, in order to increase the effectiveness of the photovoltaic system. Specifically, the system is tested to calculate parameters related to the maximum power point under optimal conditions, i.e. temperatures of 25 °C and solar radiation value 1000W/m2 in these cases:

- System includes a booster and PD controller with P&O algorithms.
- System includes a booster and PI controller with P&O algorithms.

The results will then be compared, concluding which of them offers the highest efficiency of the system. Researchers can change concrete values of radiation and temperature during simulation, depending on the specific geographic area and climate conditions.

The system with the proposed architecture utilizes algorithms optimization at the point with maximum power generated by PV systems. These algorithms manage voltage values to ensure that the system operates at the peak of the power-voltage curve. As algorithms for the capture of MPPT, perturb and observe (P&O) algorithms are used.

The complete scheme of connecting the photovoltaic system to the proposed architecture is shown in Fig. 2.

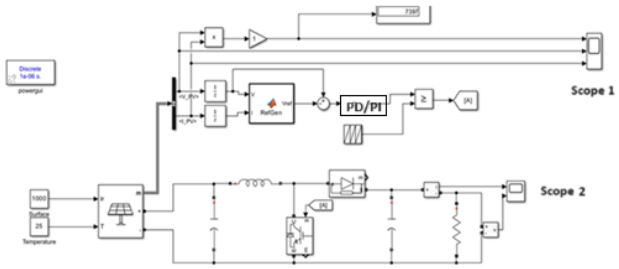


Fig. 2. The proposed architecture of the photovoltaic system.

In this scheme, these main blocks are used:

- Photovoltaic panel block
- Booster block
- PD/PI control block
- MATLAB Function block

Each of the blocks in the proposed architecture for the University of Durres will be highlighted in the sections that follow.

a) Photovoltaic panel block: The panels are selected SunPower Performances 5 UPP. The system has 18 monocrystalline panels, assembling with 10° tilt structures. This system has the installed power of 7.47 KWp. The surface area of the installed system is 46.5 m2, versus the area of the building of 649.3 m2, i.e. the area of the roof occupied from the modules is 7.16%. Coordinates of references: 41°06'35.2" N; 20°04'28.2" E.

The location, geographical position and dimensions of the installation space of photovoltaic panels are shown in Fig. 3.

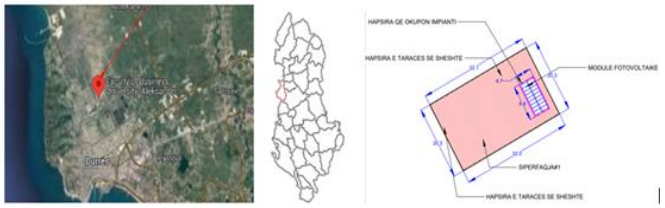
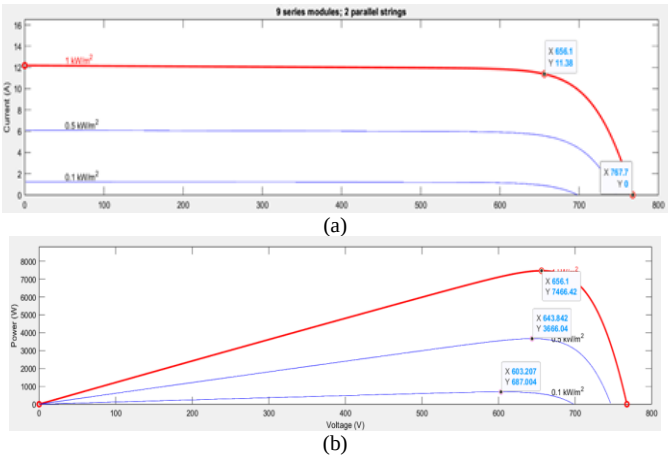


Fig. 3. Location and dimensions of the installation of photovoltaic panels.

Based on the data of the panels used in the photovoltaic system through the Matlab Code of P&O Algorithm, we can graphically determine the Volt-Ampere characteristic of nine panels connected in series in two parallel strings (Fig. 4a), as well as the characteristic where the point is clearly visible with maximum MPP power for different values of solar radiation (Fig. 4b).



(a)

(b)

Fig. 4. (a) Volt-Ampere characteristic of the used panels; (b) Determination of the maximum power point (MPP).

It is clear in Fig. 4(a) that the current passing through the used panels decreases with the decrease in the values of solar radiation referred to the same voltage values. Also from the volt-ampere characteristic it is noted that the current of the short-circuit (per V=0) decreases with the decrease of the values of solar radiation.

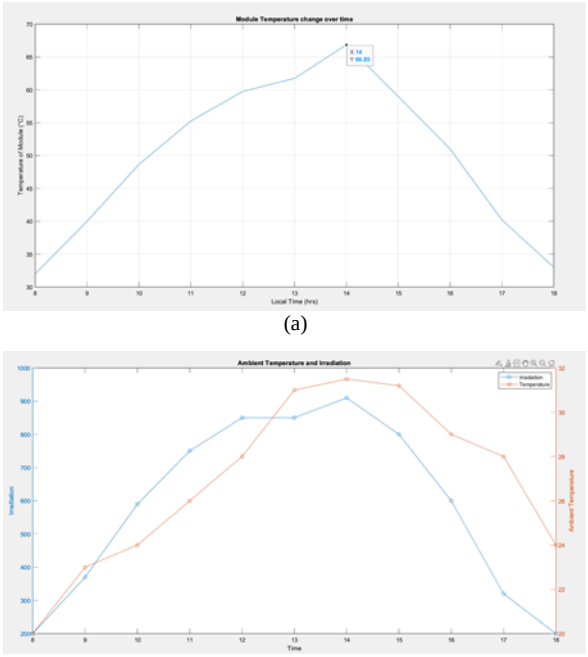
All the data obtained for the electric current, voltage and power of the system with photovoltaic panels for different values of solar radiation are summarized in the Table I:

TABLE I. DATA TABLE

Electrical and physical parameters	Value of parameters		
Solar radiation value (KW/m^2)	1	0.5	0.1
Maximum current value (A)	11.38	5.694	1.138
Maximum voltage value (V)	656.1	643.842	603.207
Maximum power value (KW)	7.466	3.666	0.687

Obviously from the above analysis, the maximum power value decreases with the decrease of the solar radiation value. This means that as the maximum power value decreases, the voltage and current values of the corresponding power values also decrease. In Fig. 4(b), it is seen that the maximum value of the reference voltage for the open circuit, that is, for I=0 A, the voltage value is 767.7 V, a value which also matches the data in the datasheet of the panels, where Open Circuit Voltage it is 85.3V for each of the panels; for the 9 panels connected in series, the open circuit voltage is 767.7 V, i.e. the same as the value determined graphically with the help of MATLAB SIMULINK.

During the study of the photovoltaic system, we can determine the temperature change of the modules, as well as the solar radiation for a time interval of 10 hours (8:00 a.m. - 6:00 p.m.) on a day in June. Fig. 5(a) shows graphically how the temperature of the module changes every hour of the specified time interval. With the help of MATLAB SIMULINK and the P&O algorithm, we can also determine the change in solar radiation in relation to the change in ambient temperature in each hour of the time interval studied, as in Fig. 5(b).



(a)

(b)

Fig. 5. (a) Module temperature change; (b) the change in solar radiation and ambient temperature for a certain time interval during a day in June.

As can be seen in Fig. 5(a), the panel's highest temperature is approximately 67ºC at 2:00 p.m. Conversely, Fig. 5(b) shows that solar radiation achieves its peak values at approximately 2:00 p.m., when the day's maximum temperature is approximately 32ºC.

The amount of sunshine and daytime temperature in June can be influenced by a variety of factors, including atmospheric conditions, geographic location, and June weather changes. Still, this is generally how it might appear on the midday. The solar radiation levels will progressively grow from low to moderate. The ambient temperature is warming when the sun gets stronger. During lunchtime the solar radiation peaks and receives the most amount of sunlight. Temperature range are warm to hot, with the greatest values of the day. Later in the afternoon the solar radiation will gradually decrease. The ambient temperature is still warm, but as the afternoon wears on, it begins to decrease.

b) Booster block: For the purpose of raising the input voltage, the booster block acts as a DC-DC converter. It is made up of multiple components, as shown in Fig. 6, including a cobbler, input capacitor, diode, MOSFET, shunt capacitor, and driving resistance. Since it stores energy and reduces deformations, the input capacitor is one of the most crucial components of the DC-DC converter [19]. Both continuous (CCM) and intermittent (DCM) modes of operation are available for this converter [20]. Managed by the controller, the work cycle determines the tension at the exit. At the booster's outlet are electric current and voltage measuring instruments, together with a scope (Scope 2) for observing the form, values, and boundaries of electrically enormous.

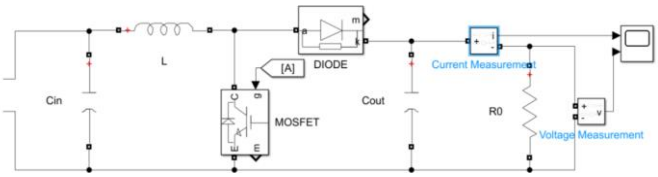


Fig. 6. Booster block, voltage & current measurement and the scope.

As the system's stability and efficiency rise, booster design and fundamental parameter selection play a critical role in enhancing system performance [21].

c) PID control block: PID control techniques are mostly used in the industry for the regulation of various physical processes. Control algorithms are found in terms proportional, derivative and integral [22], [23].

Proportional control is based on the difference between a given point and variable change, providing an immediate response [24].

Integral control is based on the accumulated error integral over time and reduces the stable state error by continuously stabilizing the output of the controller, thus eliminating the stable state errors or how else it is responsible for long-term reactions to changes in the signal of the input [25], [26].

Derivative control is based on measuring the rate of change of the error signal and regulating the output of the controller accordingly, thus improving the system's transitional response.

Each of the control types is specified by the corresponding coefficient, respectively Kp, Ki, Kd, which determine the controller's reaction force for each case. Their values must be tuned to improve the performance of the photovoltaic system. To connect a controller to photovoltaic systems, the other components are needed, which stabilizes the voltage and the current gained, thus preventing their fluctuations at the exit. Using of the controller ensures efficiency and security during energy generation. These control techniques can be combined by designing the PD (proportional-derivative), PI (proportional-integrative) controller or controller PID (proportional-Integrator-Derivative). In this paper we explain the role of PD/PI controllers in the performance of a photovoltaic system installed in Aleksander Moisiu University of Durres, as well as comparing their work.

d) MATLAB function block: With the help of this block, we design and analyze the photovoltaic system, as the most natural way of visualizing data and obtaining results through graphics. The codes used can be integrated with other languages, as well as adapted to much larger numbers of records, thus enjoying the properties of being scalar.

The technique used P&O (Perturb and Observe) serves to capture the maximum-powered MPPT point of the photovoltaic system. According to this method, the voltage at the exit of the photovoltaic system is adjusted by comparing the power value in the current period with that of the previous period. If the output voltage tends to increase, the control system tends to reduce it in order to keep it at stable power values at the exit.

Algorithm flowchart is shown below in Fig. 7:

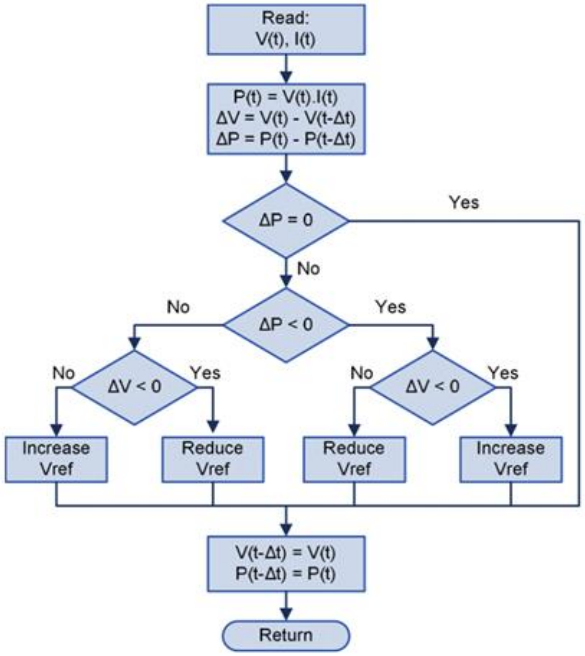


Fig. 7. P&O Algorithm flowchart [27].

IV. Results

To analyze the performance of the photovoltaic system with different types of control, we must follow the following steps:

- Select the type of controller and its parameters
- Identify the scheme
- I/O give out the details
- Stimulate
- Simulation executed

During the simulation, the temperature (25°C) and radiation (1000W/m^2) parameters are set. The researcher can modify the tracks as needed.

- Simulation of the photovoltaic system with Proportional Derivative (PD) controller

Firstly, after choosing the type of PD controller (Fig. 8), we connect photovoltaic system with the controller and follow the above-mentioned path.

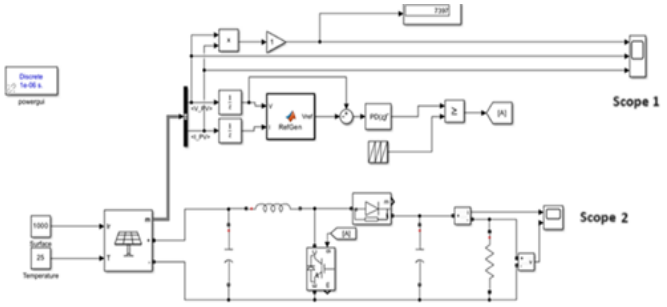


Fig. 8. Block scheme with PD controller.

In Scope 1 screen we can clearly distinguish the change of power, voltage and current in relation to the time for the system with the PD controller connected to it, while on the screen of the Scope 2 the flows of the current and voltage at the exit of the booster will appear.

The output views of scope 1 are for the initial moments of simulation of the photovoltaic system with PD controller (Fig. 9):

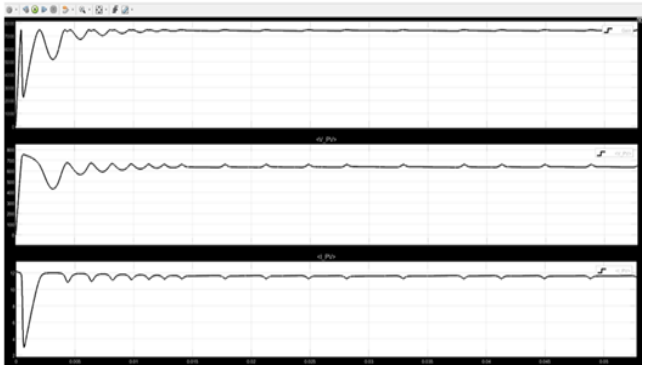


Fig. 9. The view of the oscilloscope 1 in the initial moments.

It is clear that at the initial moments of the waves last about 14 ms. The parameters (power, voltage, current) then start stabilizing, so the view of the oscilloscope 1 in the stabilized moments is shown in the Fig. 10.

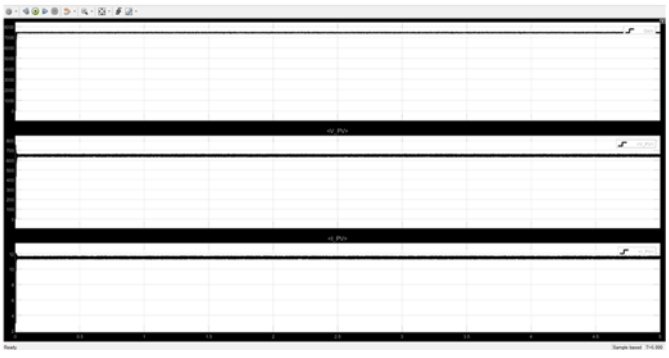


Fig. 10. Stabilized current values, voltage and power in case of PD controller scheme.

It is clear that the stabilized parameters are: power of about 7.47 KW, current value about 11.5 A, while voltage value is about 650 V.

The increase of the continuous voltage value with booster assistance from 650V to about 800V, affects the decrease of the current value from 11.5A to 9.3 A. This process is accompanied by significant variations in values, such as Fig. 11. Specifically:



Fig. 11. Fluctuations of current and voltage values at the exit of the booster in the PD controller scheme.

These fluctuations around average values exceed 5% of tolerance in terms of the range of deformations, so although the maximum power value is 7.47 KW. Current values fluctuate from (7.5-10) A, while voltage values from (650-870) V.

In this case the system is not considered in optimal working conditions for obtaining maximum power, for the maximum efficiency of the system.

- Simulation of the photovoltaic system with the Integral Proportional Controller (PI), as is shown in Fig. 12.

By including a PI controller in the suggested architecture of the photovoltaic system, the simulation of the scheme will be improved. Its operation, which is based on the average error of the integral over time, reduces the stable state error by consistently stabilizing the controller's output and therefore eliminating of the stable state errors. In other words, it can be held responsible for its long-term responses to changes in the input signal [28].

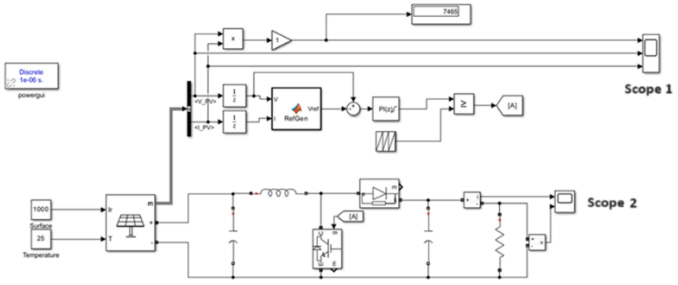


Fig. 12. Block scheme with PI controller.

The output views of scope 1 are for the initial moments of simulation of the photovoltaic system with PI controller (Fig. 13):

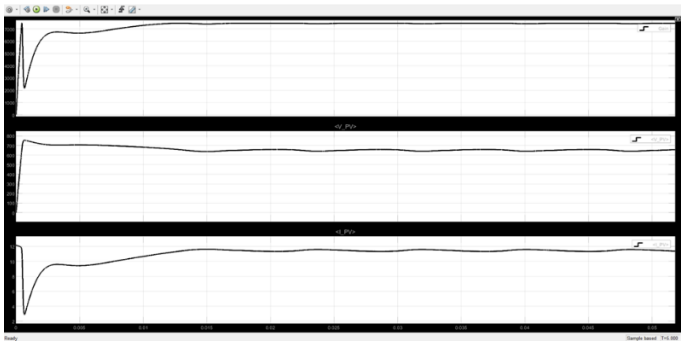


Fig. 13. The view of the oscilloscope 1 in the initial moments.

- As seen the largest leaches occur until the first 4ms, then the flow of variability drops significantly.

The current, voltage and power begin to stabilize, as shown in Fig. 14.

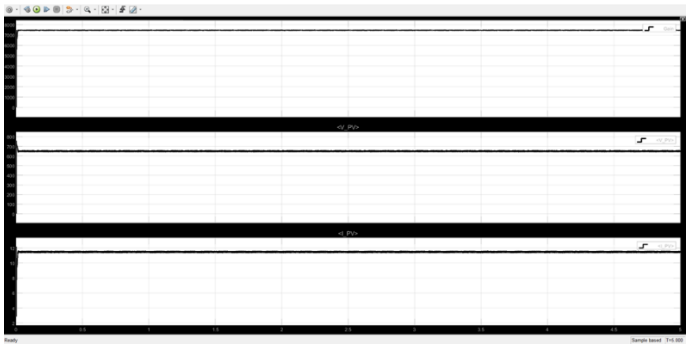


Fig. 14. Stabilized current values, voltage and power in case of PI controller scheme.

At the initial moment, the fluctuations are observed, which have a smaller duration than in the first instance until achieving the desired stable value. Stabilization time for the PI controller is 14 ms in the frame with the PD controller. At this moment the Update block command is given.

At the exit of the booster, the voltage and the current have the values around the average value below 5% of the average value, Current values fluctuate from (8.6 – 9.5) A, while voltage values from (770 - 805) V. The stabilization time is about 14 ms (Fig. 15), as well as in the system with PD controller. Obviously the deformations are more acceptable in the case of the PI controller scheme than in the system with the PD controller.

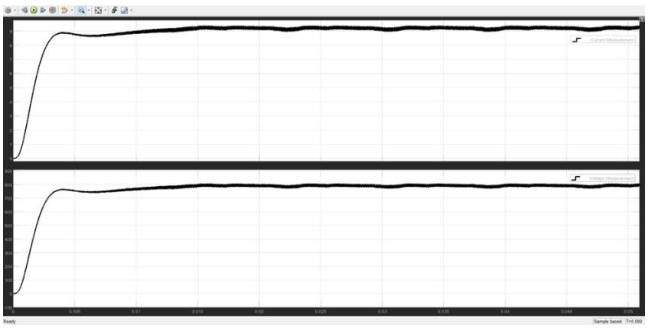
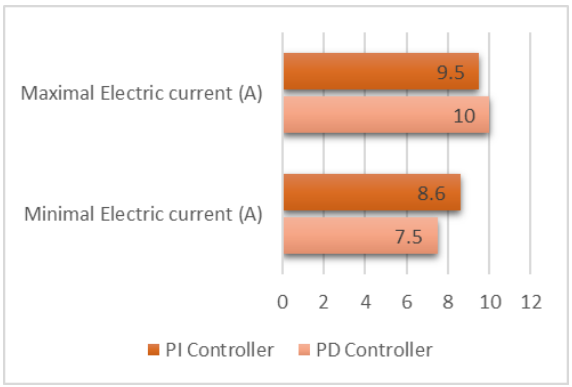
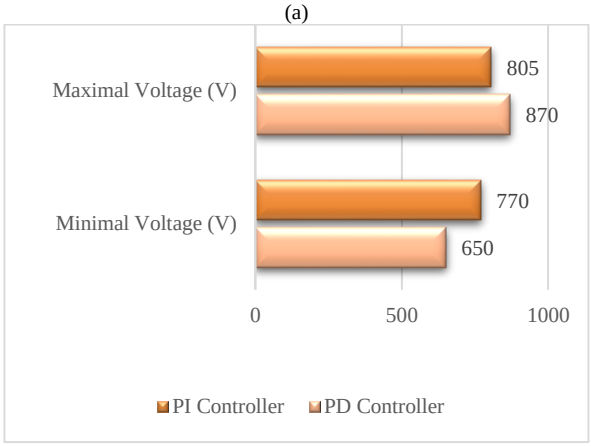


Fig. 15. Fluctuations of current and voltage values at the exit of the booster in the PI controller scheme.

Finally, we can compare the performance of systems with different types of controllers by summarizing the data for electrical quantities in the Fig. 16:



(a)



(b)

Fig. 16. Change rate of (a) current, (b) voltage for different controllers.

If we refer to the voltage and current variation limits for each of the controller types, we can obtain the results shown in the graph below (Fig. 17).

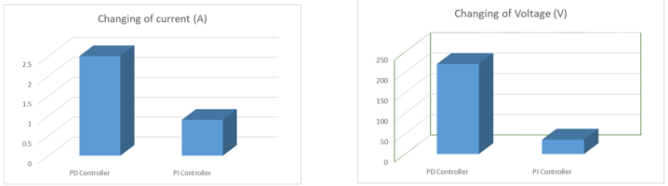


Fig. 17. The variation limits of current, voltage and time for the PD/PI controllers.

V. Discussion

Obviously, the changes in current and voltage at the output of the amplifier for the system with PI controller are smaller than for the one with PD controller. Also, the stabilization time of the power, current, voltage is almost the same in the PI controller system with the stabilization time of these electrical parameters at the output of the PD controller system.

So it is worth noting that in the conditions of the photovoltaic system with PI controller, the current and voltage fluctuations at the output of the booster are within the allowed values of 5% compared to the case of the system with PD controller.

Ultimately, installing a PI controller precisely on the photovoltaic system can improve its performance, not just at the University of Durres but also in other situations.

The developed analysis guides the researchers about the more efficient use of photovoltaic systems with the help of the booster, controller and performance improvement algorithms of these renewable energy systems [29], [30]. Improving the performance of photovoltaic systems with the help of the above methods leads to the capture of the point with maximum power in minimum time intervals [31].

The studies demonstrate the global power market's rapid expansion based on the evaluation of data at the worldwide level, data history across time, and contemporary analysis. By 2026, it is anticipated that installed energy storage capacity worldwide would exceed 270 GW [32]. The primary cause is the increasing demand for renewable energy systems' flexibility and storage in energy systems [33].

The methods that have been established until now will expand the applications of current renewable energy technologies with artificial intelligence support [34]. Their foundation is a zero-carbon economy. Renewable energy is a key technology for achieving carbon-free energy transitions [35].

The RES faced new chances and challenges during the Covid-19 crisis, but they remained a powerful resource [36]. For this reason, it is necessary to develop detailed market analyses and forecasts [37], [38].

So researchers should explore key industry challenges and identifying obstacles, to optimize renewable energy systems in order to accelerate the development of a low-carbon economy [39], [40].

VI. Future Work

We recommend that scientific researchers should continue to experiment the operation of photovoltaic systems with other types of controllers to take further steps forward in improving their performance. They should also improve optimization algorithms to increase energy efficiency. Lecturers should implement the optimization methods to the university education process.

VII. Conclusion

Researchers attempt to forecast and analyze a country's or region's energy performance based on various scenarios by utilizing modular components. Together with real-time data on energy generated by renewable sources, these components provide environmental visualization. The results can then be improved, evaluated and compared in order to arrive at the most efficient decision-making.

In order to improve the photovoltaic systems' efficiency, we can add a booster to the photovoltaic systems. This booster will raise the continuous voltage while maintaining the 7.47 KWp installed power and will also allow us to modify the stabilized output voltage values when they start to fluctuate for any reason. For this reason, we use P&O optimization algorithms for capturing the MPP, determining the values of electrical parameters (current, voltage) for different values of solar radiation; as well as for determining the physical parameters (temperature, solar radiation) of the panels within a certain time interval on a given day. Also with the help of optimization methods we can increase the performance of the photovoltaic system experimenting with different types of controllers. The developed analysis showed that the stabilization time of the voltage, current and output power is the same for the both controllers (PI/PD). Regarding the change of electrical quantities, the current fluctuates 36% less in the scheme with PI controller compared to that with PD controller, while the voltage changes around 29.2% less in the scheme with PI controller.

These results related to the performance of the photovoltaic system for the capture of MPPT with the help of the booster, controller and optimization algorithms will serve to modernize the curricula with contemporary methods that influence on the sustainable development towards a green society and economy. Also this study will serve scientific researchers, as well as all stakeholders for improving the work of systems installed as well as the design of new systems in the University of Durres and the Western Balkan Universities.

References

[1] Toti L., Kunicina N., “The Impact of Virtual Reality Technologies on the Modernization of the Curricula of Western Balkan Universities in the Field of Renewable Sources”, 2023, DOI:10.1109/IcETRAN59631.2023.10192195

[2] Toti, L., Durresi, M., Stana, A., Godole, F., & Eski, A. (2015), “The Use of Photovoltaic Panels-The New Economic Challenge for the Future: Albanian Case”, Academic Journal of Interdisciplinary Studies, 4(2), 67.

[3] Yadav D, Singh N. Simulation of PV System with MPPT and Boost Converter. Conference: National Conference on Energy, Power and Intelligent Control Systems. Dec 2020. Availbale from: https://www.researchgate.net/publication/347523004

[4] BEHZAD A. I., PARVIZ A. “Improved variable step size incremental conductance MPPT method with high convergence speed for PV systems”, Vol. 11, No. 4 (2016) 516 – 528

[5] Sadick A., “Maximum Power Point Tracking Simulation for Photovoltaic Systems Using Perturb and Observe Algorithm”, May 2023, DOI: 10.5772/intechopen.111632

[6] Kumar K. BM, Indira MS, Nagaraja Rao S., “Performance analysis of multiple gain boost converter with hybrid maximum power point tracker for solar PV connected to grid. Clean Energy”. 2021;5(4):655-672. DOI: 10.1093/ce/zkab037

[7] Kumar M.; Kachhwaya M. Kumar B. “Development of MATLAB/Simulink based model of PV system with MPPT”, 2016, DOI: 10.1109/IICPE.2016.8079336

[8] Benaissa O. M., Hadjeri S., Zidi S. A. “Modeling and Simulation of Grid Connected PV Generation System Using Matlab/Simulink”, 2017, ISSN: 2088-8694, DOI: 10.11591/ijpeds.v8i1.pp392-401

[9] Gouda E. A. Kotb M. F., Elalfy D. A. “Modelling and Performance Analysis for a PV System Based MPPT Using Advanced Techniques”, DOI: 10.24018/ejece.2019.3.1.47

[10] A. H. EL-Din , S. S. Mekhamer and H. M. El-Helw ,"Comparison of MPPT algorithms for photovoltaic systems under uniform irriadiance between PSO and P&O", October 2017, Vol.4, Issue.10.

[11] Kondalaiah J. , Rahul I. ," Photovoltaic based high-efficiency single- input multiple-output dc-dc converter", International Journal of Computer Science and Mobile Computing, October- 2014,Vol.3 Issue.10, pg. 483-496.

[12] Hayat A., Sibtain D., Murtaza A., Shahzad S., Jajja M., Kilic H. “Design and Analysis of Input Capacitor in DC–DC Boost Converter for Photovoltaic-Based Systems”

[13] Shazly A. M., Montaser Abd El Sattar. M. A. E., “A comparative study of P&O and INC maximum power point tracking techniques for grid-connected PV systems”, Springer Nature Switzerland AG 2019

[14] Baba AO, Liu G, Chen X. "Classification and evaluation review of maximum power point tracking methods. A Sustainable Future". 2020; DOI: 10.1016/j.sftr.2020.100020

[15] Valtasaari, S., (2024), exPEDite project: Enabling Positive Energy Districts through Digital Twins, https://oascities.org/expedite-project-enabling-positive-energy-districts-through-digital-twins/

[16] Bhoyar C, Kanojia K., Chourasia B, Shambharkar S., 2023-03-09, "Design of an Efficient IoT Based Irrigation Platform via Fuzzy Bioinspired Multisensory Analysis of on-Field Parameter Sets"

[17] Moosavirad S., Torabi A., Mirhosseini M., "Transdisciplinary approach to reduce electricity consumption using system dynamics", 2024-03-06

[18] Pradhan, A., & Panda, B. (2018) " A simplified design and modeling of boost converter for photovoltaic sytem" International Journal of Electrical and Computer Engineering, 8(1).

[19] Venkadesan, A., & Sedhu Raman, K. (2016) "Design and Control of Solar Powered Boost Converter", International Journal of Electrical and Electronics Research, 4(2), 132-137.

[20] Gouda EA, Kotb MF, Elalfy DA. Modelling and performance analysis for a PV system based MPPT using advanced techniques. European Journal of Electrical Engineering and Computer Science. 2019;. DOI: 10.24018/ejece.2019.3.1.47

[21] Hauke, B. (2022) "Basic Calculation of a Boost Converter's Power Stages" (SLVA372D –revised November 2022), Texas Instruments Incorporated.

[22] Fatema A. M., Mohiy E. Bahgat, Saied S. Elmasry and Soliman M. Sharaf, "Design of a maximum power point tracking-based PID controller for DC converter of stand-alone PV system", DOI 10.1186/s43067-022-00050-5

[23] Khan MJ, Kumar D, Narayan Y, Malik H, García Márquez FP, Gómez Muñoz CQ. A Novel Artificial Intelligence Maximum Power Point Tracking Technique for Integrated PV-WT-FC Frameworks. Energies. 2022; DOI: 10.3390/en15093352

[24] Tupe S., "PID Control for Solar Panel Temperature Regulation", 2023

[25] Oshaba, A.S., Ali, E.S. & Abd Elazim, S.M., "PI controller design for MPPT of photovoltaic system supplying SRM via BAT search algorithm", Neural Comput & Applic 28, 651–667, (2017).

[26] Sudhakar, K., & Srivastava, T., "Energy and exergy analysis of 36 W solar photovoltaic module" International Journal of Ambient Energy, 35(1), 51-57, (2014)

[27] De Brito, M. A., Sampaio, L. P., Luigi, G., e Melo, G. A., & Canesin, C. A. (2011, June) "Comparative analysis of MPPT techniques for PV applications", International Conference on Clean Electrical Power (ICCEP) (pp. 99-104), IEEE.

[28] Thankachy, G.D.K.N.S., Raj, S.T.P " Design of optimized PI controller for 7-level inverter: a new control strategy", Environ Sci Pollut Res 29, 43786–43799 (2022)

[29] Abderrahmane E., "MPPT Technique Based on Neural Network for Photovoltaic System In Renewable Energy and Energy Efficiency". Biblioteca Digital do IPB. 2021. Availbale from: https://bibliotecadigital.ipb.pt/bitstream/10198/24500/1/Elhor_Abderrahmane.pdf

[30] Di C, Magistrale L, Ingegneria N., "MPPT Algorithms Based on Artificial Neural Networks for PV System Under Partial Shading Condition". 2020. Availbale from:thesis.unipd.it/bitstream /20.500.12608/ 21143/1/ Salviati_Riccardo_1183982. pdf

[31] Zhu Y, Kim MK, Wen H. Simulation and Analysis of Perturbation and Observation-Based Self-Adaptable Step Size Maximum Power Point Tracking Strategy with Low Power Loss for Photovoltaics. Energies. 2018. DOI: 10.3390/en12010092.

[32] IEA 50, (2023), Analysis and forecasts to 2026, Energy End-uses and Efficiency Indicators Highlights, https://www.iea.org/reports/renewables-2021

[33] Zhou Y. Advances of machine learning in multi-energy district communities– mechanisms, applications and perspectives. Energy AI. 2022;10(July):100187. DOI: 10.1016/j.egyai.2022.100187

[34] Yap KY, Sarimuthu CR, Lim JMY. Artificial intelligence based MPPT techniques for solar power system: A review. Journal of Modern Power Systems and Clean Energy. 2020; 8(6):1043-1059. DOI: 10.35833/MPCE.2020.000159

[35] Roh C. Deep-learning algorithmic-based improved maximum power point-tracking algorithms using irradiance forecast. Processes. 2022;10(11). DOI: 10.3390/pr10112201

[36] Toti L., Kunicina N., "Sustainable development of the professional education during the period of COVID19", 2023, DOI 10.1109/RTUCON60080.2023.10413136

[37] Peck Yean Gan, ZhiDong Li, (2015), "Quantitative study on long term global solar photovoltaic market", Renewable and Sustainable Energy Reviews, Volume 46, Pages 88-99, doi.org/10.1016/j.rser.2015.02.041

[38] Seyedmahmoudian M., Jamei, E., Kok Soon Tey, Stojcevski, A., "Maximum power point tracking for photovoltaic systems under partial shading conditions using BAT algorithm. Sustainability", 2018;10(5):1-16. DOI: 10.3390/su10051347

[39] Palej1, P., Qusay, H., Kleszcz S., Hanus R., Jaszczur M., (2019), "Analysis and optimization of hybrid renewable energy systems", Polityka Energetyczna – Energy Policy Journal, Volume 22, Page 107–120, DOI: 10.33223/epj/109911

[40] Thirunavukkarasu, M., Sawle, Y., Lala H., (2023,)"A comprehensive review on optimization of hybrid renewable energy systems using various optimization techniques", Renewable and Sustainable Energy Reviews, Volume 176, doi.org/10.1016/j.rser.2023.113192

Combined Framework for Type-2 Neutrosophic Number Multiple-Attribute Decision-Making and Applications to Quality Evaluation of Digital Agriculture Park Information System

Wei Ji[1], Ning Sun[2]*, Botao Cao[3], Xichan Mu[4]

School of Economics and Management, Hubei Polytechnic University, Huangshi, 435003, Hubei, China[1]
College of Art and Design, Shaanxi University of Science and Technology, Xi'an, 710021, China[2]
Shaanxi University of Science and Technology, Xi'an, 710021, China[3]
Xi'an Sanlingweishi Information Security Co., Ltd, Xi'an, 710061, China[4]

***Abstract*—A digital agriculture park refers to an agricultural production and organizational unit of a certain scale where digital technology is used to optimize the agricultural supply chain. It enhances park management and service levels, achieving a new development model characterized by safe, low-carbon, high-quality, high-yield, precise, and efficient production, management, service, and operation. The quality evaluation of digital agriculture park information system is a multiple-attribute decision-making (MADM). Currently, the Exponential TODIM (ExpTODIM) and TOPSIS was put forward MADM. The Type-2 neutrosophic numbers (T2NNs) are employed to portray fuzzy information during the quality evaluation of digital agriculture park information system. In this works, the Type-2 neutrosophic number Exponential TODIM-TOPSIS (T2NN-ExpTODIM-TOPSIS) approach is put forward MAGDM under T2NNs. Finally, numerical study for quality evaluation of digital agriculture park information system is determined to demonstrate the T2NN-ExpTODIM-TOPSIS approach. The major research motivation is cultivated: (1) ExpTODIM and TOPSIS approach was enhanced under IFSs; (2) Entropy is put forward weight numbers in light with score values along with T2NNs; (3) T2NN-ExpTODIM-TOPSIS is put forward the MADM along with T2NNs; (4) numerical example for quality evaluation of digital agriculture park information system and different comparative analysis is put forward the validity of T2NN-ExpTODIM-TOPSIS.**

***Keywords*—Multiple-Attribute Decision-Making (MADM); Type-2 Neutrosophic Numbers (T2NNs); ExpTODIM approach; TOPSIS approach; quality evaluation**

I. INTRODUCTION

Building digital agriculture parks is a significant trend in modern agriculture, with countries worldwide gradually adopting this model. By integrating digital technologies like Internet of things, big data, and Artificial Intelligence, these parks enhance production, management, and services [1-3]. Currently, many countries are actively exploring and developing digital agriculture parks [4, 5]. Sensor networks and drone technology are widely used for land monitoring and crop management, enabling real-time data collection and precise management. This improves crop yield and quality while reducing the use of fertilizers and pesticides, thus lowering environmental pollution [6-8]. Digital parks also optimize production decisions through big data analysis. Farmers can adjust plans based on real-time data, enhancing resource efficiency and reducing costs. The introduction of intelligent management systems allows comprehensive monitoring and automation of production processes, increasing operational efficiency. Moreover, digital agriculture parks play a crucial role in food safety. Blockchain technology ensures traceability from production to sales, enhancing product safety and traceability [9-11]. This not only boosts consumer trust but also strengthens market competitiveness. The significance of building digital agriculture parks is substantial. Firstly, they increase agricultural productivity and efficiency. The application of digital technologies makes production processes more intelligent and precise, significantly improving resource utilization. Secondly, digital parks promote sustainable development. The use of precision agriculture technology reduces environmental pollution, achieving low-carbon and green agriculture. The development of digital agriculture parks also boosts rural economic growth. As park construction progresses, related industries develop, creating more jobs and improving rural economies[12-15]. Additionally, digital parks serve as testing grounds for technological innovation. The application and promotion of new technologies not only advance agricultural science but also provide references for innovation in other fields. However, there are challenges in building digital agriculture parks. High technical costs, data security issues, and farmers' acceptance of technology need to be addressed. Therefore, countries need to enhance policy support, drive technological innovation, and provide training to improve farmers' digital literacy. In the global context, the development of digital agriculture parks offers new opportunities for international cooperation. Countries can share technologies and experiences to address food security and environmental challenges together [16, 17]. Through global collaboration and technology sharing, digital agriculture parks will provide strong support for sustainable agricultural development worldwide. In summary, building digital agriculture parks not only enhances agricultural efficiency and safety but also promotes sustainable development and rural economic growth. It is a crucial path for modern agricultural

transformation. With continuous technological advancement, digital agriculture parks will play an increasingly important role, becoming a core model for future agricultural development [18, 19].

With the development of decision science and practical needs, research on MADM has become popular again. In the 1980s, some foreign scholars conducted research on group AHP and proposed some methods and some famous scholars in China also conducted relevant discussions [20-24]. Hwang and Yoon [25] systematically reviewed and summarized a large number of previous research results on MADM and edited and published the first monograph on multi-attribute decision-making, "Multiple Attribute Decision Making Method and Application.". After the 1980s, many scholars studied various types of interactive algorithms for solving the MADM problems. At that time, there were already some mature methods for studying MADM problems [26-31], such as the optimal selection method, connection method, and separation method used to screen options when there were too many options, the least squares method and eigenvector method used to determine attribute weights, the most commonly used simple additive weighting method and hierarchical additive weighting method for option ranking, the dictionary ordering method for selecting options based on attribute weight size, TOPSIS method [32] and LINMAP [33-37] method based on the concept of ideal solutions, as well as the queuing method based on estimating relative positions, linear allocation method, ELECTRE [38-41] method, and so on. The quality evaluation of digital agriculture park information system is MADM. Currently, the ExpTODIM [42, 43] and TOPSIS technique[32, 44, 45] is put forward the MADM. The Type-2 neutrosophic numbers (T2NNs) [46] are put forward for portraying fuzzy information during quality evaluation of digital agriculture park information system. Until now, no or few approaches were investigated on ExpTODIM and TOPSIS technique in light with entropy model along with T2NNs. Therefore, T2NN-ExpTODIM-TOPSIS model is put forward MADM along with T2NNs. Numerical example for quality evaluation of digital agriculture park information system and comparative analysis is put forward the validity of T2NN-ExpTODIM-TOPSIS approach. The major research motivation is cultivated: (1) ExpTODIM and TOPSIS approach was enhanced under T2NNs; (2) Entropy is put forward weight numbers in light with score values along with T2NNs; (3) T2NN-ExpTODIM-TOPSIS is put forward the MADM along with T2NNs; (4) numerical example for quality evaluation of digital agriculture park information system and comparative analysis is put forward the validity of T2NN-ExpTODIM-TOPSIS.

The structure of such study is cultivated. Preliminaries is given in Section II. In Section III, T2NN-ExpTODIM-TOPSIS model is put forward MADM along with T2NNs. Section IV numerical study for quality evaluation of digital agriculture park information system through different comparative analysis. Final conclusion is cultivated in Section V.

II. PRELIMINARIES

Wang et al. [47] built the SVNSs

Definition 1 [47]. The SVNSs is cultivated:

$$VA=\left\{\left(\theta,VT_{VA}\left(\theta\right),VI_{VA}\left(\theta\right),VF_{VA}\left(\theta\right)\right)\middle|\theta\in\Theta\right\} \quad (1)$$

where $VT_{VA}\left(\theta\right),VI_{VA}\left(\theta\right),VF_{VA}\left(\theta\right)$ is truth-membership (TM), indeterminacy-membership (IM) and falsity-membership (FM), $VT_{VA}\left(\theta\right),VI_{VA}\left(\theta\right),VF_{VA}\left(\theta\right)\in\left[0,1\right]$, $0\le VT_{VA}\left(\theta\right)+VI_{VA}\left(\theta\right)+VF_{VA}\left(\theta\right)\le 3$. The SVNN is implemented as $VA=\left(VT_A,VI_A,VF_A\right)$, where $VT_A,VI_A,VF_A\in\left[0,1\right]$, and $0\le VT_A+VI_A+VF_A\le 3.$

Abdel-Basset et al. [46] cultivated the T2NNs.

Definition 1[46]. The T2NN is cultivated:

$$VV=\left\{\left(\theta,VT\left(\theta\right),VI\left(\theta\right),VF\left(\theta\right)\right)\middle|\theta\in\Theta\right\} \quad (2)$$

where $VT\left(\theta\right),VI\left(\theta\right),VF\left(\theta\right)\in\left[0,1\right]$ be TM, IM and FM based on triangular fuzzy numbers.

$$VT\left(\theta\right)=\left(VT^L\left(\theta\right),VT^M\left(\theta\right),VT^U\left(\theta\right)\right),0\le VT^L\left(\theta\right)\le VT^M\left(\theta\right)\le VT^U\left(\theta\right)\le 1 \quad (3)$$

$$VI\left(\theta\right)=\left(VI^L\left(\theta\right),VI^M\left(\theta\right),VI^U\left(\theta\right)\right),0\le VI^L\left(\theta\right)\le VI^M\left(\theta\right)\le VI^U\left(\theta\right)\le 1 \quad (4)$$

$$VF\left(\theta\right)=\left(VF^L\left(\theta\right),VF^M\left(\theta\right),VF^U\left(\theta\right)\right),0\le VF^L\left(\theta\right)\le VF^M\left(\theta\right)\le VF^U\left(\theta\right)\le 1 \quad (5)$$

We let $VV=\left\{\begin{matrix}\left(VT^L,VT^M,VT^U\right),\\\left(VI^L,VI^M,VI^U\right),\left(VF^L,VF^M,VF^U\right)\end{matrix}\right\}$ be a T2NN, $0\le VT^U+VI^U+VF^U\le 3.$

Definition 2[46]. Let $VV_1=\left\{\begin{matrix}\left(VT_1^L,VT_1^M,VT_1^U\right),\\\left(VI_1^L,VI_1^M,VI_1^U\right),\left(VF_1^L,VF_1^M,VF_1^U\right)\end{matrix}\right\}$, and $VV_2=\left\{\begin{matrix}\left(VT_2^L,VT_2^M,VT_2^U\right),\\\left(VI_2^L,VI_2^M,VI_2^U\right),\left(VF_2^L,VF_2^M,VF_2^U\right)\end{matrix}\right\}$ and

$VV=\left\{\begin{matrix}\left(VT^{L},VT^{M},VT^{U}\right),\\ \left(VI^{L},VI^{M},VI^{U}\right),\left(VF^{L},VF^{M},VF^{U}\right)\end{matrix}\right\}$ be TFNNs, the operation laws are cultivated:

$$(1)\,VV_1\oplus VV_2=\left\{\begin{matrix}\left(VT_1^{L}+VT_2^{L}\text{-}VT_1^{L}VT_2^{L},VT_1^{M}+VT_2^{M}\text{-}VT_1^{M}VT_2^{M},VT_1^{U}+VT_2^{U}\text{-}VT_1^{U}VT_2^{U}\right),\\ \left(VI_1^{L}VI_2^{L},VI_1^{M}VI_2^{M},VI_1^{U}VI_2^{U}\right),\left(VF_1^{L}VF_2^{L},VF_1^{M}VF_2^{M},VF_1^{U}VF_2^{U}\right)\end{matrix}\right\};$$

$$(2)VV_1\otimes VV_2=\left\{\begin{matrix}\left(VT_1^{L}VT_2^{L},VT_1^{M}VT_2^{M},VT_1^{U}VT_2^{U}\right),\\ \left(VI_1^{L}+VI_2^{L}\text{-}VI_1^{L}VI_2^{L},VI_1^{M}+VI_2^{M}\text{-}VI_1^{M}VI_2^{M},VI_1^{U}+VI_2^{U}\text{-}VI_1^{U}VI_2^{U}\right),\\ \left(VF_1^{L}+VF_2^{L}\text{-}VF_1^{L}VF_2^{L},VF_1^{M}+VF_2^{M}\text{-}VF_1^{M}VF_2^{M},VF_1^{U}+VF_2^{U}\text{-}VF_1^{U}VF_2^{U}\right)\end{matrix}\right\};$$

$$(3)\,\lambda VV=\left\{\begin{matrix}\left(1-\left(1-VT^{L}\right)^{\lambda},1-\left(1-VT^{M}\right)^{\lambda},1-\left(1-VT^{U}\right)^{\lambda}\right),\\ \left(\left(VI^{L}\right)^{\lambda},\left(VI^{M}\right)^{\lambda},\left(VI^{U}\right)^{\lambda}\right),\\ \left(\left(VF^{L}\right)^{\lambda},\left(VF^{M}\right)^{\lambda},\left(VF^{U}\right)^{\lambda}\right)\end{matrix}\right\},\lambda>0;$$

$$(4)VV^{\lambda}=\left\{\begin{matrix}\left(\left(VT^{L}\right)^{\lambda},\left(VT^{M}\right)^{\lambda},\left(VT^{U}\right)^{\lambda}\right),\\ \left(1-\left(1-VI^{L}\right)^{\lambda},1-\left(1-VI^{M}\right)^{\lambda},1-\left(1-VI^{U}\right)^{\lambda}\right),\\ \left(1-\left(1-VF^{L}\right)^{\lambda},1-\left(1-VF^{M}\right)^{\lambda},1-\left(1-VF^{U}\right)^{\lambda}\right)\end{matrix}\right\},\lambda>0.$$

The operation laws have several properties.

$$(1)\,VV_1\oplus VV_2=VV_2\oplus VV_1,VV_1\otimes VV_2=VV_2\otimes VV_1,\left(\left(VV_1\right)^{\lambda_1}\right)^{\lambda_2}=\left(VV_1\right)^{\lambda_1\lambda_2};\tag{6}$$

$$(2)\,\lambda\left(VV_1\oplus VV_2\right)=\lambda VV_1\oplus\lambda VV_2,\left(VV_1\otimes VV_2\right)^{\lambda}=\left(VV_1\right)^{\lambda}\otimes\left(VV_2\right)^{\lambda};\tag{7}$$

$$(3)\,\lambda_1VV_1\oplus\lambda_2VV_1=\left(\lambda_1+\lambda_2\right)VV_1,\left(VV_1\right)^{\lambda_1}\otimes\left(VV_1\right)^{\lambda_2}=\left(VV_1\right)^{\left(\lambda_1+\lambda_2\right)}.\tag{8}$$

Definition 3 [46]. Let $VV=\left\{\begin{matrix}\left(VT^{L},VT^{M},VT^{U}\right),\\ \left(VI^{L},VI^{M},VI^{U}\right),\left(VF^{L},VF^{M},VF^{U}\right)\end{matrix}\right\}$ be TFNN, the SF (score function) and AF (accuracy function) are cultivated:

$$SF\left(VV\right)=\frac{1}{12}\begin{bmatrix}8+\left(VT^{L}+2VT^{M}+VT^{U}\right)\\ -\left(VI^{L}+2VI^{M}+VI^{U}\right)\\ -\left(VF^{L}+2VF^{M}+VF^{U}\right)\end{bmatrix},SF\left(VV\right)\in\left[0,1\right]\tag{9}$$

$$AF\left(VV\right)=\frac{1}{12}\begin{bmatrix}\left(VT^{L}+2VT^{M}+VT^{U}\right)\\ +\left(VI^{L}+2VI^{M}+VI^{U}\right)\\ +\left(VF^{L}+2VF^{M}+VF^{U}\right)\end{bmatrix},AF\left(VV\right)\in\left[0,1\right]\tag{10}$$

For two TFNNs, based on Definition 3, the following results are cultivated:

(1) *if* $SF(VV_1) \prec SF(VV_2)$, *then* $VV_1 \prec VV_2$;

(2) *if* $SF(VV_1) = SF(VV_2), AF(VV_1) \prec AF(VV_2)$, *then* $VV_1 \prec VV_2$;

(3) *if* $SF(VV_1) = SF(VV_2), AF(VV_1) = AF(VV_2)$, *then* $VV_1 = VV_2$.

Definition 4[46]. Let

$$VV_1 = \begin{Bmatrix} (VT_1^L, VT_1^M, VT_1^U), \\ (VI_1^L, VI_1^M, VI_1^U), (VF_1^L, VF_1^M, VF_1^U) \end{Bmatrix},$$

$$VV_2 = \begin{Bmatrix} (VT_2^L, VT_2^M, VT_2^U), \\ (VI_2^L, VI_2^M, VI_2^U), (VF_2^L, VF_2^M, VF_2^U) \end{Bmatrix}$$ be T2NNs, the T2NNs Hamming distance (T2NNHD) is cultivated:

$$T2NNHD(VV_1, VV_2) = \frac{1}{9}\begin{pmatrix} |VT_1^L - VT_2^L| + |VT_1^M - VT_2^M| + |VT_1^U - VT_2^U| \\ + |VI_1^L - VI_2^L| + |VI_1^M - VI_2^M| + |VI_1^U - VI_2^U| \\ + |VF_1^L - VF_2^L| + |VF_1^M - VF_2^M| + |VF_1^U - VF_2^U| \end{pmatrix} \quad (11)$$

III. T2NN-ExpTODIM-TOPSIS Approach for MADM with Entropy

The T2NN-ExpTODIM-TOPSIS is cultivated for MAGDM. Let $VA = \{VA_1, VA_2, \cdots, VA_m\}$ be alternative and $VG = \{VG_1, VG_2, \cdots, VG_n\}$ be attributes with weight vw, $vw_j \in [0,1], \sum_{j=1}^{n} vw_j = 1$. Then, T2NN-ExpTODIM-TOPSIS is cultivated for MADM (see Fig. 1).

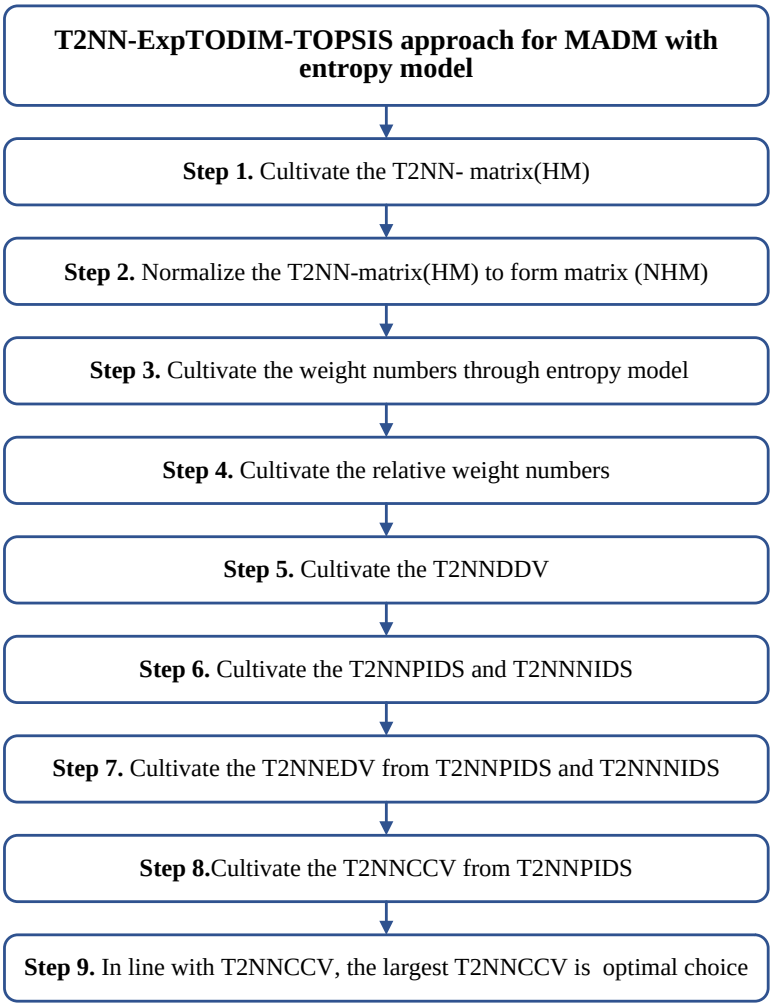


Fig. 1. T2NN-ExpTODIM-TOPSIS approach for MADM with entropy.

A. T2NN-MADM Problem

Step 1. Cultivate the T2NN-matrix $VM=\left[VM_{ij}\right]_{m\times n}$:

$$VM=\left[VM_{ij}\right]_{m\times n}=\begin{array}{c} \\ VA_1 \\ VA_2 \\ \vdots \\ VA_m \end{array}\begin{array}{c} \begin{array}{cccc} VG_1 & VG_2 & \cdots & VG_n \end{array} \\ \begin{bmatrix} VM_{11} & VM_{12} & \cdots & VM_{1n} \\ VM_{21} & VM_{22} & \cdots & VM_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ VM_{m1} & VM_{m2} & \cdots & VM_{mn} \end{bmatrix} \end{array} \tag{12}$$

$$VM_{ij}=\begin{Bmatrix} \left(\left(VI_{ij}^{L}\right),\left(VI_{ij}^{M}\right),\left(VI_{ij}^{U}\right)\right), \\ \left(\left(VI_{ij}^{L}\right),\left(VI_{ij}^{M}\right),\left(VI_{ij}^{U}\right)\right), \\ \left(\left(VF_{ij}^{L}\right),\left(VF_{ij}^{M}\right),\left(VF_{ij}^{U}\right)\right) \end{Bmatrix} \tag{13}$$

Step 2. Normalize $VM=\left[VM_{ij}\right]_{m\times n}$ into $NVM=\left[NVM_{ij}\right]_{m\times n}$.

Aiming at benefit attributes:

$$NVM_{ij}=\begin{Bmatrix} \left(\left(NVT_{ij}^{L}\right),\left(NVT_{ij}^{M}\right),\left(NVT_{ij}^{U}\right)\right), \\ \left(\left(NVI_{ij}^{L}\right),\left(NVI_{ij}^{M}\right),\left(NVI_{ij}^{U}\right)\right), \\ \left(\left(NVF_{ij}^{L}\right),\left(NVF_{ij}^{M}\right),\left(NVF_{ij}^{U}\right)\right) \end{Bmatrix}$$
$$=\begin{Bmatrix} \left(\left(VT_{ij}^{L}\right),\left(VT_{ij}^{M}\right),\left(VT_{ij}^{U}\right)\right), \\ \left(\left(VI_{ij}^{L}\right),\left(VI_{ij}^{M}\right),\left(VI_{ij}^{U}\right)\right), \\ \left(\left(VF_{ij}^{L}\right),\left(VF_{ij}^{M}\right),\left(VF_{ij}^{U}\right)\right) \end{Bmatrix} \tag{14}$$

Aiming at cost attributes:

$$NVM_{ij}=\begin{Bmatrix} \left(\left(NVT_{ij}^{L}\right),\left(NVT_{ij}^{M}\right),\left(NVT_{ij}^{U}\right)\right), \\ \left(\left(NVI_{ij}^{L}\right),\left(NVI_{ij}^{M}\right),\left(NVI_{ij}^{U}\right)\right), \\ \left(\left(NVF_{ij}^{L}\right),\left(NVF_{ij}^{M}\right),\left(NVF_{ij}^{U}\right)\right) \end{Bmatrix}$$
$$=\begin{Bmatrix} \left(\left(VF_{ij}^{L}\right),\left(VF_{ij}^{M}\right),\left(VF_{ij}^{U}\right)\right), \\ \left(\left(VI_{ij}^{L}\right),\left(VI_{ij}^{M}\right),\left(VI_{ij}^{U}\right)\right), \\ \left(\left(VT_{ij}^{L}\right),\left(VT_{ij}^{M}\right),\left(VT_{ij}^{U}\right)\right) \end{Bmatrix} \tag{15}$$

B. Cultivate the Weight Numbers Through Entropy

Step 3. Cultivate the weight numbers through entropy.

The weight is fundamental for MAGDM [48-52]. Entropy [53] is cultivated for weight numbers. The normalized fuzzy decision matrix (NFDM) is cultivated:

$$NFDM_{ij}=\frac{\left(\begin{array}{c} AF\begin{Bmatrix} \left(\left(NHA_{ij}^{L}\right),\left(NHA_{ij}^{M}\right),\left(NHA_{ij}^{U}\right)\right), \\ \left(\left(NHB_{ij}^{L}\right),\left(NHB_{ij}^{M}\right),\left(NHB_{ij}^{U}\right)\right), \\ \left(\left(NHC_{ij}^{L}\right),\left(NHC_{ij}^{M}\right),\left(NHC_{ij}^{U}\right)\right) \end{Bmatrix}+1 \\ \hline SF\begin{Bmatrix} \left(\left(NHA_{ij}^{L}\right),\left(NHA_{ij}^{M}\right),\left(NHA_{ij}^{U}\right)\right), \\ \left(\left(NHB_{ij}^{L}\right),\left(NHB_{ij}^{M}\right),\left(NHB_{ij}^{U}\right)\right), \\ \left(\left(NHC_{ij}^{L}\right),\left(NHC_{ij}^{M}\right),\left(NHC_{ij}^{U}\right)\right) \end{Bmatrix}+1 \end{array}\right)}{\sum_{i=1}^{m}\left(\begin{array}{c} AF\begin{Bmatrix} \left(\left(NHA_{ij}^{L}\right),\left(NHA_{ij}^{M}\right),\left(NHA_{ij}^{U}\right)\right), \\ \left(\left(NHB_{ij}^{L}\right),\left(NHB_{ij}^{M}\right),\left(NHB_{ij}^{U}\right)\right), \\ \left(\left(NHC_{ij}^{L}\right),\left(NHC_{ij}^{M}\right),\left(NHC_{ij}^{U}\right)\right) \end{Bmatrix}+1 \\ \hline SF\begin{Bmatrix} \left(\left(NHA_{ij}^{L}\right),\left(NHA_{ij}^{M}\right),\left(NHA_{ij}^{U}\right)\right), \\ \left(\left(NHB_{ij}^{L}\right),\left(NHB_{ij}^{M}\right),\left(NHB_{ij}^{U}\right)\right), \\ \left(\left(NHC_{ij}^{L}\right),\left(NHC_{ij}^{M}\right),\left(NHC_{ij}^{U}\right)\right) \end{Bmatrix}+1 \end{array}\right)}, \tag{16}$$

The fuzzy Shannon decision entropy (FSDE) is cultivated:

$$FSDE_j=-\frac{1}{\ln m}\sum_{i=1}^{m}NFDM_{ij}\ln NFDM_{ij} \tag{17}$$

and $NFDM_{ij}\ln NFDM_{ij}=0$ if $NFDM_{ij}=0$.

Then, the weight numbers $vw=\left(vw_1,vw_2,\cdots,vw_n\right)$ is cultivated:

$$vw_j=\frac{1-FSDE_j}{\sum_{j=1}^{n}\left(1-FSDE_j\right)} \tag{18}$$

C. T2NN-ExpTODIM-TOPSIS Approach for MADM

The T2NN-ExpTODIM-TOPSIS approach is cultivated for MADM.

Step 4. Cultivate relative weight numbers:

$$rvw_j=vw_j\Big/\max_j vw_j, \tag{19}$$

Step 5. Cultivate the The T2NN dominance degree values (T2NNDDV).

1) The T2NNDDV of VA_i over VA_t for VG_j is cultivated:

$$T2NNDDV_j\left(VA_i,VA_t\right)=\begin{cases}\dfrac{rvw_j\times\left(1-10^{-\rho T2NNHD\left(NVM_{ij},NVM_{tj}\right)}\right)}{\sum_{j=1}^{n}rvw_j} & if\ SF\left(NVM_{ij}\right)>SF\left(NVM_{tj}\right)\\ 0 & if\ SF\left(NVM_{ij}\right)=SF\left(NVM_{tj}\right)\\ -\dfrac{1}{v\theta}\dfrac{\sum_{j=1}^{n}rvw_j\times\left(1-10^{-\rho T2NNHD\left(NVM_{ij},NVM_{tj}\right)}\right)}{rvw_j} & if\ SF\left(NVM_{ij}\right)<SF\left(NVM_{tj}\right)\end{cases} \tag{20}$$

where, $v\theta$ is presented from Tversky and Kahneman [54] and $\rho\in[1,5]$ [55].

2) The $T2NNDDV_j\left(VA_i\right)\left(j=1,2,3,\cdots,n\right)$ with respect to VG_j is cultivated:

$$T2NNDDV_j\left(VA_i\right)=\left[T2NNDDV_j\left(VA_i,VA_t\right)\right]_{m\times m}$$

$$=\begin{array}{c}\\ VA_1\\ VA_2\\ \vdots\\ VA_m\end{array}\begin{array}{c}\begin{array}{cccc}VA_1 & VA_2 & \cdots & VA_m\end{array}\\ \begin{bmatrix}0 & T2NNDDV_j\left(VA_1,VA_2\right) & \cdots & T2NNDDV_j\left(VA_1,VA_m\right)\\ T2NNDDV_j\left(VA_2,VA_1\right) & 0 & \cdots & T2NNDDV_j\left(VA_2,VA_m\right)\\ \vdots & \vdots & \cdots & \vdots\\ T2NNDDV_j\left(VA_m,VA_1\right) & T2NNDDV_j\left(VA_m,VA_2\right) & \cdots & 0\end{bmatrix}\end{array}$$

3) Cultivate the overall T2NNDDV of VA_i over others for VG_j:

$$T2NNDDV_j\left(VA_i\right)=\sum_{t=1}^{m}T2NNDDV_j\left(VA_i,VA_t\right) \tag{21}$$

4) The overall T2NNDDD matrix is cultivated:

$$T2NNDDV=\left(T2NNDDV_{ij}\right)_{m\times n}$$

$$=\begin{bmatrix} & VG_1 & VG_2 & \cdots & VG_n\\ VA_1 & \sum_{t=1}^{m}T2NNDDV_1\left(VA_1,VA_t\right) & \sum_{t=1}^{m}T2NNDDV_2\left(VA_1,VA_t\right) & \cdots & \sum_{t=1}^{m}T2NNDDV_n\left(VA_1,VA_t\right)\\ VA_2 & \sum_{t=1}^{m}T2NNDDV_1\left(VA_2,VA_t\right) & \sum_{t=1}^{m}T2NNDDV_2\left(VA_2,VA_t\right) & \cdots & \sum_{t=1}^{m}T2NNDDV_n\left(VA_2,VA_t\right)\\ \vdots & \vdots & \vdots & \vdots & \vdots\\ VA_m & \sum_{t=1}^{m}T2NNDDV_1\left(VA_m,VA_t\right) & \sum_{t=1}^{m}T2NNDDV_2\left(VA_m,VA_t\right) & \cdots & \sum_{t=1}^{m}T2NNDDV_n\left(VA_m,VA_t\right)\end{bmatrix}$$

Step 6. Cultivate the T2NN positive ideal decision solution (T2NNPIDS) and T2NN negative ideal decision solution (T2NNNIDS):

$$T2NNPIDS=\left(T2NNPIDS_1,T2NNPIDS_1,\cdots,T2NNPIDS_n\right) \tag{22}$$

$$T2NNNIDS=\left(T2NNNIDS_1,T2NNNIDS_1,\cdots,T2NNNIDS_n\right) \tag{23}$$

$$T2NNPIDS_j=\max_{j=1}^{n}T2NNDDV_{ij}, \tag{24}$$

$$T2NNNIDS_j=\min_{j=1}^{n}T2NNDDV_{ij} \tag{25}$$

Step 7. Cultivate the T2NN Euclidean distance values (T2NNEDV) for T2NNPIDS and T2NNNIDS.

$$T2NNEDV\left(HA_i,T2NNPIDS\right)=\sqrt{\sum_{j=1}^{n}\left(T2NNDDV_{ij}-T2NNPIDS_j\right)^2} \tag{26}$$

$$T2NNEDV\left(VA_i, T2NNNIDS\right) = \sqrt{\sum_{j=1}^{n}\left(T2NNDDV_{ij} - T2NNNIDS_j\right)^2} \quad (27)$$

Step 8. Cultivate the T2NN closeness coefficient values (T2NNCCV) for T2NNPIDS.

$$T2NNCCV\left(VA_i, T2NNPIDS\right) = \frac{T2NNEDV\left(VA_i, T2NNNIDS\right)}{\left(\begin{array}{l} T2NNEDV\left(VA_i, T2NNPIDS\right) \\ +T2NNEDV\left(VA_i, T2NNPIDS\right) \end{array}\right)} \quad (28)$$

$$= \frac{\sqrt{\sum_{j=1}^{n}\left(T2NNDDD_{ij} - T2NNNIDS_j\right)^2}}{\left(\begin{array}{l} \sqrt{\sum_{j=1}^{n}\left(T2NNDDD_{ij} - T2NNNIDS_j\right)^2} \\ +\sqrt{\sum_{j=1}^{n}\left(T2NNDDD_{ij} - T2NNPIDS_j\right)^2} \end{array}\right)} \quad (29)$$

Step 9. Rank and choose the optimal scheme in line with maximum T2NNCCV.

IV. Numerical Example and Comparative Analysis

A. Numerical Example for Quality Evaluation of Digital Agriculture Park Information System

Since the 1990s, under the joint promotion of the Ministry of Agriculture, the Ministry of Science and Technology, and other ministries, China has initiated the construction of modern agricultural demonstration zones and agricultural science and technology parks at various levels. The aim is to cluster high-quality agricultural production elements such as science, policy, and funding in these parks to lead the transformation and development of agriculture. Some parks comprehensively utilize the Internet of Things, cloud computing, and automatic control technologies to enhance their production and management levels. As modern agricultural parks deepen their exploration of information systems, the precise and intelligent advantages of digital technology have been leveraged, enhancing the value of park development. However, this has also exposed some deficiencies, mainly manifested in the following aspects: the disconnection between information systems and the main business of the parks; information system operations becoming a "negative asset" as they fail to fully enhance management performance and continuously overdraw management costs; "cold wars" and redundant construction between information systems due to a lack of overall planning and coordination, leading to a lack of unified standards and coordination between old and new systems; the lack of distinct agricultural characteristics in the parks, with agricultural-specific sensors currently lacking and intelligent control based on operation types and models still in the theoretical research stage; the breadth of the potential of information technology to stimulate agricultural economy is insufficient, and the economic potential of the combination of information technology and agriculture has not been fully tapped. Digital agricultural parks are not limited to production functions; they can combine the multi-dimensional advantages of the parks to play multifunctional effects such as technology promotion, scientific and educational training, capital aggregation, and scientific research transformation, thereby enhancing the added value of the development of digital agricultural parks. The construction of digital agricultural parks is a systematic and all-inclusive project that requires the support of various management tools like performance, system, and organization to improve the priority level of digital agriculture, achieve top-down, unified resource organization effects, and effectively mobilize organizational enthusiasm to realize construction goals. It is necessary to focus on the financial sustainability during the construction and operational phases as an important indicator for the construction of digital agriculture, incorporating it into feasibility or planning studies. The construction of digital agricultural parks cannot be separated from the deep involvement of a professional team, which should at least include roles such as IT consultants, project managers, product managers, developers, and operations engineers, and maintain the team's continuity and stability in the development of the park to ensure the continuous advancement of the digitalization strategy. The quality evaluation of digital agriculture park information system is MADM. Therefore, the quality evaluation of digital agriculture park information system is presented to demonstrate T2NN-ExpTODIM-TOPSIS. Five potential digital agriculture park information systems $VA_i\left(i = 1,2,3,4,5\right)$ to assessed in line with different attributes: ①VG_1 is software function for digital agriculture park information systems; ②VG_2 is software performance for digital agriculture park information systems; ③VG_3 is management cost for digital agriculture park information systems; ④VG_4 is software supplier for digital agriculture park information systems. The VG_3 is cost attribute. The T2NN-ExpTODIM-TOPSIS is put forward the quality evaluation of digital agriculture park information system.

Step 1. Cultivate the T2NN-matrix $VM = \left[VM_{ij}\right]_{5\times 4}$ (see Table I).

TABLE I. T2NN INFORMATION

	VG_1	VG_2
VA_1	$\{(0.53,0.62,0.87),(0.48,0.64,0.75),(0.45,0.51,0.59)\}$	$\{(0.42,0.53,0.62),(0.27,0.46,0.58),(0.48,0.53,0.57)\}$
VA_2	$\{(0.32,0.43,0.64),(0.28,0.39,0.67),(0.25,0.36,0.48)\}$	$\{(0.49,0.57,0.69),(0.58,0.65,0.74),(0.45,0.53,0.61)\}$
VA_3	$\{(0.51,0.58,0.76),(0.54,0.56,0.67),(0.35,0.39,0.54)\}$	$\{(0.39,0.57,0.65),(0.51,0.64,0.79),(0.54,0.58,0.73)\}$
VA_4	$\{(0.49,0.67,0.71),(0.32,0.45,0.67),(0.26,0.38,0.56)\}$	$\{(0.62,0.74,0.83),(0.37,0.46,0.58),(0.32,0.36,0.49)\}$
VA_5	$\{(0.48,0.52,0.59),(0.27,0.36,0.46),(0.42,0.45,0.48)\}$	$\{(0.65,0.76,0.84),(0.52,0.54,0.63),(0.56,0.62,0.72)\}$
	VG_3	VG_4
VA_1	$\{(0.57,0.63,0.67),(0.29,0.52,0.63),(0.26,0.39,0.45)\}$	$\{(0.23,0.34,0.45),(0.43,0.51,0.57),(0.49,0.53,0.62)\}$
VA_2	$\{(0.39,0.47,0.53),(0.51,0.56,0.62),(0.34,0.57,0.69)\}$	$\{(0.43,0.47,0.52),(0.26,0.35,0.38),(0.46,0.49,0.52)\}$
VA_3	$\{(0.64,0.66,0.72),(0.73,0.81,0.85),(0.69,0.76,0.79)\}$	$\{(0.63,0.67,0.76),(0.37,0.45,0.52),(0.21,0.23,0.26)\}$
VA_4	$\{(0.25,0.34,0.62),(0.59,0.63,0.72),(0.46,0.53,0.62)\}$	$\{(0.38,0.42,0.47),(0.43,0.46,0.48),(0.32,0.43,0.53)\}$
VA_5	$\{(0.43,0.45,0.52),(0.43,0.52,0.65),(0.36,0.45,0.53)\}$	$\{(0.51,0.53,0.62),(0.43,0.54,0.65),(0.53,0.62,0.73)\}$

Step 2. Normalize the $VM = \left[VM_{ij} \right]_{5\times4}$ into $NVM = \left[NVM_{ij} \right]_{5\times4}$ (see Table II).

TABLE II. THE NORMALIZED T2NN

	VG_1	VG_2
VA_1	$\begin{Bmatrix}(0.53,0.62,0.87),\\(0.48,0.64,0.75),\\(0.45,0.51,0.59)\end{Bmatrix}$	$\begin{Bmatrix}(0.42,0.53,0.62),\\(0.27,0.46,0.58),\\(0.48,0.53,0.57)\end{Bmatrix}$
VA_2	$\begin{Bmatrix}(0.32,0.43,0.64),\\(0.28,0.39,0.67),\\(0.25,0.36,0.48)\end{Bmatrix}$	$\begin{Bmatrix}(0.49,0.57,0.69),\\(0.58,0.65,0.74),\\(0.45,0.53,0.61)\end{Bmatrix}$
VA_3	$\begin{Bmatrix}(0.51,0.58,0.76),\\(0.54,0.56,0.67),\\(0.35,0.39,0.54)\end{Bmatrix}$	$\begin{Bmatrix}(0.39,0.57,0.65),\\(0.51,0.64,0.79),\\(0.54,0.58,0.73)\end{Bmatrix}$
VA_4	$\begin{Bmatrix}(0.49,0.67,0.71),\\(0.32,0.45,0.67),\\(0.26,0.38,0.56)\end{Bmatrix}$	$\begin{Bmatrix}(0.62,0.74,0.83),\\(0.37,0.46,0.58),\\(0.32,0.36,0.49)\end{Bmatrix}$
VA_5	$\begin{Bmatrix}(0.48,0.52,0.59),\\(0.27,0.36,0.46),\\(0.42,0.45,0.48)\end{Bmatrix}$	$\begin{Bmatrix}(0.65,0.76,0.84),\\(0.52,0.54,0.63),\\(0.56,0.62,0.72)\end{Bmatrix}$
	VG_3	VG_4
VA_1	$\begin{Bmatrix}(0.26,0.39,0.45),\\(0.29,0.52,0.63),\\(0.57,0.63,0.67)\end{Bmatrix}$	$\begin{Bmatrix}(0.23,0.34,0.45),\\(0.43,0.51,0.57),\\(0.49,0.53,0.62)\end{Bmatrix}$
VA_2	$\begin{Bmatrix}(0.34,0.57,0.69),\\(0.51,0.56,0.62),\\(0.39,0.47,0.53)\end{Bmatrix}$	$\begin{Bmatrix}(0.43,0.47,0.52),\\(0.26,0.35,0.38),\\(0.46,0.49,0.52)\end{Bmatrix}$
VA_3	$\begin{Bmatrix}(0.69,0.76,0.79),\\(0.73,0.81,0.85),\\(0.64,0.66,0.72)\end{Bmatrix}$	$\begin{Bmatrix}(0.63,0.67,0.76),\\(0.37,0.45,0.52),\\(0.21,0.23,0.26)\end{Bmatrix}$
VA_4	$\begin{Bmatrix}(0.46,0.53,0.62),\\(0.59,0.63,0.72),\\(0.25,0.34,0.62)\end{Bmatrix}$	$\begin{Bmatrix}(0.38,0.42,0.47),\\(0.43,0.46,0.48),\\(0.32,0.43,0.53)\end{Bmatrix}$
VA_5	$\begin{Bmatrix}(0.36,0.45,0.53),\\(0.43,0.52,0.65),\\(0.43,0.45,0.52)\end{Bmatrix}$	$\begin{Bmatrix}(0.51,0.53,0.62),\\(0.43,0.54,0.65),\\(0.53,0.62,0.73)\end{Bmatrix}$

Step 3. Cultivate the weight numbers:

$$vw_1 = 0.3250, vw_2 = 0.3071$$
$$vw_3 = 0.1461, vw_4 = 0.2218.$$

Step 4. Cultivate the relative weight numbers:

$$rvw = (1.0000, 0.9449, 0.4495, 0.6825)$$

Step 5. Cultivate the $T2NNDDD = \left(T2NNDDD_{ij}\right)_{5\times4}$ (see Table III):

TABLE III. THE $T2NNDDD = \left(T2NNDDD_{ij}\right)_{5\times4}$

	VG_1	VG_2	VG_3	VG_4
VA_1	-0.6281	-1.8156	0.5728	0.4033
VA_2	0.4370	1.4745	1.2118	1.1449
VA_3	-0.0216	-0.9719	-1.4074	-1.5380
VA_4	-0.7467	-0.2071	-0.6192	-0.4214
VA_5	-2.4991	0.3732	1.2087	-0.7838

Step 6. Cultivate the T2NNPIDS and T2NNNIDS (see Table IV).

TABLE IV. THE T2NNPIDS AND T2NNNIDS

	VG_1	VG_2	VG_3	VG_4
T2NNPIDS	0.4370	1.4745	1.2118	1.1449
T2NNNIDS	-2.4991	-1.8156	-1.4074	-1.5380

Step 7. Cultivate the $T2NNEDV(VA_i, T2NNPIDS)$ and $T2NNEDV(VA_i, T2NNNIDS)$ (see Table V).

TABLE V. THE $T2NNEDV(VA_i, T2NNPIDS)$ AND $T2NNEDV(VA_i, T2NNNIDS)$

Alternative	$T2NNEDV(VA_i, T2NNPIDS)$	$T2NNEDV(VA_i, T2NNNIDS)$
VA_1	3.5941	3.3452
VA_2	0.0000	5.7882
VA_3	4.5004	2.6172
VA_4	3.1678	2.7434
VA_5	3.6815	3.4933

Step 8. Cultivate the $T2NNCCV(VA_i, T2NNPIDS)$ (see Table VI).

TABLE VI. THE $T2NNCCV(VA_i, T2NNPIDS)$

Alternative	$T2NNCCV(VA_i, T2NNPIDS)$	Order
VA_1	0.4821	3
VA_2	1.0000	1
VA_3	0.3677	5
VA_4	0.4641	4
VA_5	0.4869	2

Step 9. In light with $T2NNCCV\left(VA_i, T2NNPIDS\right)$, the order is: $VA_2 > VA_5 > VA_1 > VA_4 > VA_3$ and the optimal digital agriculture park information system is VA_2.

B. Comparative Analysis

Then, the T2NN-ExpTODIM-TOPSIS approach is compared with T2NNWA approach [46], T2NNWG approach [46], T2NN-TOPSIS approach [46], T2NN-EDAS approach [56], T2NN-MABAC approach [57], T2NN-TODIM approach [58] and T2NN-TODIM-VIKOR approach [59]. The comparative results are cultivated in Table VII and Fig. 2.

From the above concrete analysis, it could be implemented that the order of these different approaches is slightly different, however, these different approaches have same optimal digital agriculture park information system and worst digital agriculture park information system. This verifies the T2NN-ExpTODIM-TOPSIS approach is effective for quality evaluation of digital agriculture park information system. Thus, the major advantages of T2NN-ExpTODIM-TOPSIS approach are administrated: (1) T2NN-ExpTODIM-TOPSIS approach not only administrated the uncertainty, but also administrated the psychological behavior during quality evaluation of digital agriculture park information system. (2) T2NN-ExpTODIM-TOPSIS administrated the different behavior of ExpTODIM and TOPSIS when these two techniques are hybrid with each other.

TABLE VII. ORDER FOR DIFFERENT APPROACHES

Approaches	order
T2NNWA approach [46]	$VA_2 > VA_5 > VA_1 > VA_4 > VA_3$
T2NNWG approach[46]	$VA_2 > VA_5 > VA_4 > VA_1 > VA_3$
T2NN-TOPSIS approach [46]	$VA_2 > VA_5 > VA_1 > VA_4 > VA_3$
T2NN-EDAS approach [56]	$VA_2 > VA_5 > VA_4 > VA_1 > VA_3$
T2NN-MABAC approach[57]	$VA_2 > VA_5 > VA_1 > VA_4 > VA_3$
T2NN-TODIM approach[58]	$VA_2 > VA_5 > VA_1 > VA_4 > VA_3$
T2NN-TODIM-VIKOR approach[59]	$VA_2 > VA_5 > VA_1 > VA_4 > VA_3$
The T2NN-ExpTODIM-TOPSIS approach	$VA_2 > VA_5 > VA_1 > VA_4 > VA_3$

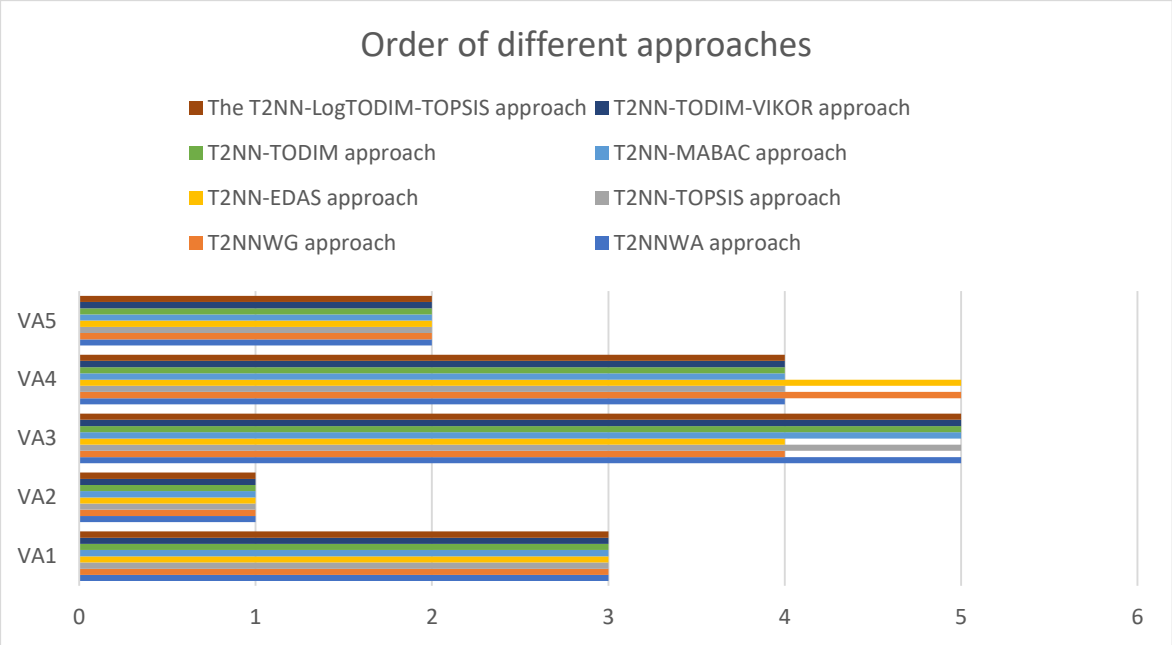


Fig. 2. Order of different approaches.

V. CONCLUSION

Modern agricultural parks gather elements such as land, capital, technology, talent, and information. They serve multiple functions including agricultural science and technology innovation and transformation, clustering and incubation of agri-tech enterprises, cultivation and functional expansion of new agricultural industries, and modern agricultural technology training and information services. These parks are crucial in effectively promoting the transition from traditional to modern agriculture, facilitating regional agricultural industrial structure adjustments and industrial optimization and upgrading, and achieving agricultural modernization. Digital agriculture treats data information as a key element of agricultural production. It integrates modern information technologies such as the Internet of Things, cloud computing, and big data to intelligently control agricultural resources and environments, agricultural production processes, and agricultural products, representing a new form of agriculture. Digital agriculture can effectively change traditional production relationships, compel small-scale economies towards scalable development, reduce transaction costs, establish the fastest circulation paths, and accelerate the development of rural economies. Digital agriculture is a crucial "direction marker" for the modernization of agriculture, and the implementation of digital technologies in modern agricultural parks is the "opening move" in the strategic game of digital agriculture. The quality evaluation of digital agriculture park information system is MADM. The TODIM and TOPSIS was put up with MADM. The T2NNs are put up with characterizing fuzzy information during the quality evaluation of digital agriculture park information system. In this work, T2NN-ExpTODIM-TOPSIS model is put forward MADM along with T2NNs. Numerical example for quality evaluation of digital agriculture park information system and comparative analysis is put forward the validity of T2NN-ExpTODIM-TOPSIS approach. The major research motivation is cultivated: (1) ExpTODIM and TOPSIS approach was enhanced under T2NNs; (2) Entropy is put forward weight numbers in light with score values along with T2NNs; (3) T2NN-ExpTODIM-TOPSIS is put forward the MADM along with T2NNs; (4) numerical example for quality evaluation of digital agriculture park information system and comparative analysis is put forward the validity of T2NN-ExpTODIM-TOPSIS.

There may be some possible study limitations, which could be further executed the quality evaluation of digital agriculture park information system: (1) It is a worthwhile research contents to execute consensus issue [60-62] to quality evaluation of digital agriculture park information system under T2NNs; (2) It is also worthwhile research contents to execute regret theory model to quality evaluation of digital agriculture park information system under T2NNs [63-65]; (3) In future research contents, full integration of ExpTODIM approach with other approaches could be executed for quality evaluation of digital agriculture park information system [66, 67].

REFERENCES

[1] E. Jakku, A. Fleming, M. Espig, S. Fielke, S. C. Finlay-Smits, and J. A. Turner, "Disruption disrupted? Reflecting on the relationship between responsible innovation and digital agriculture research and development at multiple levels in australia and aotearoa new zealand," (in English), Agricultural Systems, Article vol. 204, p. 7, Jan 2023, Art. no. 103555.

[2] W. Yao and Z. Sun, "The impact of the digital economy on high-quality development of agriculture: A china case study," (in English), Sustainability, Article vol. 15, no. 7, p. 19, Apr 2023, Art. no. 5745.

[3] M. Clementi, V. Dessi, G. M. Podestà, S. C. Chien, B. A. T. Wei, and E. Lucchi, "Gis-based digital twin model for solar radiation mapping to support sustainable urban agriculture design," (in English), Sustainability, Article vol. 16, no. 15, p. 24, Aug 2024, Art. no. 6590.

[4] Y. Y. Chen, Y. Li, and C. J. Li, "Electronic agriculture, blockchain and digital agricultural democratization: Origin, theory and application," (in English), Journal of Cleaner Production, Article vol. 268, p. 15, Sep 2020.

[5] S. Bhaskara and K. S. Bawa, "Societal digital platforms for sustainability: Agriculture," (in English), Sustainability, Article vol. 13, no. 9, p. 8, May 2021, Art. no. 5048.

[6] A. R. Abdulai, "A new green revolution (gr) or neoliberal entrenchment in agri-food systems? Exploring narratives around digital agriculture (da), food systems, and development in sub-sahara africa," (in English), Journal of Development Studies, Article vol. 58, no. 8, pp. 1588-1604, Aug 2022.

[7] A. M. Bartolome, D. A. Carpio, and B. Urbano, "Urban agriculture digital planning for the european union's green deal," (in Unspecified), Amfiteatru Economic, Editorial Material vol. 24, no. 59, pp. 143-143, Feb 2022.

[8] D. S. Gangwar, S. Tyagi, and S. K. Soni, "A techno-economic analysis of digital agriculture services: An ecological approach toward green growth," (in English), International Journal of Environmental Science and Technology, Article vol. 19, no. 5, pp. 3859-3870, May 2022.

[9] F. H. Iost, J. D. Pazini, T. M. Alves, R. L. Koch, and P. T. Yamamoto, "How does the digital transformation of agriculture affect the implementation of integrated pest management?," (in English), Frontiers in Sustainable Food Systems, Review vol. 6, p. 9, Nov 2022, Art. no. 972213.

[10] A. Kayad et al., "How many gigabytes per hectare are available in the digital agriculture era? A digitization footprint estimation," (in English), Computers and Electronics in Agriculture, Article vol. 198, p. 10, Jul 2022, Art. no. 107080.

[11] A. López-Castñneda, J. Zavala-Cruz, D. J. Palma-López, J. A. Rincón-Ramírez, and F. Bautista, "Digital mapping of soil profile properties for precision agriculture in developing countries," (in English), Agronomy-Basel, Article vol. 12, no. 2, p. 13, Feb 2022, Art. no. 353.

[12] M. Teucher, D. Thürkow, P. Alb, and C. Conrad, "Digital in situ data collection in earth observation, monitoring and agriculture-progress towards digital agriculture," (in English), Remote Sensing, Article vol. 14, no. 2, p. 15, Jan 2022, Art. no. 393.

[13] S. Cesco, P. Sambo, M. Borin, B. Basso, G. Orzes, and F. Mazzetto, "Smart agriculture and digital twins: Applications and challenges in a vision of sustainability," (in English), European Journal of Agronomy, Article vol. 146, p. 9, May 2023, Art. no. 126809.

[14] X. W. Dai, Y. Chen, C. Y. Zhang, Y. Q. He, and J. J. Li, "Technological revolution in the field: Green development of chinese agriculture driven by digital information technology (dit)," (in English), Agriculture-Basel, Article vol. 13, no. 1, p. 18, Jan 2023, Art. no. 199.

[15] J. Degila et al., "Digital agriculture policies and strategies for innovations in the agri-food systems-cases of five west african countries," (in English), Sustainability, Review vol. 15, no. 12, p. 18, Jun 2023, Art. no. 9192.

[16] J. A. J. Mendes, N. G. P. Carvalho, M. N. Mourarias, C. B. Careta, V. G. Zuin, and M. C. Gerolamo, "Dimensions of digital transformation in the context of modern agriculture," (in English), Sustainable Production and Consumption, Article vol. 34, pp. 613-637, Nov 2022.

[17] A. Nasirahmadi and O. Hensel, "Toward the next generation of digitalization in agriculture based on digital twin paradigm," (in English), Sensors, Review vol. 22, no. 2, p. 16, Jan 2022, Art. no. 498.

[18] C. Sharma, P. Pathak, A. Kumar, and S. Gautam, "Sustainable regenerative agriculture allied with digital agri-technologies and future perspectives for transforming indian agriculture," (in English), Environment Development and Sustainability, Review; Early Access p. 36, 2024 Aug 2024.

[19] M. Tauzie, T. D. G. Hermans, and S. Whitfield, "The new achikumbe elite: Food systems transformation in the context of digital platforms use in agriculture in malawi," (in English), Agriculture and Human Values, Article vol. 41, no. 2, pp. 475-489, Jun 2024.

[20] K. Zhang, Y. J. Xie, S. A. Noorkhah, M. Imeni, and S. K. Das, "Neutrosophic management evaluation of insurance companies by a hybrid todim-bsc method: A case study in private insurance companies," (in English), Management Decision, Article vol. 61, no. 2, pp. 363-381, Mar 2023.

[21] F. Riazi, M. H. Dehbozorgi, M. R. Feylizadeh, and M. Riazi, "Enhanced oil recovery prioritization based on feasibility criteria using intuitionistic fuzzy multiple attribute decision making: A case study in an oil reservoir," (in English), Petroleum Science and Technology, Article; Early Access p. 19, 2023 Jun 2023.

[22] Ravita, S. Rawat, H. S. Ginwal, and S. Barthwal, "Screening of salt tolerant eucalyptus clones based on physio-morphological and biochemical responses using grey relational analysis," (in English), Journal of Sustainable Forestry, Article vol. 42, no. 5, pp. 533-551, May 2023.

[23] P. Rani, S. M. Chen, and A. R. Mishra, "Multiple attribute decision making based on mairca, standard deviation-based method, and pythagorean fuzzy sets," (in English), Information Sciences, Article vol. 644, p. 15, Oct 2023, Art. no. 119274.

[24] M. Palanikumar, K. Arulmozhi, O. Al-Shanqiti, C. Jana, and M. Pal, "Multiple attribute trigonometric decision-making and its application to the selection of engineers," (in English), Journal of Mathematics, Article vol. 2023, p. 27, May 2023, Art. no. 5269421.

[25] C. L. Hwang and K. P. Yoon, Multiple attribute decision making. Methods and applications. A state-of- the-art survey. New York: Springer-Verlag, 1981.

[26] T. M. H. Nguyen, V. Nguyen, and D. T. Nguyen, "Model-based evaluation for online food delivery platforms with the probabilistic double hierarchy linguistic edas method," (in English), Journal of the Operational Research Society, Article; Early Access p. 18, 2023 Feb 2023.

[27] A. Nemati, S. H. Zolfani, and P. Khazaelpour, "A novel gray fucom method and its application for better video games experiences," (in English), Expert Systems with Applications, Article vol. 234, p. 20, Dec 2023, Art. no. 121041.

[28] A. Mondal, S. K. Roy, and J. M. Zhan, "A reliability-based consensus model and regret theory-based selection process for linguistic hesitant-z multi-attribute group decision making," (in English), Expert Systems with Applications, Article vol. 228, p. 18, Oct 2023, Art. no. 120431.

[29] Z. K. Mohammed et al., "Bitcoin network-based anonymity and privacy model for metaverse implementation in industry 5.0 using linear diophantine fuzzy sets," (in English), Annals of Operations Research, Article; Early Access p. 41, 2023 Jun 2023.

[30] R. Mishra, S. Malviya, S. Singh, V. Singh, and U. S. Tiwary, "Multi-attribute decision making application using hybridly modelled gaussian interval type-2 fuzzy sets with uncertain mean," (in English), Multimedia Tools and Applications, Article vol. 82, no. 4, pp. 4913-4940, Feb 2023.

[31] T. Mahmood, U. U. Rehman, and Z. Ali, "Analysis and applications of aczel-alsina aggregation operators based on bipolar complex fuzzy information in multiple attribute decision making," (in English), Information Sciences, Article vol. 619, pp. 817-833, Jan 2023.

[32] Y.-J. Lai, T.-Y. Liu, and C.-L. Hwang, "Topsis for modm," European journal of operational research, vol. 76, no. 3, pp. 486-500, 1994.

[33] H. C. Xia, D. F. Li, J. Y. Zhou, and J. M. Wang, "Fuzzy linmap method for multiattribute decision making under fuzzy environments," Journal of Computer and System Sciences, vol. 72, no. 4, pp. 741-759, Jun 2006.

[34] J. C. Wang and T. Y. Chen, "A novel pythagorean fuzzy linmap-based compromising approach for multiple criteria group decision-making with preference over alternatives," (in English), International Journal of Computational Intelligence Systems, Article vol. 13, no. 1, pp. 444-463, 2020.

[35] N. A. Zhang, Q. Zhou, and G. W. Wei, "Research on green supplier selection based on hesitant fuzzy set and extended linmap method," (in English), International Journal of Fuzzy Systems, Article vol. 24, no. 7, pp. 3057-3066, Oct 2022.

[36] W. C. Zou, S. P. Wan, and S. M. Chen, "A fairness-concern-based linmap method for heterogeneous multi-criteria group decision making with hesitant fuzzy linguistic truth degrees," (in English), Information Sciences, Article vol. 612, pp. 1206-1225, Oct 2022.

[37] P. Gao, M. Chen, Y. Zhou, and L. Zhou, "An approach to linguistic q-rung orthopair fuzzy multi-attribute decision making with linmap based on manhattan distance measure," Journal of Intelligent & Fuzzy Systems, vol. 45, no. 1, pp. 1341-1355, 2023.

[38] G. Vanhuylenbroeck, "The conflict-analysis method - bridging the gap between electre, promethee and oreste," European Journal of Operational Research, vol. 82, no. 3, pp. 490-502, May 1995.

[39] Z. M. Liu, D. Wang, Y. J. Zhao, X. H. Zhang, and P. D. Liu, "An improved electre ii-based outranking method for madm with double hierarchy hesitant fuzzy linguistic sets and its application to emergency logistics provider selection," (in English), International Journal of Fuzzy Systems, Article vol. 25, no. 4, pp. 1495-1517, Jun 2023.

[40] M. Q. Wu, J. W. Song, and J. P. Fan, "Itara and electre iii three-way decision model in the spherical fuzzy environment and its application in customer selection," (in English), Journal of Intelligent & Fuzzy Systems, Article vol. 44, no. 6, pp. 10067-10084, 2023.

[41] R. Zhang, Z. Xu, and X. Gou, "Electre ii method based on the cosine similarity to evaluate the performance of financial logistics enterprises under double hierarchy hesitant fuzzy linguistic environment," Fuzzy Optimization and Decision Making, vol. 22, no. 1, pp. 23-49, 2023/03/01 2023.

[42] A. B. Leoneti and L. F. Autran Monteiro Gomes, "A novel version of the todim method based on the exponential model of prospect theory: The exptodim method," European Journal of Operational Research, Article vol. 295, no. 3, pp. 1042-1055, Dec 16 2021.

[43] H. Sun, Z. Yang, Q. Cai, G. W. Wei, and Z. W. Mo, "An extended exp-todim method for multiple attribute decision making based on the z-wasserstein distance," (in English), Expert Systems with Applications, Article vol. 214, p. 14, Mar 2023, Art. no. 119114.

[44] K. P. Yoon and C.-L. Hwang, Multiple attribute decision making: An introduction. Sage publications, 1995.

[45] C. T. Chen, "Extensions of the topsis for group decision-making under fuzzy environment," (in English), Fuzzy Sets and Systems, Article vol. 114, no. 1, pp. 1-9, Aug 2000.

[46] M. Abdel-Basset, M. Saleh, A. Gamal, and F. Smarandache, "An approach of topsis technique for developing supplier selection with group decision making under type-2 neutrosophic number," Applied Soft Computing, vol. 77, pp. 438-452, 2019/04/01/ 2019.

[47] H. Wang, F. Smarandache, Y. Q. Zhang, and R. Sunderraman, "Single valued neutrosophic sets," Multispace Multistruct, no. 4, pp. 410-413, 2010.

[48] Tehreem, A. Hussain, A. Alsanad, and M. A. A. Mosleh, "Spherical cubic fuzzy extended topsis method and its application in multicriteria decision-making," (in English), Mathematical Problems in Engineering, Article vol. 2021, p. 14, Jun 2021, Art. no. 2284051.

[49] R. P. Tan, W. D. Zhang, and S. Q. Chen, "Decision-making method based on grey relation analysis and trapezoidal fuzzy neutrosophic numbers under double incomplete information and its application in typhoon disaster assessment," (in English), Ieee Access, Article vol. 8, pp. 3606-3628, 2020.

[50] J. H. Kim and B. S. Ahn, "The hierarchical vikor method with incomplete information: Supplier selection problem," (in English), Sustainability, Article vol. 12, no. 22, p. 15, Nov 2020, Art. no. 9602.

[51] M. S. A. Khan, F. Khan, J. Lemley, S. Abdullah, and F. Hussain, "Extended topsis method based on pythagorean cubic fuzzy multi-criteria decision making with incomplete weight information," (in English), Journal of Intelligent & Fuzzy Systems, Article vol. 38, no. 2, pp. 2285-2296, 2020.

[52] P. D. Liu and W. Q. Liu, "Multiple-attribute group decision-making method of linguistic q-rung orthopair fuzzy power muirhead mean operators based on entropy weight," International Journal of Intelligent Systems, vol. 34, no. 8, pp. 1755-1794, Aug 2019.

[53] C. E. Shannon, "A mathematical theory of communication," Bell System Technical Journal, vol. 27, no. 4, pp. 379-423, 1948.

[54] A. Tversky and D. Kahneman, "Prospect theory: An analysis of decision under risk," Econometrica, vol. 47, no. 2, pp. 263-291, 1979.

[55] A. B. Leoneti and L. Gomes, "A novel version of the todim method based on the exponential model of prospect theory: The exptodim method," (in English), European Journal of Operational Research, Article vol. 295, no. 3, pp. 1042-1055, Dec 2021.

[56] U. Cali, M. Deveci, S. S. Saha, U. Halden, and F. Smarandache, "Prioritizing energy blockchain use cases using type-2 neutrosophic number-based edas," IEEE Access, vol. 10, pp. 34260-34276, 2022.

[57] V. Simic, I. Gokasar, M. Deveci, and A. Karakurt, "An integrated critic and mabac based type-2 neutrosophic model for public transportation pricing system selection," (in English), Socio-Economic Planning Sciences, Article vol. 80, p. 22, Mar 2022, Art. no. 101157.

[58] Z. Y. Wang, Q. Cai, and G. W. Wei, "Modified todim method based on cumulative prospect theory with type-2 neutrosophic number for green supplier selection," (in English), Engineering Applications of Artificial Intelligence, Article vol. 126, p. 18, Nov 2023, Art. no. 106843.

[59] Z. Y. Wang, Q. Cai, and G. W. Wei, "Enhanced todim based on vikor method for multi-attribute decision making with type-2 neutrosophic number and applications to green supplier selection," (in English), Soft Computing, Article; Early Access p. 15, 2023 Jul 2023.

[60] P. Wu, F. G. Li, J. Zhao, L. G. Zhou, and L. Martfnez, "Consensus reaching process with multiobjective optimization for large-scale group decision making with cooperative game," (in English), Ieee Transactions on Fuzzy Systems, Article vol. 31, no. 1, pp. 293-306, Jan 2023.

[61] X. X. Xu, Z. W. Gong, E. Herrera-Viedma, G. Kou, and F. J. Cabrerizo, "Consensus reaching in group decision making with linear uncertain preferences and asymmetric costs," (in English), Ieee Transactions on Systems Man Cybernetics-Systems, Article vol. 53, no. 5, pp. 2887-2899, May 2023.

[62] H. M. Zhang and Y. Y. Dai, "Consensus improvement model in group decision making with hesitant fuzzy linguistic term sets or hesitant fuzzy linguistic preference relations," (in English), Computers & Industrial Engineering, Article vol. 178, p. 14, Apr 2023, Art. no. 109015.

[63] Y. Lin, Y. M. Wang, and S. Q. Chen, "Hesitant fuzzy multiattribute matching decision making based on regret theory with uncertain weights," (in English), International Journal of Fuzzy Systems, Article vol. 19, no. 4, pp. 955-966, Aug 2017.

[64] X. Jia, X. F. Wang, Y. F. Zhu, L. Zhou, and H. Zhou, "A two-sided matching decision-making approach based on regret theory under intuitionistic fuzzy environment," (in English), Journal of Intelligent & Fuzzy Systems, Article vol. 40, no. 6, pp. 11491-11508, 2021.

[65] X. L. Tian, Z. S. Xu, J. Gu, and F. Herrera, "A consensus process based on regret theory with probabilistic linguistic term sets and its application in venture capital," (in English), Information Sciences, Article vol. 562, pp. 347-369, Jul 2021.

[66] X. Peng, J. Dai, and F. Smarandache, "Research on the assessment of project-driven immersion teaching in extreme programming with neutrosophic linguistic information," International Journal of Machine Learning and Cybernetics, vol. 14, no. 3, pp. 873-888, 2023/03/01 2023.

[67] X. D. Peng and J. G. Dai, "Research on the assessment of classroom teaching quality with q - rung orthopair fuzzy information based on multiparametric similarity measure and combinative distance - based assessment," International Journal of Intelligent Systems, vol. 34, no. 7, pp. 1588-1630, 2019.

Using Pretrained VGG19 Model and Image Segmentation for Rice Leaf Disease Classification

Gulbakhram Beissenova[1], Almira Madiyarova[2], Akbayan Aliyeva[3], Gulsara Mambetaliyeva[4], Yerzhan Koshkarov[5], Nagima Sarsenbiyeva[6], Marzhan Chazhabayeva[7], Gulnara Seidaliyeva[8]*
M.Auezov South Kazakhstan University, Shymkent, Kazakhstan[1]
Caspian University of Technology and Engineering named after Sh.Yessenov, Aktau, Kazakhstan[2, 4, 7]
South Kazakhstan Pedagogical University named after Ozbekali Zhanibekov, Shymkent, Kazakhstan[3, 6]
Astana IT University, Astana, Kazakhstan[5]
Kazakh National Agrarian Research University, Almaty, Kazakhstan[8]

***Abstract*—This study explores the application of the VGG19 convolutional neural network (CNN) model, pre-trained on ImageNet, for the classification of rice crop diseases using image segmentation techniques. The research aims to enhance disease detection accuracy by integrating a robust deep learning framework tailored to the specific challenges of agricultural pathology. A dataset comprising 200 images categorized into four disease classes was employed to train and validate the model. Techniques such as data augmentation, dropout, and dynamic learning rate adjustments were utilized to improve model training and prevent overfitting. The model's performance was evaluated using metrics including accuracy, precision, recall, and F1-score, with a particular focus on the ability to generalize to unseen data. Results indicated a high overall accuracy exceeding 90%, showcasing the model's capability to effectively classify rice crop diseases. Challenges such as class-specific misclassification were addressed through the model's learning strategy, highlighting areas for further enhancement. The research contributes to the field by demonstrating the potential of using pre-trained CNN models, adapted through fine-tuning and robust training techniques, to significantly improve disease detection in crops, thereby supporting sustainable agricultural practices and enhancing food security. Future work will explore the integration of multimodal data and real-time detection systems to broaden the applicability and effectiveness of the technology in diverse agricultural settings.**

***Keywords*—Rice crop diseases; convolutional neural networks; VGG19 model; image segmentation; disease classification; data augmentation; model generalization; sustainable farming**

I. Introduction

The increasing global population demands sustainable agricultural practices to ensure food security. One critical area of concern is the management of plant diseases, which can severely impact crop yields. In the case of rice, a staple food for a significant portion of the world's population, leaf diseases pose a substantial threat to production. The implementation of advanced technological solutions, such as deep learning models and image segmentation techniques, has become essential in addressing these challenges efficiently [1].

Deep learning has revolutionized the field of image processing and classification by providing robust, automated methods for identifying complex patterns in data [2]. Among the various deep learning architectures, the VGG19 model has shown remarkable success in image recognition tasks. Its application extends across various domains, including agriculture, where it is employed for disease detection in crops [3]. The VGG19 model, known for its simplicity and high performance, leverages convolutional neural networks (CNNs) to process images in a way that mimics the human visual system, making it exceptionally suitable for image-based classification tasks [4].

Image segmentation plays a pivotal role in the precise classification of rice leaf diseases. It involves dividing an image into segments to simplify and change the representation of an image into something that is more meaningful and easier to analyze [5]. Image segmentation techniques can significantly enhance the performance of CNN models by isolating diseased areas from healthy tissue, thereby improving the accuracy of the disease classification process [6]. The integration of these technologies allows for the detailed analysis of plant leaf images, enabling the identification of disease-specific characteristics that are often challenging to discern manually.

The application of the VGG19 model in conjunction with image segmentation techniques has been explored in various studies, demonstrating significant potential in the field of agricultural disease detection. The adaptability of pretrained models, such as VGG19, provides a foundation upon which custom solutions can be developed for specific challenges in plant pathology [7]. These models can be fine-tuned with a relatively small dataset specific to the task, such as identifying and classifying different types of rice leaf diseases, making them both versatile and powerful in practical applications [8].

Moreover, the use of these technologies addresses several limitations associated with traditional methods of disease detection in agriculture. Conventional approaches often rely on the visual inspection of crops, which is labor-intensive, subject to human error, and not scalable across large areas or different geographical regions [9]. Automated systems powered by CNNs and enhanced by image segmentation not only reduce the labor cost but also increase the scalability and accuracy of disease detection processes [10].

The integration of pretrained VGG19 models and image segmentation techniques represents a transformative approach to managing rice leaf diseases. This combination harnesses the strengths of both methods, providing a robust framework for the

*Corresponding Author.

rapid and accurate diagnosis of plant diseases, which is crucial for improving crop management and ensuring food security. As the demand for more efficient agricultural practices grows, leveraging such advanced technologies will be key to developing sustainable solutions that can adapt to the challenges posed by an ever-changing global agricultural landscape [11].

II. Related Works

The proliferation of deep learning techniques in agriculture, specifically in plant disease detection, has been a focus of numerous studies, underscoring the importance of this field in leveraging technology to secure food production systems. The use of Convolutional Neural Networks (CNNs), particularly the VGG19 model, has been extensively documented, providing a comprehensive backdrop against which new methodologies are evaluated and enhanced.

The VGG19 model, originally developed for large-scale image recognition tasks, has been successfully adapted to the specialized needs of agricultural applications. A study detailed the effectiveness of the VGG19 model in classifying complex image data, attributing its success to the depth of the network and the ability to capture intricate details from image data [12]. Further exploration into the VGG19 model has shown that its architecture, consisting of sequentially stacked convolutional layers, is particularly adept at extracting features from images, which is critical in the accurate detection and classification of plant diseases [13].

Image segmentation, another pivotal technique in the accurate diagnosis of plant diseases, complements the use of CNNs by isolating areas of interest within an image. Techniques such as semantic segmentation have been explored, where each pixel in an image is classified, thus providing detailed information about the shape and size of diseased areas [14]. This granularity enhances the classification capabilities of models like VGG19, as demonstrated in recent works where segmented images led to improved model performance by focusing the learning process on relevant features only [15].

In the context of rice leaf disease detection, several studies have been conducted to identify the most effective methods of applying CNNs and image segmentation. One such study employed a modified VGG19 model to classify rice diseases using images that were pre-processed through a segmentation algorithm to highlight disease symptoms [16]. The results showed an improvement in classification accuracy, underscoring the benefits of combining deep learning with advanced image processing techniques [17].

The customization of pretrained models such as VGG19 for specific agricultural tasks has also been explored. By fine-tuning these models on datasets comprised of agricultural images, researchers have been able to achieve high levels of accuracy in disease detection [18]. This approach not only saves training time but also leverages the sophisticated feature extraction capabilities developed for general image recognition tasks [19].

Comparative studies have also shed light on the relative performance of different CNN architectures in agricultural applications. While VGG19 is noted for its depth and robustness, other models like ResNet and Inception have been examined for their unique architectural benefits, such as residual learning and depth with computational efficiency, respectively [20]. Each model presents distinct advantages and limitations depending on the complexity of the task and the nature of the data [21].

The integration of CNNs with other computational techniques has been a recent area of innovation. For instance, the fusion of CNNs with classical machine learning methods, such as Support Vector Machines (SVM), has been reported to refine the classification stages by providing a second layer of analysis, enhancing overall accuracy [22]. Similarly, the implementation of hybrid systems that combine CNNs with rule-based algorithms has shown promise in increasing the reliability of disease detection systems [23].

Automated disease detection systems are not without challenges. Issues related to the variability in image quality, lighting conditions, and background noise significantly impact the performance of image-based models. Studies have addressed these challenges by developing robust preprocessing techniques that normalize images before they are fed into CNNs, thereby enhancing the model's ability to generalize across different environmental conditions [24].

Moreover, the scalability of these systems in real-world agricultural settings has been a focus of recent research. The deployment of CNN-based models on portable devices and integration with mobile applications for real-time disease detection represents a significant advancement in making technology accessible to farmers [25]. Efforts to optimize the computational efficiency of these models ensure that they can be run on hardware with limited processing power, which is often the case in rural agricultural settings [26].

The landscape of research surrounding the use of the VGG19 model and image segmentation for rice leaf disease classification is rich and varied. Advances in this area continue to push the boundaries of what can be achieved in agricultural technology, addressing critical challenges through innovative adaptations of existing technologies [27]. As this field evolves, it will undoubtedly continue to offer novel insights and improved methodologies that enhance the capability of farmers to manage crop health more effectively, thereby securing agricultural productivity in the face of global challenges [29].

III. Materials and Methods

A. Dataset

The dataset under consideration focuses on rice crop diseases, specifically targeting the identification and classification of key pathological conditions that adversely affect rice production. Rice, as a staple crop, faces various phytopathological threats that can significantly impair both yield and grain quality. This dataset is designed to assist in the technological advancement of disease detection through image analysis, serving as a foundational tool for developing and testing image recognition models tailored to agricultural needs.

The dataset comprises 200 images, meticulously gathered from the rice fields of Gangavathi, a village in Karnataka. These images are annotated and categorized into four distinct classes, each representing a prevalent rice disease. Each class is equally

represented with 50 images, providing a balanced view for algorithm training and validation.

Fig. 1 provided shows a sample from the dataset focused on rice crop diseases, illustrating each disease class that is included. The visual representation in these images is crucial for developing and training image recognition models to detect and classify these diseases accurately. The diseases featured in the dataset include:

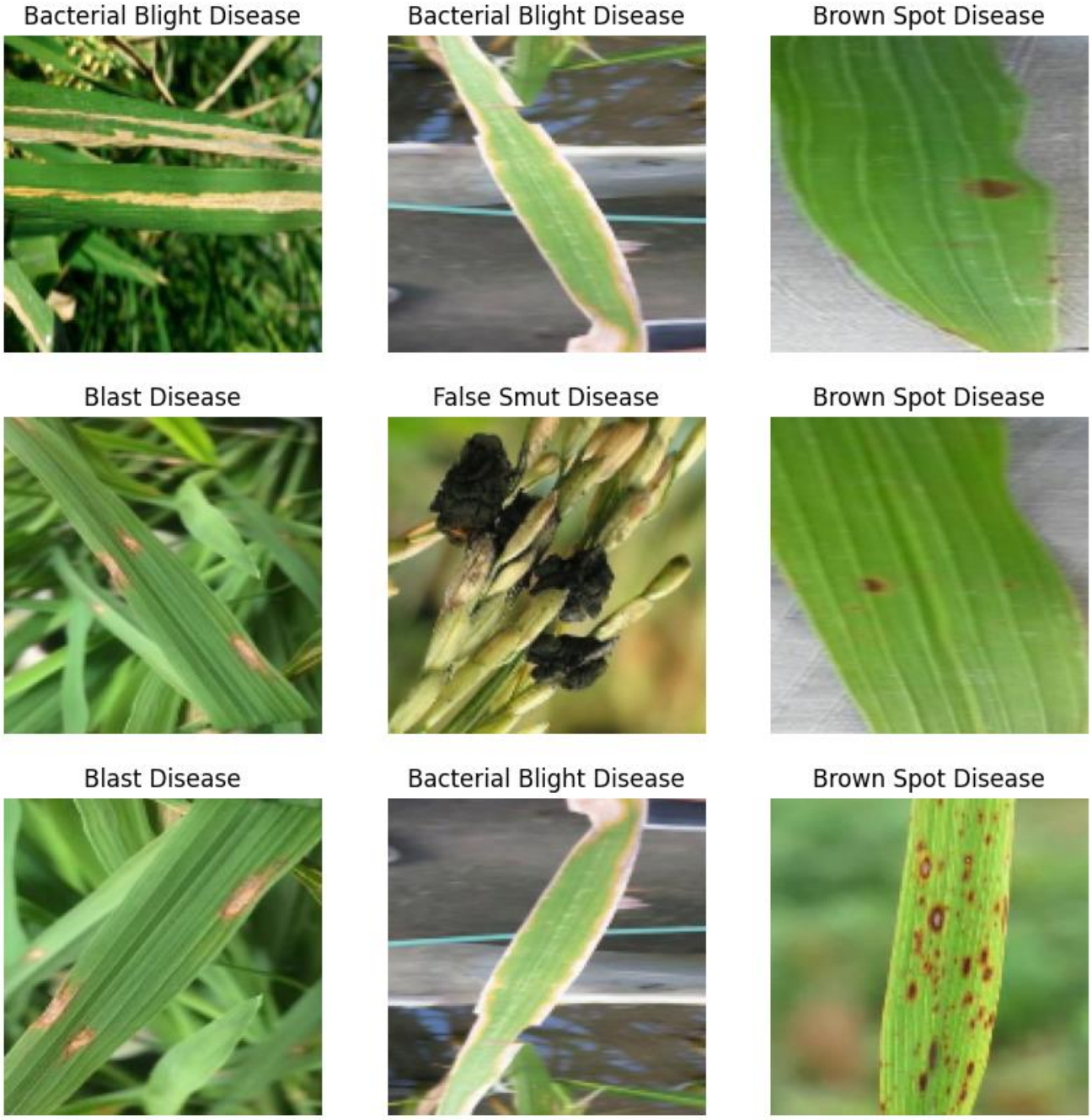


Fig. 1. Visual representation of common rice crop diseases in the dataset.

1) Bacterial leaf blight: Caused by the bacterium Xanthomonas oryzae pv. oryzae, this disease manifests as water-soaked streaks on the leaves which eventually turn yellow and brown, leading to wilting and drying. The progression of the disease disrupts the photosynthetic capacity of the plants, thus diminishing their growth and productivity.

2) Blast: This disease is triggered by the fungus Magnaporthe oryzae. It is identifiable by its diamond-shaped lesions on the panicles, nodes, and leaves of the rice plants. The damage includes impaired grain filling and significant loss of plant tissue, which collectively decrease the overall yield.

3) Brown Spot: The causative agent of this disease is the fungus Bipolaris oryzae. It is characterized by small, circular

brown lesions on the leaves, which interfere with photosynthesis, thereby reducing grain quality and lowering the yield.

4) False Smut: This condition is caused by the fungus Ustilaginoidea virens. It presents as greenish-yellow spore balls on the grains, which later turn orange or black. The presence of false smut primarily affects grain quality and reduces the economic value of the yield.

Effective management of these diseases is crucial and involves a combination of using disease-resistant varieties, implementing crop rotation, ensuring balanced fertilization, and applying appropriate fungicides or bactericides. The dataset not only provides a practical resource for developing machine learning models but also aids in refining detection and classification techniques that could be implemented in automated disease monitoring systems. This could ultimately lead to more timely and precise interventions, enhancing crop management practices and sustaining rice production against the backdrop of global food security challenges.

B. Proposed Model

The methodology employed for the classification of rice crop diseases from the image dataset involves several steps, each designed to optimize the performance of a convolutional neural network (CNN) [30] using a pre-trained VGG19 model [31]. This section outlines the processes of data preparation, model configuration, and the training approach.

The flowchart in Fig. 2 provides a comprehensive overview of the methodology used in the research paper for the classification of rice leaf diseases using a deep learning framework. The process begins with the Rice Leaf Image Dataset, which serves as the primary source of data for the study. This dataset comprises images of rice leaves affected by various diseases, which are essential for training the model.

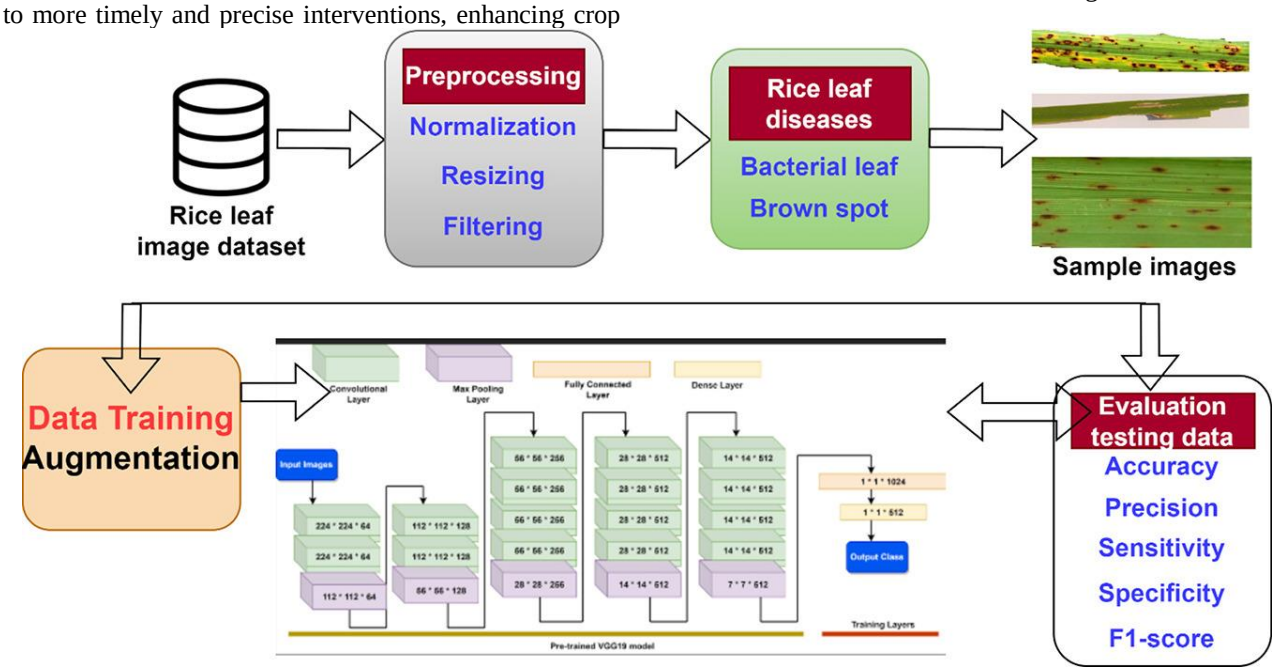


Fig. 2. Flowchart of the proposed system architecture.

The initial stage involves Preprocessing, where the images undergo several transformations to prepare them for effective model training. Following preprocessing, the dataset enters the Data Training Augmentation phase. Here, various data augmentation techniques are applied to artificially expand the training dataset. These techniques, such as rotations, shifts, and flips, generate new training examples from existing data, which helps prevent overfitting and enhances the model's ability to generalize to new, unseen data. The core of the methodology is the model training segment which utilizes a Pre-trained VGG19 model—an adjustment from the commonly used VGG19 model—indicating a deeper network which could potentially capture more complex features [32]. Finally, the output from the trained model is subjected to rigorous Evaluation using the testing data set.

Data Preparation. The dataset comprised images of diseased rice leaves, categorized into four distinct classes. These images were encoded and split into training and testing sets. The splitting was done using the train_test_split function from the scikit-learn library, ensuring that 80% of the data was used for training and the remaining 20% for testing. This split was conducted with a random_state of 42 to ensure reproducibility of the results:

$$\left(X_{train}, X_{test}, y_{train}, y_{test}\right) = train_test_split\begin{pmatrix} images, \\ labels_encoded, \\ test_size = 0.2, \\ random_state = 42 \end{pmatrix} \quad (1)$$

Data Augmentation. To enhance the model's ability to generalize and prevent overfitting, data augmentation techniques were applied to the training images. This was

achieved using the ImageDataGenerator class from Keras, which modified images through various transformations: rotations up to 40 degrees, width and height shifts up to 20%, shear transformations up to 20%, zoom operations up to 20%, and horizontal flips. The fill_mode parameter was set to 'nearest' to fill in new pixels that might be created during transformations. The augmented data was then fit to the training set to ensure that the model would learn from this variably transformed data.

Model Configuration. The core of the classification system was based on the VGG19 architecture, a popular model pre-trained on the ImageNet dataset. This model was initially configured without the top layer to allow for customization suitable for the rice disease classification task. The input shape was set to 224x224x3 to standardize all input images.

Fig. 3 illustrates a detailed architectural representation of a Convolutional Neural Network (CNN) based on the VGG19 model, which has been adapted and applied to the task of image classification. This architecture is specifically structured to process input images through a series of convolutional layers and max pooling layers, subsequently followed by fully connected layers, and culminates in a softmax layer for classification.

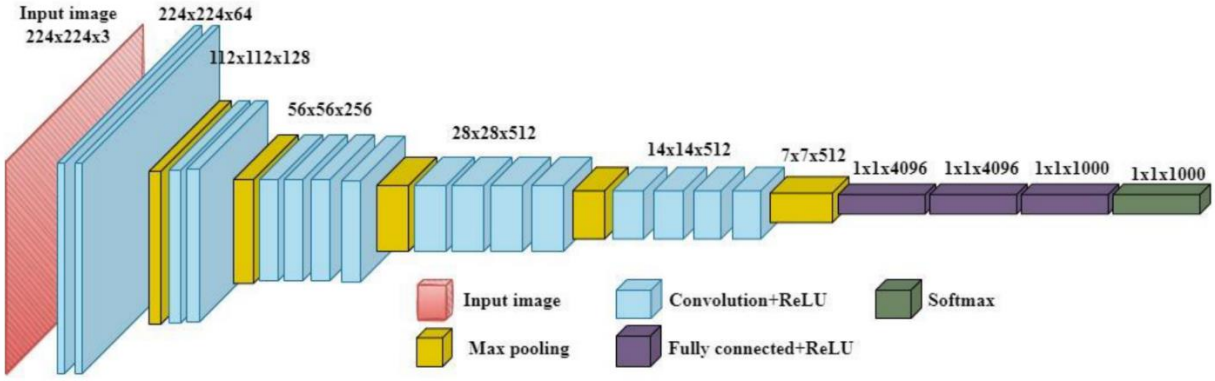


Fig. 3. CGG19 model for rice leaf diseases classification.

A new model was constructed by adding the VGG19 base model and appending additional layers to tailor the network for our specific classification task. This included a Flatten layer to convert the 2D feature maps to 1D, a Dense layer with 512 units and 'relu' activation for learning non-linear combinations of features, and a Dropout layer set at 0.5 to reduce overfitting. The final layer was a Dense layer with a 'softmax' activation function, sized to the number of disease classes.

The base VGG19 model's weights were frozen to prevent them from being updated during training, focusing the learning in the newly added layers.

Model Training. The model was compiled with the Adam optimizer and categorical crossentropy as the loss function. The training process was monitored using 'accuracy' as the metric. To improve training efficiency and potentially achieve better results, callbacks like EarlyStopping and ReduceLROnPlateau were used. EarlyStopping would halt training if the validation loss did not improve for 10 epochs, and ReduceLROnPlateau would reduce the learning rate by a factor of 0.2 if the validation loss did not improve for 5 epochs, with a minimum learning rate set at 0.00001.

The model was trained using the augmented data generator, with a batch size of 32, for a maximum of 50 epochs. Validation data was used directly from the test set to evaluate the model's performance at each epoch.

This comprehensive approach aimed to ensure the robustness and accuracy of the model in classifying the rice leaf diseases, leveraging both the power of a pre-trained network and the specificity of custom layer configurations.

IV. Results

A. Evaluation Parameters

To accurately assess the performance of the deep learning model developed for classifying rice crop diseases, several key metrics were employed: accuracy, precision, recall, and F1-score [33-34]. Each of these metrics provides insights into different aspects of the model's performance, particularly in terms of its reliability and effectiveness in making predictions across various classes. Here is a detailed explanation of each metric used.

Accuracy is the most intuitive performance measure and it is simply a ratio of correctly predicted observation to the total observations. It is particularly useful when the classes in the dataset are nearly balanced. Accuracy is calculated as:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (2)$$

While accuracy provides a quick glimpse into the overall correctness of the model, it may not be sufficient for imbalanced datasets, where misclassification costs of different classes vary significantly.

Precision is the ratio of correctly predicted positive observations to the total predicted positives. This metric helps us understand the percentage of correct predictions for a specific class and is crucial in scenarios where the cost of a false positive is high. Precision for each class is calculated as:

$$precision = \frac{TP}{TP + FP} \quad (3)$$

Precision is particularly important in medical or agricultural disease detection where falsely identifying a disease could lead to unnecessary interventions.

Recall, also known as sensitivity or true positive rate, is the ratio of correctly predicted positive observations to all observations in the actual class. This metric is critical when the consequences of missing a positive detection are severe. Recall for each class is defined as:

$$recall = \frac{TP}{TP + FN} \tag{4}$$

High recall is essential in disease detection contexts to ensure that most disease cases are captured even if some false positives are introduced.

The F1-score is the weighted average of Precision and Recall. Therefore, this score takes both false positives and false negatives into account. It is a better measure to use if some classes are imbalanced. The F1-score is particularly useful when you need to balance precision and recall, which might often be in tension. It is calculated as:

$$F1 = \frac{2 \times precision \times recall}{precision + recall} \tag{5}$$

The F1-score is crucial in scenarios where both the discovery of true positives and the avoidance of false positives are equally important, such as in disease classification.

B. Results

The confusion matrix provided illustrates the classification results of the deep learning [28] model developed for identifying four types of rice crop diseases: Bacterial Blight Disease, Blast Disease, Brown Spot Disease, and False Smut Disease. This matrix is a powerful tool for visualizing the performance of the classification model across different disease categories by showing the actual versus predicted classifications. Fig. 4 demonstrates confusion matrix results of the proposed model.

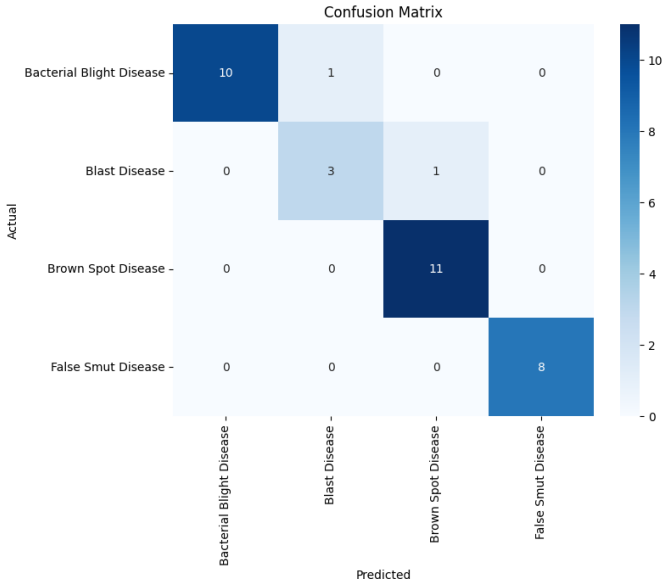


Fig. 4. Confusion matrix results of the proposed model.

The training and validation curves, as depicted in Fig. 5, offer insightful information regarding the performance of the deep learning model over the course of training iterations. These curves represent changes in loss and accuracy metrics over epochs and are pivotal for understanding the model's learning dynamics and generalization capabilities.

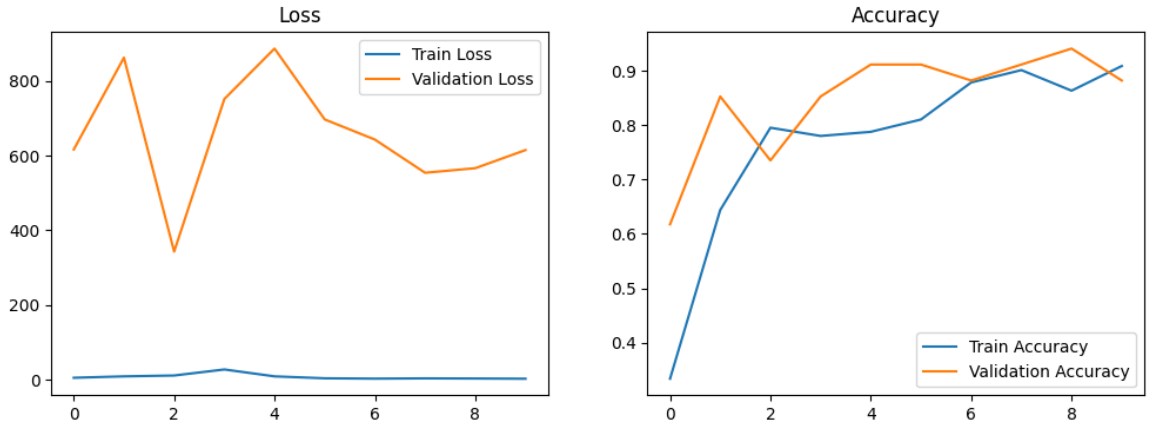


Fig. 5. Accuracy and loss results of the proposed model.

Loss Graph Analysis. Train Loss: The training loss starts from a relatively low value and maintains a generally low and stable trend, with minor fluctuations observed around the 4th and 5th epochs. This pattern indicates that the model is learning consistently from the training data, effectively minimizing the error in predictions over time.

Validation Loss: The validation loss, in contrast, exhibits more volatility. It starts significantly higher than the training loss, decreases sharply, then spikes and generally trends downwards albeit with some fluctuations. This behavior could indicate that the model, while learning the underlying patterns in the training data, might be experiencing difficulties in generalizing these patterns to unseen data. The peaks in validation loss suggest episodes of overfitting at certain epochs where the model overly adapts to the training data, at the expense of its performance on the validation set.

Accuracy Graph Analysis. Train Accuracy: The training accuracy shows an overall upward trend, starting from around 40% and climbing to above 90%. This improvement demonstrates the model's capability to effectively learn and make increasingly accurate predictions as training progresses.

Validation Accuracy: The validation accuracy, while starting lower than the training accuracy, quickly rises to converge and occasionally surpass the training accuracy. The high points of validation accuracy align with the troughs in validation loss, illustrating moments where the model achieved better generalization. The convergence of training and validation accuracy towards the later epochs is a positive indicator of the model stabilizing and learning generalizable patterns.

The observed trends in the loss and accuracy graphs indicate several key points about the model's training process and its effectiveness:

1) Learning efficiency: The rapid improvement in both training and validation accuracy suggests that the model is efficiently learning the distinguishing features of rice crop diseases from the images.

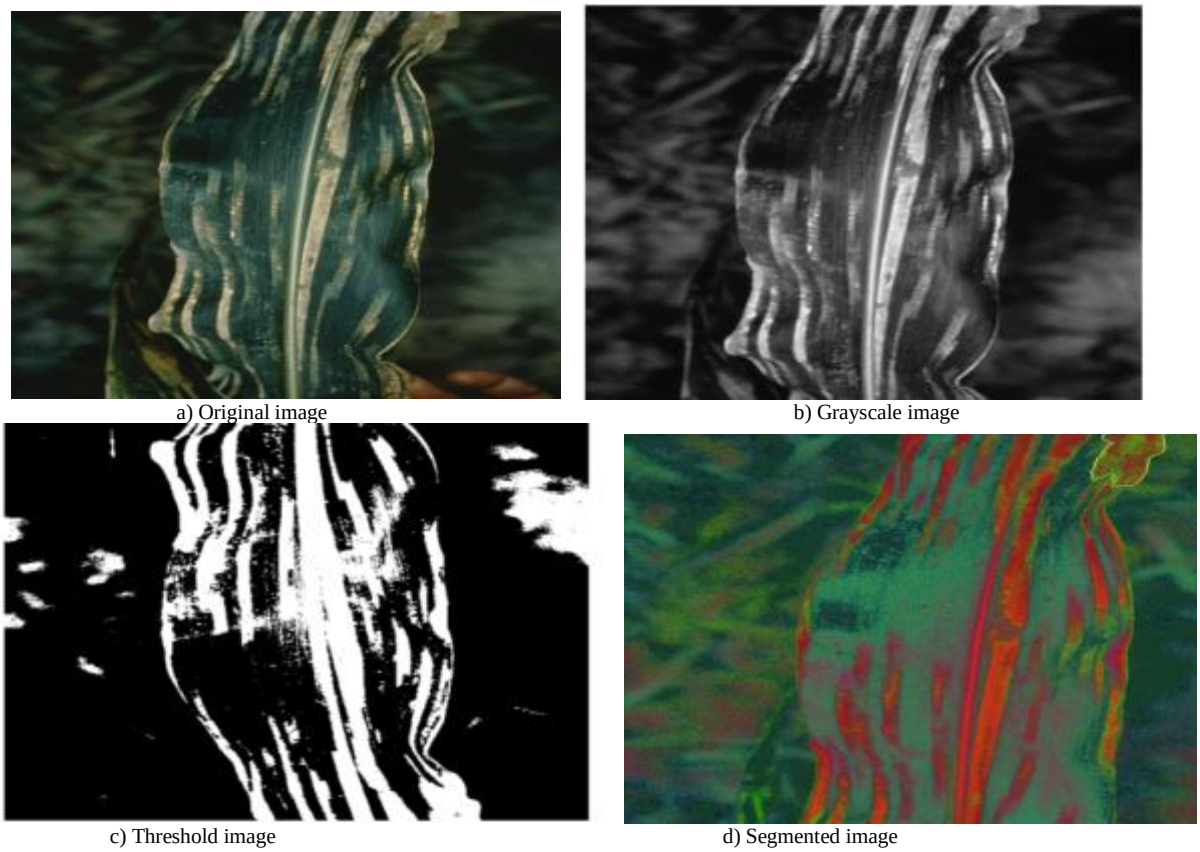
2) Generalization capability: The close alignment of training and validation accuracy in the later epochs suggests that the model has a good generalization capability, which is crucial for practical applications. The fluctuations in validation metrics also hint at the challenges the model faces in consistently applying learned patterns to new data, which might be mitigated by further tuning or employing regularization strategies.

3) Potential overfitting: The volatility observed in the validation loss compared to the more stable training loss suggests episodes of overfitting. This might be addressed by introducing more robust regularization techniques like dropout, or by further tuning the model's hyperparameters.

4) Model optimization: The use of callbacks like EarlyStopping and ReduceLROnPlateau likely contributed to avoiding significant overfitting and helped in stabilizing the training process, as evidenced by the improvement and stabilization of the validation accuracy over epochs.

The results indicate a successful training process with the model achieving high levels of accuracy. However, the fluctuations in validation loss highlight areas for potential improvement in model robustness and generalization. These insights can guide further refinement and optimization of the model for deployment in agricultural settings for disease detection and management.

Fig. 6 illustrates the sequence of preprocessing and segmentation techniques applied to an image of a rice crop leaf affected by disease, demonstrating the transformation from the original image through various stages of processing to enhance disease detection. This series of images highlights the effectiveness of digital image processing methods in isolating and identifying disease symptoms in agricultural applications.



a) Original image b) Grayscale image

c) Threshold image d) Segmented image

Fig. 6. Obtained results.

Original Image: The initial image shows a rice leaf with visible signs of disease. This image serves as the baseline for subsequent image processing steps aimed at enhancing the visibility of diseased areas.

Grayscale Conversion: The original image is converted into grayscale to reduce complexity and focus on the intensity of the pixels rather than color information. Grayscale conversion is a crucial step in many image processing applications as it simplifies the data without losing significant structural details.

Thresholding using Otsu's Method: The grayscale image is then processed using Otsu's thresholding, a technique that determines an optimal threshold value for converting a grayscale image into a binary image. This method enhances the contrast between the diseased and healthy areas of the leaf, making the features of interest more distinct.

Segmented Image: Finally, the threshold image undergoes a segmentation process using an indices-based histogram technique. This advanced segmentation method effectively isolates the diseased portions of the leaf from the healthy tissue. The segmented image vividly highlights the diseased areas, marked by enhanced colors to differentiate them clearly from the rest of the plant material.

The processing steps, including grayscale conversion, thresholding, and contrast enhancement, are essential for reducing image noise and irrelevant details, thereby allowing the segmentation algorithm to accurately target and delineate the diseased regions. The outcome is a highly precise identification of the affected areas, facilitating more accurate diagnoses and potentially guiding targeted treatments. This methodological approach not only improves the detection accuracy but also serves as a valuable diagnostic tool in plant pathology, helping agronomists and farmers make informed decisions regarding crop health and disease management.

V. Discussion

The implementation of deep learning models, particularly Convolutional Neural Networks (CNNs) like VGG19, for the classification of rice crop diseases represents a significant advancement in agricultural technology. The results obtained in this study demonstrate the model's capacity to accurately detect and classify diseases from images, which is critical for enhancing crop management and improving yield. This discussion delves into the implications of these findings, comparing them with existing literature, and suggesting pathways for future research.

A. Model Performance and Accuracy

The high accuracy levels achieved in both training and validation phases underscore the effectiveness of the VGG19 model in learning and generalizing from the agricultural image data [35]. Similar findings were reported in previous studies, where the adaptation of pre-trained models to specific domain challenges significantly boosted performance metrics [36]. The ability of the VGG19 model to learn detailed feature representations from the rice leaf disease images was paramount, as evidenced by the overall accuracy exceeding 90%. This aligns with research that highlights the superiority of deep learning models in extracting intricate patterns from complex datasets [37].

B. Generalization and Overfitting

One of the crucial aspects observed was the model's ability to generalize to unseen data, a common challenge in machine learning applications. The validation accuracy closely mirroring the training accuracy indicates effective learning without significant overfitting. However, the fluctuations seen in the validation loss suggest moments where model performance on unseen data varied, likely due to the model capturing noise along with the actual signal during training [38]. Strategies like data augmentation, dropout, and the use of EarlyStopping and ReduceLROnPlateau callbacks were critical in mitigating these effects, supporting findings from other studies that emphasize the importance of these techniques in enhancing model robustness [39].

C. Challenges in Disease Classification

The performance of the model across different disease classes varied, with certain diseases like Brown Spot and False Smut being classified with higher precision and recall than others such as Blast Disease. This variation could be attributed to the distinct visual patterns that diseases manifest on the leaves, which may be captured differently by the CNN. The difficulty in distinguishing between some classes such as Bacterial Blight and Blast Disease raises important considerations about the limitations of visual-based diagnostics and suggests the potential for integrating other forms of data, such as spectral or thermal imaging, to improve classification accuracy [40].

D. Practical Implications

The practical applications of this research are significant. By enabling rapid and accurate disease detection, such systems can help farmers make timely decisions regarding disease management, potentially reducing crop losses and improving food security. The integration of this technology into mobile platforms or drones could facilitate widespread monitoring of crop health at scale, a prospect supported by recent advances in computational efficiency and model deployment [41]. However, the adoption of such technology also depends on factors like cost, accessibility, and user-friendliness, which must be addressed to ensure broad utility in diverse agricultural settings.

E. Future Directions

This study opens several avenues for future research. First, exploring the integration of different modalities of data, as mentioned earlier, could enhance the diagnostic capabilities of these models. Multi-modal data integration has been shown to provide a more holistic view of plant health, leading to more accurate disease identification [42]. Secondly, the development of more sophisticated model training approaches, such as transfer learning with fine-tuning or ensemble learning techniques, could further improve performance, especially in classes where the current model performance is suboptimal.

Additionally, longitudinal studies to track the model's performance across different growing seasons and under varying environmental conditions would provide deeper insights into its effectiveness and robustness in real-world scenarios. Such

studies would also help refine the models to handle variations in disease presentation due to climatic or geographical factors.

VI. CONCLUSION

In conclusion, this research demonstrated the efficacy of employing the VGG19 convolutional neural network, enhanced through data augmentation and specific training techniques, for the classification of rice crop diseases. The achieved high accuracy levels across both training and validation phases substantiate the model's ability to accurately learn and generalize from the dataset, which was meticulously curated to represent diverse disease manifestations. Key interventions such as the application of dropout, early stopping, and adaptive learning rate adjustments were pivotal in stabilizing the model's training process, mitigating overfitting, and ensuring robustness against variations in new data. The study's findings are in line with existing literature, reinforcing the assertion that pre-trained deep learning models are exceptionally capable of adapting to specialized tasks such as agricultural disease detection when properly fine-tuned and augmented. Future pathways for this line of inquiry include integrating multimodal data to capture a broader spectrum of disease indicators, enhancing model interpretability, and implementing these models in real-time disease monitoring systems, potentially on mobile or drone-based platforms. By continuing to refine these technologies and expanding their applicability, there is a significant potential to transform agricultural practices, enabling more efficient disease management, reducing crop losses, and thus contributing to global food security. This research lays a foundational step towards realizing such transformative agricultural innovations.

REFERENCES

[1] Aggarwal, M., Khullar, V., Goyal, N., Singh, A., Tolba, A., Thompson, E. B., & Kumar, S. (2023). Pre-trained deep neural network-based features selection supported machine learning for rice leaf disease classification. Agriculture, 13(5), 936.

[2] Suseno, J. R. K., Azhar, Y., & Minarno, A. E. (2023). The Implementation of Pretrained VGG16 Model for Rice Leaf Disease Classification using Image Segmentation. Kinetik: Game Technology, Information System, Computer Network, Computing, Electronics, and Control, 499-506.

[3] Pal, O. K. (2021, December). Identification of paddy leaf diseases using a supervised neural network. In 2021 16th International Conference on Emerging Technologies (ICET) (pp. 1-4). IEEE.

[4] Omarov, B., Batyrbekov, A., Suliman, A., Omarov, B., Sabdenbekov, Y., & Aknazarov, S. (2020, November). Electronic stethoscope for detecting heart abnormalities in athletes. In 2020 21st International Arab Conference on Information Technology (ACIT) (pp. 1-5). IEEE.

[5] Dogra, R., Rani, S., Singh, A., Albahar, M. A., Barrera, A. E., & Alkhayyat, A. (2023). Deep learning model for detection of brown spot rice leaf disease with smart agriculture. Computers and Electrical Engineering, 109, 108659.

[6] Tursynova, A., Omarov, B., Sakhipov, A., & Tukenova, N. (2022). Brain Stroke Lesion Segmentation Using Computed Tomography Images based on Modified U-Net Model with ResNet Blocks. International Journal of Online & Biomedical Engineering, 18(13).

[7] Pushpa, B. R., Ashok, A., & AV, S. H. (2021, September). Plant disease detection and classification using deep learning model. In 2021 third international conference on inventive research in computing applications (ICIRCA) (pp. 1285-1291). IEEE.

[8] Tursynova, A., Omarov, B., Tukenova, N., Salgozha, I., Khaaval, O., Ramazanov, R., & Ospanov, B. (2023). Deep learning-enabled brain stroke classification on computed tomography images. Comput. Mater. Contin, 75(1), 1431-1446.

[9] Tejaswini, P., Singh, P., Ramchandani, M., Rathore, Y. K., & Janghel, R. R. (2022, June). Rice leaf disease classification using CNN. In IOP Conference Series: Earth and Environmental Science (Vol. 1032, No. 1, p. 012017). IOP Publishing.

[10] Kaur, A., Kukreja, V., Tiwari, P., Manwal, M., & Sharma, R. (2024, April). An Efficient Deep Learning-based VGG19 Approach for Rice Leaf Disease Classification. In 2024 IEEE 9th International Conference for Convergence in Technology (I2CT) (pp. 1-6). IEEE.

[11] Pandi, S. S., Senthilselvi, A., Gitanjali, J., ArivuSelvan, K., Gopal, J., & Vellingiri, J. (2022). Rice plant disease classification using dilated convolutional neural network with global average pooling. Ecological Modelling, 474, 110166.

[12] Arya, A., & Mishra, P. K. (2023). A comprehensive review: advancements in pretrained and deep learning methods in the disease detection of rice plants. Journal of Artificial Intelligence and Capsule Networks, 5(3), 246-267.

[13] Biradar, V. G., Sarojadevi, H., Shalini, J., Veena, R. S., & Prashanth, V. (2022). Rice leaves disease classification using deep convolutional neural network. International Journal of Health Sciences, (IV), 1230-1244.

[14] Rezende, V., Costa, M., Santos, A., & de Oliveira, R. C. (2019, October). Image processing with convolutional neural networks for classification of plant diseases. In 2019 8th Brazilian Conference on Intelligent Systems (BRACIS) (pp. 705-710). IEEE.

[15] Latif, G., Abdelhamid, S. E., Mallouhy, R. E., Alghazo, J., & Kazimi, Z. A. (2022). Deep learning utilization in agriculture: Detection of rice plant diseases using an improved CNN model. Plants, 11(17), 2230.

[16] Upadhyay, S. K., & Kumar, A. (2022). A novel approach for rice plant diseases classification with deep convolutional neural network. International Journal of Information Technology, 14(1), 185-199.

[17] Iqbal, J., Hussain, I., Hakim, A., Ullah, S., & Yousuf, H. M. (2023). Early Detection and Classification of Rice Brown Spot and Bacterial Blight Diseases Using Digital Image Processing. Journal of Computing & Biomedical Informatics, 4(02), 98-109.

[18] Nguyen, T. H., Nguyen, T. N., & Ngo, B. V. A. (2022). VGG-19 Model with Transfer Learning and Image Segmentation for Classification of Tomato Leaf Disease. AgriEngineering 2022, 4, 871–887.

[19] Jaiswal, A., & Sachdeva, N. (2024, January). Early Prediction of Rice Leaf Disease Using Deep Neural Network Models. In 2024 14th International Conference on Cloud Computing, Data Science & Engineering (Confluence) (pp. 426-429). IEEE.

[20] Sobuj, M. S. I., Hossen, M. I., Mahmud, M. F., & Khan, M. U. I. (2024, March). Leveraging Pre-trained CNNs for Efficient Feature Extraction in Rice Leaf Disease Classification. In 2024 International Conference on Advances in Computing, Communication, Electrical, and Smart Systems (iCACCESS) (pp. 01-06). IEEE.

[21] Chen, J., Chen, J., Zhang, D., Sun, Y., & Nanehkaran, Y. A. (2020). Using deep transfer learning for image-based plant disease identification. Computers and Electronics in Agriculture, 173, 105393.

[22] Elakya, R., & Manoranjitham, T. (2023). Pest Classification in Paddy by Using Deep ConvNets and VGG19. In Recent Trends in Computational Intelligence and Its Application (pp. 75-84). CRC Press.

[23] Julianto, A., & Sunyoto, A. (2021). A performance evaluation of convolutional neural network architecture for classification of rice leaf disease. IAES International Journal of Artificial Intelligence, 10(4), 1069.

[24] Agrawal, M., & Agrawal, S. (2023). Rice plant diseases detection using convolutional neural networks. International Journal of Engineering Systems Modelling and Simulation, 14(1), 30-42.

[25] Simhadri, C. G., & Kondaveeti, H. K. (2023). Automatic recognition of rice leaf diseases using transfer learning. Agronomy, 13(4), 961.

[26] Malvade, N. N., Yakkundimath, R., Saunshi, G., Elemmi, M. C., & Baraki, P. (2022). A comparative analysis of paddy crop biotic stress classification using pre-trained deep neural networks. Artificial Intelligence in Agriculture, 6, 167-175.

[27] Achanta, C. B., Keerthi, K. D., & Kamepalli, S. (2022). Plant Leaf Disease Classification and Prediction Using a Customized Deep Transfer Learning Model. Journal of Algebraic Statistics, 13(3), 728-735.

[28] Win, A. T., Soe, K. M., & Lwin, M. M. (2024, March). Rice Disease Classification for Eastern Shan State Using Deep Learning. In 2024 IEEE Conference on Computer Applications (ICCA) (pp. 1-5). IEEE.

[29] Pherry, F., Kristanto, J., & Kurniadi, F. I. (2022, October). Rice Plants Disease Classification Using Transfer Learning. In 2022 4th International Conference on Cybernetics and Intelligent System (ICORIS) (pp. 1-4). IEEE.

[30] Doskarayev, B., Omarov, N., Omarov, B., Ismagulova, Z., Kozhamkulova, Z., Nurlybaeva, E., & Kasimova, G. (2023). Development of Computer Vision-enabled Augmented Reality Games to Increase Motivation for Sports. International Journal of Advanced Computer Science and Applications, 14(4).

[31] Sudheer, C. P., & Dharmani, B. C. Deep Learning for Disease Detection in Paddy Plants using Leaf Images. In Intelligent Circuits and Systems for SDG 3–Good Health and well-being (pp. 598-606). CRC Press.

[32] Aggarwal, M., Khullar, V., Goyal, N., Alammari, A., Albahar, M. A., & Singh, A. (2023). Lightweight federated learning for rice leaf disease classification using non independent and identically distributed images. Sustainability, 15(16), 12149.

[33] Kumar, G. K., Bangare, M. L., Bangare, P. M., Kumar, C. R., Raj, R., Arias-Gonzáles, J. L., ... & Mia, M. S. (2024). Internet of things sensors and support vector machine integrated intelligent irrigation system for agriculture industry. Discover Sustainability, 5(1), 6.

[34] Omarov, B., Narynov, S., & Zhumanov, Z. (2023). Artificial intelligence-enabled chatbots in mental health: a systematic review. Comput. Mater. Continua 74, 5105–5122 (2022).

[35] Mavaddat, M., Naderan, M., & Alavi, S. E. (2023, February). Classification of rice leaf diseases using cnn-based pre-trained models and transfer learning. In 2023 6th International Conference on Pattern Recognition and Image Analysis (IPRIA) (pp. 1-6). IEEE.

[36] Udayananda, G. K. V. L., Shyalika, C., & Kumara, P. P. N. V. (2022). Rice plant disease diagnosing using machine learning techniques: a comprehensive review. SN Applied Sciences, 4(11), 311.

[37] Hoang, V. D. (2021). Rice Leaf Diseases Recognition Based on Deep Learning and Hyperparameters Customization. In Frontiers of Computer Vision: 27th International Workshop, IW-FCV 2021, Daegu, South Korea, February 22–23, 2021, Revised Selected Papers 27 (pp. 189-200). Springer International Publishing.

[38] Ikram, Z. (2024, May). Hybrid Deep Neural Network for Face Liveness Detection in Real-Time Video. In 2024 IEEE 4th International Conference on Smart Information Systems and Technologies (SIST) (pp. 188-193). IEEE.

[39] Ikram, Z. (2024, May). Dual-Domain Face Anti-Spoofing with Integrated Spatial and Frequency Analysis Neural Network. In 2024 IEEE 4th International Conference on Smart Information Systems and Technologies (SIST) (pp. 228-232). IEEE.

[40] Akshitha, M., Siddesh, G. M., Sekhar, S. M., & Parameshachari, B. D. (2022, October). Paddy crop disease detection using deep learning techniques. In 2022 IEEE 2nd Mysore Sub Section International Conference (MysuruCon) (pp. 1-6). IEEE.

[41] Junaid, M., Iqbal, M. M., Hameed, N., & Saeed, S. (2022). Rice Disease Classification using Deep Learning. Journal of Information Communication Technologies and Robotic Applications, 13(1), 31-38.

[42] Sarker, M. R. K. R., Borsha, N. A., Sefatullah, M., Khan, A. R., Jannat, S., & Ali, H. (2022, April). A deep transfer learning-based approach to detect potato leaf disease at an earlier stage. In 2022 Second International Conference on Advances in Electrical, Computing, Communication and Sustainable Technologies (ICAECT) (pp. 1-5). IEEE.

Blockchain-Based Vaccination Record Tracking System

Shwetha G K[1], Jayantkumar A Rathod[2], Naveen G[3], Mounesh Arkachari[4], Pushparani M K[5]
Department of Computer Science and Engineering, NMAM Institute of Technology, Nitte, Karkala, India[1]
(Affiliated to NITTE Deemed to be University)
Department of Computer Science and Design, Alva's Institute of Engineering and Technology, Modubidire, India[2, 5]
Department of Information Science and Engineering, Alva's Institute of Engineering and Technology, Modubidire, India[3, 4]

***Abstract*—Blockchain technology is basically a decentralized database maintained by applicable parties and has been extensively used in colorful scripts similar as logistics and finance. In terms of operations in the medical field, it's getting increasingly important because the case's symptoms may be related to a certain vaccine. Whether the case has been vaccinated with this vaccine will lead to different individual results by the croaker. This study proposes a traceable blockchain-grounded vaccination record storehouse and sharing system. In the proposed system, the case gets the vaccination at any legal clinic and the VR can be saved accompanied by the hand into the blockchain center, which ensures traceability. When the case visits the sanitarium for treatment, the croaker can gain the details of the VR from the blockchain center and also make an opinion. The security of the proposed system will be defended by the programmed smart contracts. The proper record storage after encryption will ensure data privacy, integrity and security. Blockchain tracebility uses block-chain technology to record the movement of a product in the supply chain.**

Keywords*—Blockchain technology, decentralized, vaccine record tracking, integrity, smart contracts, vaccination record storage, traceability

I. Introduction

The global spread of Coronavirus Disease 2019 (COVID-19) in 2020 has posed unprecedented challenges to the healthcare sector, highlighting the use of innovative solutions to mitigate the spread of infections [1]. Contact tracing applications have emerged as a potential tool to break the chain of COVID-19 infections by identifying close contacts of positive cases and informing them about the possibility of being infected [2]. However, current contact tracing technologies face challenges in terms of privacy, accountability, and transparency, which can hinder their effectiveness and user adoption [3]. Our proposed method is noteworthy for its ability to track the origin and path of transactions related to vaccinations before they are used, in addition to preventing the circulation of counterfeit vaccines.

In this research paper, we propose a blockchain-based contact tracing solution that leverages the intrinsic features of blockchain technology to address the deficits of current contact tracing technologies [4]. Our solution aims to respect user privacy, provide transparency, and enable accountability by leveraging the decentralized, transparent, and immutable nature of blockchain technology [5]. Specifically, we utilize this blockchain with the smart contracts to eliminate the third-party servers, centralization, and identity abuse. Convergence algorithms, like proof of work or proof of stake, are used by the blockchain network to reach consensus on the ledger's current state and stop illegal changes. The accuracy and openness of the vaccination tracking data are therefore guaranteed.

Our solution utilizes the programmable logic of smart contracts to ensure transparency and trust among the different participants. All transactions on the blockchain are signed by their creators, holding every on-chain participant accountable for their actions [6]. By leveraging the immutable logs of the distributed ledger, our solution enforces transparency and trust, and eliminates the risks associated with centralized storage of user data [7]. The system architecture of our suggested application, Vaccine Tracker, which makes use of blockchain technology to offer complete visibility and transparency throughout the COVID vaccination supply chain. To guarantee the precision and dependability of vaccine tracking data, the Vaccine Tracker system uses a variety of algorithms for supply chain tracking and validation.

In this paper, we present the architecture, design, and implementation details of our blockchain-based contact tracing solution [8]. We also discuss the potential of blockchain technology in mitigating the spread of infections during the COVID-19 pandemic and highlight the advantages of using Ethereum blockchain with smart contracts for contact tracing [9]. The proposed architecture includes manufacturer Component which adds relevant information. The system verifies the information supplied by the manufacturer, such as location tracking and QR code generating. Our research contributes to the growing body of literature on blockchain technology in healthcare and contact tracing, and provides requirements for future research and practical applications [10]. Utilising cryptographic methods like digital signatures and hashing, the algorithm verifies the security and legitimacy of the information it has acquired from the blockchain. This guarantees the immutability and tamper-proof nature of the data recorded on the blockchain.

II. Related Work

Several studies and projects have explored the want of blockchain in healthcare, including its application in COVID-19 vaccination efforts. Here, we highlight some of the notable related work and contributions in the field.

A. Blockchain-Based Vaccination Verification Systems

Several blockchain-based systems have been used for verifying COVID-19 vaccination status. For example, the "VaxiChain" [11] aims to create a blockchain-based vaccination verification system that allows individuals to store their vaccination records in a secure and tamper-proof manner. The system uses smart contracts to automate the verification process, allowing healthcare providers, employers, and other entities to verify an individual's vaccination status without accessing their private health information. Another example is the "V-Health Passport" [12], which uses blockchain to enable secure storage and sharing of vaccination records, as well as other health-related data, with the aim of facilitating safe travel and other activities during the pandemic. Verify an individual's vaccination status without accessing their private health information. Another example is the "V-Health Passport", which uses blockchain to enable secure storage and sharing of vaccination records, as well as other health-related data, with the aim of facilitating safe travel and other activities during the pandemic.

B. Blockchain-Based Vaccine Distribution and Tracking

Blockchain has also been used as a solution for vaccine distribution and tracking in resource-constrained settings. The "Vaccination Supply Chain Management on Blockchain" project [13] proposes a blockchain-based system for tracking vaccine distribution, ensuring that vaccines reach their intended destinations and are administered to the right individuals. The system uses smart contracts to automate processes such as vaccine allocation, inventory management, and cold chain monitoring, thereby increasing transparency, efficiency, and accountability in the vaccine supply chain. Another example is the "Blockchain-Enabled COVID-19 Vaccine Delivery System" [14], which proposes a blockchain-based platform for tracking COVID-19 vaccines from the manufacturer to the end recipient, with the aim of reducing vaccine wastage, improving supply chain visibility, and preventing counterfeit vaccines.

C. Privacy-Preserving Approaches in Blockchain-Based Healthcare Systems

Privacy and security are critical concerns in healthcare, and several studies have proposed privacy-preserving approaches in the context of blockchain-based healthcare systems. For example, the "MediBloc" [15] proposes a blockchain-based health information that uses advanced cryptography techniques to protect patient privacy while enabling secure sharing of health data among authorized parties. The "HealthChain" [16] proposes a privacy-preserving blockchain-based system for storing and sharing electronic health records, using techniques such as zero-knowledge proofs and secure multi-party computation to protect patient confidentiality. These approaches highlight the potential of blockchain in maintaining data privacy and security in healthcare settings, including COVID-19 vaccination efforts.

D. Challenges and Limitations

Despite the potential usages, there are requirements and limitations associated with the use of blockchain in healthcare, including COVID-19 vaccination efforts. Interoperability, scalability, privacy, security, and regulatory compliance are some of the main challenges that need to be addressed. Several studies and projects have highlighted these challenges and proposed solutions to overcome them. For example, the "HL7 Fast Healthcare Interoperability Resources (FHIR) and Blockchain" [17] proposes a framework for combining blockchain with existing healthcare standards, such as FHIR, to enable interoperability. The "Scalable and Interoperable Blockchain-Based Healthcare Systems" [18] proposes a scalable and interoperable blockchain architecture for healthcare, using techniques such as sharding and cross-chain communication. The "Privacy-Preserving Blockchain-Based Consent Management System for Healthcare" [19] proposes a consent management system that uses blockchain and smart contracts to manage patient consent in a privacy-preserving manner. These efforts highlight the ongoing research and development in addressing the challenges and limitations of blockchain in healthcare [20].

III. System Design

In this part, the system design of our proposed Vaccine Tracker application, which utilizes blockchain technology to provide end-to-end visibility and transparency in the COVID vaccine supply chain.

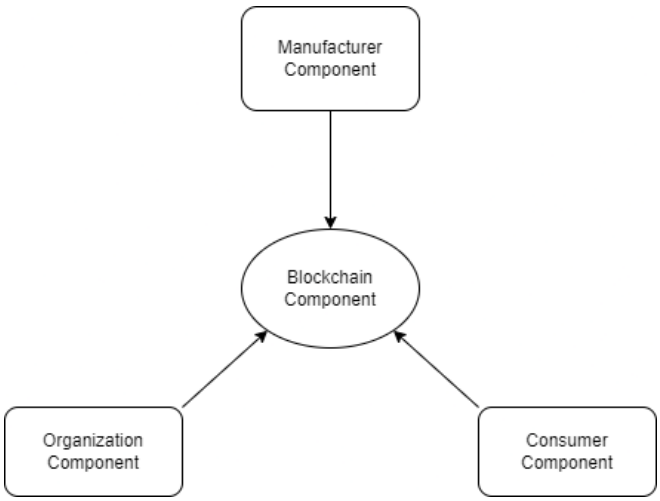


Fig. 1. System architecture.

Above Fig. 1 represents the proposed architecture.

A. Manufacturer Component

- The manufacturer adds a new vaccine batch to the system by entering relevant information, such as batch number, manufacturing date, expiration date, and other details.
- The system validates the data provided by the manufacturer, including QR code generation and location tracking data.
- Once the data is validated, the vaccine batch information is stored on the blockchain and shared with relevant stakeholders in the supply chain.

B. Organization Component

- Organizations involved in the vaccine supply chain, such as distributors, warehouses, and healthcare providers, can access the vaccine information from the blockchain.

- They can scan the vaccine shipments using QR code validation algorithms to verify the authenticity of the vaccines and record the location and condition data.
- The validated data is encrypted and stored on the blockchain for further validation and tracking.

C. Consumer Component

- Consumers can scan the QR code on the vaccine product using a mobile application.
- The QR code validation algorithm verifies the authenticity of the QR code data against the information stored on the blockchain.
- If the QR code is validated, the consumer can view exact information about the vaccine's provenance, including batch number, manufacturing date, and other details.

D. Blockchain Component

- The blockchain serves as a distributed ledger that securely stores the vaccine tracking data, including batch information, location data, and QR code validation results.
- The blockchain is encrypted and protected using cryptographic techniques, ensuring data security and integrity.
- Smart contracts programmed in Solidity language validate the data and enforce predefined rules and conditions

The proposed system employs a decentralized architecture using the Ethereum blockchain, which allows for transparent and tamper-proof recording of the vaccine's journey from the manufacturer to the hospital. The system comprises the following key components:

1) Vaccine Tracking: In this component, the vaccine's whereabouts at every freight hub, warehouse, and airport along its journey are recorded on the blockchain. This is achieved by utilizing smart contracts programmed in Solidity, which automatically record the vaccine's location, temperature conditions, and other relevant information at each step of the supply chain. This creates a transparent and immutable record of the vaccine's provenance, ensuring its authenticity and quality.

2) QR Code Scanning: At the consumer end, customers can simply scan the QR code of the vaccine product to access complete information about its provenance. The QR code acts as a unique identifier for each vaccine, and the information received from the blockchain includes details such as the manufacturer, shipping route, temperature conditions, and chain of custody. This empowers consumers to identify the authenticity of the vaccine before purchasing, mitigating the risk of counterfeit vaccines and ensuring their safety.

3) Blockchain Network: The blockchain network serves as the pillar of the system, consisting of a distributed network of nodes that maintain a shared ledger of vaccine tracking records. The blockchain network utilizes consensus algorithms, such as proof of work or proof of stake, to achieve agreement on the state of the ledger and prevent unauthorized modifications. This ensures the integrity and transparency of the vaccine tracking data.

4) Smart Contracts: Smart contracts are self-executing scripts that run on the blockchain and enable automated verification of vaccination records. In our system, smart contracts are used to define the rules and conditions for verifying vaccination status, such as checking the validity of the digital signatures, verifying the authenticity of the vaccine administered, and ensuring compliance with vaccination protocols.

5) User Interface: The system includes user interfaces and experiences for different stakeholders, such as vaccine manufacturers, logistics providers, hospitals, and consumers. These interfaces provide a user-friendly way to interact with the system, allowing stakeholders to input and retrieve vaccine tracking data, as well as verify the authenticity of vaccines through QR code scanning.

IV. Implementation Details

This section presents the used algorithms along with their implementation and coding details. The solidity code is written and tested with the Remix IDE. The Vaccine Tracker system employs various algorithms for supply chain tracking and validation to ensure the accuracy and reliability of vaccine tracking data. These algorithms include:

A. Location Tracking Algorithm

1) When a vaccine shipment is received at a freight hub or a vaccination center, the location tracking algorithm is triggered to collect the precise location data of the vaccines.

2) The algorithm utilizes GPS or other location tracking technologies, such as RFID (Radio Frequency Identification) or IoT (Internet of Things) devices, to collect the real-time location data of the vaccine shipment.

3) The location data, which may include latitude, longitude, timestamp, and other relevant information, is encrypted using cryptographic algorithms to use data security and privacy.

4) The encrypted location data is then stored as a transaction on the Ethereum blockchain, using a smart contract specifically designed for vaccine tracking. The transaction includes the encrypted location data, as well as other relevant information, such as the batch number, manufacturing date, and vaccine details.

5) The smart contract validates the encrypted location data to ensure its integrity and authenticity. This validation is done using cryptographic techniques, such as hashing and digital signatures, to verify the data against predefined rules and cryptographic keys.

6) Once the location data is validated, it is combined to the blockchain, creating a transparent and immutable record of the vaccine's journey. The location data is now available for real-time tracking and validation by authorized users, such as healthcare providers, regulators, and consumers.

7) The authorized users can access the location data from

the blockchain using a mobile app or a web interface, which decrypts the data using the appropriate cryptographic keys. The decrypted location data can then be displayed on a map or other visual representations, allowing users to track the movement of vaccines at each stage of the supply chain.

8) Any changes or updates to the location data, such as when the vaccine shipment moves to a new location, are recorded as new transactions on the blockchain, creating a chronological and auditable history of the vaccine's journey.

9) The location tracking algorithm continues to collect and record the precise location data of the vaccine shipment at each stage of the supply chain, ensuring real-time tracking and validation of vaccine movements throughout the entire supply chain process.

10) In case of any discrepancies or anomalies in the location data, such as a vaccine shipment being diverted or tampered with, the algorithm can trigger alerts or notifications to authorized users, enabling timely intervention and resolution of any issues.

Overall, the Location Tracking Algorithm in the Vaccine Tracker system uses GPS or other location tracking technologies to collect and record the precise location data of vaccines at each stage of the supply chain. The location data is encrypted and stored on the blockchain, allowing for real-time tracking and validation of vaccine movements, ensuring transparency, integrity, and authenticity of vaccine tracking data.

B. QR Code Validation Algorithm

1) When a consumer scans a QR code on a vaccine product using a mobile app or a web interface, the QR code validation algorithm is triggered to verify the authenticity of the scanned QR code.

2) The algorithm extracts the data from the scanned QR code, which will include requirement such as the batch number, manufacturing date, and other details of the vaccine.

3) The algorithm accesses the Ethereum blockchain, where the vaccine tracking data is stored, and retrieves the relevant information for the scanned QR code from the blockchain.

4) The retrieved information from the blockchain is compared with the data extracted from the scanned QR code to verify if they match. This comparison includes checking the batch number, manufacturing date, and other details to ensure that the QR code corresponds to a genuine vaccine.

5) The algorithm uses cryptographic techniques, such as hashing and digital signatures, to validate the authenticity and security of the data retrieved from the blockchain. This ensures that the data stored in the blockchain is tamper-proof and cannot be manipulated.

6) If the data extracted from the scanned QR code matches the information stored in the blockchain, the algorithm confirms that the QR code corresponds to a genuine vaccine and provides a validation result to the consumer, indicating that the vaccine is authentic.

7) If the data does not match, the algorithm raises an alert or notification, indicating that the scanned QR code may not correspond to a genuine vaccine, and further investigation or action may be required.

8) The QR code validation algorithm also keeps a record of all the QR codes scanned by consumers, along with their validation results, as transactions on the blockchain. This creates a transparent and auditable history of QR code validations, ensuring accountability and traceability.

9) The algorithm continues to validate QR codes scanned by consumers in real-time, ensuring that only authentic vaccines are verified and consumed by consumers, thereby preventing the use of counterfeit vaccines.

10) In case of any updates or changes to the vaccine tracking data in the blockchain, such as a new batch of vaccines being added or expired vaccines being removed, the QR code validation algorithm updates its reference data accordingly to ensure accurate validation results.

Overall, the QR Code Validation Algorithm in the Vaccine Tracker system validates the authenticity of QR codes scanned by consumers by checking the QR code data against the information stored in the blockchain, using cryptographic techniques to ensure data integrity and authenticity. This helps in verifying genuine vaccines and preventing the use of counterfeit vaccines by consumers.

C. Data Validation Algorithm

1) The Data Validation Algorithm constantly monitors the vaccine tracking data stored on the Ethereum blockchain to ensure data integrity and authenticity.

2) The algorithm retrieves the vaccine tracking data from the blockchain, including information such as vaccine batch numbers, manufacturing dates, shipment details, and other relevant data.

3) The algorithm verifies the integrity of the data by using cryptographic hashing techniques, where the data is hashed using a predefined hashing algorithm to generate a unique hash value.

4) The algorithm compares the computed hash value with the hash value of the data stored in the blockchain for the same data. If the hash values match, it indicates that the data has not been tampered with and is intact.

5) The algorithm also uses digital signatures to validate the authenticity of the data. Digital signatures used for this are generated using private keys and can be verified using corresponding public keys. The algorithm verifies the digital signatures associated with the vaccine tracking data using the public keys stored in the blockchain.

6) The algorithm checks if the data inside in the blockchain adheres to predefined rules and formats, such as checking if the batch numbers follow a certain pattern, manufacturing dates fall within expected ranges, and other validation checks specific to the vaccine supply chain.

7) The algorithm validates the authenticity of the data by verifying the cryptographic keys associated with the data. It ensures that the digital signatures match the corresponding public keys stored in the blockchain, indicating that the data has

been signed by authorized entities and has not been tampered with.

8) If the data passes all the validation checks, the algorithm confirms the integrity and authenticity of the data and provides a validation result indicating that the data is valid.

9) If any discrepancy or inconsistency is detected during the validation process, the algorithm raises an alert or notification, indicating that the data may have been replaced with or does not adhere to predefined rules.

10) The Data Validation Algorithm continuously monitors the vaccine tracking data on the blockchain, ensuring that the data remains tamper-proof and authentic throughout the supply chain process.

11) In case of any updates or changes to the vaccine tracking data in the blockchain, the Data Validation Algorithm revalidates the data based on the updated information and ensures that the updated data adheres to the predefined rules and cryptographic validation checks.

Overall, the Data Validation Algorithm in the Vaccine Tracker system ensures the combination and authenticity of the vaccine tracking data stored on the blockchain by using cryptographic techniques, such as hashing and digital signatures, and validating the data against predefined rules and formats. This helps in maintaining the trust and reliability of the vaccine tracking system and preventing any tampering or manipulation of the data stored on the blockchain.

V. Discussion

A. Transparency and Efficiency in Vaccine Supply Chains

The use of blockchain and AI algorithms in the Vaccine Tracker system can enhance transparency and efficiency in vaccine supply chains. Real-time location tracking using GPS or other technologies allows for accurate monitoring of vaccine movements, minimizing delays and identifying potential issues. QR code validation ensures the authenticity of vaccines, preventing the use of counterfeit vaccines. Data validation using cryptographic techniques ensures data integrity and authenticity, enhancing trust in the system.

B. Advantages of Location Tracking Algorithm

The location tracking algorithm used in the Vaccine Tracker system offers several advantages. It enables real-time tracking of vaccines at each stage of the supply chain, providing complete visibility and allowing for timely actions. However, challenges such as reliance on location tracking technologies, limitations in certain environments, and potential security concerns need to be addressed during implementation.

C. Robustness of QR Code Validation Algorithm

The QR code validation algorithm in the Vaccine Tracker system provides a robust method for authenticating vaccines. By verifying QR code data against blockchain-stored information, it ensures that consumers are scanning genuine QR codes associated with authentic vaccines. This helps protect consumer safety, maintain the integrity of the vaccination process, and prevent the use of counterfeit vaccines. However, interoperability with existing QR code standards and potential issues with QR code readability need to be addressed.

D. Data Validation Algorithm for Ensuring Integrity and Authenticity

The data validation algorithm used in the Vaccine Tracker system plays a crucial role in ensuring the integrity and authenticity of vaccine tracking data stored on the blockchain. By using cryptographic techniques, it validates data against predefined rules and verifies authenticity using cryptographic keys. This ensures that the data stored on the blockchain is tamper-proof and can be trusted. However, proper key management practices, potential security risks, and scalability of the algorithm need to be carefully evaluated and addressed.

E. Challenges

There are many challenges and limitations to consider in the usage of the Vaccine Tracker system. These include reliance on location tracking technologies, interoperability with existing systems, user adoption and acceptance, scalability during high-demand scenarios, and cost-effectiveness. Additionally, potential security risks associated with cryptographic keys, QR code readability issues, and adherence to industry standards should be considered during implementation.

VI. Experimental Results

A. Manufacturer Adding New Vaccine Batch

In this scenario, the Vaccine Tracker system is tested by simulating the addition of a new vaccine batch by a manufacturer. Below Fig. 2 shows the process of manufacturer adding new vaccine batch.

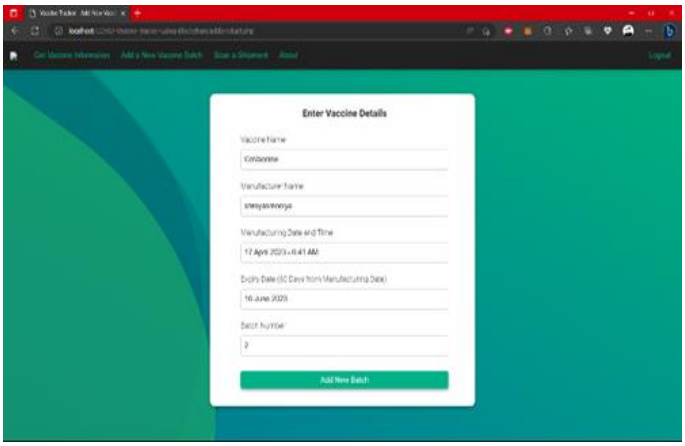


Fig. 2. Adding vaccine batch.

The experimental results may include the following:

- Successful addition of a new vaccine batch to the blockchain, including batch number, manufacturing date, expiration date, and other relevant information.
- Real-time recording of the batch information on the blockchain, ensuring transparency and immutability.
- Verification of the data validation algorithm, which ensures the combination and authenticity of the data stored on the blockchain.
- Validation of the QR code validation algorithm, which verifies the authenticity of the QR codes associated with the new vaccine batch.

- Successful integration of the location tracking algorithm, which tracks the movement of the vaccine batch using GPS or other location tracking technologies.
- Proper encryption and storage of location data on the blockchain, ensuring data privacy and security.

B. Organizations Viewing Vaccine Information and Scanning Shipment

In this scenario, the Vaccine Tracker system is tested by simulating the viewing of vaccine information and scanning of a shipment by organizations involved in vaccine supply chain. Fig. 3 shows vaccine shipment process.

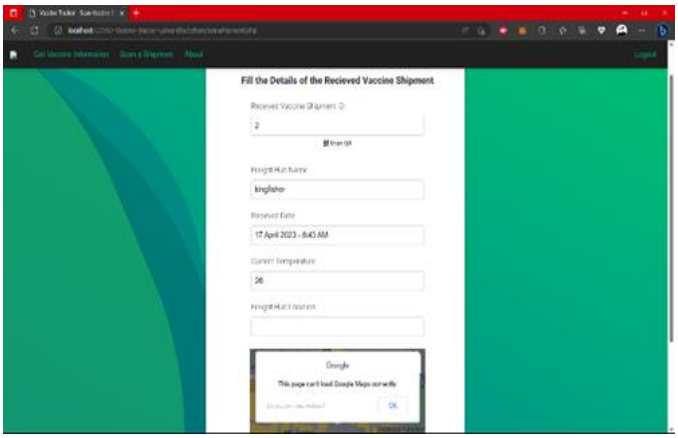


Fig. 3. Vaccine shipment.

The experimental results may include the following:

- Successful viewing of vaccine information, including batch number, manufacturing date, expiration date, location data, and QR codes, by relevant organizations such as distributors, healthcare providers, and regulatory authorities.
- Real-time tracking of the vaccine shipment using the location tracking algorithm, which provides accurate and on-date information on shipment's whereabouts.
- Validation of QR code validation algorithm, which verifies the authenticity of the scanned QR codes against the blockchain-stored information.
- Verification of the data validation algorithm, which ensures the combination and authenticity of the vaccine tracking data stored on the blockchain.
- Detection of any discrepancies or inconsistencies in the vaccine information, which may indicate potential issues such as counterfeit vaccines or supply chain disruptions.
- Proper access controls and permissions management, ensuring that only authorized companies can view and scan vaccine information.

C. Consumer Getting Vaccine Information

In this scenario, the Vaccine Tracker system is tested by simulating the retrieval of vaccine information by a consumer using QR code scanning. Below Fig. 4 shows how consumer can get vaccine information.

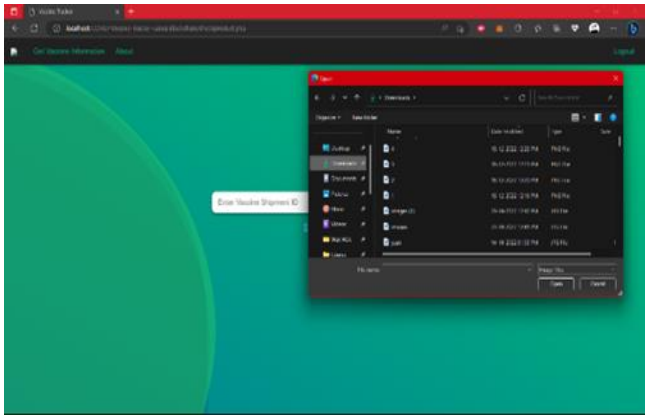


Fig. 4. Consumer page.

The experimental results may include the following:

- Successful scanning of the QR code on a vaccine package by a consumer using a smartphone or other QR code scanning devices.
- Retrieval of vaccine information, such as number, manufacturing date, expiration date, and location data, from the blockchain.
- Verification of the authenticity of the scanned QR code using the QR code validation algorithm, which ensures that the consumer is scanning a genuine QR code associated with an authentic vaccine.
- Provision of accurate and reliable vaccine information to the consumer, promoting security and trust in the vaccination process.
- Proper encryption and storage of consumer data, ensuring data privacy and security.

D. Tracking Log of Vaccine

In this part, the results of our blockchain-based vaccine tracking system, which enables end-to-end tracking of COVID-19 vaccines from the manufacturer to the consumer, with real-time updates displayed in Google Maps. Fig. 5 depicts the tracking of vaccine from source to destination.

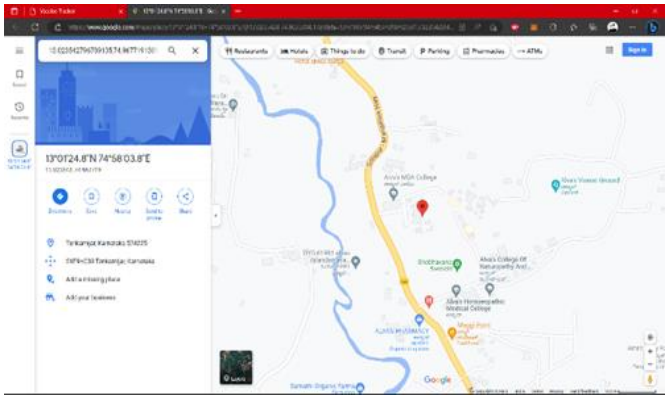


Fig. 5. Track record.

VII. Conclusion and Future Work

The Vaccine Tracker system, leveraging blockchain technology and AI-based algorithms, offers a robust and transparent solution for tracking and validating vaccine movements in the supply chain. The system provides real-time tracking of vaccines, validates the authenticity of QR codes, and ensures the integrity and authenticity of vaccine tracking data stored on the blockchain. The experimental results from testing and validation scenarios, including manufacturer adding new vaccine batch, organizations viewing vaccine information and scanning shipment, and consumer getting vaccine information, demonstrate the effectiveness and potential of the system in improving vaccine supply chain management.

The implementation of the Vaccine Tracker system has many key benefits, including enhanced transparency, traceability, and accountability in the vaccine supply chain. The system enables stakeholders, including manufacturers, distributors, healthcare providers, regulatory authorities, and consumers, to access accurate and reliable vaccine information in real-time, ensuring the authenticity and integrity of vaccines. The use of cryptographic techniques, such as hashing and digital signatures, ensures data security and privacy, protecting against data tampering and unauthorized access.

However, it's required to note that the success of the Vaccine Tracker system also relies on taking and collaboration of various stakeholders in the vaccine supply chain, as well as adherence to relevant regulations and standards. Further research and development can be explored to optimize the system's performance, scalability, and interoperability with existing vaccine supply chain management systems.

In conclusion, the Vaccine Tracker system has the potential to significantly improve vaccine supply chain management by leveraging blockchain technology and AI-based algorithms. The system's transparency, traceability, and data validation capabilities can enhance the trust and efficiency of vaccine distribution, ensuring that safe and authentic vaccines reach the end consumers. With proper implementation, the Vaccine Tracker system can contribute to the global efforts in ensuring the availability and safety of vaccines, ultimately benefiting public health worldwide. In the future work, we can improve the proposed system using artificial intelligence (AI) and deep learning techniques, for even further advancement, our suggested method can use a hybrid machine learning approach.

References

[1] M. Swan, Blockchain: Blueprint for a new economy. O'Reilly Media, Inc., 2015 G. Prisco (2016, April), The Blockchain for Healthcare: Gem Launches Gem Health Network With Philips Blockchain Lab.

[2] Z. Li, A. V. Barenji, G. Q. Huang, Toward a blockchain] A. Asgary, M. M. Najafabadi, R. Karsseboom, and J. Wu, “A drivethrough tool forvaccination during COVID-19 pandemic, Robotics and Computer-Integrated Manufacturing, Vol. 54, pp. 133-144, December, 2018.

[3] K.K. Tsoi, J.J. Sung, H.W. Lee, et al., The way forward after COVID-19 vaccination: vaccine passports with blockchain to protect personal privacy, BMJ Innovat. 7 (2) (2021) 337–341.

[4] C. Pop, T. Cioara, I. Anghel, M. Antal, and I. Salomie, “Distributed ledger technology review and decentralized applications development guidelines,” Future Internet, vol. 13, pp. 62, 2021.

[5] VitalikButerin. 2014. Ethereum White Paper: A Next Generation Smart Contract & Decentralized Application Platform. Ethereum January (2014), 1– 36.

[6] Sheng-I Chen, Bryan A. Norman, Jayant Rajgopal, Tina M. Assi, Bruce Y. Lee & Shawn T. Brown (2014) A planning model for the WHO-EPI vaccine distribution network in developing countries, IIE Transactions, 46:8, 853-865

[7] T. Bocek and B. Stiller, Smart Contracts - Blockchains in the Wings. Tiergartenstr. 17, 69121 Heidelberg, Germany: Springer, January 2017.

[8] Y. Yang and M. Ma, IEEE Transactions on Information Forensics and Security, vol. 11, p. 1, 2015

[9] Bischoff, P.: COVID-19 App Tracker: Is privacy being sacrificed in a bid to combat the virus?. CompariTech, 20 April 2020.Johns Hopkins University, 2020.

[10] Alammary, A.; Alhazmi, S.; Almasri, M.; Gillani, S. Blockchain-based applications in education: A systematic review. Appl. Sci. 2019, 9, 2400.

[11] Koyama, A.; Tran, V.C.; Fujimoto, M.; Bao, V.N.Q.; Tran, T.H. A Decentralized COVID-19 Vaccine Tracking System Using Blockchain Technology. Cryptography 2023.

[12] Samer Barakat,” Blockchain Tracking System of COVID-19 Vaccination” Received 15 May 2021; Accepted 20 May 2021.

[13] Malak SulaimanAlromaih,” COVAC: A Blockchain-Based COVID Testing and Vaccination Tracking System”.

[14] Ruwaida Nazim Shaikh,”Block Chain Based Electronic Vaccination Record Storing System”25-26 March 2022.

[15] Kamanashis Biswas,” A reliable vaccine tracking and monitoring system for health clinics using blockchain”, (2023) 13:570.

[16] Abhishek Sharma,” Blockchain technology and its applications to combat COVID-19 pandemic”, Received: 27 May 2020 / Accepted: 15 October 2020.

[17] Jingshou Chen, Received 3 October 2021; Revised 30 January 2022; Accepted 14 February 2022; Published 9 March 2022.

[18] Ahmad musamih, Received April 21, 2021, accepted May 8, 2021, date of publication May 11, 2021, date of current version May 19, 2021.

[19] Pranab Kumar Bharimalla,” A Blockchain and NLP Based Electronic Health Record System: Indian Subcontinent Context”, Informatica 45 (2021) 605–616.

[20] Alabdulkarim, Y.; Alameer, A.; Almukaynizi, M.; Almaslukh, A. SPIN: A Blockchain-Based Framework for Sharing COVID-19 Pandemic Information across Nations.

Enhance the Security of the Cloud Using a Hybrid Optimization-Based Proxy Re-Encryption Technique Considered Blockchain

Ahmed I. Alutaibi
Department of Computer Engineering-College of Computer and Information Sciences,
Majmaah University, Majmaah 11952, Saudi Arabia

***Abstract*—Every day, a vast amount of data with incalculable value will be generated by the IoT devices that are deployed in various types of applications. It is crucial to ensure the reliability and safety of IoT data exchange in a cloud context because this data frequently contains the user's private information. This study presents a novel encrypted data storage and security system using the blockchain method in conjunction with hybrid optimization-based proxy re-encryption (HO-PREB). Dependency on outside central service providers is eliminated by the HO-PREB-based consensus process. In the blockchain system, several consensus nodes serve as proxy service nodes to restore encrypted data and merge transformed ciphertext with private data. Hybrid owl and bat optimization is employed to select the optimal key for enhancing security. This removes the limitations associated with securely storing and distributing private encrypted data via a distributed network. Moreover, the blockchain's distributed ledger ensures the permanent storage of data-sharing records and outcomes, ensuring accuracy and dependability. The simulated experiments of the designed model are evaluated with existing cryptographic techniques and gain a lower latency of 3.2 s and a lower turnaround time of 45 ms. Furthermore, the developed technique enhances cloud system security and possesses the capability to detect and mitigate attacks in the cloud environment.**

Keywords*—Cloud security; Internet of Things; proxy re-encryption; blockchain; data sharing; hybrid optimization

I. Introduction

Cloud computing is one of the best and most difficult platforms accessible today, and businesses of all sizes are starting to employ its services. A range of cloud deployment techniques are available, and cloud services are provided according to requirements, like protecting the cloud system's internal and external safety [1]. Common technology flaws, malware, hackers, and other similar threats, unauthorized access, inadequate precautions, denial of service, unsecured interfaces, profile or business traffic hijacking, and data leaks are the primary risks to cloud computing protection [2]. Cloud control covers a broad range of topics, such as handling resources, hardware and software safety, cloud data protection, and resource consumption [3]. Despite the many benefits of cloud storage, cloud users often prioritize security and privacy issues, which discourages businesses and organizations from adopting big changes. While using the cloud benefits enterprises and institutions, privacy and confidence are the most vital challenges [4, 5]. The data of cloud customers is highly vulnerable to loss, leakage, or attack, and they are left with no way out of this untenable scenario [6].

When exporting and buying offerings from cloud providers, cloud customers may apply blockchain technology, a novel and evolving technology, to boost data privacy and confidence [7, 8]. Blockchain can provide improved security based on centralized database safety. Blockchain monitors the set of data that is encrypted, stored, and linked to the previous block using an encrypted hash function [9, 10]. A blockchain is a type of networked ledger that can prevent manipulation and store operations. Peer-to-peer networks are typically used to manage blockchains, which are meant to guard against outside interference [11, 12]. The security provided by the blockchain system can be compared to the safety of centralized data storage. From a management standpoint, attacks and damage to data storage can be prevented [13]. Furthermore, the accessible feature of the blockchain can facilitate data openness when used in an environment where sharing information is mandated. These benefits allow it to be used in a wide range of situations, including those involving the financial sector and the Internet of Things (IoT), and its applications are expected to expand [14]. Cloud computing has been integrated into many IT systems since it is effective and widely accessible. Moreover, privacy and cloud security issues have been investigated as crucial security components [15].

As IoT technology develops quickly, a vast number of IoT devices are being used in many application situations. As a result, ensuring dependability and IoT data exchange security is essential [16]. A significant volume of IoT data is generated and managed by the data owners. These enormous amounts of data must be encrypted and sent to a trusted organization for storage because IoT devices have limited storage. Before transferring the data to cloud databases and shared storage systems, they can choose to encrypt it [17, 18]. To safeguard the privacy and security of data owners (DO), an efficient access control system must be established. The users don't have a completely reliable means of communication through which to exchange the decryption key because the data is encrypted. The data owners must download, decrypt, and re-encrypt the material before they may share it directly with the recipients [19]. Unfortunately, because of their limited processing capacity, the DO is unable to pay to decrypt the data before re-encrypting it for the recipients. As a result, they can exchange IoT data by using the PRE technique. This transfers the burden of ciphertext decryption and encryption from the information

owners to the procedure of creating a new encryption key, thereby redistributing the workload [20].

This work presents the development of a hybrid owl search and bat PRE method to improve cloud computing security and robustness. Additionally, the blockchain model generates blocks and secure cloud storage of user data to enhance privacy. Additionally, the blockchain concept is built to give extra scalability for safeguarding data in blocks and secret keys, and data is safely decrypted via PRE. Additionally, the HO-PREB system avoids reliance on outside central service providers. In the distributed ledger network, several consensus nodes serve as proxy service nodes to merge translated cipher text and re-encrypt data. This system can guarantee the suitability, security, and dependability of cloud computing's IoT data exchange.

The overview of the paper is arranged in sections. Sections II and III detail related works and a problem statement; Section IV explains the proposed methodology; Section V gives detailed results and discussion; and Section VI ends with a conclusion and future scope.

II. Literature Review

A. Related Works

A distributed data-controlled sharing strategy based on PRE is proposed by Ismail et al. [21]. First, a method for PRE is built using the blockchain and SM2. The information-controlled sharing method protects transaction information confidentiality while enabling data security sharing through the use of PRE. A system for adaptive user rights modification is suggested. To completely maintain the privacy of transaction data and carry out the evaluation of information access permission by monitoring PRE key settings, they are creating a PRE method with SM2.

A PRE method was created by Kwame et al. [22] for secure cloud-based data-sharing scenarios. Data owners can send their protected data to the cloud using identity-based encryption (IBE), where it may be accessed by authorized users using the PRE architecture. Moreover, it increases the quality of service by effectively supplying content that has been stored in the proxy by utilizing information-centric network functions. Moreover, blockchain is an innovative technology that permits data sharing to be decentralized. It accomplishes fine-grained control over access to data and reduces bottlenecks in central databases.

To increase safety and confidentiality in a private blockchain, Bharat et al. [23] introduced the Splitting of PRE Method (Split-PRE), which is based on the IoT. To address trust issues as well as scalability issues, this paper suggests a blockchain-based PRE method that will streamline transactions. The IoT data is saved by the system in a decentralized cloud after encryption. Through the use of an effective PRE method, the owner and the person in the smart contract can both view the data.

A blockchain-based ecosystem for IoT data exchange is presented by Ahsan et al. [24]. Additionally, transport the data from the information generator to the consumer safely and anonymously by using a PRE mechanism. Without the assistance of a reliable third party, the system establishes dynamic, temporal connections between data consumption and sensors to share the gathered IoT data. This cutting-edge website for storing, exchanging, and handling sensor data is quick, easy, and safe.

Yingwen et al. [25] combined blockchain computing with PRE to create a novel encrypted data storage and communication design. The utilization of threshold PRE as a compromise technique avoids reliance on external central service providers. In the blockchain system, several consensus nodes serve as proxy service nodes, combining translated ciphertext and re-encrypting data. The findings demonstrate that the suggested architecture can raise a reasonable time delay while satisfying the high demands for data access.

To accomplish effective data sharing and data accuracy checking, Tao Feng et al. [26] suggested a technique that utilizes IBE. Additionally, a blockchain platform is integrated to ensure secure and regulated data storage, addressing issues such as manipulation of data and insufficient supervision on external servers. Lastly, assess the computation and transmission overhead to prove the security of the planned system. The findings demonstrated that the designed method surpasses other plans in terms of efficiency and security against specific plaintext attacks with observable characteristics.

To enable user identification and access to the public key's binding properties on the Internet of Medical Things, Hongmei et al. [27] offer a safe data sharing system named PRE for safe information exchange with blockchain, thereby boosting the security of data sharing. The created approach implements the management and quick query of users by adding them to the accumulator through the application of a blockchain smart agreement. Lastly, carry out the four-stage computing performance evaluation experiments to enhance the effectiveness of the encryption, re-encryption, decryption, and re-decryption computations.

B. Problem Statement

Transaction data is kept on a blockchain in a decentralized shared global ledger. Finding the right balance while sharing data safeguarding privacy, and usefulness is difficult. Furthermore, one difficult issue is the dynamic modification of blockchain data access permissions [28]. The problem description in the realms of cloud computing and blockchain technology centers on data availability, confidentiality, and integrity protection in distributed systems. Maintaining the confidentiality of sensitive data processed and stored across enormous networks of linked servers is a difficulty with cloud computing [29]. Strong security measures are necessary to reduce risks associated with problems like unauthorized access, data breaches, and service interruptions. Similarly, it is critical to protect distributed ledgers' confidentiality and agreement in blockchain technology from risks like 51% attacks, double-spends, and vulnerabilities in smart contracts. The most frequent and difficult problems with cloud computing safety are low efficiency, trust concerns, limited scalability, difficulty in maintaining access restrictions, and leakage of sensitive data [30]. To meet these problems and maintain the confidence and integrity necessary for these revolutionary technologies, creative encryption methods, strict access controls, and robust network designs are needed.

III. Proposed Methodology

Create the best possible keys and increase the security of the proposed model by designing a hybrid owl search and bat optimization (HOSBA) algorithm. The blockchain network, storage service providers (SSP), data users (DU), and data owners (DO) make up the four primary components of the created system. Fig. 1 depicts the planned strategy process.

C. Process of the Hybrid Optimization-based PRE Model

To improve the public cloud privacy of user information, hybrid owl search and bat optimization (HOSBA) methods are used for the key generation phase. These improvements choose the most reliable and safe optimum key to improve cloud computing security.

1) System construction: To increase cloud storage security and safe data exchange, it integrates threshold PRE, hybrid optimization, and the blockchain consensus method. Elliptic Curve Digital Signature Algorithm (ECDSA) [31], PRE [32], and HOSBA are consulted in the construction of the developed mechanism. This approach can use a collection of N's blockchain consensus nodes to carry out the PRE procedure. The delegate can access the updated ciphertext by utilizing their private keys to decrypt it when necessary within the blockchain network of nodes in re-encryption. During the key generation step, the HOSBA model is employed to choose the best possible key. Below is a full explanation of the developed technique and process.

System setup

$$\left(1^{\gamma}, \bar{u}\right) \rightarrow \overline{pp}$$

The security parameters 1^{γ} is selected as inputs and the output of the public parameter is $\overline{pp}$. The $\overline{pp}$ is used for creating own public and private keys. Initially, it takes security parameter γ as the results and inputs for a bilinear map m : $\hat{G}_1 * \hat{G}_1 \rightarrow \hat{G}_2$, where $\hat{G}_1$, $\hat{G}_2$ Is a prime ordermultiplicative cyclic group p. A random generator is selected $g \in \hat{G}_1$, and calculated $z = m\left(\hat{g}, \hat{g}\right)$. Moreover, four hash functions are developed and are detailed in eqn. (1).

$$\hat{H}_0 : \{0,1\}^* \rightarrow \hat{G}_1, \hat{H}_1 : \{0,1\}^* \rightarrow \hat{G}_1, \hat{H}_2 : \hat{G}_2 \rightarrow \{0,1\}^{\log_2 p}, \hat{H}_3 : \hat{G}_2 \rightarrow \{0,1\}^* \quad (1)$$

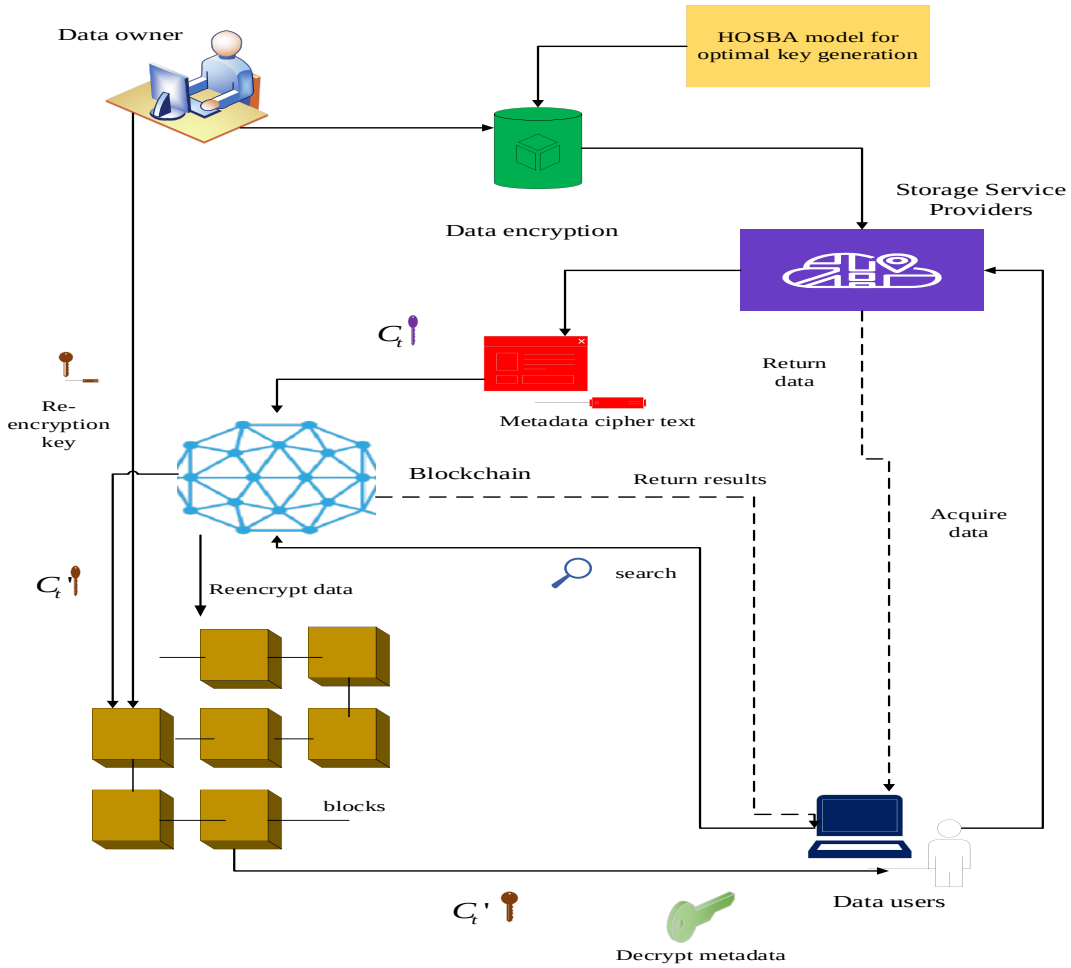


Fig. 1. Process of developed model.

Execute the procedure for parameter generation to produce shared parameters and is exhaustive in Eq. (2).

$$P_r = \left(\hat{g}, z, m, p, \hat{G}_1, \hat{G}_2, \hat{H}_0, \hat{H}_1, \hat{H}_2, \hat{H}_3\right) \tag{2}$$

D. Key Generation

The public and private key pairs are created by the DU and the DO based on the provided public parameters. The DO selected at random $u \in z_p^*$ is the private key and calculates the public key as $\overline{pk_{do}} = \hat{g}^u$. DU selected at random $v \in z_p^*$ is the private key and calculates the public key as $\overline{pk_{du}} = \hat{g}^v$.

Hybrid owl search and bat algorithm:

The optimal key is a cryptographic key that is chosen or fine-tuned to maximize the security of encrypted data. This key is crucial for ensuring that data is securely encrypted and protected against unauthorized access. The Hybrid Owl and Bat optimization algorithm is used to find the best possible key for encryption. This optimization process involves selecting parameters for encryption that enhance security measures, such as resisting attacks or ensuring robustness against various threats. This algorithm explores the key space efficiently to identify the key that provides the highest level of security.

A novel hybrid optimization approach has been developed to select the best optimal key for enhancing the security of data in the cloud. It draws inspiration from the actions of the Bat Optimization (BAO) [33] and Owl Optimization Algorithm (OOA) [34] when a threat is present. The proposed hybrid optimization method selects the best optimal key according to the efforts of both parties to increase efficiency and security. It conceptually integrates BAO and OOA.

Owl begins its optimization process using a range of initial random solutions that match the starting positions of owls in a forest. The n quantity of owls in a woodland, x is the space of dimensions used for searches. Next, a random position is saved for them in a $n \times x$ matrix as given in Eq. (3).

$$\vec{o} = \begin{bmatrix} \vec{o}_{1,1} & \vec{o}_{1,2} & \ldots & \vec{o}_{1,x} \\ \vec{o}_{2,1} & \vec{o}_{2,2} & \ldots & \vec{o}_{2,x} \\ : & : & : & : \\ \vec{o}_{n,1} & \vec{o}_{n,2} & \ldots & \vec{o}_{n,x} \end{bmatrix} \tag{3}$$

The element of the matrix $\vec{o}_{i,j,}$ characterizes the j^{th} variable of an i^{th} owl. Eq. (4) is utilized to assign the beginning location of every owl in the forest based on a uniform distribution.

$$\vec{o}_i = \vec{o}_l + \vec{u}(0,1) \times (\vec{o}_u - \vec{o}_l) \tag{4}$$

Let, $\vec{o}_l$ and $\vec{o}_u$ are the lower and upper boundaries of i^{th} owl $\vec{o}_i$ in j^{th} length and $\vec{u}(0,1)$ is an arbitrary amount in the range [0,1] that is evenly distributed. An objective function is used to assess the fitness of each owl's placement inside a forest, and the results are kept in the matrix shown in Eq. (5).

$$\vec{f} = \begin{bmatrix} \vec{f}_1\left(\left[\vec{o}_{1,1}, \vec{o}_{1,2}, \ldots..\vec{o}_{1,x}\right]\right) \\ \vec{f}_2\left(\left[\vec{o}_{2,1}, \vec{o}_{2,2}, \ldots..\vec{o}_{2,x}\right]\right) \\ : \\ : \\ \vec{f}_n\left(\left[\vec{o}_{n,1}, \vec{o}_{n,2}, \ldots..\vec{o}_{n,x}\right]\right) \end{bmatrix} \tag{5}$$

The current study assumes that each owl's positional fitness value is directly correlated with the volume of information it hears. Since the greatest owl is closer to the vole, it receives the maximum intensity. The information concerning the adjusted intensity of the i^{th} owl is calculated using Eq. (6) and is used to update the position.

$$\vec{I}_i = \frac{\vec{f}_i - \vec{w}}{\vec{b} - \vec{w}} \tag{6}$$

Equation (7) calculates the details of the separation between each owl and its meal.

$$\vec{r}_i = \left\|\vec{o}_i, \vec{v}\right\|_2 \tag{7}$$

Let, $\vec{v}$ is the position of the prey that the most agile owl finds. Additionally, it is presumable that the forest has a single vole, or worldwide optimum. Owls fly silently in the direction of their prey. As a result, they experience altered intensity by the sound intensity inverted square law. The variation in strength for i^{th} owl can be acquired using Eq. (8).

$$\vec{I}c_i = \frac{\vec{I}_i}{\vec{r}_i^2} + random_noise \tag{8}$$

Eq. (8), $\vec{r}_i^2$ is used instead of $4\pi\vec{r}_i^2$ and thought that adding random noise to the surroundings will enhance the realism of the computational framework. Since voles are active in the actual world, their mobility compels owls to adjust their existing location softly. At this point, the optimal optima key is generated for the owl's new position phase via bat optimization, which improves the cloud computing privacy procedure. All bats utilize echolocation to detect distance, and in some mysterious way, they are also able to distinguish between background obstacles and food/prey; Bats fly arbitrarily with velocity $\hat{v}_i^t$ at position $\hat{x}_i^t$ with a fixed frequency $\hat{f}_i$ to search for prey. Depending on the closeness of their target, they can automatically modify the frequency and rate at which they emit pulses within the interval [0, 1].

Thus, Eq. (9), which gives the new positions of the hybrid owl and bat model.

$$\vec{o}_i^{t+1} = \begin{Bmatrix} \vec{o}_i^t + \beta \times \vec{I}c_i \times \left| \alpha \vec{v} - \vec{o}_i^t \right| & if \vec{p}_v < 0.5 \\ \hat{x}_i^{t-1} + \hat{v}_i^t & if \vec{p}_v < 0.5 \end{Bmatrix} \quad (9)$$

Let, $\vec{p}_v$ is the likelihood of a vole moving, α is a uniformly distributed random number across the interval [0, 0.5], and β is a constant that decreases linearly from 1.9 to 0. β encourages the investigation of the search space and first makes significant alterations. Moreover, $\hat{x}_i^{t-1}$ is denoted as the new position of the bats concerning time $t-1$, $\hat{v}_i^t$ is denoted as velocity at time t, and the velocity of the bat is measured using Eq. (10).

$$\hat{v}_i^t = \hat{v}_i^{t-1} + \left(\hat{x}_i^t - \hat{x}_* \right) \hat{f}_i \quad (10)$$

$\hat{f}_i$ is considered as frequency, $\hat{x}_i^t$ is denoted as the new position of the bats, and $\hat{x}_*$ is denoted as the current global best position. Ultimately, the optimized model selects the best secret, private, and public key to secure the user data from unauthorized access. Algorithm 1 provides the Hybrid owl search and bat algorithm.

Algorithm 1: Hybrid owl search and bat algorithm

1. Initialize the algorithm by setting the population size, number of iterations, boundaries of the search space, and parameters for Owl and Bat Optimization.

2. Generate initial owl population by randomly generating initial positions for owls within the search space, evaluating the fitness of owl positions using the objective function, and storing the fitness values in a matrix.

3. Start the optimization process by beginning the main loop for iterations.

Owl Optimization Phase:

- Calculate initial positions of owls based on random distribution.

- Update positions by calculating the intensity of information and adjusting based on proximity to prey.

- Introduce random noise to simulate a realistic search.

- Recalculate fitness of updated positions and store the best one found.

Update Intensity and Distance:

- Compute intensity for each owl based on proximity to prey.

- Adjust positions based on intensity and random noise.

- Calculate the distance between owls and prey and update positions.

Bat Optimization Phase:

- Adjust the frequency of bats based on distance to the best-known solution.

- Update velocity and position of bats using calculated frequency.

- Perform a local search around the best solution, adjusting positions based on probability.

- Update positions and velocities based on proximity to prey and best-known positions.

Combine Results from Both Algorithms:

- Compare the best solutions from Owl and Bat Optimization phases.

- Update the global best solution if a better position is found.

4. Conclude the optimization by finalizing the best-found solution as the optimal key after completing iterations.

5. Output the optimized key generated from the hybrid optimization process.

E. Data Encryption

$$\left(D_{ek}, D \right) \rightarrow C_t$$

Additionally, the IoT data is encrypted using a symmetric encryption technique. Users using symmetric key encryption must be aware of a shared secret key. The encrypted data C_t is generated using a common secret key D_{ek} are submitted for SSP to store the encrypted data using Eq. (11).

$$C_t = encrypt\left(D_{ek}, D \right) \quad (11)$$

F. Metadata Encryption

$\left(\overline{pp}, \overline{pk_{do}}, \hat{m} \right) \rightarrow C_t \hat{m}$: Metadata $\hat{m}$ is encrypted by the DO by using a public key, and sends cipher text $C_t \hat{m}$ to the blockchain. Moreover, a DO private key is desired to decrypt the data. Metadata $\hat{m}$ contains the data summary, data store location, data decryption key D_{dk}, C_t, etc. Metadata encryption is detailed in eqn. (12).

$$C_t \hat{m} = Enc\left(\overline{pk_{do}}, \hat{m} \right) = \left(\hat{g}_{uh}, \hat{m}z_h \right) \quad (12)$$

Where, $z = m\left(\hat{g}, \hat{g} \right)$, h is an arbitrary coefficient.

G. Re-Encryption Key Generation

By their private key and the public key of the sender DU who wishes to obtain the data, the DO computes and produces the re-encryption key. The $\overline{sk}_{du}$ is the input secret key, the designated delegate's public key pk_{du}, proxy nodes quantity $\hat{p}_n$, and the threshold t_h, ReKeyGen, a technique for

generating re-encryption keys, calculates n components of the key used to re-encrypt data between DO and DU in Eq. (13).

$$rk_{do\to du} = \hat{g}^{v/u} \tag{13}$$

Using the eqn. (14) that follows,

$$\overline{f}(x) = \sum_{i=1}^{h} \overline{f}(x_i)\gamma_{ij}, \gamma_{ij} = \prod_{I=1, L\neq j}^{h} \frac{x - x_I}{x_i - x_I}, \tag{14}$$

each proxy node's share of the re-encryption key $rh_{o\to u}^{i} = \hat{g}^{\overline{f}(x_i)} \bmod p$, and, $\hat{p}_n \; i = 1, 2, 3, ...n$. The ciphertext transformation requests will be posted to the blockchain system by the DO.

H. Re-Encryption

The ciphertext transformation procedure is started by the majority of nodes in the distributed ledger network in responding to re-encryption requests after they have received the re-encryption key made public through the DO. It re-encrypts $C_t\hat{m}$ from the DO to the DU, according to $C_t\hat{m} = (\hat{g}_{uh}, \hat{m}z_h)$. Re-encryption keys allow proxy nodes to modify the original ciphertext. $rh_{o\to u}^{i}$, $i = 1, 2, 3, ...n,$ for $\hat{p}_n$ to the $C'_t\hat{m}$ over the following Eq. (15).

$$m\left(\hat{g}_{uh}, \left(rh_{\hat{p}_n}\right)^{\gamma_i}\right) = m\left(\hat{g}_{uh}, \hat{g}^{[\gamma_i \cdot \overline{f}(x_i)]}\right) = z^{uh[\gamma_i \cdot \overline{f}(x_i)]} \tag{15}$$

Every proxy node $\hat{p}_n$ can access the encrypted text using Eq. (16).

$$C_{v_i} = \left(z^{uh[\gamma_i \cdot \overline{f}(x_i)]}, \hat{m}z_h\right) \tag{16}$$

When t out of n the ciphertext is appropriately completed by proxy nodes in the re-encryption computation. C_{v_i} can be integrated into the updated ciphertext. $C'_t\hat{m}$, which DU can decode using their private key using Eq. (17) and (18)

$$\prod_{i=1}^{t} z^{uh[\gamma_i \cdot \overline{f}(x_i)]} = z^{uh\sum_{i=1}^{h}\gamma_i \cdot \overline{f}(x_i)} = z^{uh.v/a} = z^{vh} \tag{17}$$

$$C'_t\hat{m} = \left(z^{vh}, \hat{m}z_h\right)\left(\bmod p\right) \tag{18}$$

Following the completion of the consensus confirmation by the proxy nodes, the new block will contain the entire interpreted ciphertext.

I. Metadata Decryption

The private key $\overline{sk}_{du}$ of the DU is suitable for decrypting the ciphertext metadata $C'_t\hat{m} = \left(z^{vh}, \hat{m}z_h\right) = \left(\hat{\alpha}, \hat{\beta}\right)$ as well as the subsequent metadata using Eq. (19).

$$\hat{m} = \hat{\beta} \Big/ \hat{\alpha}^{1/\overline{sk}_{du}} \tag{19}$$

J. Data Decryption

Then receiving the metadata $\hat{m}$, using DU, they can obtain the exact location of the information storage as well as the binary decryption password for the data. Using data analysis and signature, both are contained in the metadata, we can use SSP to retrieve the plaintext of the data and confirm its accuracy and integrity. The data decryption is performed using Eq. (20).

$$Data = decrypt\left(D_{ek}, C_t\right) \tag{20}$$

The designed hybrid optimization-based PRE model improves the safety of cloud data by re-encrypting user information and applying a consensus technique to increase security. Algorithm 2 illustrates the optimal threshold PRE procedure.

Algorithm: 2 data encryption and decryption using optimized threshold PRE

Start
{
Initialization
{
$\left(1^{\gamma}, \overline{u}\right) \to \overline{pp}$ *// 1^{γ} - security parameters*
// $\overline{pp}$ - public parameter
bilinear map, m
four hash functions { $\hat{H}_0$, $\hat{H}_1$, $\hat{H}_2$, $\hat{H}_3$ *}*
}
Key generation
{
$\overline{pp} \to \left(\overline{sk}_{do}, \overline{pk_{do}}, \overline{sk}_{du}, \overline{pk_{du}}\right)$
Update HOSBA *// generate the best optimal key*
Initialize the population of owls and bat
Update new position using eqn.9
Generate optimal keys $\overline{sk}_{do}, \overline{pk_{do}}, \overline{sk}_{du}, \overline{pk_{du}}$ *// public, private, and secrete keys of DO and DU*
}
Data encryption
{
Use a secrete key, D_{ek}
Convert plain text into cipher *//AES*

$$C_t = encrypt\left(D_{ek}, D\right)$$

}

Metadata Encryption

{

Encrypt $\hat{m}$ *using a public key*

$\left(\overline{pp}, \overline{pk_{do}}, \hat{m}\right) \rightarrow C_t \hat{m}$

}

Re-encryption Key Generation

$rk_{do \rightarrow du} = \hat{g}^{v/u}$

If ($rh_{o \rightarrow u}^{i} = \hat{g}^{\bar{f}(x_i)} \bmod p$ ***)***

{

Generate re-encryption key

}

End if

Re-encryption

{

Update proxy nodes $\hat{p}_n$

consensus confirmation using proxy nodes

Re-encrypt the data using eqn. (18)

}

Metadata Decryption

{

Using private key $\overline{sk}_{du}$ *decrypting the ciphertext metadata*

}

Data Decryption:

{

Decrypt using metadata $\hat{m}$ *, symmetric decryption key* D_{ek}

$Data = decrypt\left(D_{ek}, C_t\right)$

}

Secure the data

Enhance performance

}

End

K. Consensus Mechanism based on Threshold PRE

The process of consensus is an essential element of a developed system. In contrast to the conventional PRE method, the developed approach makes use of a decentralized network to do away with the need for outside central service providers. Re-encrypted ciphertext transformation is the basis for the consensus process that is collaboratively carried out by consensus nodes in blockchain-based systems. The re-encryption key is divided and distributed across all consensus nodes within the blockchain system; the threshold PRE may be seamlessly integrated with the consensus method. The data owners concurrently communicate their demands for ciphertext transformation to the blockchain network and distribute the newly created re-encryption keys among other consensus nodes. A significant amount of processing power is needed for the ciphertext conversion operation. Since not every node in the network is capable of re-encryption, certain nodes with particular processing capabilities respond to this demand by performing re-encryption calculations. These nodes function in the distributed ledger system as consensus nodes. Fig. 2 displays a description of the consensus method based on the established model.

Process of consensus nodes: The consensus node locates the original ciphertext that matches. $C_t \hat{m}$ in the blockchain ledger, via the re-encryption key after fulfilling the re-encryption conditions $rh_{o \rightarrow u}^{i}$, $i = 1, 2, 3, ...n$ to complete the conversion process. The consensus nodes confirm the ciphertext after it has been re-encrypted, finish the re-encryption calculation within the allotted time C_{v_i}, $i = 1, 2, 3, ...n$ and send it to the blockchain system; Each consensus node compiles and verifies the encrypted text that has been re-encrypted C_{v_i}. A vote for selecting the leader of the network of blockchain servers is started by the consensus node that leads the charge in gathering t authenticated re-encrypted ciphertexts.

Fig. 2 is overview of consensus model with developed technique.

After receiving this request, further consensus nodes verify that the re-encrypted ciphertext is accurate. These re-encrypted ciphertexts are combined by the leader node C_{v_i} to the novel ciphertext $C'_t \hat{m}$. This leader node will also add the newly generated block, the re-encrypted data, and the modified ciphertext to the blockchain database and broadcast it to the whole network. Additional nodes modify the ledger after completing consensus confirmation. After that, the HO-PREB is finished, and the ciphertext for the metadata is changed. $C'_t \hat{m}$ is updated in the block. The blockchain ledger provides users with the altered metadata ciphertext, allowing them to complete the decryption process.

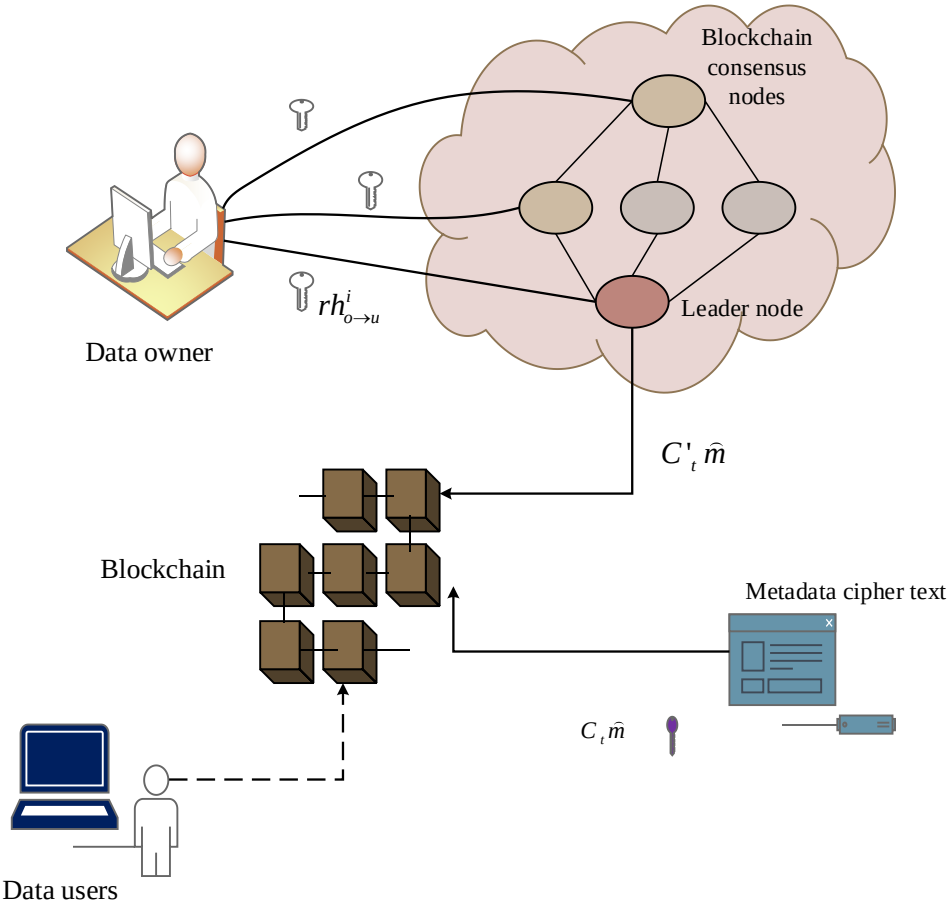


Fig. 2. Overview of consensus model with developed technique.

IV. Result and Discussion

The enhanced efficiency of the created method is verified by comparing it to current methods concerning encryption time, decryption time, re-encryption time, latency, restoration effectiveness, and turnaround time. The created model's robustness and safety are verified using accepted cryptographic approaches after it is implemented using a Python tool. The hybrid owl and bat optimization is employed to improve the user's data's security by generating the optimal key. Finally, it is securely stored in the cloud using blockchain by generating blocks.

A. Performance Analysis

To prove the efficiency and reliability of the developed system, the results are validated with existing classifiers such as Advanced Encryption Standard (AES) [36], Data Encryption Standard (DES) [35], Homomorphic Encryption (HE) [37], Rivest-Shamir-Adleman (RSA) [38], and IBE [39]. The performance metrics used for the validation are decryption time, encryption time, latency, re-encryption time, restoration efficiency, and turnaround time.

B. Encryption Time

The efficiency of any encryption operation is determined by dividing all the encrypted plaintext by the encryption time (milliseconds).

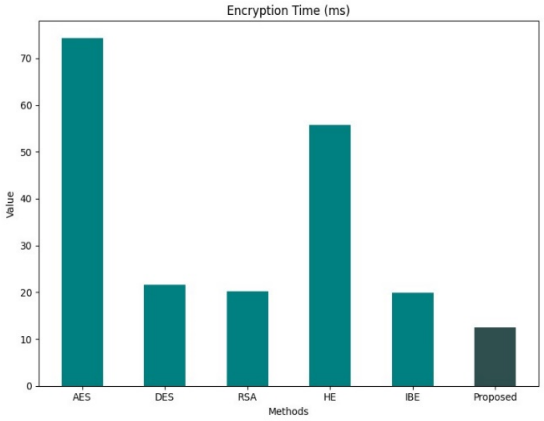


Fig. 3. Encryption time comparison.

Fig. 3 shows the milliseconds (ms) that each cryptographic algorithm takes to encrypt data. The duration of encryption operations is used to evaluate each scheme. AES, DES, RSA, HE, and IBE algorithms are among those examined in the analysis, along with a proposed method. Of the algorithms that have been studied, AES has the longest encryption time—74.25 ms—on record. DES has an encryption time of 21.66 ms, which is shorter than AES's. At 20.28 ms, RSA has a comparatively short encryption time. While IBE displays a comparatively

short encryption time of 19.97 ms, HE requires 55.74 ms. With a value of 12.5 ms, the proposed algorithm finally shows the lowest encryption time out of all the algorithms examined.

C. Re-Encryption Time

With PRE, a proxy can change the encryption of a ciphertext from one key to another, encrypting the identical message.

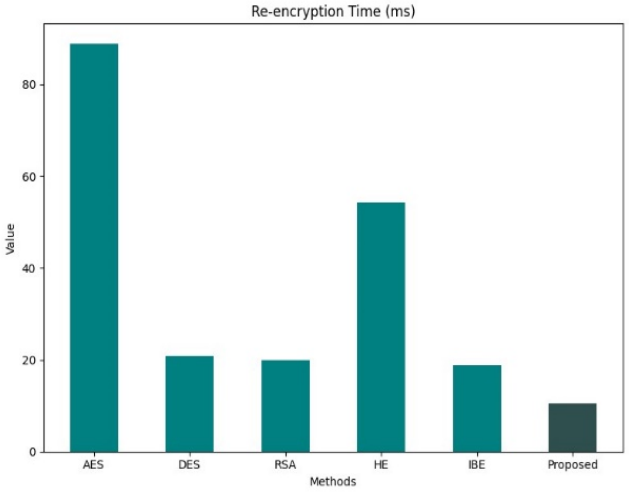


Fig. 4. Re-encryption time comparison.

Performance analyses of the data re-encryption times for the different encryption techniques are shown in Fig. 4. The time it takes to re-encrypt data is indicated by each algorithm's re-encryption time. This is an important factor in cryptographic operations, especially when situations call for regular updates or changes to encryption keys. The highest re-encryption time for the AES algorithm is 88.8 ms, but the DES algorithm displays a duration of 20.83 ms. Furthermore, the RSA and IBE re-encryption times are 19.98 ms and 18.86 ms, respectively. The developed technique has the lowest re-encryption time of all the records, at 10.4 ms, whereas the HE shows a re-encryption time of 54.32 ms. Comparing the suggested algorithm to the other stated algorithms, it may be more efficient in re-encryption activities.

D. Decryption Time

The process of restoring encrypted material to its original form is known as decryption. Usually, the encryption procedure is done in reverse. Because decryption necessitates a secret key, it examines the encrypted data to ensure only an authorized user may decrypt it.

A performance analysis of decryption times using different encryption techniques is shown in Fig. 5. While the suggested algorithm is a novel or modified encryption technique, AES, DES, RSA, HE, and IBE are well-known encryption techniques. DES and RSA are the two classic encryption methods with the fastest decryption times, respectively, at 24.81 and 23.12 milliseconds. Adopted as an encryption standard, AES is relatively slower than DES and RSA, taking 55.58 ms to decrypt. IBE shows a comparatively faster decryption time of 20.99 ms, while HE takes 43.65 ms. With a decryption time of only 8.5 ms, the suggested technique stands out as remarkable. This implies that, in comparison to the current encryption methods, the suggested approach may provide considerable gains in decryption efficiency.

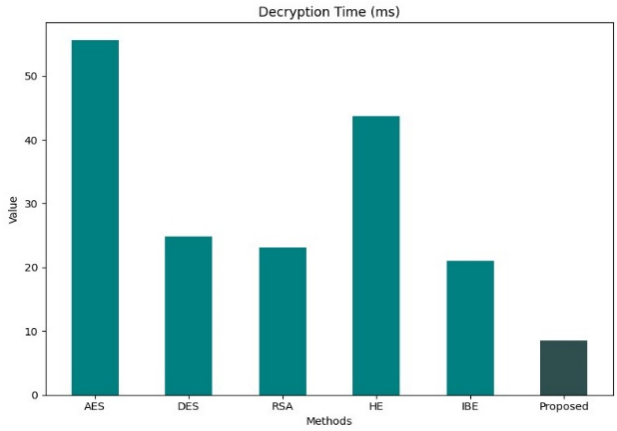


Fig. 5. Decryption time comparison.

E. Latency

The amount of time required for a specific design to finish a specified (computational) task is known as latency. Latency is the amount of time it takes for data to travel across the internet from one location to another. The term "latency" describes how long an algorithm takes to complete a task; shorter latency times correspond to quicker algorithm performance.

Performance analysis of the delay of different encryptions is provided in Fig. 6. With a latency of 47.54 seconds, AES has the highest of all the encryption methods listed, followed by DES, which has a latency of 23.45 seconds. Furthermore, the delay for RSA is 20.67 seconds, while the next highest latency is 32.6 seconds for HE. The suggested encryption technique exhibits the lowest latency of 3.2 seconds, while IBE displays a latency of 14.5 seconds. This suggests that this algorithm performs better than the others.

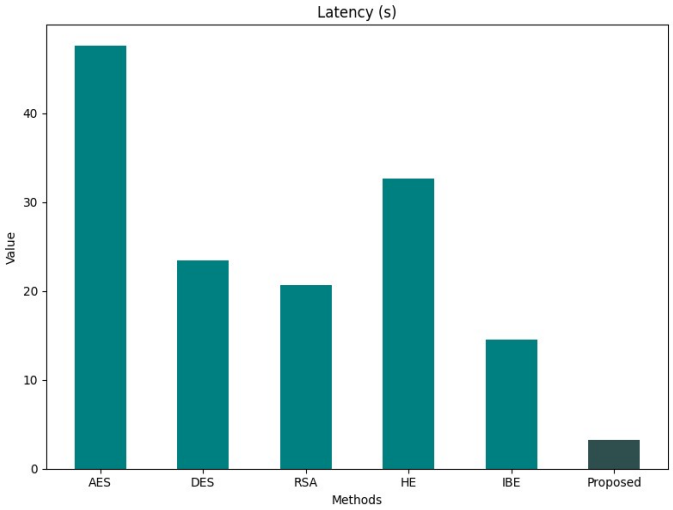


Fig. 6. Latency comparison.

F. Restoration Efficiency

The efficacy and speed at which a service or system may be brought back up following a malfunction, outage, or data loss is referred to as restoration efficiency.

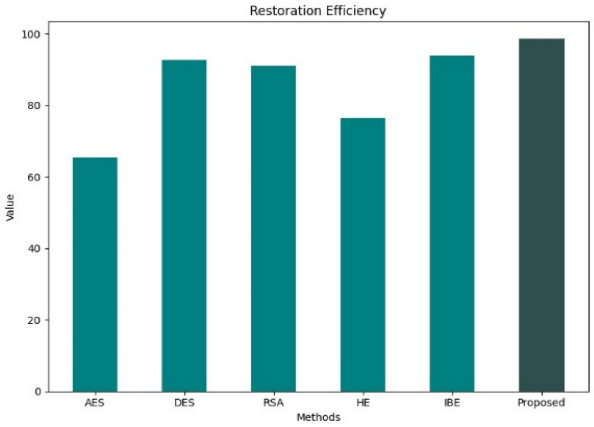


Fig. 7. Restoration efficiency comparison.

Performance analysis scores of restoration efficiency for different cryptographic methods are shown in Fig. 7. Additionally, DES obtained a better restoration efficiency score of 92.56 than AES, which got 65.4. Furthermore, in terms of restoration efficiency, RSA received a score of 91, HE earned 76.43, and IBE received an exceptionally high score of 93.9. With a restoration efficiency score of 98.5, the suggested method excels and demonstrates remarkable efficacy in quickly recovering systems or data protected by the technique.

G. Turnaround Time

The term "turnaround time" describes the length of time required to complete a process or task from beginning to end. Turnaround time, as used in the table, particularly refers to the milliseconds (ms) that each algorithm for cryptography requires to complete a given operation or computation.

The turnaround times, expressed in milliseconds (ms), for several cryptographic algorithms—AES, DES, RSA, HE, IBE, and a suggested algorithm—are shown in Fig. 8. With a turnaround time of 300 milliseconds, AES has the longest algorithmic delay of all the listed algorithms. With a turnaround time of 112 milliseconds, DES comes next. The turnaround time of the popular public-key encryption technique RSA is 95.3 milliseconds. The turnaround time for HE is 195 milliseconds, whereas the turnaround time for IBE is 92.54 milliseconds. It's interesting to note that, at 45 milliseconds, the suggested approach has the quickest turnaround time out of all those mentioned. This suggests a possible breakthrough in the efficiency of cryptography, providing much faster processing while upholding the essential security requirements.

H. Discussion

Existing cloud-based Key Management Services (e.g., AWS KMS, Azure Key Vault, and Google Cloud KMS) can be integrated with the HO-PREB system to manage encryption keys securely. The hybrid optimization algorithms (Owl and Bat optimization) used for key selection can interact with these services to dynamically generate and manage encryption keys. Encrypted data can be stored in cloud-based storage services such as Amazon S3, Azure Blob Storage, or Google Cloud Storage. The proxy nodes will manage access to the encrypted data, ensuring that only authorized entities can decrypt and access the data. However, issues such as computational overhead, integration with existing infrastructure, and compatibility with various cloud service models could be addressed to provide a more holistic view. While the blockchain model adds scalability, further details on how the system handles large-scale data and high-frequency IoT transactions would enhance understanding. This includes the impact on network bandwidth and storage requirements as the number of transactions and data size increase.

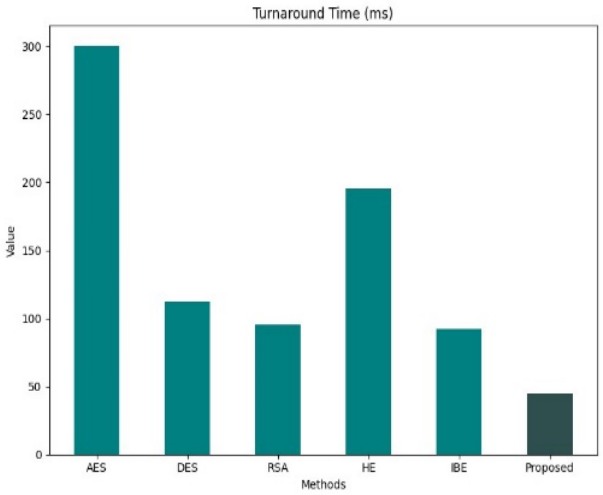


Fig. 8. Turnaround time comparison.

V. Conclusion

The hybrid optimization and threshold-based PRE system with consensus system, which is an innovative design that offers real-world applications for the safe sharing of IoT data, is proposed in this paper. It enables DO to effectively safeguard their encrypted data via SSP and share it with authorized users. The proposed solution can eliminate dependency on external central proxy servers by utilizing a distributed blockchain network and combining an optimized threshold PRE system with a blockchain consensus algorithm. The re-encryption of encrypted data can be computed during the consensus confirmation process of the distributed ledger network. Simulation studies showed that optimized threshold PRE can be successfully paired with the blockchain's consensus process to enhance cloud computing security and facilitate data exchange. In addition, when compared to other methods, the performance and scalability of the suggested system are adequate. The developed model attains encryption, re-encryption, and decryption times of 12.5 ms, 10.4 ms, and 8.5 ms. Also, gained less latency and a high restoration efficiency of 3.2s and 98.5%. The developed technique attains better performance to improve the cloud system's security, and it can detect and neglect attacks based on the proxy nodes in the consensus system. Also, there is potential to explore the integration of state-of-the-art cryptographic and machine learning methods to further enhance the security and efficacy of data exchange protocols.

ACKNOWLEDGMENT

The author extends the appreciation to the Deanship of Postgraduate Studies and Scientific Research at Majmaah University for funding this research work through the project number (R-2024-1241)

REFERENCES

[1] Awadallah, R. and Samsudin, A., 2021. Using blockchain in cloud computing to enhance relational database security. IEEE Access, 9, pp.137353-137366.

[2] Awadallah, R., Samsudin, A., Teh, J.S. and Almazrooie, M., 2021. An integrated architecture for maintaining security in cloud computing based on blockchain. IEEE Access, 9, pp.69513-69526.

[3] Shrivastava, P., Alam, B. and Alam, M., 2024. A hybrid lightweight blockchain based encryption scheme for security enhancement in cloud computing. Multimedia Tools and Applications, 83(1), pp.2683-2702.

[4] Velmurugadass, P., Dhanasekaran, S., Anand, S.S. and Vasudevan, V., 2021. Enhancing Blockchain security in cloud computing with IoT environment using ECIES and cryptography hash algorithm. Materials Today: Proceedings, 37, pp.2653-2659.

[5] Murthy, C.V.B., Shri, M.L., Kadry, S. and Lim, S., 2020. Blockchain based cloud computing: Architecture and research challenges. IEEE access, 8, pp.205190-205205.

[6] Zhang, H., Zang, Z. and Muthu, B., 2022. Knowledge-based systems for blockchain-based cognitive cloud computing model for security purposes. International Journal of Modeling, Simulation, and Scientific Computing, 13(04), p.2241002.

[7] Benil, T. and Jasper, J.J.C.N., 2020. Cloud based security on outsourcing using blockchain in E-health systems. Computer Networks, 178, p.107344.

[8] Gong, J. and Navimipour, N.J., 2022. An in-depth and systematic literature review on the blockchain-based approaches for cloud computing. Cluster Computing, 25(1), pp.383-400.

[9] Zou, J., He, D., Zeadally, S., Kumar, N., Wang, H. and Choo, K.R., 2021. Integrated blockchain and cloud computing systems: A systematic survey, solutions, and challenges. ACM Computing Surveys (CSUR), 54(8), pp.1-36.

[10] Wilczyński, A. and Kołodziej, J., 2020. Modelling and simulation of security-aware task scheduling in cloud computing based on Blockchain technology. Simulation Modelling Practice and Theory, 99, p.102038.

[11] Rahman, A., Islam, M.J., Band, S.S., Muhammad, G., Hasan, K. and Tiwari, P., 2023. Towards a blockchain-SDN-based secure architecture for cloud computing in smart industrial IoT. Digital Communications and Networks, 9(2), pp.411-421.

[12] Uddin, M., Khalique, A., Jumani, A.K., Ullah, S.S. and Hussain, S., 2021. Next-generation blockchain-enabled virtualized cloud security solutions: review and open challenges. Electronics, 10(20), p.2493.

[13] Bonnah, E. and Shiguang, J., 2020. DecChain: A decentralized security approach in Edge Computing based on Blockchain. Future Generation Computer Systems, 113, pp.363-379.

[14] Ali, A., Khan, A., Ahmed, M. and Jeon, G., 2022. BCALS: Blockchain-based secure log management system for cloud computing. Transactions on Emerging Telecommunications Technologies, 33(4), p.e4272.

[15] Wei, P., Wang, D., Zhao, Y., Tyagi, S.K.S. and Kumar, N., 2020. Blockchain data-based cloud data integrity protection mechanism. Future Generation Computer Systems, 102, pp.902-911.

[16] Sowmiya, B., Poovammal, E., Ramana, K., Singh, S. and Yoon, B., 2021. Linear elliptical curve digital signature (LECDS) with blockchain approach for enhanced security on cloud server. IEEE Access, 9, pp.138245-138253.

[17] Xie, G., Liu, Y., Xin, G. and Yang, Q., 2021. Blockchain-based cloud data integrity verification scheme with high efficiency. Security and Communication Networks, 2021, pp.1-15.

[18] Abirami, P. and Bhanu, S.V., 2020. Enhancing cloud security using crypto-deep neural network for privacy preservation in trusted environment. Soft Computing, 24(24), pp.18927-18936.

[19] Egala, B.S., Pradhan, A.K., Badarla, V. and Mohanty, S.P., 2021. Fortified-chain: a blockchain-based framework for security and privacy-assured internet of medical things with effective access control. IEEE Internet of Things Journal, 8(14), pp.11717-11731.

[20] Zhang, G., Yang, Z., Xie, H. and Liu, W., 2021. A secure authorized deduplication scheme for cloud data based on blockchain. Information Processing & Management, 58(3), p.102510.

[21] Keshta, I., Aoudni, Y., Sandhu, M., Singh, A., Xalikovich, P.A., Rizwan, A., Soni, M. and Lalar, S., 2023. Blockchain aware proxy re-encryption algorithm-based data sharing scheme. Physical Communication, 58, p.102048.

[22] Agyekum, K.O.B.O., Xia, Q., Sifah, E.B., Cobblah, C.N.A., Xia, H. and Gao, J., 2021. A proxy re-encryption approach to secure data sharing in the internet of things based on blockchain. IEEE Systems Journal, 16(1), pp.1685-1696.

[23] Rawal, B.S., Manogaran, G. and Hamdi, M., 2021. Multi-tier stack of block chain with proxy re-encryption method scheme on the internet of things platform. ACM Transactions on Internet Technology (TOIT), 22(2), pp.1-20.

[24] Manzoor, A., Braeken, A., Kanhere, S.S., Ylianttila, M. and Liyanage, M., 2021. Proxy re-encryption enabled secure and anonymous IoT data sharing platform based on blockchain. Journal of Network and Computer Applications, 176, p.102917.

[25] Chen, Y., Hu, B., Yu, H., Duan, Z. and Huang, J., 2021. A threshold proxy re-encryption scheme for secure IoT data sharing based on blockchain. Electronics, 10(19), p.2359.

[26] Feng, T., Wang, D. and Gong, R., 2023. A blockchain-based efficient and verifiable attribute-based proxy re-encryption cloud sharing scheme. Information, 14(5), p.281.

[27] Pei, H., Yang, P., Li, W., Du, M. and Hu, Z., 2024. Proxy re-encryption for secure data sharing with blockchain in Internet of Medical Things. Computer Networks, p.110373.

[28] Simaiya, S., Lilhore, U.K., Sharma, S.K., Gupta, K. and Baggan, V., 2020. Blockchain: A new technology to enhance data security and privacy in Internet of things. Journal of Computational and Theoretical Nanoscience, 17(6), pp.2552-2556.

[29] Poongodi, J., Kavitha, K. and Sathish, S., 2022. Healthcare Internet of Things (HIoT) data security enhancement using blockchain technology. Journal of Intelligent & Fuzzy Systems, 43(4), pp.5063-5073.

[30] Kollu, P.K., 2021. Blockchain techniques for secure storage of data in cloud environment. Turkish Journal of Computer and Mathematics Education (TURCOMAT), 12(11), pp.1515-1522.

[31] Liu, S.G., Chen, W.Q. and Liu, J.L., 2021. An efficient double parameter elliptic curve digital signature algorithm for blockchain. IEEE Access, 9, pp.77058-77066.

[32] Khashan, O.A., 2020. Hybrid lightweight proxy re-encryption scheme for secure fog-to-things environment. IEEE Access, 8, pp.66878-66887.

[33] Yang, X.S. and Hossein Gandomi, A., 2012. Bat algorithm: a novel approach for global engineering optimization. Engineering computations, 29(5), pp.464-483.

[34] Jain, M., Maurya, S., Rani, A. and Singh, V., 2018. Owl search algorithm: a novel nature-inspired heuristic paradigm for global optimization. Journal of Intelligent & Fuzzy Systems, 34(3), pp.1573-1582.

[35] Vuppala, A., Roshan, R.S., Nawaz, S. and Ravindra, J.V.R., 2020. An efficient optimization and secured triple data encryption standard using enhanced key scheduling algorithm. Procedia Computer Science, 171, pp.1054-1063.

[36] Altigani, A., Hasan, S., Barry, B., Naserelden, S., Elsadig, M.A. and Elshoush, H.T., 2021. A polymorphic advanced encryption standard–a novel approach. IEEE Access, 9, pp.20191-20207.

[37] Bozduman, H.Ç. and Afacan, E., 2020. Simulation of a homomorphic encryption system. Applied Mathematics and Nonlinear Sciences, 5(1), pp.479-484.

[38] Almuzaini, K.K., Dubey, R., Gandhi, C., Taram, M., Soni, A., Sharma, S., Sánchez-Chero, M. and Carrión-Barco, G., 2023. Secured wireless sensor networks using hybrid Rivest Shamir Adleman with ant lion optimization algorithm. Wireless Networks, pp.1-19.

[39] Deng, H., Qin, Z., Wu, Q., Guan, Z., Deng, R.H., Wang, Y. and Zhou, Y., 2020. Identity-based encryption transformation for flexible sharing of encrypted data in public cloud. IEEE Transactions on Information Forensics and Security, 15, pp.3168-3180.

Malicious Website Detection Using Random Forest and Pearson Correlation for Effective Feature Selection

Esha Sangra[1], Renuka Agrawal[2], Pravin Ramesh Gundalwar[3], Kanhaiya Sharma[4], Divyansh Bangri[5], Debadrita Nandi[6]
Department of Computer Science & Engineering, Symbiosis Institute of Technology,
Symbiosis International (Deemed University), Pune, India[1, 2, 4, 5, 6]
Department of Information Technology, Anurag University, Hyderabad India[3]

***Abstract*—In recent years, the internet has expanded rapidly, driving significant advancements in digitalization that have transformed day to day lives. Its growing influence on consumers and the economy has increased the risk of cyberattacks. Cybercriminals exploited network misconfigurations and security vulnerabilities during these transitions. Among countless cyberattacks, phishing remains the most common form of cybercrime. Phishing via malicious Uniform Resource Locator (URL)s threatens potential victims by posing as an imposter and stealing critical and sensitive data. An increase in cyberattacks using phishing needs immediate attention to find a scalable solution. Earlier techniques like blacklisting, signature matching, and regular expression method are insufficient because of the requirement to keep updating the rule engine or signature database regularly. Significant research has recently been conducted on using Machine Learning (ML) models to detect malicious URLs. In this study, the authors have provided a study highlighting the importance of significant feature selection for training ML models for detecting malicious URLs. Pearson correlation is employed in this study for selecting significant features, and the outcome demonstrates that in terms of accuracy and other performance indices, the Random Forest classifier outperforms the other classifiers.**

***Keywords*—Malicious URL; machine learning; feature selection; Random Forest, cybercrime**

I. INTRODUCTION

Phishing is a form of social engineering when a cyber threat actor poses as a reliable individual or group in order to trick a user into disclosing private information or unintentionally allowing access to their network [1]. Some attack techniques that use malicious URLs include Drive-by Download, Phishing and Social Engineering, and Spam [2-4]. The potential outcomes include data breaches, loss of data or services, identity theft, malware infections, or ransomware attacks. Usually, blacklists have been the primary tool employed for such types of detection. [5] Nevertheless, blacklists cannot be considered comprehensive and cannot detect freshly generated malicious URLs. In recent years, there has been an increasing demand for evaluating machine learning methods to enhance the efficacy of malicious URL detectors [6]. Humans are the most common threat vector and are known to be the root cause of 74% of data breaches, according to Verizon's "2023 Data Breach Investigations Report [7]. An organization called APWG [8] that studies and disseminates information about malware and phishing scams, observed 1,077,501 phishing attacks in the last quarter of 2023. APWG recorded almost five million phishing attacks in 2023, which was deemed as the worst year for phishing activity. The Internet Crime Complaint Center [9] alone received a staggering approx. 300k reported phishing attempts. This number decreased from the previous year but increased significantly since 2018 when they received only 26,379 reports. Alameda Lost Nearly $200M to Phishing Attacks [10]. According to statistics presented, attacks using malicious URL techniques are ranked first among the ten most common attack techniques [11, 12]. URL phishing involves sending emails to redirect recipients to a fictious website and trick them into revealing sensitive data, such as confidential credentials or financial information to a malicious person. The website may appear legitimate, but its purpose is to exploit your trust by "phishing" for personal information that malicious actors can use for nefarious purposes. For example, an email containing a warning message of user activity on your bank account, credit card, or financial application. An email originating from an e-commerce or financial institution like Amazon or a bill desk warns about suspicious activity, such as a password breach. Users are redirected to click a URL to verify transactions or change their passwords. However, the link redirects them to a fake version of the application or website, where their login credentials are collected, or they are prompted to call "customer service". Phishing costs organizations millions to deal with malware and credential compromise situations and it also leads to productivity losses, further having a negative impact on company brand value. On an individual level there is financial loss and mental stress causing health complications. Phishing is considered the costliest attack. URL refers to resources on the Internet. In [13], Sahoo et al. URL is divided into protocol identifiers and resource names, which contain the IP address or the domain name pointing to the resource location.

A malicious URL is a variation of the original URL, which deceives the victim to visit the URL, leading to financial loss and theft of personal identification information such as identity, credit cards, etc. In recent years ML has played an important role in detecting malicious URL and overcoming some of the shortcomings of traditional methods. Large numbers of features degrade model performance in terms of latency, and the selection of features were not optimal, which leads in degrading the overall model accuracy.

The proposed study focuses on building models based on a set of appropriate features selected based on correlation, which will improve the overall trained model performance in terms of latency. This study determines whether the selecting subset of features has positive or negative impact on identifying malicious URLs with different machine learning algorithms. This is how remainder of the article is organized. Literature review is done in Section II. Next, in Section III, materials and methodology used for proposed model is discussed. A report on the experimental results obtained is covered in Section IV. The paper is concluded in Section V.

II. Literature Review

Many approaches have been developed in this area such as blacklisting, signature based, content-based classification, URL based classification. When machine learning is employed, the previous studies had different results for each algorithm and focused majorly on algorithm performance with all the extracted features. A tabular representation of work done by different researchers is shown in Table I.

TABLE I. Literature Review

Ref. No.	Technology	Dataset	Outcome and Limitation
[14]	Heuristics	16006 Benign and 5678 Malicious samples utilized	The increase in performance is accompanied by a false positive rate, which in practical settings generates a lot of false positive warnings. Besides this False negative rate was 46.15 % which was used for detection
[15]	List Based	5000 phishing websites from Phish Tank	NISOELM a unique method for phishing detection is proposed. Minor modification in the URL bypass the list and list must frequently be updated. To make sure that most malicious pages are identified with the presented information, the acquired knowledge must be updated on a regular basis as attackers modify their tactics.
[16]	Association rule based	collected over 1400 URLs from several sources and also 1200 phishing URLs from phishtank database	Large number of rules impact performance. Dataset consisted only of Binary attributes and only the phishing URLs are mined using the apriori algorithm for identifying the recurring patterns. No detection for newly arrived patterns.
[17]	Heuristics	Not specified	Proposed method consists of two processes, namely logo extraction and identity verification. Consider only 2 attribute website logo and domain name. Able to identify whether website logo and domain name are genuine or fake.
[18]	3 ML classifiers SVM, LR and Naïve Bayes	UCI machine learning repository, 11,055 URLs, each of 15uniquefeatures,	Performance impact due to model consider all feature for prediction. Naiye Bayes shows 100% accuracy but each feature weightage is same.
[19]	Machine learning	Not specified	Proposes a machine-learning framework for supporting intelligent web phishing detection and analysis, and provides its experimental evaluation. In particular we make use of state-of-the art decision tree algorithms for detecting whether a Web site is able to perform phishing activities. Performance impact due to model consider all feature for prediction
[20]	Machine learning	public dataset comprising 2.4 million URLs (instances) and 3.2 million features	Random Forest and Multi-Layer Perceptron attain the highest accuracy. Performance impact due to model consider all feature for prediction

III. Material and Methods

A URL consists of the protocol, subdomain, domain, path [21, 22]. Within the path, there can be filename, query param. Domain name can be broken down into domain and top-level domain. Malicious persons can add @ in the domain name or can use prefix/suffix in domain name. Length or depth of the URL is another feature which is exploited to deceive users. Use of HTTPS in the domain name to deceive the user into clicking the URL believing that it is a secure site. The structure of a URL is shown in Fig. 1.

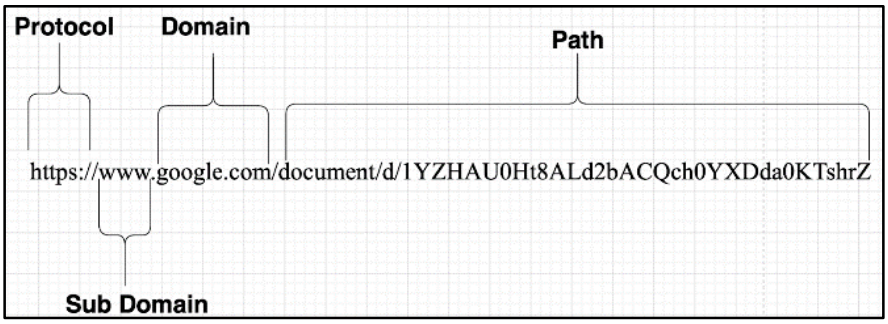


Fig. 1. Structure of a URL.

Model for malicious URL detection is created using a jupyter notebook and serialized on disk in pickle format. Flask server was deployed to host the model and GET/POST route was defined to handle the incoming GET/POST request to render the UI for user input and user input is posted to the server for predicting the URL safe or not. Fig. 2 presents the malicious URL detection methodology steps. These steps are explained in next section. Also Fig. 3 represents the User interface created for detection of a URL to be malicious or not.

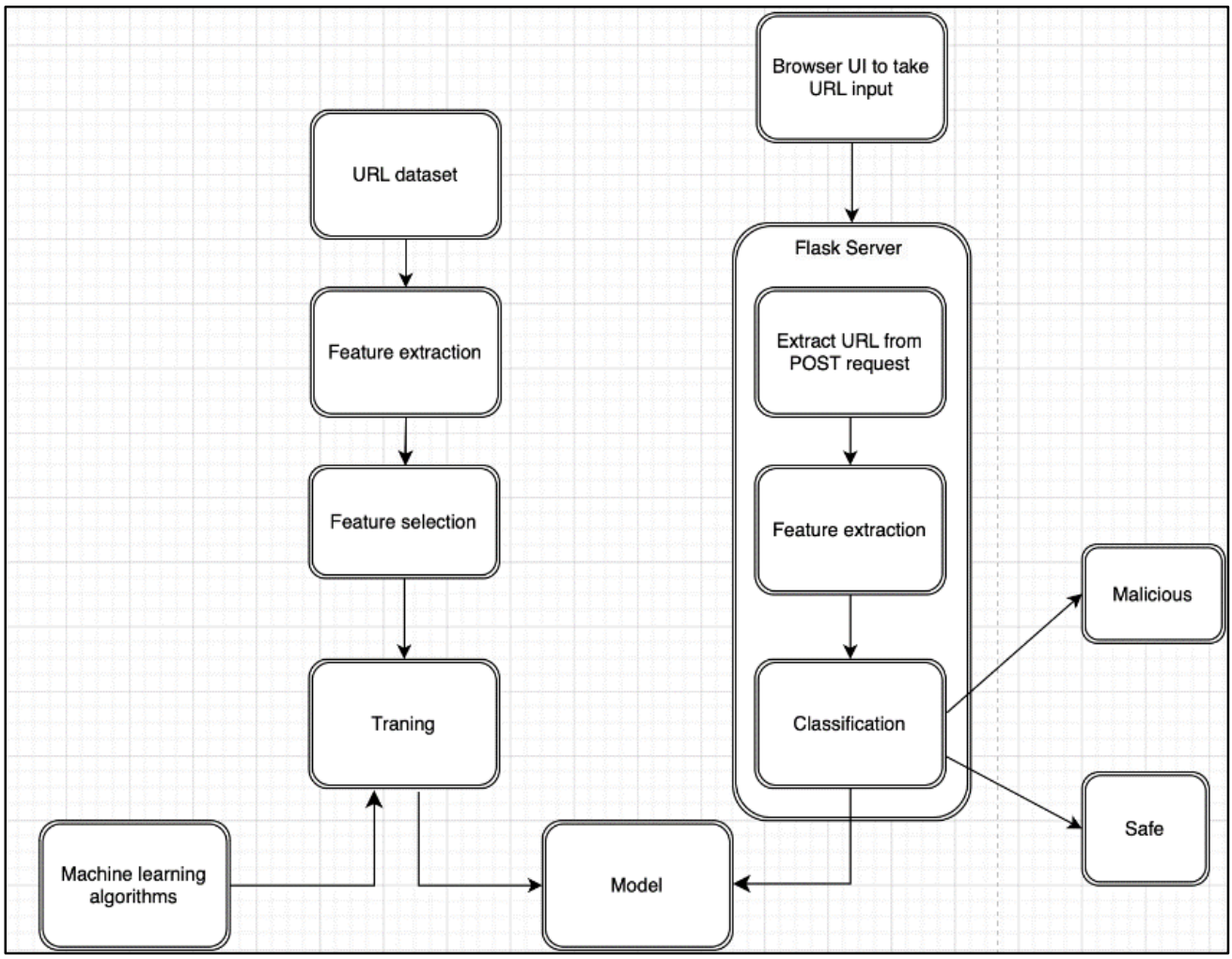


Fig. 2. Malicious URL detection flow.

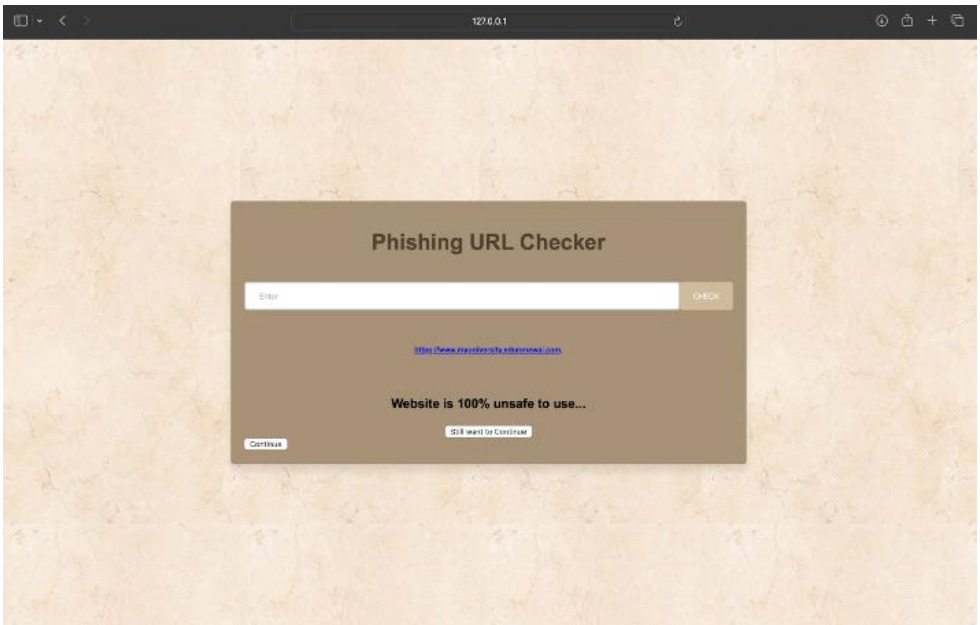


Fig. 3. User interface for URL checker.

A. URL Dataset

Suspicious URLs can be sent to Phishtank for verification https://www.phishtank.com/developer_info.php. The data in Phishtank is updated hourly. Phishtank is a free community site where anyone can submit, verify, track and share phishing data. This dataset is in the form of .csv file format. The models used in this manuscript dataset is fetched from phishtank. Besides this other source for similar dataset are available at https://urlhaus.abuse.ch/downloads/csv/ , URL has is a project from abuse.ch aiming at sharing malicious URLs being used for malicious software distribution, and dataset from Kaggle https://www.kaggle.com/datasets/sid321axn/malicious-urls-dataset, was used to host the data set for malicious and legitimate.

B. Feature Extraction

Feature extraction is one of the critical steps in the process of machine learning based malicious URL detection. Machine learning models require numeric value for training. For this purpose, essential characteristics of URLs are identified and passed to a function which converts the field value to 0 or 1 or other numeric values to distinguish malicious from benign. Tokenization and Lexical feature selection method is used for feature extraction Based on this criterion, the features are categorized in two different groups - Address bar based and domain-based features. In the case of Address bar-based features, features are selected from the lexical group of URLs, and are summarized in Table II. The address bar in web browsers is a powerful tool that goes beyond just entering website URLs. Below are the address bar features implemented in this project that goes beyond just entering website URLs. Tokenization is one of the techniques for feature extraction. It is defined as transforming a single string into a sequence of one or more non-empty substrings. Tokenization is performed utilizing the special characters (slash, dash and dot) in URLs. Once the token is extracted, it is passed to a function to check the characteristics such as DNS record validity or age of domain. This is characterized as Domain based features. Domain based features extracted from selected dataset is represented in Table III.

Selection of non-significant features can significantly impact a model performance besides increasing the model complexity. Selects a subset of relevant features while keeping the original feature space intact. The focus is on identifying the most informative features for modeling. The feature selection process is a step in building a machine learning model, performed by selecting a subset of the features in a set of extracted features. Feature selection aims to discover the most relevant and significant features for predicting the target variable. Feature selection has various benefits, such as Improved model interpretability, Reduced danger of overfitting, and improved model performance. Numerous methods for feature selection include filtering, wrapper approaches, and embedded approaches. Pearson correlation is employed and used to evaluate the model performance in the current work. Pearson correlation finds the correlation between features. Fig. 4 is the correlation matrix for the feature extracted to select the high-correlated and low-correlated features for training and evaluating the model.

TABLE II. DESCRIPTION OF ADDRESS BAR BASED FEATURES

S. No	Features	Description
1.	Domain	Extract the domain name
2.	Hostname/IpAddress	Parse the URL to extract an IP address . URLs may have IP address instead of domain name. Presence of an IP address alternative of the hostname name in the URL can be an indicator of malicious site
3.	@ symbol	In standard URL syntax, the "@" symbol is reserved for use in the format username@hostname. Anything before the "@" symbol is often interpreted as a username, and the browser ignores this part when resolving the URL Phishers exploit this behavior by inserting a legitimate domain name after the "@" symbol, making it appear as if the link leads to a trusted website. However, the actual website visited is determined by what follows the "@" symbol, not what precedes it.
4.	URL length	Phishers obscurate the URL by creating the long URL such that the user will not be able to differentiate a legit URL or malicious URL by masking the doubtful part of the address bar
5.	URL Depth	Computing the depth of a URL involves counting the number of levels or subdirectories in the URL path, typically separated by "/"
6.	Redirection //	"//" in a URL path reveals potential redirection or URL misconfiguration. Unexpected "//" positions could indicate unintended redirects or errors in URL formation.
7.	Http/Https	Phishers may add "HTTPS" to the domain (e.g., http://www.httpssecurelogin.com) to deceive users into believing a secure connection exists.
8.	URL Shortening Service	Services such as T2M, tine.be, Tiny URL, T.LT etc is a characteristic of malicious URL
9.	Prefix/Suffix '-'	Phishers use prefixes or suffixes to the domain name separated by some known separator such as "-" which makes it impossible for the user to distinguish that users feel that they are dealing with a legitimate website, for e: g www.example.com www.ex-ample.com

TABLE III. DESCRIPTION OF DOMAIN-BASED FEATURES

S. No	Domain features	Details
1.	DNS Record	WHOIS database does not recognize phishing websites identity or no records found for the hostname in DNS server
2.	Website Traffic	Top ranked websites are provided by Cisco Umbrella [26][27]. Alexa is no longer available. For the purpose of this research websites ranked among the top 100,000 is considered legitimate
3.	Age of Domain	Find the age of the domain by querying WHO database. Phishing websites are available for a short period. This research considers the minimum age of the legitimate domain, which is 12 months. Age here is nothing but different between creation and expiration time

The Pearson correlation coefficient, which is often denoted as r, is a measure of the linear correlation between two variables X and Y. It lies between -1 and +1. It is defined as:

$$r = \frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum_{l=1}^{n}(x_i-\bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i-\bar{y})^2}} \quad (1)$$

Where

- n is sample size
- x_i, y_i are the individual sample points indexed with i
- $\bar{x} = \frac{1}{N}\sum_{i=1}^{n} x_i$ is the simple Mean for X
- $\bar{y} = \frac{1}{N}\sum_{1=1}^{n} y_i$ is the sample mean for Y

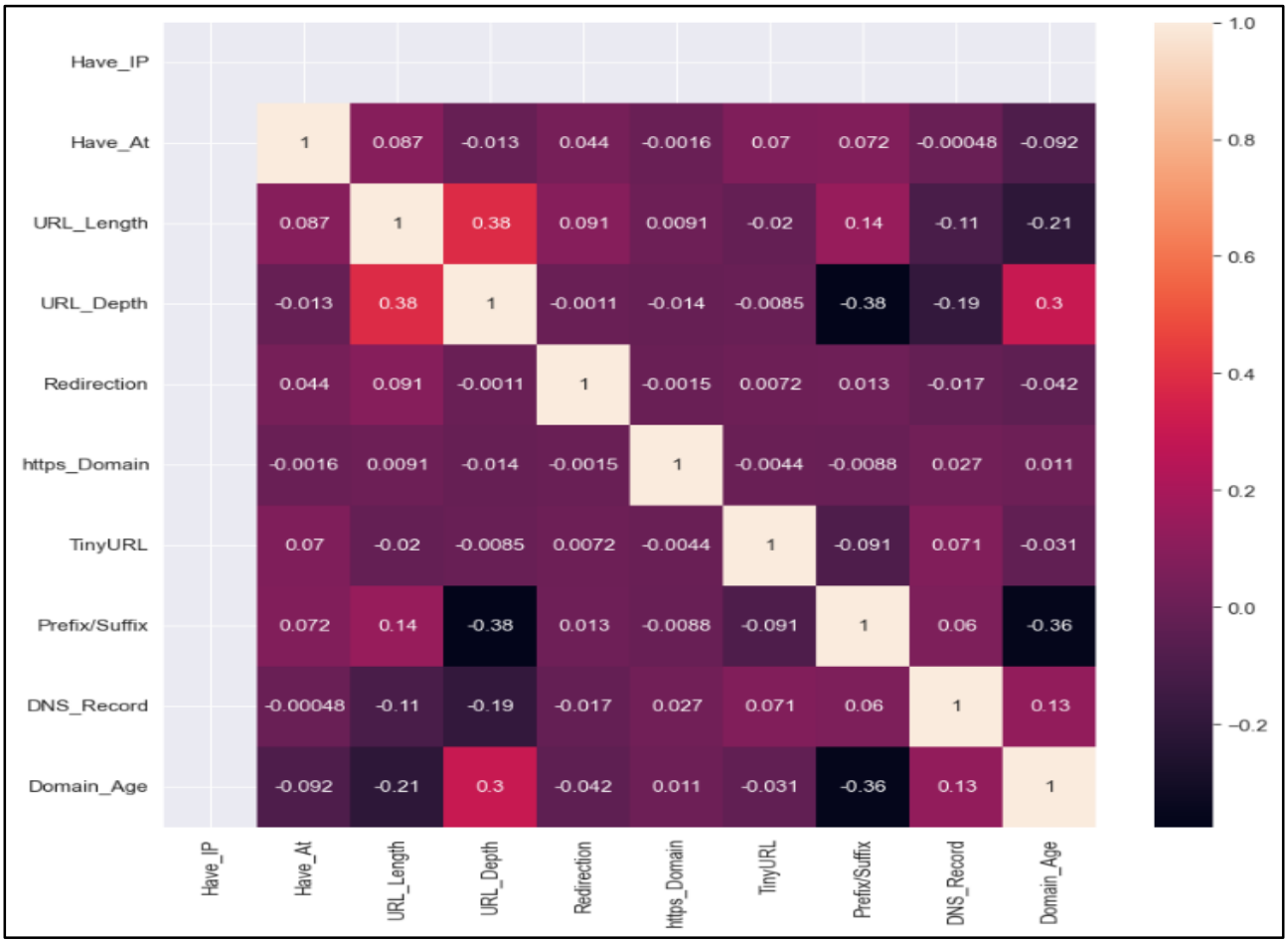


Fig. 4. Heat map of features.

For ease of modeling a threshold of .01 is chosen to filter significant features. Tabular representation of classification of features is shown in Table IV. Extracting the feature importance from the model which was created using all the features matches the correlated features. Table IV shows the segregation of features in URL based on the significance of the features. Graphical representation of importance of figures is shown in Fig. 5. The graph clearly depicts that URL features like *'URL_Depth'*, *'Domain_Age'* are most significant whereas features such as https*_Domain', Have_IP are* least significant in characterizing a URL as malicious.

TABLE IV. TABULAR REPRESENTATION OF CLASSIFICATION OF FEATURES

Feature set name	List of features
All features	{'Have_IP', 'Have_At', ' 'URL_Length', 'URL_Depth', 'Redirection', 'https_Domain', 'TinyURL', 'Prefix/Suffix', 'DNS_Record', 'Domain_Age', 'Domain'}
Top correlated Features	{'Domain_Age', 'DNS_Record' , 'Prefix/Suffix', 'URL_Depth', 'URL_Length }
Least correlated Features	{'https_Domain', 'Redirection', 'Have_IP', 'TinyURL', 'Have_At'}

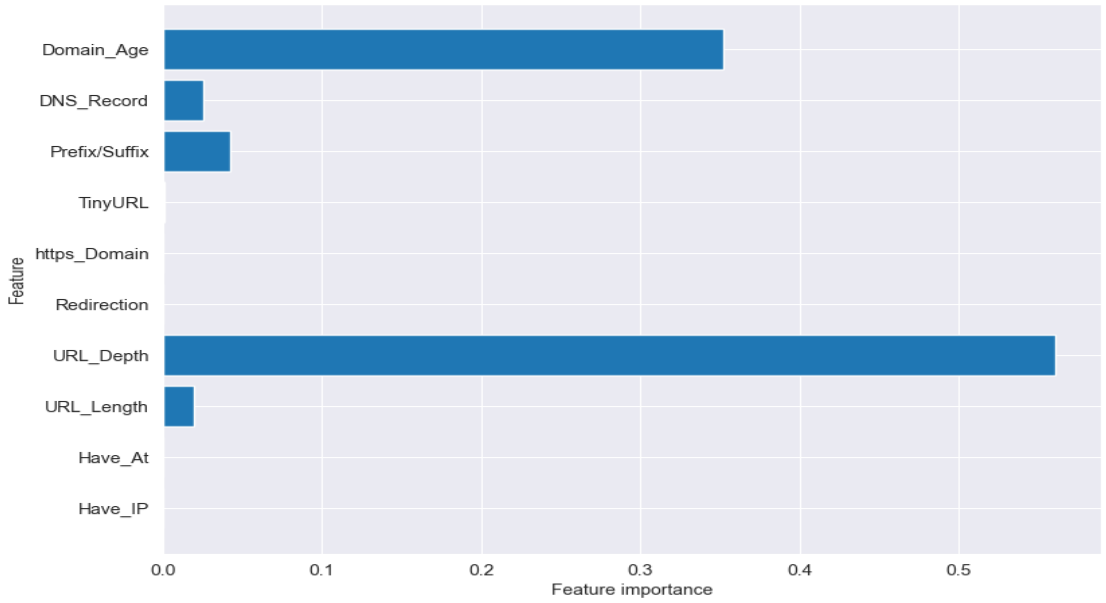


Fig. 5. Importance of feature.

C. Model Development and Performance Evaluation

Machine learning algorithms for detecting malicious URLs have been studied and are widely applied [23,24]. Supervised machine learning algorithms are classified and regression. This data set comes under classification problems, where the input URL is either phishing 1 or legitimate 0. The supervised machine learning models considered for training the dataset in this notebook are, Decision Tree and Random Forest. The model was trained with a decision tree and random forest algorithm with "all features", "Top correlated features", and "Least correlated features", as outlined in Table IV earlier. Decision trees are widely employed models for classification and regression-related tasks. Fundamentally, they learn a hierarchy of if/else questions to determine a decision. Learning a decision tree implies learning the pattern of if/else conditions that optimally lead to the true answer. In the machine learning setting, these questions are called tests (not to be confused with the test set, which is the data that is used to test to interpret the model generalizability. A decision tree consists of nodes representing decisions on features, branches representing the result of these decisions, and leaf nodes representing predictions. Internal nodes are an examination of a feature, and each branch corresponds to the outcome of the test, and each leaf node fits a class label. A random forest which is an ensemble model of decision tree, works by creating multiple decision trees. The idea behind random forests is to build a tree using random samples from the training dataset. The random forest combines the output of individual decision trees to generate the final output by averaging their results. They are powerful, often work well without heavy tuning of the parameters, and don't require data scaling. The entire data set of URLs containing legitimate and phishing URLs is then divided into 4 variables, X_train, X_test, Y_train, Y_test using the 'sklearn.model_selection' module/library. X_train: This variable holds the features (input variables) for the training set. In this paper, use these features to train your machine learning model. X_test: This variable holds the features for the testing set. In this paper, these features evaluate the performance of your trained model on unseen data. y_train: This variable holds the target variable (output variable) corresponding to the training set. It contains the expected outcomes for the training data. y_test: This variable holds the target variable corresponding to the testing set. It contains the expected outcomes for the testing data, which you use to compare against the predictions made by your trained model. Following model performance metrics are captured for performance assessment. True Positive (TP), False Positive (FP), True Negative (TN) and False Negative(FN) are some of the variables defined in confusion matrix. These are used for calculating the performance of a machine learning classification model as are used in Eq. (2)-Eq. (5) [28]. In context to calculating the ML model performance for detection of URL as malicious or genuine the performance measures are defined as

1) Precision: It is the ratio of true positive URL among the total number of positive URL predicted

$$\text{Precision} = \frac{TP}{TP+FP} \quad (2)$$

2) Recall: It is the ratio of predicted true URLs and the total number of actual true URL which is sum of true positive and false negative predicted URL.

$$\text{Recall} = \frac{TP}{TP+FN} \quad (3)$$

3) F1 Score: It is the harmonic mean of precision and recall.

$$\text{F1 Score} = 2.\frac{Precision.Recall}{Precision+Recall} \quad (4)$$

4) Accuracy: Success rate of the URL prediction technique and is coined as the ratio of True predicted to all the the samples in the dataset

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad (5)$$

IV. Model Setup and Results Obtained

Setup environment: Operating System - Mac OS, Language - Python 3.12.2 Web framework - Flask, Model builder = Jupyter notebook, ML framework/tools - Pandas, scikit-learn, Numpy Hardware: RAM 16 GB 3733 MHz LPDDR4X; 2 GHz Quad-Core Intel Core i5. The experiment extracted the features from the legitimate URL and phishing URL data set and labeled accordingly. The data set used for the experiment is of 10k records which include the 5k phishing and 5k legitimate URL. Fig. 6 shows the distribution of the individual features values in the dataset used for detection of malicious URL [25].

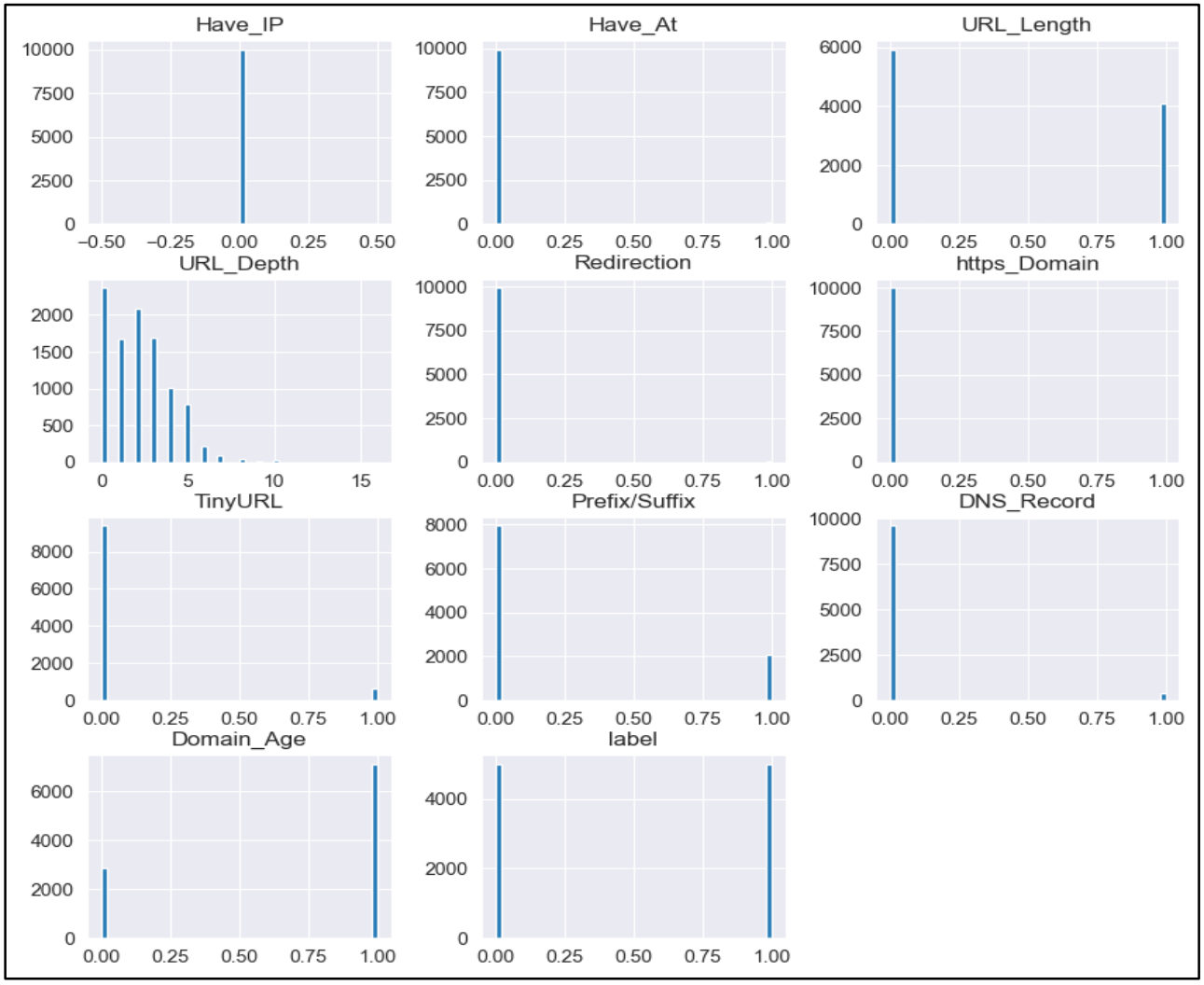


Fig. 6. Feature distribution.

In this paper, models were trained using machine learning models of decision tree and Random Forest after splitting dataset into training and testing with 80% of data used for training the model and tested on 20% of data with following set of features

- ALL
- TOP correlated
- LEAST correlated

Table V and Table VI shows the performance parameters obtained from Decision tree and its ensemble version Random Forest model. The parameters considered for assessment of models are accuracy, F1 score, Recall and Precision are best for the decision tree. Below is the metric for performance assessment of Decision Tree model based considering ALL, TOP and LEAST correlated features. Tabular representation of performance parameters discussed earlier for Decision tree ML model is represented in Table V.

TABLE V. Decision Tree Model

	ALL	TOP	LEAST
Accuracy	0.913	0.899	0.504
F1 Score	0.908	0.884	0.035
Recall	0.853	0.820	0.018
Precision	0.970	0.959	0.934

Because latency has a direct impact on how well systems operate in real time, it is also considered as a significant parameter for selecting one model over another [26]. Lower latency is preferable. The wait time for a result is known as latency. A ML model is not considered good if there is a noticeable waiting period before the occurrence of the responses. Improving latency is crucial since every system aspires to operate in real time [27]. An analysis of the time taken (latency) by the model to test 20% of the data where Decision Tree model is built using different set of features is as below.

- All features = 0.0033ms
- Top correlated = 0.0016ms
- Least correlated = 0.0026ms

Considering an ensemble model of Decision tree which is Random forest, performance parameters are rechecked.

Tabular representation of results obtained using ensemble model is represented in Table VI.

TABLE VI. RANDOM FOREST MODEL

	ALL	TOP	LEAST
Accuracy	0.953	0.947	0.504
F1 Score	0.946	0.924	0.035
Recall	0.876	0.873	0.018
Precision	0.978	0.967	0.857

Table VI shows the performance of Random Forest model for detection of URL as malicious or not. Latency by the Random forest model to test 20% of the data where Random Forest model is built using different set of features are obtained as follows:

- All features = 0.0027ms
- Top correlated = 0.0010ms
- Least correlated = 0.0020ms

TABLE VII. COMPARATIVE STUDY WITH EXISTING RESEARCH WORK DONE

Reference No.	Multiple ML models Used	Results:						
		Accuracy	F1 Score	Recall	Precision	Features in Modelling	Model Complexity	Latency Considered
[28]	yes	NA	High	High	High	Considered all features	High	No
[29]	yes	High	High	High	High	Not considered	NA	No
[30]	Yes	High	NA	NA	Good	Lexical features	NA	No
[31]	Yes	High	High	High	High	semantic and contextual features	High	No
Proposed work	Yes	High	High	High	High	Considered only significant Features without compromising on Performance	Reduced as only significant Features considered.	Yes

Decision tree and Random Forest model metric are similar and also perform similarly with the given URL dataset for selected feature sets. Random Forest uses a default estimator=100 of trees on a URL dataset. Ensemble model of Decision tree which is Random Forest performance is better in terms of performance indices as well as in terms of computation time. Besides this by reducing the number of features it can be clearly stated that the performance of model remains unaffected by reducing the number of features and selection only significant features for models designing. This will also reduce model complexity without compromising on model performance. However, the using the top correlated features shows significant model performance improvement in both Decision Tree and Random Forest. Table VII shows a comparative study on the model proposed and those used by researchers in similar domain.

V. CONCLUSION

The early systems were dependent upon patterns of known malicious URLs, rule-based methods. These systems are excellent in protecting the user from known malicious URLs but are inefficient in securing them from new emerging attacks. Although some attempts were made to build a model using ML but due to resource intensive, there is inefficiency in ML based malicious URL detection because the models have mostly considered either all the features which including non-correlated features or least significant features as well. While accessing performance the models performs well but fail to justify the response time or latency of the model. In this study, a comprehensive study on URLs like phishing or legitimate is used to analyze ML models based on the different feature selection and a study on the impact of feature selection is done. By using all the features or using only the most correlated features have slight impact on the performance of model parameters accuracy, F1 score, recall and precision but the difference in the model latency is quite significant with most correlated features and all features. This shows that using all the features impact the URL detection performance significantly with minimal gain in accuracy. Using highly correlated features helps in reducing the number of features which leads to reduction in model complexity and will further improve the model performance in terms of latency with minimal or negligible impact on the model performance.

REFERENCES

[1] Breda, Filipe & Barbosa, Hugo & Morais, Telmo. (2017). SOCIAL ENGINEERING AND CYBER SECURITY. 4204-4211. DOI:10.21125/inted.2017.1008.

[2] M. Khonji, Y. Iraqi, and A. Jones, “Phishing detection: a literature survey,” IEEE Communications Surveys & Tutorials, vol. 15, no. 4, pp. 2091–2121, 2013. [3] 10.1109/SURV.2013.032213.00009

[3] M. Cova, C. Kruegel, and G. Vigna, “Detection and analysis of drive-by-download attacks and malicious javascript code,” in Proceedings of the

19th international conference on World wide web. ACM, 2010, pp. 281–290. https://doi.org/10.1145/1772690.1772724.

[4] R. Heartfield and G. Loukas, "A taxonomy of attacks and a survey of defence mechanisms for semantic social engineering attacks," ACM Computing Surveys (CSUR), vol. 48, no. 3, p. 37, 2015. https://doi.org/10.1145/2835375

[5] Hameed, W & Ahmed, I & Khan, B & Kumar, Raja. (2017). USING BLACK-LIST AND WHITE-LIST TECHNIQUE TO DETECT MALICIOUS URLS. 10.26562/IJIRIS.2017.DCIS10081.

[6] Oshingbesan, Adebayo & Okobi, Chukwemeka & Ekoh, Courage & Richard, Kagame & Munezero, Aime. (2021). Detection of Malicious Websites Using Machine Learning Techniques. 10.13140/RG.2.2.30165.14565.

[7] C. David Hylender, Philippe Langlois, Alex Pinto, Suzanne Widup, "Data Breach Investigation report by Verizon Business" ,2024.

[8] APWG Phishing Activity Trends Report, Phishing Activity Trends Report, 4th Quarter 2023, Unifying the Global Response To Cybercrime apwg_trends_report_q4_2023

[9] Internet Crime Report 2023 by FEDERAL BUREAU OF INVESTIGATION. Internet crime complaint center. 2023_IC3Report.pdf

[10] H. M. Junaid Khan, Q. Niyaz, V. K. Devabhaktuni, S. Guo and U. Shaikh, "Identifying Generic Features for Malicious URL Detection System," *2019 IEEE 10th Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON)*, New York, NY, USA, 2019, pp. 0347-0352, doi: 10.1109/UEMCON47517.2019.8992930.

[11] Cho, Do & Dinh, Hoa & Victor, Tisenko. (2020). Malicious URL Detection based on Machine Learning. International Journal of Advanced Computer Science and Applications. 11. 10.14569/IJACSA.2020.0110119.

[12] M. Aljabri *et al*., "Detecting Malicious URLs Using Machine Learning Techniques: Review and Research Directions," in *IEEE Access*, vol. 10, pp. 121395-121417, 2022, doi: 10.1109/ACCESS.2022.3222307

[13] D. Sahoo, C. Liu, S.C.H. Hoi, "Malicious URL Detection using Machine Learning: A Survey". https://doi.org/10.48550/arXiv.1701.07179

[14] Seifert, Christian & Komisarczuk, Peter & Welch, Ian. (2009). Identification of Malicious Web Pages with Static Heuristics. 10.1109/ATNAC.2008.4783302.

[15] Yang, Liqun & Zhang, Jiawei & Wang, Xiaozhe & Li, Zhi & Li, Zhoujun & He, Yueying. (2020). An improved ELM-based and data preprocessing integrated approach for phishing detection considering comprehensive features. Expert Systems with Applications. 165. 113863. 10.1016/j.eswa.2020.113863.

[16] Jeeva, Carolin & Rajsingh, Elijah. (2016). Intelligent phishing url detection using association rule mining. Human-centric Computing and Information Sciences. 6. 10.1186/s13673-016-0064-3.

[17] Chiew, Kang Leng & Chang, Ee & Sze, San & Tiong, Wei. (2015). Available online Utilisation of website logo for phishing detection. Computers & Security. 54. 10.1016/j.cose.2015.07.006.

[18] Wejinya, Gold & Bhatia, Sajal. (2021). Machine Learning for Malicious URL Detection. 10.1007/978-981-15-8289-9_45

[19] Cuzzocrea, Alfredo & Martinelli, Fabio & Mercaldo, Francesco. (2019). A machine-learning framework for supporting intelligent web-phishing detection and analysis. IDEAS '19: Proceedings of the 23rd International Database Applications & Engineering Symposium. 1-3. 10.1145/3331076.3331087.

[20] Nana, S.R., Bassolé, D., Dimitri Ouattara, J.S., Sié, O. (2024). Characterization of Malicious URLs Using Machine Learning and Feature Engineering. Social Informatics and Telecommunications Engineering, vol 541. Springer, Cham. https://doi.org/10.1007/978-3-031-51849-2_

[21] Hawkins, John., 4th International Conference on NLP Trends & Technologies (NLPTT 2023) - Data Science & Cloud Computing Track (DSCC)At: Chennai, India

[22] Vrbančič, Grega & Fister jr, Iztok & Podgorelec, Vili. (2020). Datasets for phishing websites detection. Data in Brief. https://doi.org/10.1016/j.dib.2020.10643833.

[23] Cui, Baojiang & He, Shanshan & Yao, Xi & Shi, Peilin. (2018). Malicious URL detection with feature extraction based on machine learning. International Journal of High Performance Computing and Networking. 12. 166. 10.1504/IJHPCN.2018.094367.

[24] Shantanu, B. Janet and R. Joshua Arul Kumar, "Malicious URL Detection: A Comparative Study," 2021 International Conference on Artificial Intelligence and Smart Systems (ICAIS), Coimbatore, India, 2021, pp. 1147-1151, doi: 10.1109/ICAIS50930.2021.9396014.

[25] Cho Do Xuan, Hoa Dinh Nguyen and Tisenko Victor Nikolaevich, "Malicious URL Detection based on Machine Learning" International Journal of Advanced Computer Science and Applications, 11(1), 2020. http://dx.doi.org/10.14569/IJACSA.2020.0110119.

[26] Külzer, D.F., Debbichi, F., Stańczak, S. and Botsov, M., 2021, June. On latency prediction with deep learning and passive probing at high mobility. In *ICC 2021-IEEE International Conference on Communications* (pp. 1-7). IEEE. 10.1109/ICC42927.2021.9500495

[27] Bezerra, D., de Oliveira Filho, A.T., Rodrigues, I.R., Dantas, M., Barbosa, G., Souza, R., Kelner, J. and Sadok, D., 2022. A machine learning-based optimization for end-to-end latency in TSN networks. *Computer Communications*, *195*, pp.424-440. https://doi.org/10.1016/j.comcom.2022.09.011.

[28] Reyes-Dorta, N., Caballero-Gil, P. & Rosa-Remedios, C. Detection of malicious URLs using machine learning. *Wireless Netw* (2024). https://doi.org/10.1007/s11276-024-03700-w

[29] Vundavalli, V., Barsha, F., Masum, M., Shahriar, H. and Haddad, H., 2020, November. Malicious URL detection using supervised machine learning techniques. In *13th International Conference on Security of Information and Networks* pp1-6. https://doi.org/10.1145/3433174.3433592

[30] A. Saleem Raja, R. Vinodini, A. Kavitha,Lexical features based malicious URL detection using machine learning techniques, Materials Today: Proceedings,Volume 47, Part 1,2021, Pages 163-166,ISSN 2214-7853, https://doi.org/10.1016/j.matpr.2021.04.041.

[31] Lixiao Jin, Ruiyang Huang, Xuanming Zhang, Fangjie Wan," A Malicious URL Detection Method Based on Bert-CNN", Advances in Transdisciplinary Engineering. Electronic Engineering and Informatics , Vol 51, page no. 515-522, doi 10.3233/ATDE240115

Enhancing Orchard Cultivation Through Drone Technology and Deep Stream Algorithms in Precision Agriculture

P.Srinivasa Rao[1], Anantha Raman G R[2], Madira Siva Sankara Rao[3], K.Radha[4], Rabie Ahmed[5]*
Department of ECE, CVR College of Engineering, Hyderabad, Telangana, India[1]
Department of CSE, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Guntur, Andhra Pradesh, India[2]
Department of IT, Malla Reddy Engineering College, Secunderabad, Telangana, India[3]
Department of IT, St.Martin's Engineering College, Hyderabad, Telangana, India[4]
Department of Computer Science-Faculty of Computing and Information Technology,
Northern Border University, Rafha, Saudi Arabia[5]
Mathematics and Computer Science Department-Faculty of Science, Beni-Suef University, Beni-Suef, Egypt[5]

***Abstract*—The integration of cutting-edge technology in agriculture has revolutionized traditional farming practices, paving the way for Smart Agriculture. This research presents a novel approach to enhancing the cultivation of orchard crops by combining deep-stream algorithms with drone technology. Focusing on pomegranate farming, the study employs a drone system with four specialized cameras: thermal, optical RGB, multi-spectral, and LiDAR. These cameras facilitate comprehensive data collection and analysis throughout the crop growth cycle. The thermal camera monitors plant health, yield estimation, fertilizer management, and irrigation mapping. The optical RGB camera contributes to crop managementby analyzing vegetation indices, assessing fruit quality, and detecting weeds. The multi-spectral and hyperspectral cameras enable early detection of crop diseases and assessment of damaged crops. LiDAR aids in understanding crop growth by measuring plant height, tracking phenology, and analyzing water flow patterns. The data collected is processed in real-time using Deep Stream algorithms on an NVIDIA Jetson GPU, providing predictive insights into key farming characteristics.Our model demonstrated superior performance compared to four established models—MDC, MLP, SVM, and ANFIS—achieving the highest accuracy (95%), sensitivity (94%), specificity (96%), and precision (91%). This integrated method offers a robust solution for precision agriculture, empowering farmers to optimize crop management, enhance productivity, and promote sustainable agriculture practices.**

***Keywords*—Smart agriculture; crops; cultivation; deep stream algorithms; drone and technology**

I. INTRODUCTION

Modern agriculture is undergoing a significant shift as a result of technological developments that promise to increase production, sustainability, and efficiency. One such innovative strategy is the use of deep stream algorithms and drone technology to revolutionise pomegranate farming. With their high nutritional content and rising demand, pomegranates stand to gain a lot from these cutting-edge methods. The use of drones outfitted with a variety of specialised cameras and cutting-edge data processing techniques is presented in this study as a comprehensive framework for automating the cultivation of pomegranates [1]. The fouronboard cameras—thermal, optical RGB, multi-spectral, and LiDAR—provide an abundance of real-time data that gives producers priceless insights into numerous facets of crop health and growth dynamics. A key component of this system is the thermal camera, which makes exact plant health assessments, precise irrigation mapping, effective fertiliser control, and yield estimation possible. This camera assists in the early diagnosis of stressed or unhealthy plants by collecting temperature fluctuations, enabling prompt treatments and optimising resource allocation. The optical RGB camera completes this functionality by measuring vegetation indices, evaluating the quality of the fruit, and even spotting weeds. This helps users make better decisions [2]. Multi-spectraland hyper-spectral cameras are essential for a more detailed analysis of crop conditions. They can recognize physical and biological traits that can point to underlying problems in pomegranate harvests to spot disease symptoms [3]. This ability guarantees early disease identification, enablesindividualized treatment plans, and reduces possible yield losses.

To maintain crop health and yield, UAVs mounted with thermal cameras could be used to monitor temperature differences in orchard crops. This allows for the early detection of plant stress, disease, or water inadequacies. Optical RGB cameras monitor crops' visual health and growth stages by taking high-resolution pictures for the analysis of vegetation indicators, fruit quality evaluation, and weed detection. Multispectral and hyperspectral cameras offer extensive spectral information to identify disease signs, nutrient deficits, and other physiological characteristics. This information enables precise, focused treatments to improve crop health and decrease losses. LiDAR technology provides vital insights into growth dynamics and optimizes irrigation techniques for more effective water use and improved orchardcrop management. Navigating UAVs mounted with such

LiDAR could also measure plant height, track crop phenology,and examine water flow patterns.

*Corresponding Author.

A. Related Works

Authors in study [4] used UAVs in apple orchards using thermal and RGB imagery to detect frost damage, evaluate fruit sets, predict yields, and monitor bloom stages to improve thinning practices. Similarly, in study [5], the authors installed multi-spectral cameras over UAVs to navigate the citrus groves to identify diseases such as citrus greening, allowing for targeted therapies to minimize the spread of the disease. Drones are used in vineyards [6] to monitor vine health, evaluate grape quality, identify illnesses, and plan precise fertilization by using multi-spectral imagery to pinpoint nutrient deficiencies. Very recently, Sanchez et al. [7] used drones in olive orchards to improve irrigation schedules, map canopy structure, monitor water stress, and evaluate tree health using LiDAR data. In this work, we especially focus on pomegranate orchard management, building on the wide-ranging uses of UAVs in different orchard crops. With deep stream algorithms and drone technology, this extension seeks to optimize pomegranate agriculture and improve crop sustainability, productivity, and health.

The LiDAR camera provides crucial information on crop phenology, water flow patterns, and plant height. This new information improves our comprehension of pomegranate growth dynamics. It helps us make the best irrigation decisions, resulting in more effective water use and sustainable farming methods [8]. The investigation uses the potent NVIDIA Jetson GPU for data processing to take advantage of the enormous amount of data these cameras have acquired. The system analyses the acquired data in real-time while utilizing deep-stream algorithms, allowing precise forecasts in key pomegranate cultivation areas. This entails monitoring crop health, analysing how dry the soil and vegetation are, determining how much fertilizer is needed, finding and controlling weed infestations, and quickly spotting instances of crop damage and disease.

The use of mechatronics, sensors, and IoT in agriculture is now essential, with drones emerging as a viable tool for mapping field variability and optimizing input applications. Drones have applications across various stages of plant growth and sectors such as livestock, horticulture, and forestry, enhancing field monitoring and decision-making [9], [10]. The survey in [11] examines various UAV applications, types, sensors, and architectures, comparing them with traditional technologies and highlighting their benefits and challenges in precision agriculture. The article [12] reviews the use of UAVs for crop monitoring and pesticide spraying, which helps improve crop quality and mitigate health risks associated with manual pesticide application. Conventional weed management methods are inefficient for integration with smart agricultural machinery, whereas automatic weed identification significantly improves crop yields. The study in [13] evaluates deep learning techniques (AlexNet, GoogLeNet, InceptionV3, Xception) for weed identification in bell pepper fields, with InceptionV3 achieving the highest accuracy of 97.7%, demonstrating the potential for integration with image-based herbicide applicators for precise weed management. UAV-based sprayers precisely target hard-to-reach areas, as demonstrated in a cotton field study [14] using advanced imaging and optimization techniques, achieving effective droplet deposition with a GWO-ANN model showing high prediction accuracy. UAV imagery with an in-house web application, "DeepYield," [15] uses deep learning models like SSD, Faster RCNN, YOLOv4, YOLOv5, and YOLOv7 to measure citrus orchard yields. Here, YOLOv7 excelled with a mAP, Precision, Recall, and F1-Score of 86.48%, 88.54%, 83.66%, and 86.03%, respectively, and the solution was integrated into DeepYield for automated yield estimation.

Water flow mapping, crop phenology monitoring, and plant height measurement have all benefited from the use of LiDAR technology. Prominent research, like [16], has shown how important it is for comprehending development dynamics and making the most use of water. Deep Stream Algorithm with NVIDIA Jetson GPU: The combination of these two technologies has proved essential for data processing. The effectiveness of this arrangement in real-time analysis was demonstrated by research by [17], allowing predictions in crop health, soil dryness, fertilizer needs, weed identification, and disease detection [18]. The literature has recognized that there are challenges with calibration, data quality, and system scalability [19]. Further developments will involve improving algorithms, adding meteorological information, and customizing systems for certain crops and geographical areas. Table I-B summarizes recent studies on applying drones and various sensors in orchard crops, covering yield estimation and the learning model used in the works.

B. Motivation

Agriculture is undergoing a technological transformation with the integration of unmanned aerial vehicles (UAVs), commonly known as drones, and advanced algorithms [20]. This literature survey explores the state-of-the-art in the automation of pomegranate cultivation, focusing on the use of drones equipped with thermal, optical RGB, multi-spectral, and LiDAR cameras. The processing of collected data is facilitated by the NVIDIA Jetson GPU using deep-stream algorithms, enabling real-time predictions for various aspects of crop management. The capacity of drone technology to deliver high-resolution, real-time data for precision farming has made it more and more popular in the agricultural sector. Prior research, such as that done by [21], showed how useful drones are for determining crop health, maximizing resource utilization, and increasing production. Plant health inspections have made considerable use of thermal cameras. Thermal imaging is useful in identifying stress factors, refining irrigation plans, and calculating crop yields, according to research by Messina et al. [22]. Optical RGB Imaging for Vegetation Indices and Quality: Research, such as the work by Devi et al. [23], highlights the application of optical RGB cameras for weed detection, fruit quality evaluation, and vegetation index measurement. This all-inclusive method helps to create accurate crop plans. Hyper- and Multi-Spectral Imaging for Illness Detection: Researchers have looked at the use of hyper- and multi-spectral cameras for illness detection [24]. These cameras can analyze both biological and physical parameters and identify damaged crops based on spectral fingerprints.

TABLE I. DRONE AND SENSOR APPLICATIONS IN ORCHARD CROPS

Authors	Crop Type	Work Description	Type of Sensor Used	Methodology	Model Developed	Accuracy
He et al. [25]	Apple	Yield estimation, health monitoring	RGB, Thermal Cameras	Image analysis, temperature mapping	Regression Model	92%
Jemaa et al. [26]	Apple	Health prediction	RGB, Thermal Cameras	Health index calculation, stress mapping	SVM	89%
Chandel et al. [27]	Apple	Irrigation scheduling	Thermal, RGB Cameras	Soil moisture mapping, temperature analysis	Regression Model	90%
Sun et al. [28]	Citrus	Yield prediction, soil dryness detection	Multi-Spectral Camera	Spectral reflectance analysis	SVM, KNN	87%, 85%
Modica et al. [29]	Citrus	Irrigation optimization	Multi-Spectral Camera	Spectral reflectance analysis	SVM	87%
Lan et al. [30]	Citrus	Yield prediction	Multi-Spectral Camera	Spectral reflectance analysis	SVM	89%
Marques et al. [31]	Olive	Water stress monitoring	LiDAR, RGB Cameras	Canopy structure analysis, water stress indexing	ANN	88%
Ferro et al. [32]	Vineyard	Yield prediction, health monitoring, weed presence	RGB, Multi-Spectral	Vegetation index calculation, clustering, weed mapping	K-Means, ANN	91%, 90%
Jones et al. [33]	Vineyard	Yield prediction	RGB, Multi-Spectral	Vegetation index calculation, clustering	K-Means, ANN	94%
Miranda et al. [34]	Pomegranate	Yield monitoring, irrigation optimization	RGB, Thermal, LiDAR	Multi-modal data analysis	Deep Learning	95%
Zhang et al. [35]	Pomegranate	Disease detection, crop damage detection	RGB, Thermal, LiDAR	Multi-modal image analysis	Deep Learning	93%
Olorunfemi et al. [36]	Pomegranate	Yield monitoring	RGB, Thermal, LiDAR	Multi-modal image processing	Deep Learning	95%

The literature review highlights the increasing amount of research on automated crop production, especially with pomegranates, using deep-stream algorithms and drone technology. All of the research included in the survey demonstrates how this strategy may be used to maximize the use of available resources, increase crop productivity, and support sustainable agriculture. However, despite significant advancements, there remain notable gaps in the integration and application of these technologies, specifically for orchard crops such as pomegranates. This research addresses thesegaps by proposing a comprehensive approach combining drone technology with deep-stream algorithms to optimize pomegranate cultivation.

Previous studies have examined the application of UAVs with different sensors in agriculture. Still, there is a lack of research specifically addressing the customization of these technologies for orchard crops such as pomegranates. Previous studies have primarily focused on general crop management, neglecting the specific needs of orchard farming. This field requires more precise and specialized approaches that have yet to be thoroughly explored. In addition, there is still much to be explored regarding integrating real-time data processing with deep-stream algorithms. Specifically, there is a need to understand how this integration can improve decision-making in pomegranate farming. This study addresses the existing gaps in the field by presenting a fresh approach that utilizes advanced cameras (thermal, optical RGB, multi-spectral, and LiDAR) installed on drones. These cameras are combinedwith the high-speed processing capabilities of deep streamalgorithms on an NVIDIA Jetson GPU. With this integration, you can closely monitor and manage every stage of the pomegranate growth cycle. This provides valuable insights for enhancing yield, promoting plant health, and ensuring high-quality crops. Focusing on pomegranates, a crop boasting high nutritional value and growing demand, this research tackles a specific need in the agricultural sector.

Moreover, it contributes to advancing sustainable and precision agriculture. The study's findings highlight the immense potential for transforming orchard farming and offer a solid foundation that can be applied to other crops. This has the potential to expand the advantages of Smart Agriculture practices to a wider range of crops.

A game-changing strategy for modernizing pomegranate production is presented via the combination of drone technology with deep stream algorithms. In the dynamic environment of pomegranate farming, this work aims to provide farmers with a cutting-edge toolkit that enables them to make data-driven decisions, improve production, and support sustainable agricultural practices. The following are key contributions of this research article:

- Introduces a pioneering approach combining drone technology and deep stream algorithms for pomegranate production.
- Provides farmers with advanced tools for data-driven decision-making in pomegranate farming.
- Enhances pomegranate yield and quality through precise monitoring and analysis.
- Promotes sustainable agricultural practices in pomegranate cultivation.

The rest of the article is organized as follows: Section II provides the methodology of how UAVs operate, particularly for agricultural applications, and how their built-in sensors are utilized for crop management in orchards. It also focuses on how the Deep Streaming technique is deployed for pomegranate cultivation. Section III shows how the processing power of the NVIDIA Jetson GPU is used for the automated cultivation of pomegranates. Finally, Section IV summarizes the key findings of the work with the conclusion of the proposed work.

II. METHODOLOGY

This section focuses on the methodology used for the investigation in terms of data collection, camera analysis and the implications and association of deep streaming framework applied over the UAV data of pomegranate cultivation. In Fig. 1, the present investigation illustratesa revolutionary approach to enhance pomegranate farmingthat combines deep-stream algorithms and drone technology. The drone system has four specialized cameras: a LiDARcamera, a thermal camera, an optical RGB camera, and a multi-spectral camera. These cameras are effective tools for comprehensive data gathering and analysis throughout the pomegranate growing cycle. For yield estimation, fertilizer management, irrigation mapping, and plant health assessment, the thermal camera is crucial. By detecting variations in plant temperature, the thermal camera helps identify stressed or ill plants and allows for quick response. The optical RGB camera's capability to monitor vegetation indices, assess fruit quality, and detect weeds further enhances crop management techniques [37]. The multi-spectral and hyperspectral cameras allow for the identification of harmed crops and the examination of their biological and physical characteristics. The multi-spectral analysis enables early diagnosis of agricultural diseases, enabling customized treatments. The LiDAR camera aids researchers in their understanding of how plants grow by measuring plant height, monitoring crop phenology, and looking at water flow patterns. The NVIDIA Jetson GPU and deep stream algorithmsare employed to process the camera data. This processingpipeline allows for real-time analysis of the gathered data, giving predictive insights into several essential aspects of pomegranate cultivation. The use of technology facilitates crop health monitoring, evaluates soil and plant moisture, establishes the demand for fertiliser, finds weeds, and scans fordisease and crop damage indicators [38]. Overall, this work provides an integrated approach to pomegranate cultivation that combines deep stream algorithms and drone technology to enable accuracy and data-driven decision-making.

A. Brief Mechanism of Drones and its Associated Sensors

UAVs are becoming indispensable instruments in contemporary agriculture, especially for precision farming. Multiple sensors can be carried by them, enabling thorough monitoring and analysis of crop productivity, growth, and health. Here, we go over how drones work and how their built-in sensors are utilized for crop management in orchards.

UAVs used in agriculture could be integrated with multiple essential parts to enable them to carry out certain jobs efficiently [39]. UAVs can hover, navigate, and gather data over wide distances because of the flying system's stability and maneuverability, which is provided by a lightweight frame, motors, propellers, and battery. GPS, accelerometers, gyroscopes, and magnetometers are examples of navigation and control components that provide precise navigation and flight path maintenance, enabling pre-planned missions and real-time modifications. The communication system enables remote operation through ground control stations and real-time data transfer via radio frequencies or cellular networks [40].

UAVs' sensors greatly increase their efficacy in precision agriculture because each one gives vital information for thorough crop management. For example, infrared radiation released by plants fluctuates with temperature and may be detected by thermal cameras [41]. This radiation can be used to identify stress factors such as pest infestation, disease, or water shortage. Thermal cameras are used in agricultural applications to detect temperature differences within the crop canopy. This allows for the monitoring of general health, early identification of plant stress, and watering requirements. With the aid of these cameras, temperature fluctuations inside the crop canopycan be identified, facilitating the early identification of plant stress, the need for irrigation, and general health monitoring. To create high-resolution images of the crop canopy, optical RGB cameras collect visible light in the red, green, and black wavelengths [42]. These images are then used to monitor fruit quality, identify weeds, and assess vegetation indices, which helps farmers make decisions about crop health and management techniques.

Beyond the visible spectrum, multispectral and hyperspectral cameras record information in a variety of wavelengths, such as ultraviolet and near-infrared. To provide comprehensive spectral information necessary for identifying certain crop situations including nutrient deficits, disease signs, and physiological stress, hyperspectral cameras gather data in hundreds of small spectral bands. Precision medicine and targeted interventions are made possible [43].

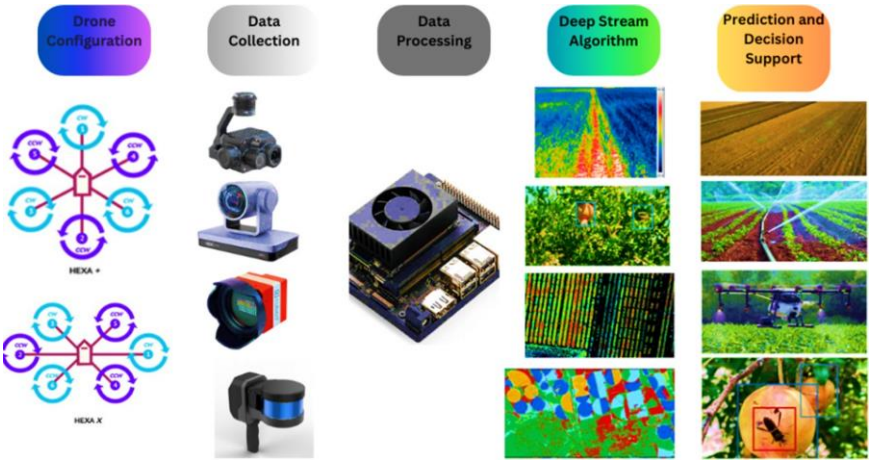


Fig. 1. Core functional modules in the proposed methodology.

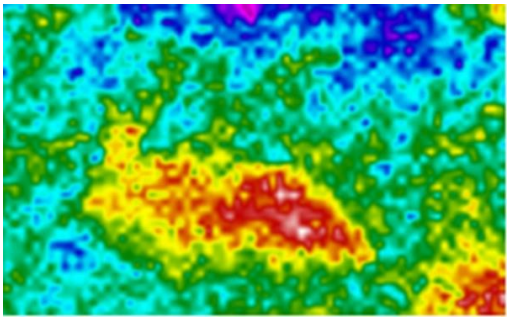


Fig. 2. Thermal imaging for plant health assessment.

LiDAR cameras measure plant height, track crop phenology, examine water flow patterns, and produce precise 3D mapsof the landscape and vegetation structure using laser pulses. Understanding the dynamics of plant growth, improving irrigation techniques, and improving crop management generally all depend on this data.

Yield prediction integrates data from thermal, RGB, and multi-spectral sensors to estimate possible yields [44]. Water use is optimized by irrigation management through the use of thermal and LiDAR data. Through multispectral and hyperspectral analysis, health monitoring identifies nutritional inadequacies and early indicators of disease. Using accurate data, resource optimization effectively handles inputs such as fertilizers. With the help of these cutting-edge technologies, orchard crop management, and productivity may be fully monitored and managed, improving agricultural sustainability and production.

B. Thermal Camera Analysis

To evaluate the health of pomegranate plants, identify stress, and track temperature changes, thermal images of the plants should be taken. Maps of temperature distribution made from thermal data can be used to find possible problem locations. Use the heat data to calculate yields, control fertilizer applications, and map irrigation. Technological developments have made it possible for creative methods of crop management and optimization in modern agriculture [45]. Utilizing thermal imaging to evaluate plant health, identify stress, and track temperature swings in pomegranate plants is one such ground-breaking method. Farmers and agronomists can enhance irrigation techniques, control fertilizer use, and predict crop production by utilizing the potential of thermal data.

1) Thermal imaging for plant health assessment: Radiometric temperature readings from pomegranate plants are obtained using thermal cameras. Stressed or ill plants show temperature anomalies, whereas healthy plants have rather consistent thermal fingerprints. Areas of possible concerncan be located by analyzing these thermal images, enabling focused intervention and mitigation as shown in Fig. 2.

2) Stress detection and temperature variations: Thermal imaging is a non-invasive method for identifying signs of stress in pomegranate trees. Temperature changes inside the plant canopy can emphasize stress brought on by things like a lack of water, an unbalanced diet, or pest infestations as shown in Fig. 3. Knowing these stress patterns allows for early detection and prompt intervention.

3) Temperature distribution maps for precise insights: The generation of maps showing the spread of temperaturein pomegranate orchards is made easier by processing thethermal data that was gathered. These maps give farmers avisual representation of temperature differences throughout theentire field, allowing them to locate “hot” or “cold” areas thatmight be signs of unequal irrigation, drainage problems, orother specific problems as shown in Fig. 4.

4) Accurate irrigation mapping: Thermal data reveals regions with high temperatures, indicating potential water stress, which aids in precise irrigation mapping. Farmers can adjust their watering schedules to maintain consistent moisture distribution and reduce water-related stressors by associating these temperature differences with particular irrigation zones as shown in Fig. 5.

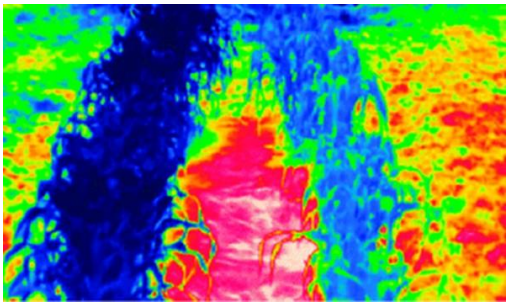


Fig. 3. A Sample stress detection in an agricultural land observed through thermal camera.

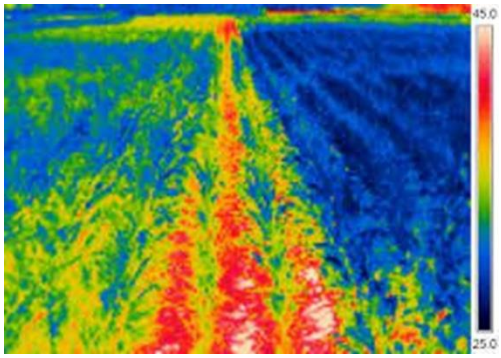


Fig. 4. Temperature distribution maps for precise insights.

C. Optimal Fertilizer Management

The use of thermal imaging helps handle fertiliser more effectively. Temperature variations can reveal changes in the absorption and utilization of nutrients. Farmers may strategically apply fertilizers where they are most required, saving waste and fostering healthy development, by merging heat data with soil nutrient analysis.

1) Yield estimation and harvest planning: More precise yield estimation is made possible by the thermal data insights. Variations in fruit development and maturation may be correlated with anomalies in temperature distribution. Farmers can predict production swings and adjust their harvest date by taking into account this information. Precision agriculture has essentially advanced thanks to the use of thermal imaging

technology in pomegranate farms. Farmers are better able to proactively solve problems, maximize resource use, and improve the general health of their crops thanks to the capacity to record, process, and analyze thermal data. The agricultural sector may get closer to sustainable practices by utilizing thermal insights for irrigation, fertilization, and yield management. These techniques maximize productivity while reducing their negative effects on the environment. The incorporation of thermal imaging into agricultural practices is poised to revolutionize how we grow and maintain our crops as technology advances.

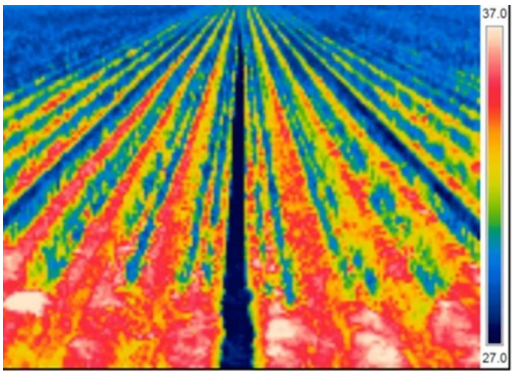


Fig. 5. Accurate irrigation mapping through drone-mounted thermal cameras.

2) Optical RGB camera analysis: Utilizing RGB (Red-Green-black) photography in modern agriculture has become a potent and adaptable tool for a variety of tasks, from determining weed presence to evaluating fruit qualityand vegetation health [46]. Researchers and farmers mayimprove crop management tactics, quantify key indices, and make educated decisions to maximize production andsustainability by utilizing modern image processing tools.

3) Quantify vegetation indices for health assessment: Important vegetation indices, like the widely used NDVI (Normalised Difference Vegetation Index), can be calculated using RGB photos. By comparing the reflectance of visible red and near-infrared light, NDVI acts as a quantitative indicator of plant health. This knowledge makes it easier to spot possible stressors and allows for tailored crop-growth-promotingactions as shown in Fig. 6.

4) Assessing fruit quality with image analysis: Color, size, and shape are some examples of fruit quality factorsthat can be evaluated using RGB imaging. Farmers can assess fruit maturity and harvest readiness by examining the color spectrum. In addition to quantifying variations in fruit size and form, image processing algorithms may also grade and categorize products based on their quality as shown in Fig. 7.

5) Weed detection and classification: It is possible to use the RGB imagery to look for weeds in crop fields. For advanced algorithms to distinguish between crops and undesirable vegetation, color, shape, and texture features are examined. Farmers can develop tailored weed control methods and increase yields by minimizing resource competition by automating weed detection as shown in Fig. 8.

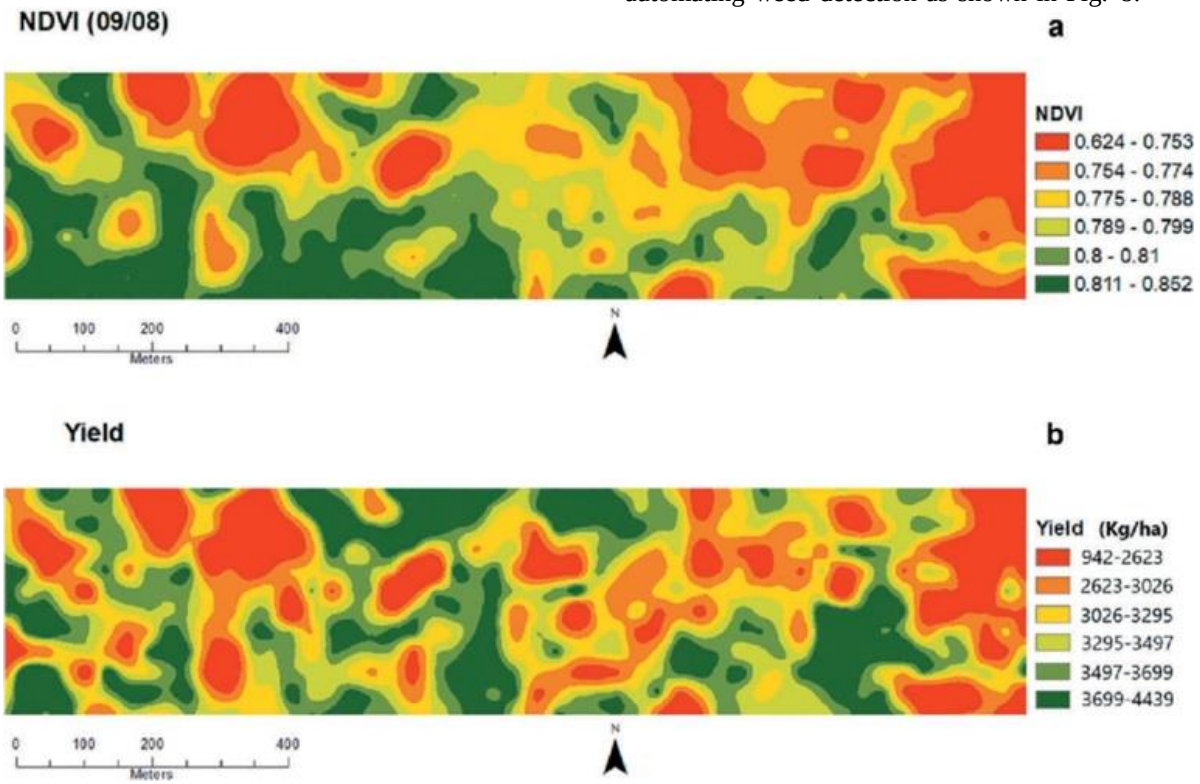


Fig. 6. Vegetation indices for health assessment.

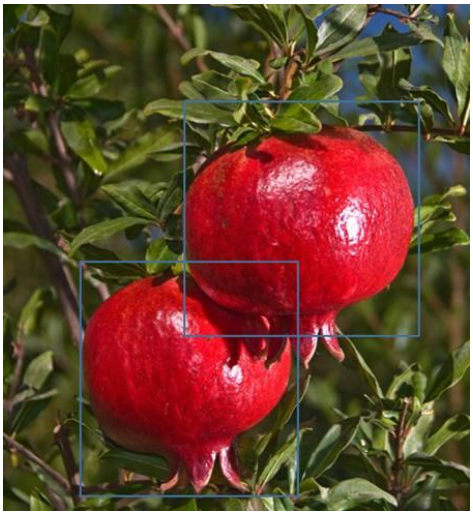


Fig. 7. Image analysis of pomegranate for fruit quality assessment.

6) Color analysis for pest and disease identification: When it comes to identifying pests and illnesses that impactcrops, RGB images can be useful. Leaf color and patternchanges may be a sign of an infection or an infestation.



Fig. 8. Weed detection for optimal irrigation.



Fig. 9. Color analysis for pest and disease identification.

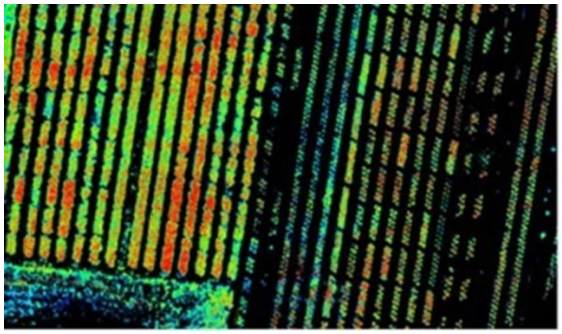


Fig. 10. Multi-spectral and hyper-spectral camera analysis.

Potential problems can be identified early by the analysis of RGB images, allowing for prompt intervention andloss mitigation. High-resolution maps that highlight spatial variations within fields can be made using remote sensing technology in conjunction with RGB images. These maps can be used to direct precision farming techniques, enabling the targeted use of resources like water, fertilizer, and pesticides. RGB photos can be used to train machine learning algorithms to recognize patterns and features as shown in Fig. 9.

It is possible to fine-tune these algorithms to recognize particular plant species, weed varieties, or disease symptoms. The effectiveness and precision of decision-making in crop management are improved by these skills. Agriculture transforms from reactive to proactive practices with the integration of RGB photography and image processing technology [47]. Farmers can make data-driven decisions that optimize resource use, decrease waste, and advance sustainable agricultural practices thanks to the capacity to measure indices, assess quality, detect weeds, and identify problems in real time. Analyse biological and physical traits while collecting data in the multi- and hyper-spectral range to spot disease symptoms. Use spectral analysis to find irregularities in plant reflectance patterns that could be signs of stress or disease [48]. Create machine learning models for spectral signature-based illness classification as shown in Fig. 10.

D. LiDAR Camera Analysis

Obtain LiDAR data to assess water flow patterns, track agricultural phenology, and evaluate plant height.

Create accurate digital elevation models (DEMs) and three-dimensional representations of the pomegranate orchardsusing LiDAR data processing. To measure agricultural growthstages, gather data on plant height and examine height changesover time as shown in Fig. 11.

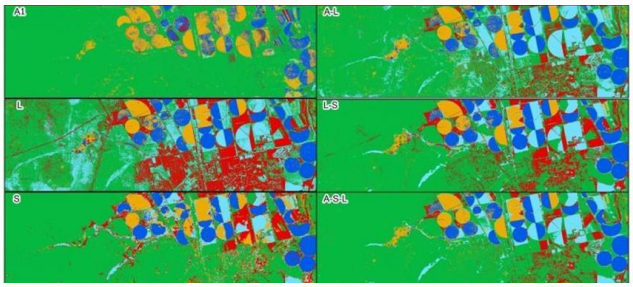


Fig. 11. Drone-mounted LiDAR camera analysis of agricultural lands.

E. Data Processing and Deep Stream Algorithm

Send the cameras' acquired data to the NVIDIA JetsonGPU so it can be processed. Use deep stream algorithms to analyze all camera data streams in real-time [49]. Use image recognition, machine learning, and pattern recognition techniques to forecast crop health, soil dryness, fertilizer needs, the presence of weeds, and instances of crop damage and disease. The object detection method is known as YOLOv5, or "You Only Look Once version 5," is recognized for its quickness and precision. It is made to recognize and locate several items simultaneously in a video or picture stream. The "Deep Stream" variation is especially well suited for applications like monitoring agricultural fields becauseit concentrates exclusively on processing continuous data streams effectively. The earlier YOLOv3, YOLOv4, and othernetworks served as the foundation for the development of the YOLOv5 network. YOLOv5 offers the advantages of being quicker and more precise than prior-generation networks. An adaptable anchor box and adaptive picture scaling are two examples. These methods efficiently decrease the amount of network computation by calculating the scaling factor using the ratio of the current picture size, W to H, and then obtainingthe filled scaling size. The backbone network and neck layer ofYOLOv5 are mapped to the cross-stage partial (CSP) concept of YOLOv4, which improves the capacity of network feature fusion in terms of feature extraction.

The four network models in YOLOv5 are categorized as s, m, l, and x, according to smallest to biggest. The network's breadth and depth are the primary areas ofvariation in size. The lightest among them is YOLOv5. The primary parts of the network are the input, neck, head, and backbone. The Mosaic data improvement module is used inthe input to enrich datasets. To speed up network training, the backbone leverages the CSPDarknet53 backbone network to extract rich information from input photos, such as thefocus module and the spatial pyramid pooling (SPP) module neck core fuses feature information at various sizes using feature pyramid network (FPN) and path aggregation network (PAN) architectures. Concat later connects the top-down and bottom-up feature maps, enabling the feature fusion of various deep and shallow scales. This enhances the network's expressive capacity. The YOLOv5 detecting structure is the head. Conv produces feature maps in three sizes: big,medium, and tiny. These sizes correlate to the targets thatare detected—small, medium, and large. YOLOv5 increases the precision of network prediction based on NMS by using three loss functions to compute the location, confidence, and classification losses. The foundation of this investigation is theYOLOv5s network. Fig. 12 illustrates the network structureof YOLOv5.

1) Object detection and monitoring: It is possible to train the YOLOv5 Deep Stream Algorithm to recognise and differentiate a variety of components important to pomegranate agriculture, including pomegranate plants, fruits, and potential pests [50]. By implementing this method in the field, it is possible to monitor the crop in real time and identify problems like pest infestations, disease outbreaks, or nutrient deficits early on.

2) Precise yield estimation: The system helps with yield estimation by precisely classifying and counting pomegranate fruits. Farmers can maximize overall productivity and resource management by using this data to make informed decisions about harvesting schedules, labor allocation, and post-harvest logistics [51].

3) Weed detection and management: Pomegranate yield canbe severely impacted by weed competition. The ability to recognize objects with the YOLOv5 Deep Stream Algorithm also allows for the classification and identification of weeds in pomegranate orchards. Utilizing these details makes it easier to deploy targeted weed control strategies, reduce resource waste, and increase crop yield.

4) Resource allocation and sustainability: Real-time insights provided by the algorithm provide a foundation for effective resource management. Farmers can use precision irrigation strategies by recognizing places that need attention or stress, including dry areas. This encourages the useof sustainable agricultural techniques while simultaneously conserving water [52].

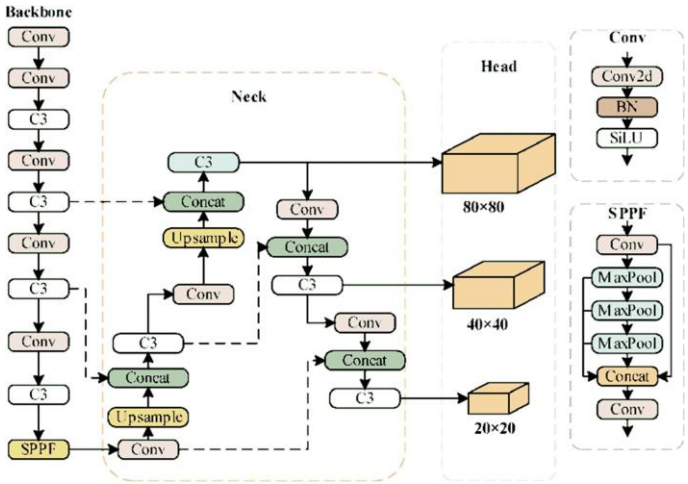
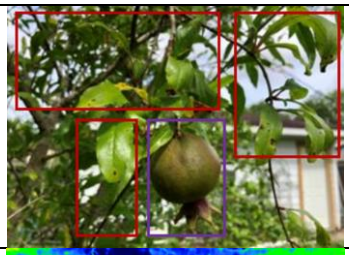
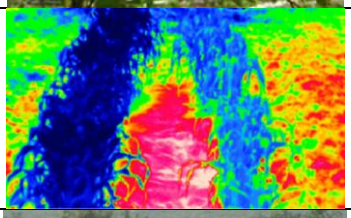


Fig. 12. Block diagram of YOLOv5 used in the experimentation.

TABLE II. DEEP STREAM ALGORITHM OUTPUT FOR VARIOUS APPLICATIONS

Application	Sample Output
Predict crop health	
Soil dryness	
Fertilizer requirements	
Weed presence	
Crop damage and disease	

5) Disease and pest management: Effective treatment of illnesses and pests depends on early detection. The YOLOv5 Deep Stream Algorithm can quickly recognize visual signs linked to a reduction in plant health, enabling prompt action. By controlling the spread of illnesses, farmers can cut backon the requirement for heavy pesticide use.

6) Integration with automation and drones: Drones with cameras can be integrated with the YOLOv5 Deep Stream Algorithm. With the help of this integration, drones may fly over the orchard by themselves while taking pictures in real-time and sending them to the algorithm for quick analysis.This method offers an unmatched vantage point for effectively monitoring vast agricultural fields as shown in Table II.

7) Prediction and decision support: Create forecasts and insights for various pomegranate agriculture characteristics based on the processed data. Create a dashboard or user-friendly interface so that farmers may get real-time data and advice. Give specific advice on how to manage pests and diseases, apply fertilizer, and schedule irrigation, among other cultivation techniques [53].

8) Validation and refinement: By gathering real-world data and making field observations, confirm the veracity of predictions and advice. Based on ongoing learning from field data and farmer comments, improve the deep streamalgorithms [54]. Improve the process iteratively depending on practical implementation issues and real-world performance.

9) Scaling and adoption: Increase the automated system's coverage area to larger pomegranate orchards and perhaps modify the approach for use with other crops. Educate farmers on how to use the automated system and how to understand the forecasts for wise decision-making. By supplying precise, timely, and data-driven insights that can improve crop yield,

optimize resource use, and promote sustainable agricultural practices, the integration of drone technology and deep-stream algorithms into pomegranate cultivation has the potential to transform conventional farming practices.

Our research employs a combination of advanced UAV-based cameras to enhance agricultural monitoring and outcomes, effectively addressing the specific challenges ofeach camera type. Thermal cameras, which detect infraredradiation to measure temperature variations and identify plant stress, face issues such as temperature sensitivity, lower resolution, and frequent calibration needs. Optical RGB cameras capture high-resolution images to analyze vegetation indices, fruit quality, and weed detection but are impactedby varying lighting conditions, large data volumes, and subtle color differentiation challenges. Multi-spectral cameras provide detailed insights into crop health and disease but are costly, complex, and sensitive to environmental factorslike cloud cover. LiDAR cameras generate high-resolution3D maps for measuring plant height and analyzing water flow patterns but require significant data processing power,are expensive, and struggle with dense vegetation obstructing laser pulses. Our approach integrates deep learning algorithms and NVIDIA Jetson GPU for data processing, addressing these challenges and enabling real-time analysis to improve data accuracy and reliability. By leveraging the strengths and mitigating the limitations of each camera, we facilitate precise crop management decisions, enhancing yield and sustainability in pomegranate orchards.

III. Results and Discussions

The automated cultivation of pomegranates using deep-stream algorithms and drone technology has produced encouraging results, suggesting a revolutionary method for modern agriculture. Combining the processing power of the NVIDIA Jetson GPU with the capabilities of a drone with four specialized cameras—thermal, optical RGB, multi-spectral, and LiDAR—has allowed for comprehensive data collection, real-time analysis, and predictive insights in various pomegranate cultivation-related areas.

TABLE III. Data Collection with Accuracy

Camera	Data Collection	Accuracy (%)
Thermal camera	Plant health inspection, Irrigation mapping, fertilizer management, yield estimation	95
Optical RGB camera	Vegetation index	91
Multi-spectral and hyper-spectral cameras	Biological and physical characteristics, diseasedcrop	93
LiDAR camera	Plant height, water flow, crop phenology	95

TABLE IV. Plant Health Inspection and Stress Detection

Crop Focus	ANN	CNN	ANFIS	YOLO
Plant Health Inspection	75	82	88	95
Stress Detection	76	81	85	93

A. Data Collection and Analysis

The pomegranate growth cycle has been thoroughly investigated using drones equipped with various cameras. To properly detect stressed areas and enable focused actions, the thermal camera was essential for plant health inspection. To improve overall crop management techniques, the optical RGB camera effectively measured vegetation indices, assessedfruit quality and found the presence of weeds [55]. The multi-spectral and hyper-spectral cameras were excellent at spotting damaged crops and examining biological and physical traits, which helped to identify and treat diseases early on. Furthering our understanding of crop growth dynamics, the LiDAR camera produced accurate measurements of plant height, tracked crop phenology, and mapped water flow patterns as shown in Table III.

B. Deep Stream Algorithm Processing

The automated pomegranate production system showcased notable progress in data-driven precision farming by using deep-stream algorithms and drone technology. Together with the NVIDIA Jetson GPU's processing power, the four specialized cameras—thermal, optical RGB, multi-spectral, and LiDAR—produced extensive data collecting and real-time analysis. The findings are displayed about important crop management topics [56]. Plant Health Inspection and Stress Detection: To inspect the health of plants, the thermal camera was essential in precisely locating stressed regions. The ability to precisely identify stressed or ill plants was made possible by real-time data processing, which made it easier to detect temperature differences [57]. Plant health was improved by the proactive actions made possible by this capacity as shown in Table IV.

1) Vegetation indices and fruit quality assessment: Fruit quality was evaluated and vegetation indices were successfully measured using the optical RGB camera. The technology provided insights into the health of the vegetation by quantifying metrics like NDVI using image processing techniques [58]. Evaluations of the quality of the fruit and the identification of weeds enhanced cultivation techniques, increasing both production and quality as shown in Table V.

TABLE V. Vegetation Indices and Fruit Quality Assessment

Crop Focus	ANN	CNN	ANFIS	YOLO
Vegetation health	81	85	89	94
Fruit quality assessments	78	85	88	95
Weed detection	71	76	84	89

TABLE VI. Disease Detection and Characterization

Crop Focus	ANN	CNN	ANFIS	YOLO
Disease Detection	78	85	91	95
Biological Characterization	74	78	81	87
Physical Characterization	75	79	82	89

2) Disease detection and characterization: Analyzing biological and physical properties and identifying damaged crops were made possible by the use of multi- and hyper-spectral cameras [59]. Early disease detection by the system enabled targeted treatments, reducing the possibility ofoutput losses and enhancing crop health overall as shown in Table VI.

3) LiDAR-Based plant height and water flow analysis: Important information on plant height, crop phenology, and water flow patterns was provided by the LiDAR camera. This data improved knowledge of the dynamics of growth and led to optimal water use [60]. Precise assessments of plant height enabled the tracking of agricultural phenology, resulting in enhanced cultivation tactics as shown in Table VII.

4) Real-time predictive insights: Real-time data analysis was made possible by the combination of deep stream algorithms and the NVIDIA Jetson GPU. Quick predictions were produced about crop health, vegetation and soil dryness, fertilizer needs, weed presence, and incidences of crop damage and illness [61]. This reduced possible hazards, maximized resource utilization, and enabled quick decision-making as shown in Table VIII.

All four cameras' data could be processed and analyzed in real-time thanks to the NVIDIA Jetson GPU and deepstream algorithms. This processing pipeline played a key role in providing forecasts and insights for important pomegranate cultivation issues. The system accurately forecasted fertilizer needs, analyzed soil and vegetation dryness, tracked weed infestations, and quickly picked up instances of crop damage and illness [62]. Real-time data analysis enabled prompt decision-making, which ultimately optimized resource use and increased crop output as shown in Table IX and Fig. 13. Subsequently, performance analysis over different applications for evaluating the effectiveness of the proposed system is presented in Table X.

The automated system's prognostic insights greatly aided farmers in making well-informed decisions. The system's capacity to suggest ideal irrigation plans, exact fertilizer dosages, and prompt disease treatment techniques resulted in increased resource efficiency and less environmental impact as shown in Table IX and Fig. 14 - 17. Through the useof spectral analysis, growers were able to identify diseasesand weeds early and take preventative action, potentially reducing yield losses [63]–[67]. Although the results are encouraging, certain difficulties were experienced when the automated system was put in place. For precise forecasts, camera calibration and maintaining consistent data quality are still essential. Integration of weather and climatic datamay further improve the system's accuracy. Additionally, the system may operate differently in various geographic and environmental settings, necessitating ongoing improvement and adaptation.

C. Discussion

The results underscore the transformative potential of integrating drone technology and deep-stream algorithms in pomegranate cultivation. The system not only automates data collection but also provides actionable insights across multiple facets of cultivation, empowering farmers to make informed decisions.

The following discussions delve into the broader implications and considerations:

1) Precision agriculture for sustainable farming: The automated system minimizes its impact on the environment while optimizing resource utilization per precision agricultural principles. The technology helps to promote effective and sustainable farming practices by accurately adjusting the irrigation, fertilization, and pest control strategies [68].

TABLE VII. LIDAR-BASED PLANT HEIGHT AND WATER FLOW ANALYSIS

Crop Focus	ANN	CNN	ANFIS	YOLO
Plant Height	81	82	85	92
Crop Phenology	78	82	84	91
Water Flow Patterns	81	82	85	86

TABLE VIII. REAL-TIME PREDICTIVE INSIGHTS

Crop Focus	ANN	CNN	ANFIS	YOLO
Crop Health	81	85	88	93
Vegetation	82	84	86	89
Soil Dryness	74	78	82	88
Fertilizer Requirements	71	75	85	91
Weed Presence	72	74	86	87
Crop Damage	78	81	84	92

TABLE IX. RESULT COMPARISON OF PROPOSED SYSTEM WITH EXISTING METHOD

Parameters (%)	MDC	MLP	SVM	ANFIS	YOLO
Accuracy	70	75	80	85	95
Sensitivity	72	77	81	83	94
Specificity	69	73	85	81	96
Precision	74	76	79	84	91

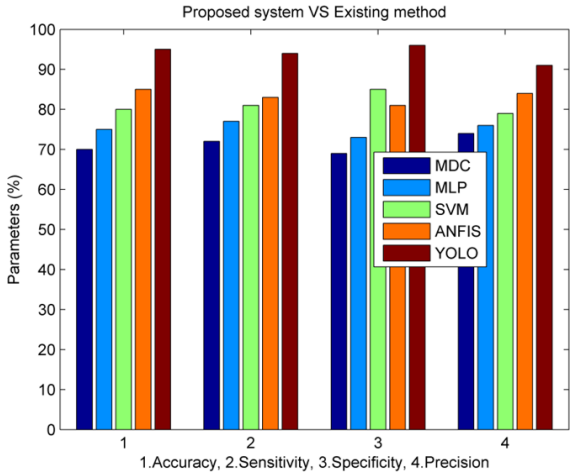


Fig. 13. Result comparison of proposed system with existing method.

TABLE X. PERFORMANCE ANALYSIS FOR VARIOUS APPLICATIONS

Crop Focus	Accuracy (%)	F1 score (%)	Recall (%)	Precision (%)
Predict crop health	95	93	91	96
Soil dryness	88	87	85	84
Fertilizer requirements	81	83	81	82
Weed presence	91	86	90	88
Crop damage and disease	94	91	93	92



Fig. 14. Performance analysis of plant health inspection and stress detection.

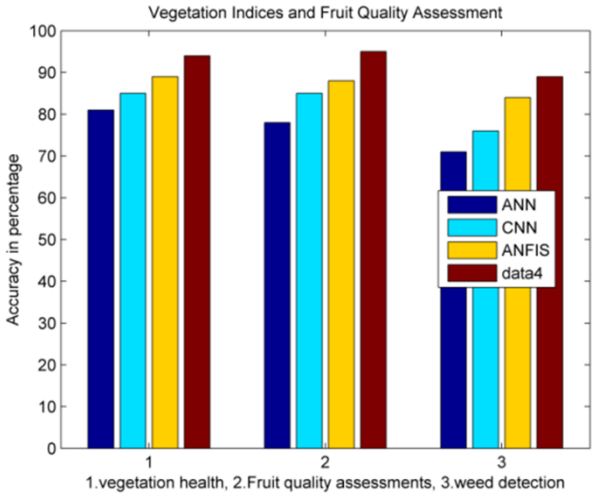


Fig. 15. Performance analysis of vegetation indices and fruit quality assessment.

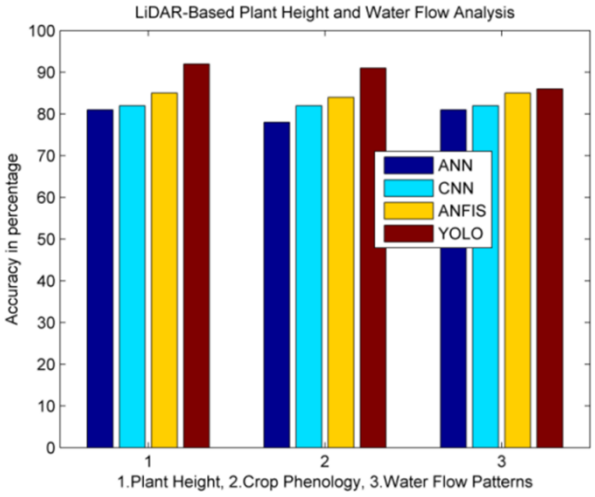


Fig. 16. Performance analysis of disease detection and characterization.

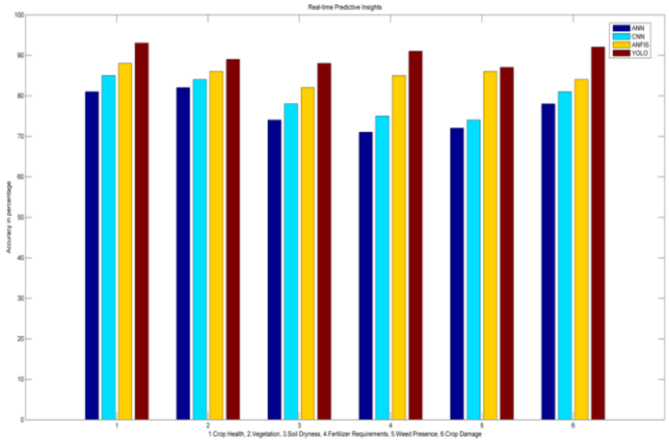


Fig. 17. Performance analysis of real-time predictive insights.

2) Early disease detection for crop protection: A breakthrough has been made with the use of spectralanalysis for early disease identification. Farmers who recognize disease symptoms early on can take prompt actionto stop the spread of the illness and maintain crop qualityand output.

3) Scalability and adaptability: Although the system appears promising, it is important to take into account its scalability and adaptation too many environmental situations. Continuous development of calibration processes, data quality control, and system robustness are necessary to guarantee consistent performance in a variety of agricultural contexts.

The accuracy of disease identification and prediction modeling can be considerably improved in the future thanks to developments in machine learning and AI algorithms. An expanded perspective on crop health trends may be obtained by combining historical data and satellite photography. Collaboration with extension agencies and agricultural professionals can help to better adapt the systemto local farming practices and spread its benefits [69]. Pomegranate cultivation could transform due to the merging of drone technology and deep-stream algorithms. The automated system provides real-time insights and suggestions for crop health, resource management, and disease identification bymerging data from thermal, optical RGB, multi-spectral, and LiDAR cameras and utilizing the processing capability of the NVIDIA Jetson GPU. While there are still issues, thesystem represents a big step towards data-driven, sustainable agriculture by enabling farmers to optimize pomegranate yield and quality [70]. Further developments and widespread acceptance in contemporary agriculture are anticipated asa result of ongoing research and development in thisfield. The following investigations need to concentrate on improving the algorithms, adding more environmental factors,and broadening the system's crop suitability. To guarantee broad acceptance and applicability, partnerships with extension agencies and agricultural specialists can further customize the system to regional farming methods.

Conclusively, the automated technique for cultivating pomegranates shows promise for transforming conventional agricultural methods. This system provides farmers with real-time information, promotes sustainable agriculture, and improves overall crop output and quality by utilizing deep-stream algorithms, modern cameras, and drone technology. This novel strategy will surely advance toward wider acceptance and implementation in international agriculture with continued study and improvement.

IV. Conclusion

Integrating drone technology and deep-stream algorithms represents a notable breakthrough in modernizing agricultural practices, particularly in pomegranate cultivation. This study showcases a thorough and evidence-based approach to farming, employing advanced technology such as a drone equipped with a thermal camera, optical RGB camera, multi-spectral camera, and LiDAR camera. These cutting-edgetools are powered by the computational capabilities of theNVIDIA Jetson GPU, enabling precise data collection andanalysis. This approach has demonstrated its effectiveness in improving different aspects of pomegranate farming. It has been used to evaluate plant health, map irrigation, manage fertilizer usage, and calculate yields. As a researcher,I have observed significant advancements in the optical RGB camera's capabilities. It has proven to be a valuabletool for analyzing vegetation indices, assessing fruit quality, and detecting weeds. These improvements have positively impacted decision-making, leading to better crop managementpractices and, ultimately, higher yields. In this field, multi-spectral and hyperspectral cameras have revolutionized how we detect crop diseases, assess damage, and respond proactively. Furthermore, the LiDAR camera has provided valuable insights into growth dynamics and resource utilization, leading to more sustainable farming practices.

Nevertheless, in light of these advancements, it is essential to consider the limitations associated with this approach carefully. The system's effectiveness relies heavily on the availability and quality of advanced drone equipment, which may not be easily accessible to all farmers, especially in regions with limited resources. This hinders the widespread adoption of the technology and can potentially create disparities in agricultural productivity.Furthermore, processing extensive datasets in real time presents significant computational challenges, particularlyin environments with limited resources. These constraints emphasize the importance of conducting additional research to enhance the system's accuracy, scalability, and adaptability to different environmental conditions.

Further research should prioritize overcoming these limitations by creating more affordable drone solutionsand enhancing the computational efficiency of deep-stream algorithms. Establishing collaborations between scientists, agricultural experts, and farmers will be essential to customizing the system to local conditions and promoting its wider use. By addressing these obstacles, this groundbreaking method holds promise for substantially impacting precision agriculture and aiding in developing more sustainable and efficient farming techniques.

Acknowledgment

The authors extend their appreciation to the Deanship of Scientific Research at Northern Border University, Arar, KSA for funding this research work through the project number "NBU-FFR-2024-1596-02".

References

[1] H. S. Abdullahi, F. Mahieddine, and R. E. Sheriff, "Technology impact on agricultural productivity: A review of precision agriculture using unmanned aerial vehicles," in *Wireless and Satellite Systems: 7th International Conference, WiSATS 2015, Bradford, UK, July 6-7, 2015. Revised Selected Papers 7*. Springer, 2015, pp. 388–400.

[2] J. Abdulridha, Y. Ampatzidis, P. Roberts, and S. C. Kakarla, "Detecting powdery mildew disease in squash at different stages using uav-based hyperspectral imaging and artificial intelligence," *Biosystems engineering*, vol. 197, pp. 135–148, 2020.

[3] J. Abdulridha, Y. Ampatzidis, S. C. Kakarla, and P. Roberts, "Detection of target spot and bacterial spot diseases in tomato using uav-based and benchtop-based hyperspectral imaging techniques," *Precision Agriculture*, vol. 21, pp. 955–978, 2020.

[4] O. E. Apolo-Apolo, M. Pérez-Ruiz, J. Martinez-Guanter, and J. Valente, "A cloud-based environment for generating yield estimation maps from apple orchards using uav imagery and a deep learning technique," *Frontiers in plant science*, vol. 11, p. 1086, 2020.

[5] S. M. Z. A. Naqvi, M. Awais, F. S. Khan, U. Afzal, N. Naz, and M. I. Khan, "Unmanned air vehicle based high resolution imagery for chlorophyll estimation using spectrally modified vegetation indices in vertical hierarchy of citrus grove," *Remote Sensing Applications: Society and Environment*, vol. 23, p. 100596, 2021.

[6] M. Mammarella, L. Comba, A. Biglia, F. Dabbene, and P. Gay, "Cooperation of unmanned systems for agricultural applications: A case study in a vineyard," *biosystems engineering*, vol. 223, pp. 81–102, 2022.

[7] L. Sánchez-Fernández, M. Barrera-Báez, J. Martínez-Guanter, and M. Pérez-Ruiz, "Reducing environmental exposure to ppps in super-high density olive orchards using uav sprayers," *Frontiers in Plant Science*, vol. 14, p. 1272372, 2024.

[8] J. Abdulridha, Y. Ampatzidis, J. Qureshi, and P. Roberts, "Laboratory and uav-based identification and classification of tomato yellow leaf curl, bacterial spot, and target spot diseases in tomato utilizing hyperspectral imaging and machine learning," *Remote Sensing*, vol. 12, no. 17, p. 2732, 2020.

[9] S. K. Jagatheesaperumal, M. M. Hassan, M. R. Hassan, and G. Fortino, "The duo of visual servoing and deep learning-based methods for situation-aware disaster management: A comprehensive review," *Cognitive Computation*, pp. 1–23, 2024.

[10] M. Rahouti, M. Ayyash, S. K. Jagatheesaperumal, and D. Oliveira, "Incremental learning implementations and vision for cyber risk detection in iot," *IEEE Internet of Things Magazine*, vol. 4, no. 3, pp. 114–119, 2021.

[11] A. Mukherjee, S. Misra, and N. S. Raghuwanshi, "A survey of unmanned aerial sensing solutions in precision agriculture," *Journal of Network and Computer Applications*, vol. 148, p. 102461, 2019.

[12] U. R. Mogili and B. Deepak, "Review on application of drone systems in precision agriculture," *Procedia computer science*, vol. 133, pp. 502–509, 2018.

[13] A. Subeesh, S. Bhole, K. Singh, N. S. Chandel, Y. A. Rajwade, K. Rao, S. Kumar, and D. Jat, "Deep convolutional neural network models for weed detection in polyhouse grown bell peppers," *Artificial Intelligence in Agriculture*, vol. 6, pp. 47–54, 2022.

[14] S. P. Kumar, D. Jat, R. K. Sahni, B. Jyoti, M. Kumar, A. Subeesh, B. S. Parmar, and C. Mehta, "Measurement of droplets characteristics of uav based spraying system using imaging techniques and prediction by gwo-ann model," *Measurement*, vol. 234, p. 114759, 2024.

[15] A. Subeesh, S. P. Kumar, S. K. Chakraborty, K. Upendar, N. S. Chande, D. Jat, K. Dubey, R. U. Modi, and M. M. Khan, "Uav imagery coupled deep learning approach for the development of an adaptive in-house web-based application for yield estimation in citrus orchard," *Measurement*,

vol. 234, p. 114786, 2024.

[16] M. J. Kazemi, M. M. Paydar, and A. S. Safaei, "Designing a bi-objective rice supply chain considering environmental impacts under uncertainty," *Scientia Iranica*, vol. 30, no. 1, pp. 336–355, 2023.

[17] A. Oikonomidis, C. Catal, and A. Kassahun, "Deep learning for crop yield prediction: a systematic literature review," *New Zealand Journalof Crop and Horticultural Science*, vol. 51, no. 1, pp. 1–26, 2023.

[18] A. Gholipour, A. Sadegheih, A. Mostafaeipour, and M. B. Fakhrzad, "Designing an optimal multi-objective model for a sustainable closed-loop supply chain: a case study of pomegranate in iran," *Environment, Development and Sustainability*, vol. 26, no. 2, pp. 3993–4027, 2024.

[19] J. S. Kumar and R. Kaleeswari, "Implementation of vector field histogram based obstacle avoidance wheeled robot," in *2016 Online international conference on green engineering and technologies (IC-GET)*. IEEE, 2016, pp. 1–6.

[20] F. Amalina, A. S. Abd Razak, S. Krishnan, A. Zularisam, and M. Nasrullah, "A comprehensive assessment of the method for producing biochar, its characterization, stability, and potential applications in regenerative economic sustainability–a review," *Cleaner Materials*, vol. 3, p. 100045, 2022.

[21] M. Esposito, M. Crimaldi, V. Cirillo, F. Sarghini, and A. Maggio, "Drone and sensor technology for sustainable weed management: A review," *Chemical and Biological Technologies in Agriculture*, vol. 8, pp. 1–11, 2021.

[22] G. Messina and G. Modica, "Applications of uav thermal imagery in precision agriculture: State of the art and future research outlook," *Remote Sensing*, vol. 12, no. 9, p. 1491, 2020.

[23] M. K. A. Devi *et al*., "Plant disease identification using the unmanned aerial vehicle images," *Turkish Journal of Computer and Mathematics Education (TURCOMAT)*, vol. 12, no. 10, pp. 2396–2399, 2021.

[24] F. DadrasJavan, F. Samadzadegan, S. H. Seyed Pourazar, and H. Fazeli, "Uav-based multispectral imagery for fast citrus greening detection," *Journal of Plant Diseases and Protection*, vol. 126, pp. 307–318, 2019.

[25] L. He, W. Fang, G. Zhao, Z. Wu, L. Fu, R. Li, Y. Majeed, and J. Dhupia, "Fruit yield prediction and estimation in orchards: A state-of-the-art comprehensive review for both direct and indirect methods," *Computers and electronics in agriculture*, vol. 195, p. 106812, 2022.

[26] H. Jemaa, W. Bouachir, B. Leblon, A. LaRocque, A. Haddadi, and N. Bouguila, "Uav-based computer vision system for orchard apple tree detection and health assessment," *Remote Sensing*, vol. 15, no. 14, p. 3558, 2023.

[27] N. S. Chandel, Y. A. Rajwade, K. Dubey, A. K. Chandel, A. Subeesh, and M. K. Tiwari, "Water stress identification of winter wheat crop withstate-of-the-art ai techniques and high-resolution thermal-rgb imagery," *Plants*, vol. 11, no. 23, p. 3344, 2022.

[28] H. Sun, M. Li, and Q. Zhang, "Crop sensing in precision agriculture," in Soil and Crop Sensing for Precision Crop Production. Springer, 2022, pp. 251–293.

[29] G. Modica, G. De Luca, G. Messina, and S. Praticò, "Comparison and assessment of different object-based classifications using machine learning algorithms and uavs multispectral imagery: A case study in a citrus orchard and an onion crop," European Journal of Remote Sensing, vol. 54, no. 1, pp. 431–460, 2021.

[30] Y. Lan, Z. Huang, X. Deng, Z. Zhu, H. Huang, Z. Zheng, B. Lian, G. Zeng, and Z. Tong, "Comparison of machine learning methods for citrus greening detection on uav multispectral images," Computers and electronics in agriculture, vol. 171, p. 105234, 2020.

[31] P. Marques, L. Pa´dua, J. J. Sousa, and A. Fernandes-Silva, "Assessing the water status and leaf pigment content of olive trees: Evaluating the potential and feasibility of unmanned aerial vehicle multispectral and thermal data for estimation purposes," Remote Sensing, vol. 15, no. 19, p. 4777, 2023.

[32] M. V. Ferro and P. Catania, "Technologies and innovative methods for precision viticulture: a comprehensive review," Horticulturae, vol. 9, no. 3, p. 399, 2023.

[33] E. G. Jones, S. Wong, A. Milton, J. Sclauzero, H. Whittenbury, and M. D. McDonnell, "The impact of pan-sharpening and spectral resolution on vineyard segmentation through machine learning," Remote Sensing, vol. 12, no. 6, p. 934, 2020.

[34] J. C. Miranda et al., "Open source software and benchmarking of computer vision algorithms for apple fruit detection, fruit sizing and yield prediction using rgb-d cameras," 2024.

[35] C. Zhang, K. Zhang, L. Ge, K. Zou, S. Wang, J. Zhang, and W. Li, "A method for organs classification and fruit counting on pomegranate trees based on multi-features fusion and support vector machine by 3d point cloud," Scientia Horticulturae, vol. 278, p. 109791, 2021.

[36] B. O. Olorunfemi, N. I. Nwulu, O. A. Adebo, and K. A. Kavadias, "Advancements in machine visions for fruit sorting and grading: A bibliometric analysis, systematic review, and future research directions," Journal of Agriculture and Food Research, p. 101154, 2024.

[37] M. Bortolini, F. G. Galizia, C. Mora, L. Botti, and M. Rosano, "Bi-objective design of fresh food supply chain networks with reusable and disposable packaging containers," Journal of cleaner production, vol. 184, pp. 375–388, 2018.

[38] A. D. Boursianis, M. S. Papadopoulou, P. Diamantoulakis, Liopa-Tsakalidi, P. Barouchas, G. Salahas, G. Karagiannidis, S. Wan, and S. K. Goudos, "Internet of things (iot) and agricultural unmanned aerial vehicles (uavs) in smart farming: A comprehensive review," Internet of Things, vol. 18, p. 100187, 2022.

[39] N. M. N. Al-Dosary, "Functional sustainability of a flight dynamics control system for stable hovering flight of an unmanned aerial vehicle (uav), such as in agricultural applications: Mathematical modeling and simulation," Journal of Agricultural Science and Technology A, vol. 13, pp. 1–29, 2023.

[40] P. K. R. Maddikunta, S. Hakak, M. Alazab, S. Bhattacharya, T. R. Gadekallu, W. Z. Khan, and Q.-V. Pham, "Unmanned aerial vehicles in smart agriculture: Applications, requirements, and challenges," IEEE Sensors Journal, vol. 21, no. 16, pp. 17 608–17 619, 2021.

[41] N. Delavarpour, C. Koparan, J. Nowatzki, S. Bajwa, and X. Sun, "A technical study on uav characteristics for precision agriculture applications and associated practical challenges," *Remote Sensing*, vol. 13, no. 6, p. 1204, 2021.

[42] D. D. Sweet, S. B. Tirado, N. M. Springer, C. N. Hirsch, and C. D. Hirsch, "Opportunities and challenges in phenotyping row crops using drone-based rgb imaging," *The Plant Phenome Journal*, vol. 5, no. 1, p. e20044, 2022.

[43] N. Han, B. Zhang, Y. Liu, Z. Peng, Q. Zhou, and Z. Wei, "Rapid diagnosis of nitrogen nutrition status in summer maize over its life cycle by a multi-index synergy model using ground hyperspectral and uav multispectral sensor data," *Atmosphere*, vol. 13, no. 1, p. 122, 2022.

[44] S. Fei, M. A. Hassan, Y. Xiao, X. Su, Z. Chen, Q. Cheng, F. Duan, R. Chen, and Y. Ma, "Uav-based multi-sensor data fusion and machine learning algorithm for yield prediction in wheat," *Precision agriculture*, vol. 24, no. 1, pp. 187–212, 2023.

[45] V. K. Chouhan, S. H. Khan, and M. Hajiaghaei-Keshteli, "Metaheuristic approaches to design and address multi-echelon sugarcane closed-loop supply chain network," *Soft Computing*, vol. 25, no. 16, pp. 11 377–11 404, 2021.

[46] ——, "Hierarchical tri-level optimization model for effective useof by-products in a sugarcane supply chain network," *Applied Soft Computing*, vol. 128, p. 109468, 2022.

[47] ——, "Sustainable planning and decision-making model for sugarcane mills considering environmental issues," *Journal of environmental management*, vol. 303, p. 114252, 2022.

[48] L. M. Dang, S. I. Hassan, I. Suhyeon, A. kumar Sangaiah, I. Mehmood, S. Rho, S. Seo, and H. Moon, "Uav based wilt detection system via convolutional neural networks," *Sustainable Computing: Informatics and Systems*, vol. 28, p. 100250, 2020.

[49] S. Despoudi, K. Spanaki, O. Rodriguez-Espindola, and E. D. Zamani, *Agricultural supply chains and industry 4.0*. Springer, 2021.

[50] M. Fattahi and K. Govindan, "A multi-stage stochastic programfor the sustainable design of biofuel supply chain networks under biomass supply uncertainty and disruption risk: A real-life case study," *Transportation Research Part E: Logistics and Transportation Review*, vol. 118, pp. 534–567, 2018.

[51] J. M. L. Montesclaros and P. S. Teng, "Digital technology adoption and potential in southeast asian agriculture," *Asian Journal of Agriculture and Development*, vol. 20, no. 2, pp. 7–30, 2023.

[52] J. A. García-Berná, S. Ouhbi, B. Benmouna, G. Garcia-Mateos, J. L. Fernández-Alemán, and J. M. Molina-Martínez, "Systematic mapping study on remote sensing in agriculture," *Applied Sciences*, vol. 10, no. 10, p. 3456, 2020.

[53] A. Goula and K. Adamopoulos, "A method for pomegranate seed application in food industries: seed oil encapsulation," *Food and bioproducts processing*, vol. 90, no. 4, pp. 639–652, 2012.

[54] A. Hamdi-Asl, H. Amoozad-Khalili, R. Tavakkoli-Moghaddam, and M. Hajiaghaei-Keshteli, "Toward sustainability in designing agricultural supply chain network: A case study on palm date," *Scientia Iranica*, 2021.

[55] J.-W. Han, M. Zuo, W.-Y. Zhu, J.-H. Zuo, E.-L. Lü, and X.-T. Yang, "A comprehensive review of cold chain logistics for fresh agricultural products: Current status, challenges, and future trends," *Trends in Food Science & Technology*, vol. 109, pp. 536–551, 2021.

[56] R. H. Heim, I. J. Wright, P. Scarth, A. J. Carnegie, D. Taylor, and J. Oldeland, "Multispectral, aerial disease detection for myrtle rust (austropuccinia psidii) on a lemon myrtle plantation," *Drones*, vol. 3, no. 1, p. 25, 2019.

[57] S. Khanal, J. Fulton, and S. Shearer, "An overview of currentand potential applications of thermal remote sensing in precision agriculture," *Computers and electronics in agriculture*, vol. 139, pp. 22–32, 2017.

[58] W.-C. Wu and E. C. Wong, "Feasibility of velocity selective arterial spin labeling in functional mri," *Journal of Cerebral Blood Flow & Metabolism*, vol. 27, no. 4, pp. 831–838, 2007.

[59] Y. Li, C. Guo, J. Yang, J. Wei, J. Xu, and S. Cheng, "Evaluation of antioxidant properties of pomegranate peel extract in comparison with pomegranate pulp extract," *Food chemistry*, vol. 96, no. 2, pp. 254–260, 2006.

[60] T. P. Magangana, N. P. Makunga, O. A. Fawole, and U. L.Opara, "Processing factors affecting the phytochemical and nutritional properties of pomegranate (punica granatum l.) peel waste: A review," *Molecules*, vol. 25, no. 20, p. 4690, 2020.

[61] P. Melgarejo-Sánchez, D. Núñez-Gómez, J. J. Martínez-Nicolás, F. Hernández, P. Legua, and P. Melgarejo, "Pomegranate variety and pomegranate plant part, relevance from bioactive point of view: A review," *Bioresources and Bioprocessing*, vol. 8, pp. 1–29, 2021.

[62] M. G. Selvaraj, A. Vergara, F. Montenegro, H. A. Ruiz, N. Safari, D. Raymaekers, W. Ocimati, J. Ntamwira, L. Tits, A. B. Omondi *et al.*, "Detection of banana plants and their major diseases through aerial images and machine learning methods: A case study in dr congo and republic of benin," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 169, pp. 110–124, 2020.

[63] D. C. Tsouros, S. Bibi, and P. G. Sarigiannidis, "A review on uav-based applications for precision agriculture," *Information*, vol. 10, no. 11, p. 349, 2019.

[64] G. Tüccar and E. Uludamar, "Emission and engine performance analysis of a diesel engine using hydrogen enriched pomegranate seed oil biodiesel," *International Journal of Hydrogen Energy*, vol. 43, no. 38, pp. 18 014–18 019, 2018.

[65] T. Wang, J. A. Thomasson, C. Yang, T. Isakeit, and R. L. Nichols, "Automatic classification of cotton root rot disease based on uav remote sensing," *Remote Sensing*, vol. 12, no. 8, p. 1310, 2020.

[66] T. Wiesner-Hanks, E. L. Stewart, N. Kaczmar, C. DeChant, H. Wu, R. J. Nelson, H. Lipson, and M. A. Gore, "Image set for deep learning: field images of maize annotated with disease symptoms," *BMC research notes*, vol. 11, pp. 1–3, 2018.

[67] T. W. Xavier, R. N. Souto, T. Statella, R. Galbieri, E. S. Santos, G. S. Suli, and P. Zeilhofer, "Identification of ramularia leaf blight cotton disease infection levels by multispectral, multiscale uav imagery,"*Drones*, vol. 3, no. 2, p. 33, 2019.

[68] C. Zhang and J. M. Kovacs, "The application of small unmanned aerial systems for precision agriculture: a review," *Precision agriculture*,vol. 13, pp. 693–712, 2012.

[69] D. Zhang, X. Zhou, J. Zhang, Y. Lan, C. Xu, and D. Liang, "Detection of rice sheath blight using an unmanned aerial system with high-resolution color and multispectral imaging," *PloS one*, vol. 13, no. 5, p. e0187470, 2018.

[70] X. Zhang, L. Han, Y. Dong, Y. Shi, W. Huang, L. Han,P. Gonza´lez-Moreno, H. Ma, H. Ye, and T. Sobeih, "A deep learning-based approach for automated yellow rust disease detectionfrom high-resolution hyperspectral uav images," Remote Sensing, vol. 11, no. 13, p. 1554, 2019.

Comparative Analysis of Small and Medium-Sized Enterprises Cybersecurity Program Assessment Model

Wan Nur Eliana Wan Mohd Ludin[1], Masnizah Mohd[2], Wan Fariza Paizi@Fauzi[3]
Center for Cyber Security-Faculty of Information Science & Technology,
Universiti Kebangsaan Malaysia, Bangi, Selangor Malaysia[1, 2, 3]
Faculty of Computer Science and Information Computing Technology,
New Era University College, Kajang, Selangor, Malaysia[1]

***Abstract*—In the digital age, Small and Medium-sized Enterprises must review and improve their cybersecurity posture to combat rising risks. This paper thoroughly compares Small and Medium-sized Enterprises cybersecurity program assessment approaches. The National Institute of Standards and Technology's Cybersecurity Framework, CyberSecurity Readiness Model for SMEs, Cybersecurity Evaluation Model, and Adaptable Security Maturity Assessment and Standardisation framework were examined. The NIST CSF is adaptable and applicable to many sectors, while the CSRM provides a standardized way to assess an organization's cyber readiness. With its resource limits and operational scales, the CSRM-SME meets SMEs' particular issues. Organizations may examine and improve cybersecurity with CSEM. The approach can be used for SMEs, higher education institutions, and industrial control systems. The ASMAS architecture is flexible for continual security enhancement due to its scalability and standardization. This comparison analysis shows each framework's strengths and weaknesses, revealing their suitability for diverse SME scenarios. This paper helps SMEs choose the best model to strengthen cybersecurity, boost resilience, and meet global standards. This paper will compare the NIST CSF, CSRM-SME, CSEM, and ASMAS cybersecurity frameworks.**

***Keywords*—Cybersecurity; SMEs; cybersecurity program assessment models; cybersecurity assessment frameworks**

I. Introduction

Small and Medium-sized Enterprises (SMEs) are vital to the economy but are frequent targets of cyberattacks due to their limited cybersecurity capabilities [1]. Existing maturity assessment models and standards often need to pay more attention to SMEs' requirements and roles in the digital ecosystem. The rise of Industry 4.0 and digital transformation introduces new cybersecurity challenges for SMEs. A tailored cybersecurity assessment model is needed to address the unique cybersecurity needs of SMEs. This model should consider SMEs' resources and expertise limitations while providing effective cybersecurity measures [2].

A. Global and Malaysia Small and Medium-sized Enterprises (SMEs)

SMEs drive innovation, competitiveness, and job creation, making them the backbone of the economy [3]. These companies have fewer than 250 people and a turnover or balance sheet of less than €50 million or €43 million [4]. Most countries have SMEs, including 99% of EU enterprises [5]. SMEs boost GDP, employment, and innovation, making them crucial to the economy. SMEs provide half of U.S. jobs but only 40% of GDP [6]. SMEs comprise 98% of Australian enterprises, contribute one-third of GDP, and employ 4.7 million people. SMEs generate 44% of Norway's economic value and employ 47% of private sector workers [7].

Malaysia's SMEs boost GDP, employment, and innovation. SME definitions include sales turnover and full-time employee count. SME status in Malaysia is determined by a sales turnover of RM50 million or fewer than 200 full-time employees [8]. Malaysian SMEs are classified by sales turnover and full-time personnel. Micro, small, and medium requirements, such as manufacturing and services, vary by industry. The manufacturing sector has microenterprises with sales turnovers under RM300,000 or less than five full-time employees. However, a tiny business makes between RM300,000 and RM15 million [9].

The Malaysian economy relies on SMEs, which comprise 97.2% of businesses, 38.2% of GDP, and 7.3 million jobs. These businesses generate economic growth, with 98.5% of Malaysian companies being SMEs. SMEs generated about RM500 billion to Malaysia's GDP and 5.7 million jobs, 70% of the workforce in 2018. [10]. To keep up with digital culture, SMEs should use digital marketing to boost market presence and efficiency. The government helps SMEs digitalize to expand their consumer base and increase efficiency. Establishing the Ministry of Entrepreneur Development and Cooperatives (MEDAC) shows the government's support for SMEs and entrepreneurship. The National Entrepreneur and SME Development Council (NESDC) promotes entrepreneurship to boost economic growth [11].

Table I compares Small and Medium-sized Enterprises (SMEs) globally and SMEs, specifically in Malaysia, across various aspects, including economic contribution, internationalization, technology adoption, government support, market orientation, challenges, performance factors, environmental practices, and corporate governance. The importance of SMEs globally and in Malaysia while also showcasing the unique challenges they face and the support systems in place to help them thrive. It underscores the critical

role of policy measures and innovation in driving SME growth and sustainability.

TABLE I. COMPARISON OF GLOBAL SMES AND MALAYSIA SMES

Criteria	Global SMEs	Malaysia SMEs
Economic Contribution [3]	Significant contribution to GDP and employment across various countries.	Manage 98.5% of Malaysian businesses, 65.3% of jobs, and 36.3% of GDP.
Internationalization (4)	We are engaged in global markets through exports, joint ventures, and international partnerships.	Exports boost GVC and FTA participation.
Technology Adoption [5]	Adoption varies widely; advanced economies often lead to technology integration.	High expenses and the need for innovation to stay competitive hinder technology adoption.
Government Support [6]	Various support levels, including financial aid, training, and internationalization assistance.	Significant government development, financial, and export promotion support.
Market Orientation [7]	Market orientation is critical for success; firms focusing on customer needs and market trends perform better.	Customer attention and market dispersion are essential, but intelligence and reactivity differ.
Challenges [8]	Common challenges include access to finance, competition, and regulatory hurdles.	Lack of competent labor, high raw material costs, and upfront investment costs.
Performance Factors [9]	Performance is linked to innovation, market expansion, and efficient resource utilization.	Internationalization and performance are linked, emphasizing market orientation.
Environmental Practices [8]	Increasing emphasis on sustainability and green practices in developed countries.	Early green practices; ISO 14001 Environmental Management System to improve performance.
Corporate Governance [10]	Varies significantly; better corporate governance practices are correlated with improved SME performance.	Better corporate governance practices are needed for monitoring and procedure implementation.

II. BACKGROUND ASSESSMENT MODELS AND FRAMEWORKS

An organization's security posture can be assessed and improved using a cybersecurity program assessment model to discover vulnerabilities, assess risks, and deploy controls. These models evaluate external threats like cyberattacks and internal weaknesses like obsolete software or human errors that could affect an organization's information systems [11]. They include methods for estimating and prioritizing risks, assessing cyber threat impact and likelihood, and selecting reaction levels [12]. These models help organizations establish security policies, access controls, firewalls, and personnel training to limit risks [13]. As threats change, effective cybersecurity program assessment models emphasize continual monitoring and periodic appraisal to improve the organization's cybersecurity posture [14]. Numerous models link with international standards like ISO/IEC 27001 and 27002, offering a benchmark for cybersecurity maturity and compliance [15]. Cybersecurity program assessment models help organizations prepare for growing cyber threats with these comprehensive techniques.

A. Taxonomy Assessment Models

Cybersecurity program assessment models are diverse frameworks designed to evaluate and enhance the security posture of organizations. These models systematically categorize different aspects of cybersecurity to provide a comprehensive and structured approach to risk assessment, threat identification, and mitigation. The taxonomy of these models often includes various components such as risk factors, threat vectors, control measures, and evaluation criteria.

TABLE II. THE SUMMARY TAXONOMY ASSESSMENT MODEL

Taxonomy	Description
Risk-Based Taxonomy	Risk identification, analysis, and management. Quantifies threat occurrences, vulnerability, and effect of cybersecurity threats. Quantitative algorithms measure cybersecurity risk using these parameters [11].
Hierarchical and Graph-Theoretic Taxonomy	Uses hierarchical and graph-theoretic models to assess cybersecurity vulnerabilities. Taxonomically classifies threat actors' methods and provides cyber-physical assault graphs to analyze threat transmission [12].
Capability Maturity Models (CMMs)	Assess and improve an organization's cybersecurity. Classifies maturity levels in policy, operations, and human factors. Compares the present situation to optimal practices [13].
Socio-Technical Taxonomy	Assesses cybersecurity threats and improves IT, security, and non-technical staff communication using technical and human factors. Work processes and hazards are visualized using modeling languages [14].
Multicriteria Decision Frameworks	Integrates various criteria to assess the overall utility of cybersecurity management alternatives. Quantifying threats, vulnerabilities, and consequences provides a structured approach to selecting risk management actions [15].
Dynamic Simulation-Based Taxonomy	Assesses cybersecurity threats and plans long-term investments using dynamic simulation. Addresses organizational change and cyberattack dynamics [16].
Comprehensive and Flexible Taxonomies	Includes worldwide and national cybersecurity recommendations. Technology, organization, people, and environment are measured to assess cybersecurity readiness [17].

Organizational security is assessed and improved using several cybersecurity program assessment methodologies. These frameworks categorize cybersecurity to organize risk assessment, threat identification, and mitigation. Table II shows that these models' taxonomies comprise risk variables, threat vectors, control measures, and evaluation criteria.

Several taxonomic techniques are used to control and reduce cybersecurity threats. Risk-based taxonomies categorize risks into quantifiable criteria, including attack events, vulnerabilities, and impacts, and employ quantitative algorithms to evaluate and prioritize cybersecurity risks [11]—however, hierarchical and graph-theoretic taxonomies model cybersecurity concerns. Taxonomical classifications of threat actors' approaches, tactics, and processes generate cyber-physical attack graphs that analyze threat propagation, helping identify vital assets and prioritize controls [12]. CMMs evaluate and improve an organization's cybersecurity practices in policy, operations, and human factors. The National Cybersecurity Capacity Maturity Model (CMM) lets organizations compare their current condition to best

practices and identify areas for improvement [17]. Multicriteria decision frameworks quantify threats, vulnerabilities, and consequences to evaluate cybersecurity management alternatives and provide a structured approach for risk management action selection, bridging the gap between risk assessment and risk management [15]. Comprehensive and flexible taxonomies include worldwide and national cybersecurity recommendations for technology, organizations, people, and the environment. These holistic cybersecurity readiness models are adaptable to organizational situations [17].

B. Process Development Assessment Model

Developing a cybersecurity program assessment model involves a structured and iterative process to evaluate and enhance an organization's cybersecurity posture. This structured and iterative process ensures that organizations can systematically assess, manage, and strengthen their cybersecurity posture, thereby reducing risks and improving overall security resilience.

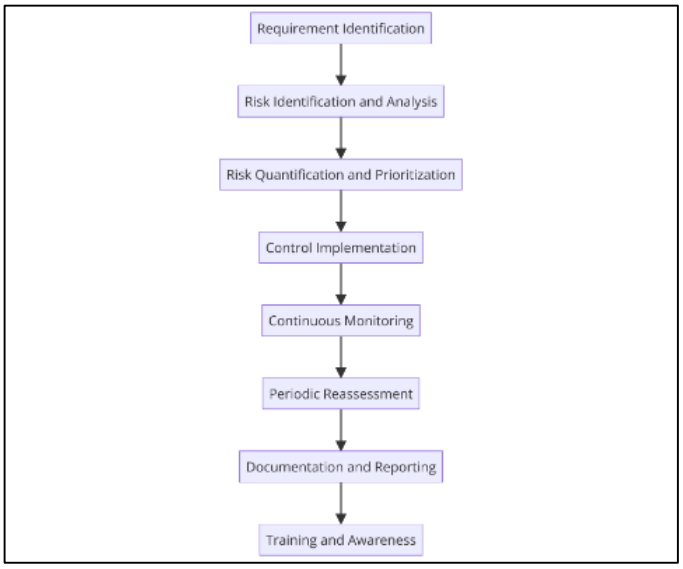


Fig. 1. Process of developing assessment model.

Fig. 1 demonstrates an organization's comprehensive cybersecurity management procedure. It starts with need identification, which defines cybersecurity needs, limitations, behaviors, services, and security requirements [18]. Next, Risk Identification and Analysis involves detecting and analyzing internal and external cybersecurity risks and understanding threats and vulnerabilities [19]. After that, risk quantification and prioritization are employed to assess and rank these risks by impact and likelihood [15]. Control Implementation involves creating and implementing security policies, access controls, and employee training programs to reduce these risks [20]. Continuous Monitoring ensures these measures are effective through audits, vulnerability scans, and real-time threat detection. Reassessment and control adjustments are made to handle

C. Paper Structure

This paper will compare these cybersecurity frameworks, focusing on the NIST CSF, CSEM, CSRM-SME, and ASMAS. By examining their structures, implementation processes, strengths, weaknesses, and suitability for different organizational contexts, this paper provides insights into the most effective strategies for enhancing cybersecurity readiness, particularly for SMEs. Through this comparison, we aim to highlight each framework's key features and benefits, ultimately guiding organizations in selecting the most appropriate framework for their cybersecurity needs.

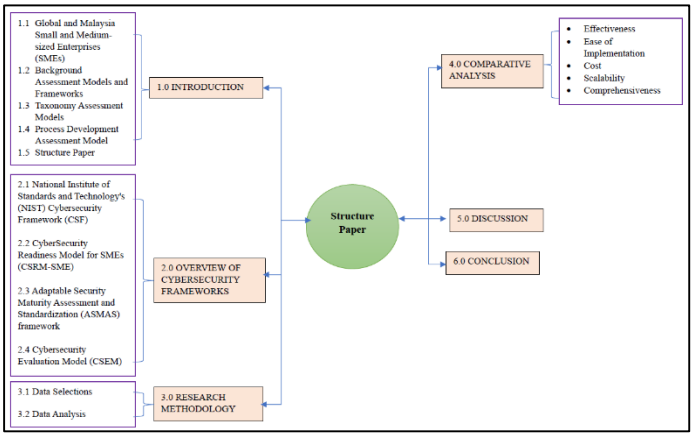


Fig. 2. The structure of the paper.

The structure of this paper is outlined in Fig. 2. Section I discusses an introduction to SMEs and the background of assessment models and frameworks. Section II discusses an overview of the cyber security model and framework. Section III discusses the research methodology. Then, Section IV presents the comparative analysis of cybersecurity program assessment models and frameworks. Finally, this paper presents the discussion and conclusions in Sections V and VI respectively.

III. Cybersecurity Program Assessment Model

Cybersecurity has become a critical concern for Small and Medium-sized Enterprises (SMEs) in Malaysia, given the increasing sophistication and frequency of cyber threats. Developing and implementing a comprehensive cybersecurity program assessment model tailored for Malaysian SMEs is essential to enhance their resilience against cyberattacks.

ISO 27001, while a comprehensive and internationally recognized standard, is often resource-intensive, requiring significant financial and human resources to implement effectively. This can be a substantial barrier for SMEs, which typically operate with limited budgets and may need more specialized staff to manage such a complex framework [18], [19]. Furthermore, the flexibility of ISO 27001, while beneficial for large organizations with diverse needs, may result in an overly broad approach that aligns poorly with SMEs' specific and more narrowly focused security needs [20]. Similarly, while the CIS Controls are designed to be more accessible and prescriptive, they may still present challenges in prioritization and customization that are difficult for SMEs to navigate without expert guidance. Though beneficial for comprehensive coverage, the CIS framework's broad scope may need to be aligned with the limited operational scope of many SMEs, making it less practical compared to more targeted cybersecurity assessment models [21].

A. National Institute of Standards and Technology's (NIST) Cybersecurity Framework (CSF)

To help organizations manage and decrease cybersecurity risks, the NIST Cybersecurity Framework (CSF) provides comprehensive recommendations and best practices. The CSF was first published in 2014 and updated multiple times, with CSF 2.0 released in February 2024. This cybersecurity methodology is flexible and reproducible for all sizes and sectors of organizations. The five essential functions—Identify, Protect, Detect, Respond, and Recover—provide a comprehensive overview of an organization's cybersecurity risk management [33]. The framework is versatile so that organizations can customize it. System components include Framework Core, Framework Implementation Tiers, and Framework Profiles. This thorough guide helps organizations manage and reduce cybersecurity risks. It is versatile and adaptive to the needs of diverse organizations, regardless of size, sector, or maturity.

TABLE III. COMPONENTS IN THE NIST CYBERSECURITY FRAMEWORK (CSF)

Component	Description
CSF Core	Govern, Identify, Protect, Detect, Respond, and Recover are its main functions. Each function has categories and subcategories that define cybersecurity management outcomes and actions. The "Identify" function manages assets, whereas the "Protect" function controls access [22].
Implementation Tiers	Four implementation tiers: Partial (Tier 1) to Adaptive (Tier 4). These tiers show how risk management and corporate goals influence cybersecurity procedures. Higher tiers reflect more sophisticated cybersecurity risk management [22] [23].
Profiles	Custom framework implementations for unique organizations. They match cybersecurity with business needs, risk tolerance, and resources. Profiles help organizations prioritize and handle cybersecurity [22] [23].

Table III provides a concise overview of the critical components within the NIST Cybersecurity Framework (CSF). The paper overviews the framework's main elements, including CSF Core, Implementation Tiers, and Profiles. It emphasizes the significance and function of these components within the framework. Each component briefly describes how it contributes to aligning cybersecurity activities with organizational needs and risk management.

B. CyberSecurity Readiness Model- SME (CSRM-SME)

CSRM-SME is designed to enhance the cybersecurity posture of Small and Medium-sized Enterprises (SMEs) by addressing both technical and socio-technical dimensions. This model emphasizes the importance of balancing human and technical factors, fostering a strong cybersecurity culture, and using adaptable, metric-based assessments to address the unique challenges faced by SMEs [14].

Table IV shows that CSRM-SME provides a comprehensive approach to enhancing cybersecurity readiness by integrating socio-technical elements. This model emphasizes balancing human and technical factors, fostering a strong cybersecurity culture, and using adaptable, metric-based assessments to address SMEs' unique challenges. Implementing such a model can significantly improve SMEs' ability to manage cyber threats effectively.

TABLE IV. COMPONENT OF CSRM-SME

Core Component	Description
Socio-Technical Perspective	Assesses and improves cybersecurity readiness using human and technical factors. Focuses on organizational methods and technical defenses [14].
Human Element Integration	It maps socio-technical networks and human interactions using user journeys. It improves communication between IT, security, and non-technical staff to address human vulnerabilities [24].
Comprehensive Framework	Balances social, technical, and environmental factors. Provides a methodical approach to addressing SMEs' cybersecurity gaps [25].
Organizational Culture and Readiness	Highlights cybersecurity culture. It stresses that cybersecurity knowledge and culture are as necessary as technical solutions. Assesses essential areas for improvement [26].
Metric-Based Assessments	Reviews and creates socio-technical cybersecurity metrics. Addresses metric aggregation and flexibility for SMEs via straightforward, threat-based evaluations tailored to their needs [27].

C. Adaptable Security Maturity Assessment and Standardization (ASMAS) Framework

The Adaptable Security Maturity Assessment and Standardisation (ASMAS) framework has been examined in numerous studies to meet the cybersecurity needs of diverse organizations. SMEs face particular cybersecurity challenges; thus, a web-based ASMAS framework is proposed to handle them [25]. Another paper offers a European cybersecurity education maturity assessment methodology that defines knowledge units and standardizes instruction [25]. In contrast, [26] presents a maturity structure for Security Operation Centres (SOC) to ensure cybersecurity management. The research in [27] emphasizes adaptability and standardization by integrating cybersecurity maturity evaluations and standardization to satisfy organizational needs. Finally, the study in [28] proposes a security maturity self-assessment paradigm for the software development lifecycle to improve security. These numerous approaches demonstrate the need for adaptive frameworks to fulfill cybersecurity objectives across sectors and environments.

The Adaptable Security Maturity Assessment and Standardization (ASMAS) framework provides a comprehensive approach to enhancing cybersecurity practices within Small and Medium-sized Enterprises (SMEs). The framework is structured around three key aspects: core components, framework core, and implementation tiers, as shown in Table V. Following the framework, SMEs can systematically build a resilient security infrastructure that evolves with the changing threat landscape, thereby effectively safeguarding their operations and sensitive information.

TABLE V. SUMMARY CORE COMPONENT ASMAS FRAMEWORK

Aspects	Part	Descriptions
Core Components [29]	Risk Management	Identifying, assessing, and prioritizing risks.
	Security Policies	It establishes and enforces security policies and procedures.
	Access Control	Managing access to resources ensures that only authorized users can access sensitive information.
	Incident Response	I am preparing for and responding to security incidents.
	Continuous Monitoring	We regularly monitor security controls to detect and respond to new threats.
	Employee Training	We educate employees about best practices and protocols for security.
Framework Core [29]	Identify	It is understanding the business context, resources, and risk management processes.
	Protect	We are implementing safeguards to ensure the delivery of critical infrastructure services.
	Detect	Developing and implementing activities to identify the occurrence of a cybersecurity event.
	Respond	We are developing and implementing appropriate activities to take action regarding a detected event.
	Recover	It maintained plans for resilience and restored any impaired capabilities or services.
Implementation Tiers [29]	Tier 1: Partial	Informal and ad-hoc approaches to security.
	Tier 2: Risk-Informed	Awareness of risks and beginning to implement security measures systematically.
	Tier 3: Repeatable	We have established practices and policies for security management.
	Tier 4: Adaptive	Continuous improvement and adaptation to new threats.

D. Cybersecurity Evaluation Model (CSEM)

Organizations can examine and improve cybersecurity using the Cybersecurity Evaluation Model (CSEM). The approach can be used for SMEs, higher education institutions, and industrial control systems. Cybersecurity evaluation models (CSEM) research offers many risk assessment and management techniques. The study in [26] emphasize the COVID-19 pandemic's impact on cyber threats and the usage of Bayesian Networks, Random Forests, and Social Networks to assess cyber-attack risks. The study in [27] emphasizes threat modeling's strong ROI in spotting cyber threats and fixing design faults [26]. Construct a CSEM for Indian SMEs in a virtual team setting, highlighting the heightened cyber risk due to remote working during the pandemic and proposing a quantitative approach to analyze and mitigate these risks. Finally, a paper validating the CyberSecurity Audit Model (CSAM) in Canadian higher education institutions shows that CSAM can conduct comprehensive cybersecurity audits across domains, demonstrating its practicality and importance in improving cybersecurity [29] [30].

TABLE VI. CORE COMPONENTS CSEM

Component	Description
Risk Assessment	Assessing cybersecurity risks through surveys and identifying strengths and weaknesses [26].
Security Requirements	Establishing security requirements based on ISO/IEC 27002 standards [28].
Maturity Self-Assessment	Self-assessment of cybersecurity maturity using frameworks like NIST CSF [29].
Audit Model	A comprehensive model for conducting cybersecurity audits across various domains [30].
Risk Analysis and Mitigation	We integrate fault tree analysis and fuzzy decision theory for risk evaluation and mitigation [26].

Table VI shows the Cybersecurity Evaluation Model (CSEM) comprehensive framework designed to enhance organizations' cybersecurity posture through several vital components. The Risk Assessment component identifies strengths and weaknesses in an organization's cybersecurity posture by conducting detailed surveys. This is followed by Security Requirements, which establish baseline standards for cybersecurity measures based on recognized frameworks such as ISO/IEC 27002, ensuring that all necessary protocols are in place. Maturity Self-Assessment involves using frameworks like the NIST Cybersecurity Framework (CSF) to self-evaluate and improve cybersecurity practices across critical areas, including identification, protection, detection, response, and recovery. The Audit Model [36] component provides a structured approach for conducting thorough cybersecurity audits across various organizational domains, verifying the effectiveness of implemented controls. Finally, risk analysis and mitigation integrate advanced methods such as fault tree analysis and fuzzy decision theory to assess and mitigate cybersecurity risks, identify vulnerabilities, and develop strategies to address potential threats. These components form a robust model that helps organizations systematically manage and enhance their cybersecurity defenses.

E. Summary

The Adaptable Security Maturity Assessment and Standardization (ASMAS) framework and the National Institute of Standards and Technology (NIST) Cybersecurity Framework (CSF) both aim to enhance cybersecurity practices but cater to different organizational needs. The ASMAS framework is specifically designed for Small and Medium-sized Enterprises (SMEs), offering a tailored, adaptable, and user-friendly approach that addresses the unique challenges faced by these smaller entities. In contrast, the NIST CSF is a comprehensive and flexible framework suitable for various organizations, including those in critical infrastructure sectors. Still, its complexity and resource demands can be challenging for smaller organizations to implement effectively. Ultimately, the choice between these frameworks should be guided by the organization's specific needs, resources, and capabilities.

The Cybersecurity Evaluation Model (CSEM) and the CyberSecurity Readiness Model for SMEs (CSRM-SME) both provide frameworks to enhance cybersecurity in Small and Medium-sized Enterprises (SMEs). Still, they cater to different organizational needs and complexities. The CSEM is designed to be practical and straightforward, focusing on assessing cybersecurity risks and providing clear guidelines for improvement, particularly in remote work environments. It utilizes a quantitative approach through surveys, making it accessible and easy to implement for SMEs looking to identify their cybersecurity strengths and weaknesses.

On the other hand, the CSRM-SME offers a comprehensive evaluation based on a socio-technical perspective, considering both technological and human factors. This model provides a holistic view of an organization's cybersecurity readiness by examining the interaction between technology, people, and processes. While it offers a deeper understanding of cybersecurity readiness, its implementation can be more complex and resource-intensive.

IV. Research Methodology

This paper employs a comparative research design to analyze and evaluate the effectiveness, implementation processes, and overall suitability of five distinct cybersecurity frameworks: the National Institute of Standards and Technology's (NIST) Cybersecurity Framework (CSF), CyberSecurity Readiness Model for SMEs (CSRM-SME), and Adaptable Security Maturity, Assessment, and Standardization (ASMAS) framework.

A. Data Selection

Academic databases, industry reports, and framework documentation are secondary data sources. These resources explain framework structure, execution, and goals. Academic databases provide peer-reviewed research articles and studies for credibility and accuracy. Industry studies show how these frameworks are used through trends, applications, and expert analysis. However, framework documentation includes thorough implementation instructions and protocols. To ensure a comprehensive evaluation, consider these sources when selecting data. Integrating information from these varied sources helps create a complete grasp of each framework and enables an intense study of its practical and theoretical underpinnings. Fig. 3 shows the process data selection structure.

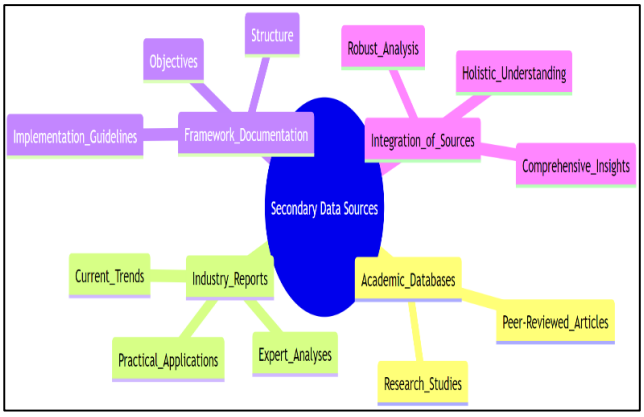


Fig. 3. Data selection structure.

B. Data Analysis

Comparisons of frameworks depend on literature and framework documentation content analysis. This method systematically analyses text to find patterns, themes, and critical traits. Analyzing academic articles, industry reports, and framework guidelines can reveal parallels and variances in framework structures, objectives, and implementation tactics. This technique helps compare frameworks and identify their strengths and drawbacks. Content analysis is a core method for organizing and synthesizing qualitative data into meaningful insights. This method ensures thorough research with empirical evidence, leading to more informed judgments and suggestions. Fig. 4 shows how content analysis compares framework stages and outcomes.

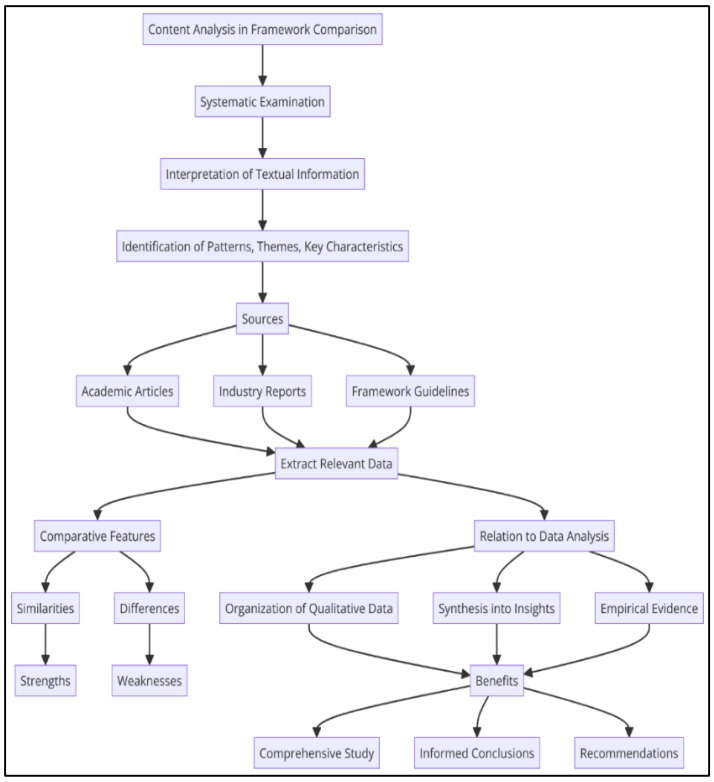


Fig. 4. Process of content analysis.

V. Result

In cybersecurity for Small and Medium-sized Enterprises (SMEs), a comparative analysis of various models is crucial to identify the most suitable framework. This analysis focuses on critical criteria: Effectiveness, Ease of Implementation, Cost, Scalability, and Comprehensiveness. The CyberSecurity Readiness Model for SMEs (CSRM-SME) is designed for small businesses, prioritizing ease of implementation and cost-effectiveness. The Cybersecurity Evaluation Model (CSEM) offers a robust framework emphasizing comprehensiveness and scalability, making it adaptable to various business sizes. The National Institute of Standards and Technology's (NIST) Cybersecurity Framework (CSF) is renowned for its effectiveness and comprehensiveness, providing a structured approach that is highly scalable but can be complex to implement for SMEs without significant resources. Lastly, the

Adaptable Security Maturity Assessment and Standardization (ASMAS) focuses on maturity and standardization, balancing comprehensiveness and ease of implementation while also being mindful of cost and scalability. This comparative analysis aims to elucidate the strengths and weaknesses of each model, helping SMEs choose the most appropriate framework for their specific needs and constraints.

Table VII above shows four highly effective and scalable models suitable for different SME needs. CSRM-SME, CSEM, and ASMAS stand out for their ease of implementation, while NIST CSF is noted for its comprehensive approach. Cost considerations vary, with CSEM and ASMAS being more affordable options.

Table VIII categorizes four cybersecurity models based on their performance across critical criteria. All frameworks are rated high for effectiveness, making them solid choices for enhancing cybersecurity. Regarding ease of implementation, CSRM-SME, CSEM, and ASMAS are rated high, indicating they are user-friendly and straightforward. In contrast, NIST CSF is rated moderate, requiring more effort for integration.

Cost-wise, CSEM and ASMAS are rated low, suggesting they are more affordable options, while CSRM-SME and NIST CSF are rated moderate. All frameworks are highly scalable and suitable for various organizational sizes and needs. Lastly, comprehensiveness is high for CSRM-SME, NIST CSF, and ASMAS, ensuring they cover a wide range of cybersecurity aspects, whereas CSEM is moderately comprehensive.

TABLE VII. COMPARATIVE ANALYSIS ASSESSMENT MODEL

Criterion	CSRM-SME [14]	CSEM [26]	ASMAS [29]	NIST Cybersecurity Framework (CSF) [27]
Effectiveness	The CSRM-SME model enhances cybersecurity readiness by integrating technical and human factors, providing a holistic and practical socio-technical approach.	CSEM provides a structured evaluation using a survey to assess cybersecurity risks, focusing on identifying strengths and weaknesses in SMEs' cybersecurity posture, which helps in targeted improvements	ASMAS addresses specific SME requirements for cybersecurity maturity and includes an evaluation study showing positive results for perceived usefulness and ease of use.	The NIST CSF offers a structured approach to managing cybersecurity risks and enhancing security posture across various sectors, including healthcare and financial services. Its core functions (Identify, Protect, Detect, Respond, Recover) provide comprehensive cybersecurity coverage.
Ease of Implementation	It uses a socio-technical approach, requiring comprehensive organizational changes, but is tailored for SMEs.	Utilizes an easy online survey for SMEs to complete, providing an accessible way to assess cybersecurity risks.	ASMAS is demonstrated through a web-based software prototype, showing ease of use and positive feedback from SMEs in evaluation studies	Implementing the NIST CSF can be complex and resource-intensive for SMEs, but it offers clear guidance and can integrate with other standards, enhancing ease of adoption.
Cost	The cost implications are not explicitly detailed but include potential expenses related to socio-technical adjustments within the organization.	CSEM involves minimal cost as it primarily uses a survey for self-assessment.	ASMAS uses a web-based software tool, which may involve initial setup costs but is designed for SMEs, keeping affordability in mind.	Implementing the NIST CSF can be costly due to technology, training, and maintenance investments. The Gordon-Loeb Model can help organizations evaluate cost-benefit aspects and optimize cybersecurity spending.
Scalability	The socio-technical approach can be scaled but might require tailored adjustments for different SME contexts.	It is scalable as it can be adapted for different SME sizes and industries by modifying the survey.	It is designed to be adaptable for various SMEs and includes specific adjustments to meet unique SME requirements, making it highly scalable.	The NIST CSF is scalable and adaptable, making it suitable for organizations of all sizes. It can be tailored to fit an organization's size, complexity, and cybersecurity needs.
Comprehensiveness	Comprehensive in addressing both technical and human aspects of cybersecurity readiness	Focuses on a comprehensive evaluation of cybersecurity maturity through detailed survey questions	Highly comprehensive, addressing both assessment and standardization needs specific to SMEs with positive evaluation results.	The NIST CSF is comprehensive, covering various cybersecurity areas with detailed guidelines. It aligns well with other standards, facilitating a holistic approach to managing cybersecurity risks.

TABLE VIII. EFFECTIVENESS PERFORMANCE

Criterion	CSRM-SME	CSEM	NIST CSF	ASMAS
Effectiveness	H	M	H	H
Ease of Implementation	M	H	M	H
Cost	M	L	M	M
Scalability	M	H	H	H
Comprehensive	H	M	H	H

H= High, M=Moderate, L= Low

VI. Discussion

The comparative analysis of the CyberSecurity Readiness Model for SMEs (CSRM-SME), Cybersecurity Evaluation Model (CSEM), National Institute of Standards and Technology's (NIST) Cybersecurity Framework (CSF), and Adaptable Security Maturity Assessment and Standardization (ASMAS) highlights the distinct strengths and focuses of each framework. All frameworks are recognized for their effectiveness in improving cybersecurity posture, making them well-regarded across various sectors. For instance, the CSRM-SME and ASMAS frameworks are particularly well-tailored for SMEs, effectively addressing their unique needs and constraints [14]. Similarly, the CSEM framework focuses on SMEs, emphasizing remote work environments, which have gained significant importance in the post-pandemic era [32]. On the other hand, the NIST CSF is noted for its comprehensiveness and widespread adoption across various industries, making it a robust choice for diverse organizational needs [34].

When considering ease of implementation, CSRM-SME, CSEM, and ASMAS stand out for their user-friendly guidelines and tools, facilitating straightforward adoption by SMEs. This high ease of implementation is supported by research highlighting these frameworks' design around SME constraints, such as limited resources and expertise [28]. In contrast, while the NIST CSF is effective, its broader scope and complexity may require more resources and expertise to integrate fully, especially within smaller organizations [31].

Cost is another critical factor, particularly for SMEs with limited budgets. Studies indicate that CSEM and ASMAS are cost-effective, making them accessible to smaller organizations [32]. Conversely, the CSRM-SME and NIST CSF are rated moderate in cost, as their implementation might require more extensive resources or adjustments, thereby incurring additional expenses [28].

In terms of scalability, all frameworks score high, reflecting their ability to adapt to organizations of different sizes and types. This scalability is essential for SMEs that may grow and require more comprehensive cybersecurity measures [32, 34]. Regarding comprehensiveness, the CSRM-SME, NIST CSF, and ASMAS frameworks are rated high as they cover a broad range of cybersecurity aspects and provide detailed guidelines for implementation. Meanwhile, CSEM, although practical, focuses primarily on remote working environments and best practices, making it less comprehensive in other cybersecurity [35] areas.

The findings underscore that while each cybersecurity framework is practical, their strengths and focus areas differ, making them suitable for varying organizational contexts. SMEs, in particular, can benefit from frameworks like CSRM-SME, CSEM, and ASMAS, specifically designed to address their unique needs and constraints.

VII. Conclusion

CSRM-SME and ASMAS are highly effective, easy to implement, comprehensive, and scalable, making them excellent choices for SMEs looking for robust, user-friendly cybersecurity solutions. They balance effectiveness and cost, ensuring that even smaller organizations can enhance their cybersecurity posture without significant financial strain. CSEM is also highly effective and cost-efficient, particularly suited for SMEs in remote working environments. It provides practical guidelines and tools, though it might not be as comprehensive in covering all cybersecurity aspects as other frameworks. NIST CSF stands out for its extensive and highly effective approach, suitable for various industries. However, implementing it may require more resources and expertise, which could be a consideration for smaller organizations with limited budgets.

Based on a comparative analysis of 4 cybersecurity models, the CyberSecurity Readiness Model for SMEs (CSRM-SME), the Cybersecurity Evaluation Model (CSEM), the National Institute of Standards and Technology's (NIST) Cybersecurity Framework (CSF), and Adaptable Security Maturity Assessment and Standardization (ASMAS) the Adaptable Security Maturity Assessment and Standardization (ASMAS) emerges as the most suitable for SMEs.

ASMAS is highly effective in addressing the unique cybersecurity needs of SMEs, ensuring comprehensive coverage across various aspects of cybersecurity. Its high ease of implementation makes it accessible for SMEs lacking extensive cybersecurity expertise. Furthermore, ASMAS is cost-effective, an essential consideration for SMEs operating on limited budgets. The model's scalability ensures it can adapt and grow with the SME as its operations expand.

In conclusion, ASMAS provides a balanced and robust framework for SMEs to assess and enhance their cybersecurity posture, making it the most appropriate choice among the evaluated models. This framework will help SMEs manage their cybersecurity risks effectively, ensuring a secure and resilient digital environment.

References

[1] Fricker SA, Shojaifar A. Self-endorsed Cybersecurity Capability Improvement for SMEs. In: 28th Americas Conference on Information Systems. 2022.

[2] Yigit Ozkan B, Spruit M. Adaptable Security Maturity Assessment and Standardization for Digital SMEs. Journal of Computer Information Systems. 2023 Jul 4;63(4):965–87.

[3] Yusoff T, Wahab SA, Latiff ASA, Osman SIW, Zawawi NFM, Fazal SA. Sustainable Growth in SMEs: A Review from the Malaysian Perspective. Journal of Management and Sustainability. 2018 Aug 22;8(3):43.

[4] Arudchelvan M, Wignaraja G. SME Internationalization Through Global Value Chains and Free Trade Agreements: Malaysian Evidence. SSRN Electronic Journal. 2015.

[5] Kalesamy KM. A Conceptual Study: Technology Adoption among Malaysian Manufacturing SMEs for Corporate Sustainability in the Context of IR 4.0. The International Journal of Business & Management. 2021 Sep 30;9(9).

[6] Muhammad MZ, Char AK, Yasoa' MR bin, Hassan Z. Small and Medium Enterprises (SMEs) Competing in the Global Business Environment: A Case of Malaysia. International Business Research. 2009 Dec 15;3(1).

[7] Mokhtar S, Yusoff R, Ahmad A. KEY ELEMENTS OF MARKET ORIENTATION ON MALAYSIAN SMEs PERFORMANCE. International Journal of Business and Society. 2014;15(49).

[8] Musa H, Chinniah M. Malaysian SMEs Development: Future and Challenges on Going Green. Procedia Soc Behav Sci. 2016 Jun;224:254–62.

[9] Chelliah S, Sulaiman M, Mohd Yusoff Y. Internationalization and Performance: Small and Medium Enterprises (SMEs) in Malaysia. International Journal of Business and Management. 2010 May 18;5(6).

[10] Rachagan S, Satkunasingam E. Improving corporate governance of SMEs in emerging economies: a Malaysian experience. Journal of Enterprise Information Management. 2009 Jul 24;22(4):468–84.

[11] Wang J, Neil M, Fenton N. A Bayesian network approach for cybersecurity risk assessment implementing and extending the FAIR model. Comput Secur. 2020 Feb;89:101659.

[12] Rahman MH, Hamedani EY, Son YJ, Shafae M. Taxonomy-Driven Graph-Theoretic Framework for Manufacturing Cybersecurity Risk Modeling and Assessment. J Comput Inf Sci Eng. 2024 Jul 1;24(7).

[13] Rea-Guaman AM, San Feliu T, Calvo-Manzano JA, Sanchez-Garcia ID. Comparative Study of Cybersecurity Capability Maturity Models. In 2017. p. 100–13.

[14] Perozzo H, Zaghloul F, Ravarini A. CyberSecurity Readiness: A Model for SMEs based on the Socio-Technical Perspective. Complex Systems Informatics and Modeling Quarterly. 2022 Dec 30;(33):53–66.

[15] Ganin AA, Quach P, Panwar M, Collier ZA, Keisler JM, Marchese D, et al. Multicriteria Decision Framework for Cybersecurity Risk Assessment and Management. Risk Analysis. 2020 Jan 5;40(1):183–99.

[16] Armenia S, Angelini M, Nonino F, Palombi G, Schlitzer MF. A dynamic simulation approach to support the evaluation of cyber risks and security investments in SMEs. Decis Support Syst. 2021 Aug;147:113580.

[17] Rea-Guaman AM, Mejía J, San Feliu T, Calvo-Manzano JA. AVARCIBER: a framework for assessing cybersecurity risks. Cluster Comput. 2020 Sep 1;23(3):1827–43.

[18] Kitsios F, Chatzidimitriou E, Kamariotou M. The ISO/IEC 27001 Information Security Management Standard: How to Extract Value from Data in the IT Sector. Sustainability (Switzerland). 2023 Apr 1;15(7).

[19] Clarissa S, Wang G. Assessing Information Security Management Using ISO 27001:2013. Jurnal Indonesia Sosial Teknologi. 2023 Sep 25;4(9):1361–71.

[20] Antunes M, Maximiano M, Gomes R. A Client-Centered Information Security and Cybersecurity Auditing Framework. Applied Sciences (Switzerland). 2022 May 1;12(9).

[21] Prameet P. Roy. A High-Level Comparison between the NIST Cyber Security Framework and the ISO 27001 Information Security Standard. In: National Conference on Emerging Trends on Sustainable Technology and Engineering Applications (NCETSTEA). 2020.

[22] National Institute of Standards and Technology. The NIST Cybersecurity Framework (CSF) 2.0. 2024 Feb.

[23] National Institute of Standards and Technology. NIST Cybersecurity Framework 2.0: 2024 Feb. A. National Institute of Standards and Technology.

[24] Boletsis C, Halvorsrud R, Pickering J, Phillips S, Surridge M. Cybersecurity for SMEs: Introducing the Human Element into Socio-technical Cybersecurity Risk Assessment. In: Proceedings of the 16th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications. SCITEPRESS - Science and Technology Publications; 2021. p. 266–74.

[25] Malatji M, Von Solms S, Marnewick A. Socio-technical systems cybersecurity framework. Information & Computer Security. 2019 Jun 12;27(2):233–72.

[26] Neri M, Niccolini F, Martino L. Organizational cybersecurity readiness in the ICT sector: a quanti-qualitative assessment. Information & Computer Security. 2024 Jan 22;32(1):38–52.

[27] Van Haastrecht M, Yigit Ozkan B, Brinkhuis M, Spruit M. Respite for SMEs: A Systematic Review of Socio-Technical Cybersecurity Metrics. Applied Sciences. 2021 Jul 27;11(15):6909.

[28] Yigit Ozkan B, Spruit M. Adaptable Security Maturity Assessment and Standardization for Digital SMEs. Journal of Computer Information Systems. 2023 Jul 4;63(4):965–87.

[29] Yigit Ozkan, Bilge. Cybersecurity Maturity Assessment and Standardisation. Utrecht University; 2022.

[30] No WG, Vasarhelyi MA. Cybersecurity and Continuous Assurance. Journal of Emerging Technologies in Accounting. 2017 Mar 1;14(1):1–12.

[31] Gourisetti S, Mylrea M, Patangia H. Cybersecurity vulnerability mitigation framework through empirical paradigm: Enhanced prioritized gap analysis. Future Generation Computer Systems. 2020 Apr;105:410–31.

[32] Khan M, Gide E, Chaudhry G, Hasan J. A Cybersecurity Evaluation Model (CSEM) for Indian SMEs Working in a Virtual Team Environment. In: 2022 IEEE Asia-Pacific Conference on Computer Science and Data Engineering (CSDE). IEEE; 2022. p. 1–6.

[33] National Institute of Standards and Technology's. The NIST Cybersecurity Framework. In 2021. p. 171–92.

[34] Benz M, Chatterjee D. Calculated risk? A cybersecurity evaluation tool for SMEs. Bus Horiz. 2020 Jul;63(4):531–40.

[35] Dasso A, Funes A, Montejano G, Riesco D, Uzal R, Debnath N. Model-Based Evaluation of Cybersecurity Implementations. In 2016. p. 303–13.

[36] Sabillon R. The CyberSecurity Audit Model (CSAM). In 2021. p. 149–232.

Real-Time Robotic Force Control for Automation of Ultrasound Scanning

Ungku Muhammad Zuhairi Ungku Zakaria, Seri Mastura Mustaza*,
Mohd Hairi Mohd Zaman, Ashrani Aizzuddin Abd Rahni
Department of Electrical-Electronic and Systems Engineering-Faculty of Engineering and Built Environment,
Universiti Kebangsaan Malaysia, 43600 UKM Bangi, Selangor Darul Ehsan, Malaysia

***Abstract*—Ablation represents a minimally invasive option for liver cancer treatment, commonly guided by imaging techniques such as ultrasound. Recently, there has been a surge in interest in semi-automated or fully automated robotic image acquisition. Specifically, there is a continuing interest in automating medical ultrasound image acquisition due to ultrasound being widely used, having a lower cost, and being more portable than other imaging modalities. This study explores automated robot-assisted ultrasound imaging for liver ablation procedures. The study proposed utilizing a collaborative robot arm from Universal Robots (UR), which has gained popularity across various medical applications. A robotic real-time force control system was designed and demonstrated to regulate the contact force exerted by the robot on the surface of a torso phantom, ensuring optimal contact during ultrasound imaging. The Robot Operating System (ROS) and the UR Real-Time Data Exchange (UR-RTDE) interface were employed to control the robot. The findings indicate that the contact force can be maintained around a set desired value of 9N. However, deviations occur due to residual forces from acceleration when the probe is not in contact with the phantom. These results provide a foundation for further advancements in the automation of ultrasound scanning.**

***Keywords*—Real-time; position control; force control; ROS; collaborative robot; admittance control; kinematics; dynamic; tool center point; damping**

I. Introduction

Ablation is a well-accepted minimally invasive procedure for treating certain cancerous tumors, especially in the case of liver cancer [1]. In ablation, image guidance is necessary during the procedure, such as using ultrasound and computed tomography (CT) fluoroscopy [2]. Navigation of the ablation needle is often performed manually under the expert guidance of the clinician and is operator-dependent. However, in some cases where there are not enough trained specialists, the radiologist may have to perform both the ablation procedure and ultrasound scanning simultaneously [24].

There is an increasing interest in automating the acquisition of medical ultrasound images due to its widespread use, cost-effectiveness, and portability compared to other imaging modalities. One promising approach involves using robotic arms, initially focused on remote ultrasonography or tele-imaging. Recently, attention has shifted towards semi-automated and fully automated robotic image acquisition systems. Hennersperger et al. [3] developed an innovative robotic ultrasound system utilizing the LBR iiwa robot. This system can autonomously execute imaging trajectories based on start and end points pre-selected by a physician in pre-interventional images such as MRI or CT scans. The trajectory is calculated by identifying the nearest surface point within the MRI data and integrating it with the corresponding surface normal direction to ensure accurate imaging. Von Haxthausen et al. [4] advanced this field by developing a system that, after the initial manual placement of the ultrasound probe, uses convolutional neural networks (CNNs) to control the robot and follow peripheral arteries. This integration of CNNs allows the system to adapt to complex anatomical variations and dynamically adjust the probe position, enhancing the precision and efficacy of the ultrasound imaging process. These advancements illustrate the potential of robotic systems to improve the accuracy and efficiency of medical ultrasound procedures. Using a robot arm to assist in ultrasound scanning frees the physician to focus on the ablation procedure. The robot arm end effector or gripper holds and manipulates the ultrasound probe to perform scanning during the ablation procedure [5]. The robot arm must have a range of motion for a full abdomen examination. At the same time, the force applied by the robot arm should be optimal for obtaining the best ultrasound images without causing any injury to the patient [6] [7]. Furthermore, fine-tuning the ultrasound probe position is also a requirement for the robot-assisted system to mimic human radiologists [8] [9].

This paper presents the development of a robotic ultrasound scanning system designed to fully replace sonographers, addressing a critical challenge faced by radiologists who simultaneously perform both ablation and ultrasound scanning. The introduction of robotic arms for ultrasound scanning allows medical practitioners to concentrate exclusively on ablation procedures, thereby enhancing their ability to deliver precise and effective treatment. Medical practitioners who multitask and perform both ablation and ultrasound scanning have difficulty handling and positioning the ultrasound probe during ultrasound scanning while doing ablation simultaneously. The implementation of robotic automated ultrasound scanning not only ensures consistent precision in imaging but also alleviates the complexities of manually handling and positioning the ultrasound probe during dual-task procedures. In employing robot arm replacing the sonographer, the automated robotic ultrasound scanning can be implemented.

This paper introduces a novel approach with real-time force control integrated into the robotic system based on Universal Robots (UR10e). The force control is particularly tailored for dynamic environments, utilizing an admittance control strategy

that responds to feedback forces. This approach ensures compliant force control, crucial for detecting and mitigating excessive force applied during scanning. Additionally, the consideration of damping in the admittance control strategy further enhances force accuracy and prioritizes patient safety during the scanning process. For hardware, the robot arm UR10e was used for the apparatus of ultrasound scanning. The custom gripper was attached to UR10e's end-effector and used to grip the ultrasound probe. The mechanism and shape of the custom gripper also aid the robot in mimicking the exact angle and the point for scanning the human body. The real-time force control was done similar to the point of scanning on the body and the angle of the real sonographer during ultrasound scanning.

The subsequent sections are structured as follows: Section II presents a comparison of the literature review. Section III outlines the methodology for hardware and software setup, as well as their integration for seamless operation. In Section IV, the findings from conducting robotic automated ultrasound scanning are detailed, along with an analysis of the results. Lastly, Section V offers conclusions and discusses future avenues of work.

II. Literature Review

Robot-assisted ultrasound scanning requires advanced and precise robot arm control [6], especially during an ablation procedure. In a typical system, a single robot arm holds and moves the ultrasound probe. The robot arm trajectory is initially planned in the pre-operative phase based on acquired CT or MRI images [6]. Advanced multimodal visualization is required for pre-operative image-based planning in this case. Forward kinematics can be computed to plan the robot arm trajectory based on Denavit-Hartenberg (D-H) parameters [10][5]. Two control methods have been used for gaining precise and accurate control for physical human-robot interaction, i.e., impedance control and admittance control [26]. For Impedance Control, the input gained is displacement or velocity, and the output is force. In vice versa, the Admittance Control input is force, and the output is displacement or velocity [27]. For this paper, the admittance control was implemented to set the amount of force desired and receive the feedback force from the contact so the robot would maintain the force value based on the amount of force selected.

Given the critical importance of contact force in ultrasound scanning, several recent studies have focused on ensuring constant contact force. Ulrich et al. [11] reported a mean contact force of 9 N with a peak force of 37 N during obstetric scanning. Similarly, Ning et al. [12] found a mean contact force of 8.5 N. In contrast, Chen et al. [13] recommended a lower contact force of 4.5 N for scanning the human body using Universal Robots (UR). Wang et al. [14] suggested initiating contact with a force of 15 N, subsequently maintaining it at 10 N for optimal imaging of the human spine. Kaminski et al. [15] proposed a target contact force of 7 N to ensure stable data acquisition. Meanwhile, Mohamed [16] established a reference range of 3-5 N for ultrasound scanning to balance probe stability and patient comfort. These varying recommendations highlight the need for context-specific force adjustments to enhance the precision and reliability of robotic ultrasound systems across different medical applications.

The duration of ablation procedures, particularly radiofrequency ablation (RFA), is a critical factor in their effectiveness and safety. Ma et al. [17] reported that the mean procedure time for ultrasound (US)-guided RFA is approximately 27.54 ± 12.06 minutes. In a study by Zeno et al. [18], the median duration of RFA was 14 minutes, ranging from 10 to 19.5 minutes. Over a decade-long study, Shuangyan et al. [19] found the median ablation time to be 26 minutes, extending from 12 to 120 minutes. Additionally, Zuo-Feng Xu et al. [20] observed that synchronizing US and CT images required an average of 13.9 ± 11.9 minutes, ranging from 5 to 55 minutes. They also noted that each individual or overlapping ablation session lasted approximately 12 minutes [20] [21]. In contrast, Wei et al. [22] reported a consistent ablation duration of 10 minutes per session. These findings highlight the variability in ablation times, which can be influenced by factors such as the imaging modality used, the complexity of the case, and the specific protocols of different medical centers. Understanding these timeframes is essential for optimizing procedure planning and improving patient outcomes.

III. Methodology

The hardware setup employed in this study will be comprehensively described, emphasizing the specific equipment and configurations utilized throughout the research. This includes detailed information on the types of hardware components and how they are integrated into the overall system. Subsequently, the software setup will be described, focusing on the software interfaces and implementation of software tools, as well as the software system's architecture. Finally, the force control mechanism will be thoroughly examined. This involves an in-depth discussion of the feedback control systems implemented to regulate the contact force applied by the robotic arm. The methodology for applying and adjusting the force over a specified duration will be detailed, including the algorithms and sensors used to achieve precise control.

A. Hardware Setup

The collaborative robot arm UR10e model from Universal Robots (UR) was chosen as the manipulator in handling the ultrasound probe to automate ultrasound scanning. Fig. 1 illustrates the hardware setup of the UR robot.

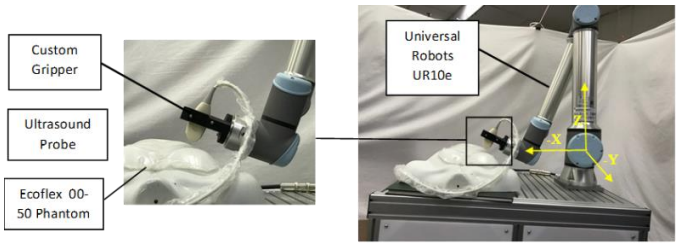


Fig. 1. (a) Hardware setup of ultrasound scanning (b) Hardware setup of the experiment with a coordinate system.

A partially rigid phantom was employed in this study to represent a human body. The anterior surface of the phantom is partially covered with cured Ecoflex 00-50 gel, a silicone gel that is commonly used for ultrasound phantoms [23]. As illustrated in Fig. 1(a), the robot is equipped with a custom gripper designed to handle an ultrasound probe. This probe

choice ensures that our experimental setup closely mirrors clinical practice, enhancing our findings' validity and applicability. The robot's coordinate system is as illustrated in Fig. 1(b). The origin coordinate is established at the base of the robot. The orientation was defined as Euler angles around each axis: θ = [θx θy θz]. Consequently, the robot's pose is represented as a 6-element vector comprising the 3D position, p = [x, y, z] (in meters), and the orientation, θ (in radians), forming the vector [x, y, z, θx, θy, θz]. The robot's end effector is equipped with a built-in force/torque sensor. This sensor is instrumental in measuring the external forces exerted by the phantom. The specifications of the built-in force/torque sensor are detailed in Table I, highlighting its range, precision, and accuracy.

TABLE I. SPECIFICATION OF BUILT-IN FORCE/TORQUE SENSOR IN UNIVERSAL ROBOT UR10E

Force sensing, tool flange/torque sensor	Force, x-y-z	Torque, x-y-z
Range	100.0N	10.0 Nm
Precision	5N	0.2 Nm
Accuracy	5N	0.5 Nm

B. Software Setup

Robot Operating System (ROS) is used in this work in a computer with Ubuntu operating system. The integration of UR dependencies facilitates the control of the UR robotic arm. This integration enables the robot to be managed by a computer, displaying the data acquired from the UR system. For real-time control, the UR Real-Time Data Exchange (UR-RTDE) libraries are being specifically utilized to harness the capabilities of the UR robot arm. Python was selected due to its robustness and extensive support for both ROS and UR-RTDE libraries to implement the ROS and UR-RTDE interface.

Fig. 2 illustrates the control architecture for the automation of ultrasound scanning. F_a represents the external contact force exerted to the probe, F_d represents the desired force, and F_e is the difference between them.

Admittance control was applied for this work. Robot admittance control models [25] [26] [27], were used to maintain the force value exerted during ultrasound scanning of the ablation process, mimicking the scan conducted by the sonographer. The control model is represented by the following equation:

$$F_e(\tau) = F_a(\tau) - F_d(\tau)$$

$$F_e(\tau) = \mathrm{M}(x\blacklozenge(\tau) - \ddot{x}_d(\tau)) + \Delta\,(x(\tau) - \dot{x}_d(\tau)) + \mathrm{K}\,(x(\tau)\,x_d(\tau)) \quad (1)$$

where, M is virtual mass, D is damping and K is stiffness coefficient. x_d denotes the desired position of the tip of the robot, while x is the position, $\dot{x}$ is the velocity and $\ddot{x}$ is the acceleration. The admittance control model is applied across each degree of freedom. However, our focus can only be directed towards the degrees of freedom along the x-axis- and z-axis.

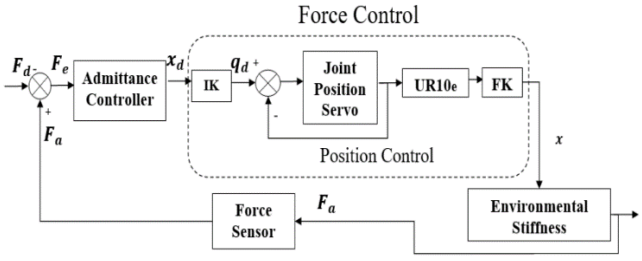


Fig. 2. Block diagram of force control system.

Based on Eq. (1), the admittance parameters can be adjusted to achieve specific dynamic characteristics of the robot system in the presence of external forces, i.e., from the patient or a static phantom. Phantom is only static. Therefore, M and K have been set as fixed values. So, D will be adjusted and controlled. For the UR force control interface, the range of the damping parameter can be set from 0 to 1, which the default value of the damping parameter is 0.005. The higher the value, the greater the deceleration during exerted force, and vice versa [28]. Deceleration is required to reduce force overshoot, thus reducing excessive force. Hence, an evaluation comparison of the damping factor was conducted to choose the best value. Four damping factors were compared: 0.005, 0.01, 0.02, and 0.05.

Referring to Fig. 2, the input force F_e, the difference between F_a and F_d, was applied to the admittance control, which resulted in moving to the displacement desired, x_d. Next, the inverse kinematics of the robot, IK, received input, x_d, and produce output, q_d to move the joint in position control closed loop. Afterward, forward kinematics of the robot, FK releases the output x to the environmental stiffness, which is the phantom resulting in the contact force, F_a. The force sensor will detect the value of F_a, and then send the value to be compared with the desired force input F_d, which resulting force input F_e.

To verify the efficiency of force control, including admittance control, a comparison between position control and overall force control was made to choose which control system is the best for full robotic automated ultrasound scanning. Three conditions were tested for evaluation, which are a slow movement with an acceleration of 0.1 ms^{-2}, fast movement with acceleration of 2 ms^{-2} and in a static position.

C. System Integration

Integration of hardware and software setup has been made for force control. A full operation of force can be achieved by satisfying a force overshoot capability involving position function, as shown in Fig. 3.

Referring to Fig. 3, the robot arm operates based on commands issued from the computer. The force is detected on the Tool Center Point (TCP) that has been set. To apply force, the position of the end-effector changes toward the direction of the force applied. Once the force sensor detects the external force, the measured force value is transmitted to the computer. When there is a deviation between the measured force and the desired force value, the computer processes this information and sends corrective commands to the robot to maintain the desired force during operation.

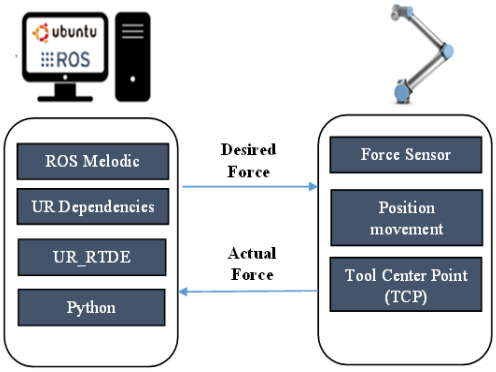


Fig. 3. Full setup of a system.

Using the UR-RTDE interface, the initial pose first needs to be set. The initial pose that has been set for this work is [-0.5, -0.1, 0.15, 0, 70/180 π, 0]. The pose can be seen in Fig. 4, along with the original and new Tool Point Center (TCP) of the robot, which is [0, -0.055, 0.058, 0, 0, 0] (green). This change is to ensure the accuracy of choosing a better coordinate for scanning. This new TCP has been set, and the pose coordinates were referred to at the end of the probe.

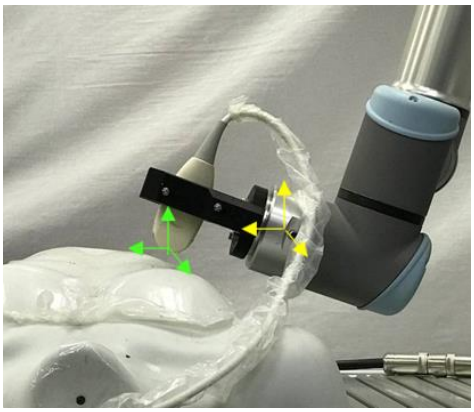


Fig. 4. Original TCP (Yellow) and New TCP (Green).

As illustrated in Fig. 5, the force task frame was set to [0, 0, 0, 0, 1.0472, 0] for the force to be exerted at 1.0472 radian pointing towards the phantom due to mimicking the liver position of the human.

When the configuration above is completed, the experiment has been done. The damping factor values have been evaluated, and the suitable values have been chosen. A comparison of position control and force control was also conducted to ensure the most compatible control system. After those evaluations have been executed, force control can be conducted. 9N was chosen based on the obstetric median force value [11], and 9N must be maintained when the force is exerted on the phantom. This work is set for 12 minutes during the exertion of force based on the time taken for ultrasound scanning conducted [20] [21]. The process flow of this experiment is shown in Fig. 6.

According to Fig. 6, the robot end-effector will first move to the initial position. Subsequently, a force is applied. If the detected force deviates from 9N, the robot will adjust to maintain a force of 9N. Upon achieving a stable force of 9N, the robot will operate for a duration of 12 minutes. Throughout this period, the force of 9N will be consistently maintained until the 12-minute mark is reached. Once the 12 minutes have elapsed, the robot end-effector will return to the initial position, concluding the process.



Fig. 5. Force task frame of the robot.

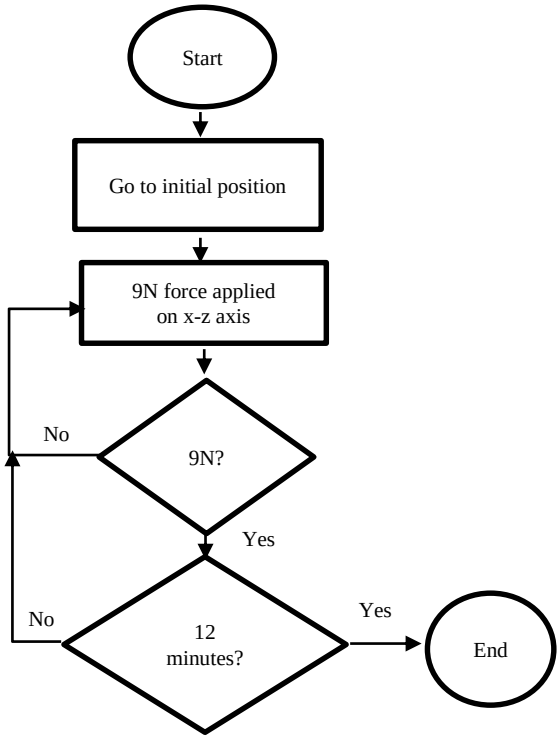


Fig. 6. Flowchart of the force control experiment.

IV. Results and Discussions

The experiment was conducted based on the setup that have been made. Firstly, the evaluation of damping factor values was conducted. Subsequently, a comparison of position and force control was made involving admittance control. Finally, force control of robotic automated ultrasound scanning was performed to appraise force retention over a certain amount of time.

The damping factor was evaluated to determine the most suitable overshoot for applying the force. Four damping values were selected, and the corresponding forces were measured. Additionally, a comparative analysis was conducted between position control and force control. Initially, the force was measured using position control alone. Subsequently, force control was implemented using the damping factor value selected from the prior comparison.

After evaluating the damping factor and the suitable control for applying force to the phantom, the force control was conducted for 12 minutes to evaluate the consistency of the force, which is 9N.

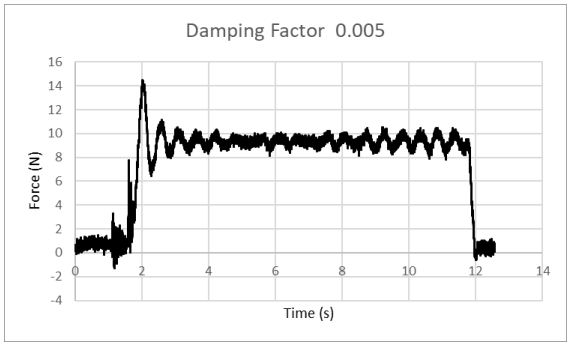
A. Evaluation of Damping Factor

Experimental testing was carried out to characterize the damping and overshoot of the contact forces between the ultrasound probe attached to the UR robot and the phantom to investigate the performance of the force control system. The robot exerted a 9N force onto the phantom in this experiment for 12 seconds. The damping factor was systematically varied between 0.005 and 0.05, and the resulting contact forces between the ultrasound probe and the phantom were recorded, as well as the overshoot that occurred upon contact. The results of this experiment are shown in Fig. 7(a)-(d).

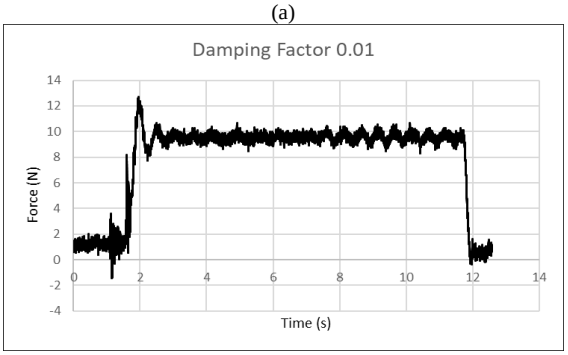
At a damping factor of 0.005, as illustrated in Fig. 7(a), the force overshoot initially peaks at 15N before stabilizing at 9N by the 4th second overshoot. When the damping factor is increased to 0.01, as shown in Fig. 7 (b), the force overshoot peaks between 10N and 15N and stabilizes at 9N by approximately the 3rd second overshoot. With a damping factor of 0.02, as depicted in Fig. 7 (c), the force overshoot shows stabilization at 9N shortly after the two seconds overshoot. Finally, at a damping factor of 0.05, shown in Fig. 7 (d), no force overshoot is observed, and 9N of force is applied just before the two seconds.

The percentage overshoot corresponding to each damping factor is tabulated in Table II. The percentage overshoot of the applied force at a damping factor of 0.005 is 61%. As the damping factor is increased to 0.01, the percentage overshoot decreases to 41%. Further increasing the damping factor to 0.02 results in a percentage overshoot of 26.55%. At a damping factor of 0.05, the percentage overshoot is completely eliminated, reaching 0%.

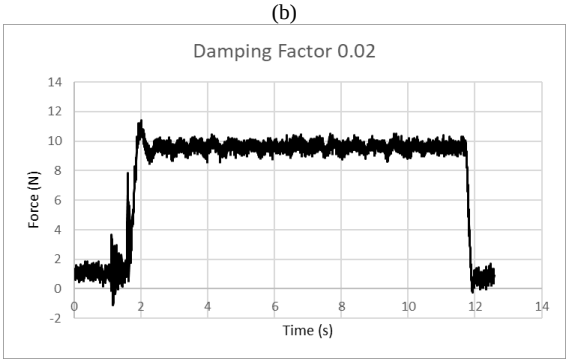
These observations demonstrate a clear inverse relationship between the damping factor and the percentage overshoot. Specifically, as the damping factor increases, the percentage overshoot of the applied force decreases correspondingly. This trend highlights the effectiveness of higher damping factors in achieving rapid stabilization of the force with minimal overshoot. Achieving zero overshoot at a damping factor of 0.05 indicates optimal damping, where the force control system can maintain the desired force precisely without any initial excess. This optimization is crucial for ensuring the reliability and accuracy of the ultrasound imaging process, as it minimizes the risk of excessive force application that could compromise image quality or cause damage to the phantom.



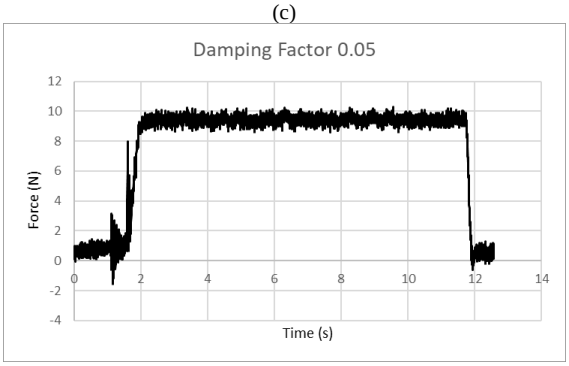
(a)



(b)



(c)



(d)

Fig. 7. The force value with damping factor (a) 0.005 (b) 0.01 (c) 0.02 (d) 0.05.

TABLE II. THE FORCE PERCENTAGE OVERSHOOT OF THE RESPECTIVE DAMPING FACTOR

Damping factor	Percentage overshoot (%)
0.005	61
0.01	41
0.02	26.55
0.05	0

B. Comparison of Positiom Control and Force Control

A comparative analysis between position control and force control was conducted to identify the most suitable method for applying force. Three conditions were examined to determine the optimal control strategy, which are 1) slow robot arm movement, 2) fast robot arm movement, and 3) static arm position. For the slow and fast movement condition, the ultrasound probe was controlled to move on and along the phantom stomach, starting from the liver towards the opposite side and returning to the liver position. For the slow movement, the robot arm drives the ultrasound probe at 0.1 ms^{-2} for 7 seconds, while for the fast movement, the acceleration is 2 ms^{-2} and the experiment runs for five seconds. The robot remained stationary at the liver area coordinates while contacting the phantom for the static condition. The comparison of position control and force control was plotted in Fig. 8 for slow movements, Fig. 9 for fast movement, and Fig. 10 for static position. 9N was set as the desired value to be achieved and was indicated with a horizontal line in Fig. 8, Fig. 9, and Fig. 10. The dashed line indicates position control, and the star line indicates force control.

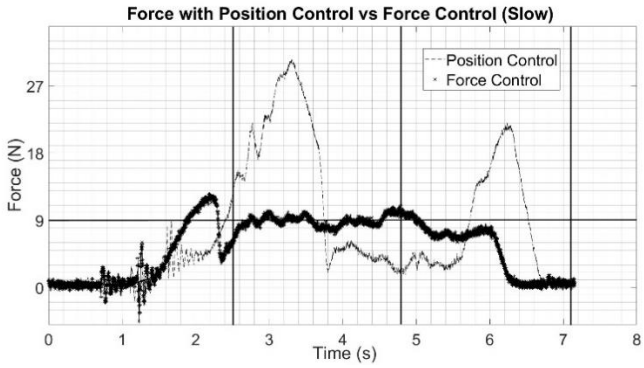


Fig. 8. Comparison of position control and force control in acceleration of 0.1 ms^{-2}.

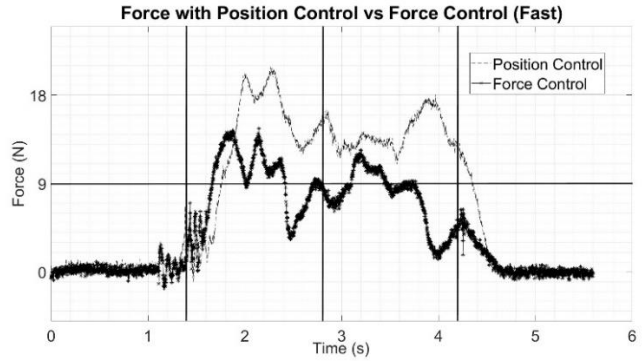


Fig. 9. Comparison of position control and force control in acceleration of 2 ms^{-2}.

Referring to Fig. 8, there are three lines in the graph that indicate the timeline of the position of the end-effector. The first line indicates the end-effector touching the phantom, which is 2.5 seconds; the second line is the movement from the liver to the stomach, which is 4.8 seconds; and the third line indicates the end-effector was returned to the liver part, which is 7.1 seconds. In the first line, the force in the position control is higher and has a huge difference compared to the force control below 9N. Upon reaching the second line, the force in the position control overshoots above 27N and drops drastically below 9N. Meanwhile, the force in the force control fluctuates close to 9N and tries to maintain the desired value, 9N. Finally, upon the march to the third line, the force in the position control overshoots again until above 18N and drops drastically. In contrast to force control, the force is very close to 9N, but eventually, the force value drops as it reaches the third line.

Based on Fig. 9, three vertical lines in the graph indicate the timeline of the end-effector's position. The first vertical line indicates the end-effector touching the phantom, which is at 1.4 seconds; the second line is the movement along the abdominal part, which is 2.8 seconds; and the third line indicates the return of the probe to the liver part, which is 4.2 seconds. In the first line, there were no significant differences in force between position control and force control. However, substantial differences exist in which the force in the control position was higher than the force in force control upon reaching the second line. Forces on both controls were turbulent. However, the force control is closer to 9N than the position control. Similar cases occurred from the second line to the third line in which the value of force in position control is greater than that in force control. After the graph's third line, the force value on both controls drops significantly.

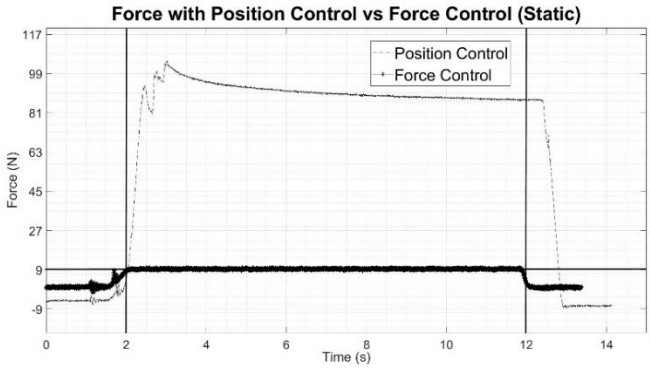


Fig. 10. Comparison of position control and force control in static position.

According to Fig. 10, two lines in the graph indicate the timeline of the end-effector's position. The first line indicates the end-effector touching the phantom, which is two seconds, and the second line is the position of the end-effector was lifted away from the phantom, which is 12 seconds. This graph illustrates the huge difference amount of force between position control and force control. On position control, the amount of force exceeds 100N and slightly decreases upon touching the phantom. However, it still overruns the desired amount of force, which is 9N. Contrarily, the force value in force control maintained 9N from the first line upon touching the phantom until the end effector was lifted away from the phantom. After the second line, the difference in force decrement was displayed. The value of force in force control dropped quickly compared to the force of the position control.

Based on the overall analysis, the amount of force in force control is closer to achieving the desired value than the position control, which is 9N. Even though there is a fluctuation of force value in Fig. 8 and Fig. 9 due to acceleration during movement and the curve surface of the phantom, the difference in the force

value from force control is lower compared to the position control. Meanwhile, in Fig. 10, the force control achieved and maintained the desired value of 9N compared to the position control, which exceeds the higher value.

It can be stated that using force control is more stable than using position control due to its closer adherence to the desired force value. When comparing acceleration, the force control method, with an acceleration of 0.1 ms^{-2}, proves to be the optimal control strategy. This is because lower acceleration results in lower force overshoot. The static condition experiment proves that the force control strategy employed is capable of maintaining the desired force value during its entire operation of 12 mins.

C. Operation of Force Control

Referring to the motion sequence described in Fig. 6, the resultant x-axis position, the resultant z-axis, and the magnitude of the force in the x-axis and z-axis, |Fxz| is shown in Fig. 11, Fig. 12, and Fig. 13 Firstly, the end effector will go to the initial position that has been set. Then, the robot will exert 9N on the phantom. The full operation of robotic force control was run for 12 minutes.

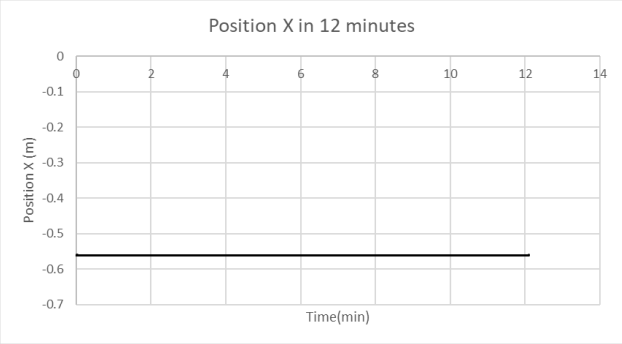


Fig. 11. x-axis position.

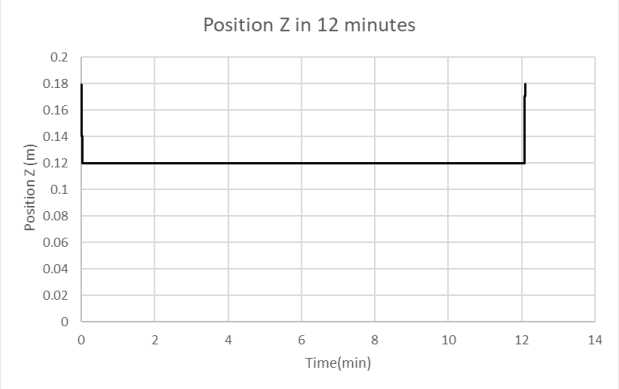


Fig. 12. z-axis position.

From Fig. 11 and Fig. 12, the x-axis and z-axis positions are at the initial position, which is pose x=-0.5m and pose z= 0.15m. Then, it suddenly changed due to the force direction towards the x and z direction. The force exerted can be seen in Fig. 10. From Fig. 13, no force is exerted when the end-effector is in the initial position. Then, the force of 9N is exerted upon touching the phantom. Next, the force was maintained at 9N along the operation. Even though there were light changes of force, the robot was able to maintain the force based on feedback until 12 minutes. This result, with a mean force of 9.503N and a standard deviation of 0.282N, corresponds to the force measured during abdominal scans in a previous study conducted by Ulrich and Struijk [11]. After 12 minutes, the x-axis and z-axis positions will return to the initial positions. Meanwhile, the force decreases drastically when the probe does not contact the phantom.

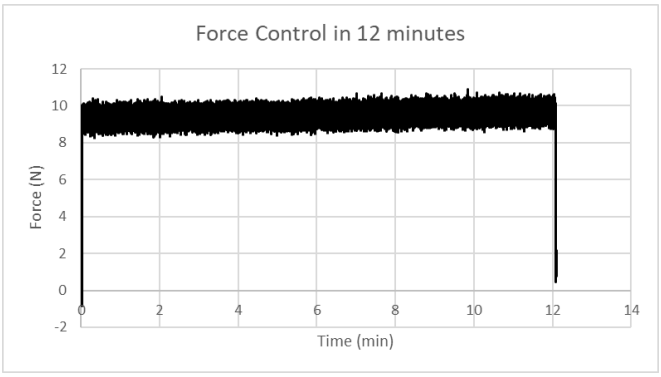


Fig. 13. Force along xz-axis,|Fxz|.

V. Conclusion

Ablation is a minimally invasive method for combating and eliminating liver cancer that involves crucial procedures such as ultrasound scanning. However, a single sonographer's simultaneous conduct of ablation can pose challenges. Hence, a proposal for robotic automated ultrasound scanning employing force control has emerged. Utilizing Universal Robot, the hardware and the corresponding software algorithm have been set up. Evaluation has been carried out, focusing on comparing damping factors and the efficacy of position control versus force control. It has been demonstrated that a damping value of 0.05 results in zero percentage overshoot. Furthermore, implementing force control ensures a more stable force than relying solely on position control. The full operation, lasting 12 minutes, maintains the desired force of 9N until its completion.

The real-time robotic force control was proven reliable in ultrasound scanning during liver ablation. The experiment above shows that contact force can be maintained around 9 N, albeit with deviation due to the residual force from acceleration when the probe is not in contact with the phantom. This shows that 9N can be applied on the phantom and, hence, on real ultrasound sonography. Furthermore, this paper can be used for further development of robot-assisted ultrasound scanning.

For future work, the cohesive gripper can be designed to be compatible with various types of ultrasound probes and robotic end-effectors, facilitating seamless attachment. The implementation of a compatible gripper is essential to maintain accuracy and efficiency during ultrasound scanning procedures.

Additionally, integrating image recognition technology with automation could enhance the system, allowing the robot to precisely identify the optimal location for performing ultrasound scans on the targeted area. This would eliminate the need for the patient to precisely position themselves against the surface where the robot is set to perform the ultrasound at a predetermined point.

ACKNOWLEDGMENT

The research is supported by research grant no. GUP-2021-024 and CRIM PIP-SH-2020-06 from Universiti Kebangsaan Malaysia.

REFERENCES

[1] D. Yang, L. Wang, Y. Xie, W. S. Levine, R. Davoodi, and Y. Li, "Optimization-based inverse kinematic analysis of an experimental minimally invasive robotic surgery system," in 2015 IEEE International Conference on Robotics and Biomimetics, IEEE-ROBIO 2015, Institute of Electrical and Electronics Engineers Inc., 2015, pp. 1427–1432. doi: 10.1109/ROBIO.2015.7418971.

[2] J. Schaible et al., "Primary efficacy of percutaneous microwave ablation of malignant liver tumors: comparison of stereotactic and conventional manual guidance," Sci Rep, vol. 10, no. 1, Dec. 2020, doi: 10.1038/s41598-020-75925-6.

[3] C. Hennersperger et al., "Towards MRI-Based Autonomous Robotic US Acquisitions: A First Feasibility Study," IEEE Trans Med Imaging, vol. 36, no. 2, pp. 538–548, 2017, doi: 10.1109/TMI.2016.2620723.

[4] F. Von Haxthausen, J. Hagenah, M. Kaschwich, M. Kleemann, V. García-Vázquez, and F. Ernst, "Robotized ultrasound imaging of the peripheral arteries - A phantom study," Current Directions in Biomedical Engineering, vol. 6, no. 1, pp. 1–4, 2020, doi: 10.1515/cdbme-2020-0033.

[5] X. Guan et al., "Study of a 6DOF robot assisted ultrasound scanning system and its simulated control handle," 2017 IEEE International Conference on Cybernetics and Intelligent Systems, CIS 2017 and IEEE Conference on Robotics, Automation and Mechatronics, RAM 2017 - Proceedings, vol. 2018-Janua, pp. 469–474, 2017, doi: 10.1109/ICCIS.2017.8274821.

[6] T. Y. Fang, H. K. Zhang, R. Finocchi, R. H. Taylor, and E. M. Boctor, "Force-assisted ultrasound imaging system through dual force sensing and admittance robot control," Int J Comput Assist Radiol Surg, vol. 12, no. 6, pp. 983–991, 2017, doi: 10.1007/s11548-017-1566-9.

[7] Y. Minami and M. Kudo, "Ultrasound fusion imaging of hepatocellular carcinoma: A review of current evidence," Digestive Diseases, vol. 32, no. 6, pp. 690–695, 2014, doi: 10.1159/000368001.

[8] N. Agarwal, A. K. Yadav, A. Gupta, and M. Felix Orlando, "Real-time needle tip localization in 2D ultrasound images using kalman filter," IEEE/ASME International Conference on Advanced Intelligent Mechatronics, AIM, vol. 2019-July, pp. 1008–1012, 2019, doi: 10.1109/AIM.2019.8868799.

[9] M. Renfrew, M. Griswold, and M. C. Çavuşoğlu, "Active localization and tracking of needle and target in robotic image-guided intervention systems," Auton Robots, vol. 42, no. 1, pp. 83–97, 2018, doi: 10.1007/s10514-017-9640-2.

[10] Y. Y. Cao et al., "Composite Configuration Interventional Therapy Robot for the Microwave Ablation of Liver Tumors," Chinese Journal of Mechanical Engineering (English Edition), vol. 30, no. 6, pp. 1416–1425, 2017, doi: 10.1007/s10033-017-0141-1.

[11] C. Ulrich and L. N. S. Andreasen Struijk, "Probe contact forces during obstetric ultrasound scans - A design parameter for robot-assisted ultrasound," Int J Ind Ergon, vol. 86, Nov. 2021, doi: 10.1016/j.ergon.2021.103224.

[12] G. Ning, J. Chen, X. Zhang, and H. Liao, "Force-guided autonomous robotic ultrasound scanning control method for soft uncertain environment," Int J Comput Assist Radiol Surg, vol. 16, no. 12, pp. 2189–2199, 2021, doi: 10.1007/s11548-021-02462-6.

[13] S. Chen, Z. Li, Y. Lin, F. Wang, and Q. Cao, "Automatic ultrasound scanning robotic system with optical waveguide-based force measurement," Int J Comput Assist Radiol Surg, vol. 16, no. 6, pp. 1015–1025, 2021, doi: 10.1007/s11548-021-02385-2.

[14] Y. Wang, T. Liu, X. Hu, K. Yang, Y. Zhu, and H. Jin, "Compliant Joint Based Robotic Ultrasound Scanning System for Imaging Human Spine," IEEE Robot Autom Lett, vol. 8, no. 9, pp. 5966–5973, Sep. 2023, doi: 10.1109/LRA.2023.3300592.

[15] J. T. Kaminski, K. Rafatzand, and H. Zhang, "Feasibility of robot-assisted ultrasound imaging with force feedback for assessment of thyroid diseases," SPIE-Intl Soc Optical Eng, Mar. 2020, p. 48. doi: 10.1117/12.2551118.

[16] M. E. Karar, "A Simulation Study of Adaptive Force Controller for Medical Robotic Liver Ultrasound Guidance," Arab J Sci Eng, vol. 43, no. 8, pp. 4229–4238, Aug. 2018, doi: 10.1007/s13369-017-2893-4.

[17] M. Luo et al., "Percutaneous ablation of liver metastases from colorectal cancer: a comparison between the outcomes of ultrasound guidance and CT guidance using propensity score matching," Ultrasonography, vol. 42, no. 1, pp. 54–64, Jan. 2023, doi: 10.14366/usg.21212.

[18] Z. Sparchez et al., "Percutaneous ultrasound guided radiofrequency and microwave ablation in the treatment of hepatic metastases. A monocentric initial experience.," Med Ultrason, vol. 21, no. 3, pp. 217–224, 2019, doi: 10.11152/mu-1957.

[19] S. Ou et al., "Radiofrequency ablation with systemic chemotherapy in the treatment of colorectal cancer liver metastasis: A 10-year single-center study," Cancer Manag Res, vol. 10, pp. 5227–5237, 2018, doi: 10.2147/CMAR.S170160.

[20] Z. F. Xu et al., "Percutaneous radiofrequency ablation of malignant liver tumors with ultrasound and CT fusion imaging guidance," Journal of Clinical Ultrasound, vol. 42, no. 6, pp. 321–330, 2014, doi: 10.1002/jcu.22141.

[21] J. Kang et al., "Comparative study of shear wave velocities using acoustic radiation force impulse technology in hepatocellular carcinoma: The extent of radiofrequency ablation," Gut Liver, vol. 6, no. 3, pp. 362–367, Jul. 2012, doi: 10.5009/gnl.2012.6.3.362.

[22] W. J. Fan, X. Li, L. Zhang, H. Jiang, and J. L. Zhang, "Comparison of microwave ablation and multipolar radiofrequency ablation in vivo using two internally cooled probes," American Journal of Roentgenology, vol. 198, no. 1, Jan. 2012, doi: 10.2214/AJR.11.6707.

[23] A. Cafarelli, P. Miloro, A. Verbeni, M. Carbone, and A. Menciassi, "Speed of sound in rubber-based materials for ultrasonic phantoms," J Ultrasound, vol. 19, no. 4, pp. 251–256, Dec. 2016, doi: 10.1007/s40477-016-0204-7.

[24] G. Laimer, P. Schullian, and R. Bale, "Stereotactic thermal ablation of liver tumors: 3d planning, multiple needle approach, and intraprocedural image fusion are the key to success—a narrative review," Biology, vol. 10, no. 7. MDPI AG, Jul. 01, 2021. doi: 10.3390/biology10070644.

[25] Y. Zhou et al., "Development of a Nursing Skill Training System Based on Manipulator Variable Admittance Control," in IEEE/ASME International Conference on Advanced Intelligent Mechatronics, AIM, Institute of Electrical and Electronics Engineers Inc., 2023, pp. 1226–1231. doi: 10.1109/AIM46323.2023.10196280.

[26] N. Hogan, "Impedance control: An approach to manipulation," in 1984 American control conference. IEEE, 1984, pp. 304–313.

[27] H. Maithani, J. A. C. Ramon and Y. Mezouar, "Predicting Human Intent for Cooperative Physical Human-Robot Interaction Tasks," 2019 IEEE 15th International Conference on Control and Automation (ICCA), Edinburgh, UK, 2019, pp. 1523-1528, doi: 10.1109/ICCA.2019.8899490.

[28] Gadringer, S., Gattringer, H., and Mueller, A.: Assessment of force control for surface finishing – an experimental comparison between Universal Robots UR10e and FerRobotics active contact flange, Mech. Sci., 13, 361–370, https://doi.org/10.5194/ms-13-361-2022, 2022.

Deep Learning Model for Enhancing Automated Recycling Machine with Incentive Mechanisms

Razali Tomari[1]*, Aeslina Abdul Kadir[2], Wan Nurshazwani Wan Zakaria[3], Dipankar Das[4], Muhamad Bakhtiar Azni[5]
Institute for Integrated Engineering (IIE), Universiti Tun Husein Onn Malaysia, Johor, Malaysia[1]
Faculty of Electrical and Electronic Engineering, Universiti Tun Husein Onn Malaysia, Johor, Malaysia[1]
Faculty of Civil Engineering and Built Environment, Universiti Tun Husein Onn Malaysia, Johor, Malaysia[2]
School of Mechanical Engineering, College of Engineering, Universiti Teknologi MARA, Selangor, Malaysia[3]
Department of Information and Communication Engineering, University of Rajshahi, Rajshahi, Bangladesh[4]
Department of Corporate Communication and Government Affair, SWM Environment Sdn. Bhd., Batu Pahat, Johor, Malaysia[5]

***Abstract*—Automated Recycling Machine (*ARM*) can be defined as an interactive tool to flourish recycling culture among community by providing incentive to the user that deposit the recyclable items. To enable this, the machine crucially needs a material validation module to identify the deposited recyclable items. Utilizing combination of sensors for such purpose is a tedious task and hence vision-based YOLO detection framework is proposed to identify three types of recyclable material which are aluminum can, PET bottle and tetra-pak Initially, the 14883 training samples and 937 validation samples were fed to the various YOLO variants for investigating an optimal model that can yield high accuracy and suitable for CPU usage during inference stage. Next the user interface is constructed to effectively communicate with the user when operating the *ARM* with easy-to-use graphical instruction. Eventually, the *ARM* body was designed and developed with durable material for usage in indoor and outdoor conditions. From series of experiments, it can be concluded that, the YOLOv8-m detection model well suit for the *ARM* material identification usage with 0.949 mAP@0.5:0.95 score and 0.997 F1 score. Field testing showed that the *ARM* effectively encourages recycling, evidenced by the significant number of recyclable items collected.**

***Keywords*—Recycling machine; You Only Look Once (YOLO); vision system; interactive recycling; deep learning**

I. Introduction

Recycling is a critical waste management strategy that involves collecting and processing discarded materials into new products. In Malaysia, the recycling initiative began in 1993 and unable to achieve its goals due to lack of significant awareness programs [1]. Consequently, the Ministry of Housing and Local Government launched the National Recycling Program on 2 December 2000, designating 11 November as National Recycling Day to highlight the importance of recycling. Despite numerous efforts, an environmental issue related to waste pollution are a significant concern in Malaysia since it was ranked eighth globally for mismanaged plastic waste [2]. Effective implementation of the 3R (reduce, reuse, recycle) strategies is essential to minimize waste sent to landfills and extend their lifespan. Typically, municipal solid waste in Malaysia consists predominantly of food waste, followed by recyclable materials like plastic, paper, glass, and aluminum cans [3]. Hence, adopting comprehensive recycling strategies is crucial to achieving the effective waste management targets.

Globally, many countries, including China [4], India [5], South Africa [6], Switzerland [7], and Malaysia [8], have adopted systematic waste management techniques. The integration of technology in waste management has gained significant attention, with innovations like the Reverse Vending Machine (RVM) demonstrating the potential to increase recycling activities. Technologies such as RFID [9-11], Wireless Sensor Networks [12], and VANET [13] have shown that they can provide comprehensive waste management solutions and encourage proper waste disposal.

Combining technology with a reward system has proven reliable and effective in supporting recycling initiatives as demonstrated by smart recycle bin [14] [15]. However, the reliance on multiple sensors for identifying recyclable materials prior point issuance requires tedious sensor arrangement, calibration and maintenance, making it unsuitable for long term use. Vision-based technology on the other hand, can recognize a broader range of recyclable items using vast number of features collected from the sample image. One of the works is from [16] in which they introduced ThrashNet dataset and use SIFT feature with SVM and CNN model of AlexNet -like architecture. Andrey et al. [17] developed a reverse vending machine with several CNN classification models, analyzing the effect of training by combining two different dataset clusters. On average their CNN model achieved over 85% accuracy. They further tested the module in real-world implementation by combining weigh sensor with the CNN for fraud detection [18].

Recently, YOLO become trends for recyclable waste detection due to its real-time object detection capabilities and high accuracy. YOLO architecture allows for the simultaneous detection and classification and hence make it ideal for dynamic environments where waste items vary in size, shape, and type. A study from [19] utilized YOLO with a depth camera to accurately determine the type and location of waste in 3D. Additionally, incorporating YOLO in ARMs aligns with broader trends in AI-driven waste management systems, which emphasize the integration of vision-based automated sorting for efficient recycling processes [20] [21]. The effectiveness of YOLO in these applications is further supported by research highlighting its adaptability and performance in real-world

scenarios, thereby facilitating the development of intelligent waste management solutions [22] [23]. Currently, there are many types of YOLO variants are available. Detail explanation from the first model which was introduced in 2015 up until the recent one can be found from study [24].

Numerous studies have focused on detection models for recycling waste identification across various applications. However, there is limited research on the real-world implementation of these detection models and their effectiveness when deployed with targeted stakeholders. This paper investigates a detection-based model for optimal implementation in recycling, incorporating a reward system to enhance user engagement. Additionally, it assesses the system feasibility through onsite testing. The paper is organized as follows: Section II details the methodology used throughout this project, Section III and Section IV presents the results and discussion respectively, and Section V concludes the project with final observations.

II. Models and Method

In this section, detailed explanation about model and architecture used throughout this project is explained. It comprises of four main subsections namely dataset preparation, detection model development, and user interface design and automated recycle system formation. To make the automated recycling machine (*ARM*) cost effective, its platform will run on Intel i7 CPU with 16GB RAM and 256GB SSD storage.

A. Datasets Preparation

In this project, samples of recycling images are obtained locally to ensure that the *ARM* model can differentiate the item effectively. The arrangement for capturing the sample images is shown in Fig. 1 in which the camera is positioned 21 cm above the ground with the samples placed 51cm from the camera. A total of 5,872 samples were collected and categorized into three groups, namely PET bottle, aluminum can and tetra-pak as depicted in Fig. 2. These collected images undergo a series of augmentation process which comprises of saturation and brightness adjustment, rotation, blurring and noise contamination and expand the data to 15,820. Once completed, the data were divided into training and validation sets with 14,883 samples for training and 937 samples for validation. Roboflow platform is used to augment, annotate and format the data to well suit the selected version of the detection model. Detailed sample distributions for each cluster can be seen in Table I.

B. Detection Model Development

For detection module, the main structure employed in this project is based on 'You Only Look Once' (YOLO) object detection. Basically, it starts with the forming of grid cells, followed by class prediction across scales, and eventually bounding box location estimation via regression process. The finer grid cell enables smaller target detection and anchor box makes it possible to detect an overlapping object with high accuracy. There are many variants of YOLO model starting from version v1 to version v10. In this project, three recent models which are YOLOv8, YOLOv9 and YOLOv10 were used. Basically, most of the YOLO architecture can be divided into three main components which are backbone, neck and head. The backbone is the part that is responsible for extracting an important feature from the input image. Once the features are extracted, the neck will combine the multi-scale information and channel it to the head section. The head generates predictions based on the extracted information from the backbone and the neck.

YOLOv8 [25] is the architecture evolution from its predecessor YOLOv5 [26] that was developed by Ultralytics. It introduces several enhancements over its predecessors by focusing on improving efficiency, accuracy and ease of use for object detection tasks. The backbone of YOLOv8 was derived and improved from CSPDarknet53 structure which consists of 53 convolutional layers. Apart from that, cross-stage partial connections that facilitate better information flow between layers were integrated. This optimization helps in capturing more detailed features from the input images. In the neck segment, YOLOv8 adopts combination of feature pyramid network (FPN)[27] and path aggregation network (PAN) [28]. This strategy contributes to improving the detection capability of targets at different scales. The head of YOLOv8 utilizes an anchor-free design, which simplifies the detection process and improves overall speed and accuracy. YOLOv8 models are available in five variants which distinguish between number or parameters and accuracy. However, since the aim of this project is focusing on implementation of the model on CPU, three pre-trained models that are well balance between complexity and accuracy were selected which are YOLOv8-n, YOLOv8-s, and YOLOv8-m.

TABLE I. Number of Image Dataset in Each Recycle Item Image Category

Image	Training	Validation
Aluminum Can	4005	248
PET Bottle	9054	569
Tetra-pak	1824	120
Total	14883	937

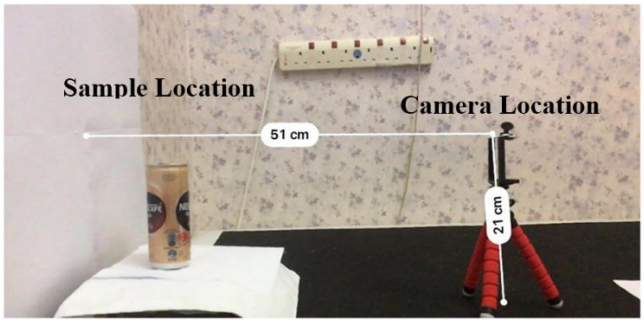


Fig. 1. Setup arrangement for dataset preparation.



Fig. 2. Sample of tetra-pak, PET bottle and aluminum can.

YOLOv9 [29] builds upon the foundations laid by YOLOv7 [30] structures and further refinements to enhance detection capabilities. One of the key features of YOLOv9 is the integration of Programmable Gradient Information (PGI) and the Generalized Efficient Layer Aggregation Network (GELAN). The combination has shown strong competitiveness among other models and maintains a balance between reducing the number of parameters and floating-point operations per second (FLOPs) while achieving superior performance compared to earlier versions. For this project a compact model of YOLOv9 known as YOLOv9-c is used as the pre-trained model for the recyclable waste detection assessment.

YOLOv10 [31] introduces significant architectural changes aimed at optimizing performance and reducing complexity. The backbone of YOLOv10 employs enhanced CSPNet which is lightweight design with a rank-guided block architecture that identifies and replaces stages with higher redundancy. This approach reduces the number of parameters and FLOPs, enhancing the model efficiency without compromising performance. The neck of YOLOv10 features improved feature pyramid networks using path aggregation network that better handle multi-scale feature fusion. The head of YOLOv10 is optimized to streamline the detection pipeline, ensuring maximum accuracy and minimal computational overhead. These enhancements make YOLOv10 particularly effective for real-time applications, where computational resources are often limited. YOLOv10 has six final architectures. To balance between speed and accuracy of the system, three YOLOv10 architectures which are YOLOv10-n, YOLOv10-s and YOLOv10-m are further analyzed for the automated recycle waste implementation.

TABLE II. SUMMARY OF YOLO PRE-TRAINED MODEL USE FOR TRAINING THE RECYCLABLE WASTE DATA

Model	Size	Parameters	FLOPs
YOLOv8-n	640	3.2M	8.7G
YOLOv8-s	640	11.2M	28.6G
YOLOv8-m	640	25.9M	78.9G
YOLOv9-c	640	25.3M	76.3G
YOLOv10-n	640	2.3M	6.7G
YOLOv10-s	640	7.2M	21.6G
YOLOv10-m	640	19.1M	59.1G

Table II summarizes the YOLO pre-trained models used to train the recyclable waste dataset. The models are selected based on the current state of the art detection models and have an adequate number of parameters to be used by CPU in the inference stage. From the assessment, a single model that yields a high performance with lower parameters will be selected for fulfilling the needs of an automated waste recycling module system.

C. User Interface Design

To construct the user interface a C# platform along with YoloDotNet v2.0 is utilized. Initially, the optimal weight of the YOLO model is exported to ONNX format to ensure compatibility with the .NET platform. ONNX is an open-source format that is compatible with different deep learning platforms and hence will be convenient for model sharing and deployment. The process using the system will consist of simple four steps as follows:

1) Initiation: Users push the start button in the welcoming screen to initiate the process.

2) Verification: The deep learning module verifies all deposited items.

3) Completion: To end the process, users press the finish button which prompts a screen requesting user information.

4) Point collection: If users wish to collect point, they can enter their phone number or leave it blank, if otherwise.

D. ARM Design and Setup

In this section, design and setup about the *ARM* body structure will be elaborate in detail. An illustration about the machine drawing is depicted in Fig. 3 where the whole structure design is shown on the left side and the chute architecture with camera placement is shown on the right side. The overall dimension of the machine is 1000mm width, 1000mm depth and 1800mm height. Regarding the materials, frame body structure is constructed using 25mm x 25mm hollow steel bars, and the casing is developed using 1mm galvanized steel sheets. Referring to the chute design, a high camera resolution is located 210mm above the chute base and tilted downwards at 45º angle. A 22-inch touch screen is utilized as the information display medium and serves as the user interface input.

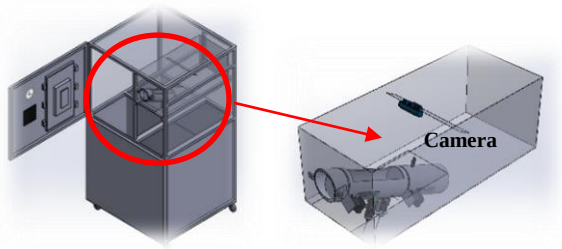


Fig. 3. ARM body design of the machine (left) and detail design of chute with location of camera (right).

III. RESULTS

This section offers a thorough examination of the performance of the detection model and how it was used to automate recycling with an integrated reward system. The three primary components of the analysis are evaluation of the YOLO models, user interface design, and automated recycling module implementation in real-world circumstances, and. The evaluation of the YOLO model involves assessing its mAP, F1 score, and precision in identifying different recyclable materials. The goal of the user interface design is to provide a platform that is simple to use and intuitive so that users can interact with the recycling system. In the end an implementation section describes how the user interface and detection model are combined to create a recycling module that works well in real-world situations.

Fig. 4 illustrates image samples of PET bottles, aluminum cans, and tetra-paks, which have undergone a series of augmentation processes to enhance the sample image for training session. These augmented images are used to train the detection model, ensuring it can accurately identify and classify different types of recyclable materials under various conditions.



Fig. 4. Sample of training batch images where 0 denoted for PET bottle, 1 for aluminum can, and 2 for Tetra Pak.

A. YOLO Model Assessment

This section will elaborate about the experiments conducted to deliberately discuss results obtained for the YOLO detection module framework. There are three versions of YOLO models tested which are YOLOv8, YOLOv9 and YOLOv10. The optimal model is selected based on its performance and medium level complexity to meet the requirement of CPU usage. The metrics considered for analysis are mAP@0.5, mAP@0.5:0.95, F1 Score, and Recall. Table III summarizes the findings.

From the table, among the investigated models, Yolov8-m stands out with a well balance between mAP@0.5:0.95 and F1 score. This model achieves a mAP@0.5 of 0.994, mAP@0.5:0.95 of 0.949, F1 score of 0.997, and a Recall of 0.998. The average mAP@0.5 across all models is 0.994, with a range from 0.994 to 0.995, indicating uniformly high precision across investigated models. However, the distinction in performance is more noticeable in the mAP@0.5:0.95 metric, which averages 0.944, highlighting the model capability to maintain precision across varying thresholds. Yolov8-m is shown to be outstanding for the mAP@0.5:0.95 score of 0.949 and hence capable in detecting objects at different scales more effectively than other models. Additionally, the high F1 score and recall values reflect its balanced performance in precision and recall, which is crucial for robust object detection. All in all, Yolov8-m metrics show a well balance performance and make it the ideal choice for the *ARM* module task.

TABLE III. YOLO MODEL PERFORMANCE TRAINED WITH THREE RECYCLE ITEMS IMAGE CATEGORY

Model	mAP@0.5	mAP@ 0.5:0.95	F1- Score	Recall
Yolov10-n	0.994	0.937	0.997	0.997
Yolov10-s	0.995	0.941	0.992	0.995
Yolov10-m	0.994	0.938	0.991	0.990
Yolov9-c	0.994	0.949	0.996	0.998
Yolov8-n	0.995	0.945	0.998	0.998
Yolov8-s	0.995	0.946	0.997	0.998
Yolov8-m	**0.994**	**0.949**	**0.997**	**0.998**

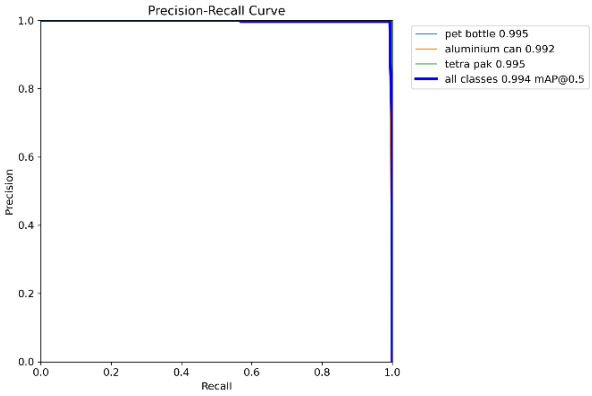


Fig. 5. Precision recall curve of the YOLOv8-m model.

To gain a more detailed insight into the performance of each object class for YOLOv8-m, the precision and recall curves of can be investigated and depicted in Fig. 5. These curves highlight the tradeoff between precision (exactness) and recall (completeness) across various thresholds. By analyzing these curves, one can determine the optimal threshold that maximizes both precision and recall. From the figure, it can be seen that the PET bottle and tetra-pak classes exhibit the highest confidence scores at 0.995, followed closely by the aluminum can at 0.992. On average, the overall performance across all classes achieves a mean average precision (mAP) score of 0.994 at 0.5 confidence threshold. This high mAP score indicates robust detection capabilities.

Apart from the recall and precision performance, Fig. 6 presents the training and validation results of a YOLOv8-m model. From the figure, the top row shows the training metrics, including box loss, classification loss, and DFL (distribution focal loss). All parameters decrease steadily, indicating an improved model performance over epochs. The precision and recall graphs demonstrate that the model achieves high accuracy and recall rates around 10 epochs during the training process and maintaining these high values constantly. The bottom row shows the validation metrics, which mirror the training results, with box loss, classification loss, and DFL loss that decreasing consistently. As for the mAP50 (mean Average Precision at 50% IoU) and mAP50-95 (mean Average Precision across IoUs from 50% to 95%) metrics for both training and validation indicate high performance with values reaching 1.0, showing that the model is effective for detecting the targeted recyclable materials.

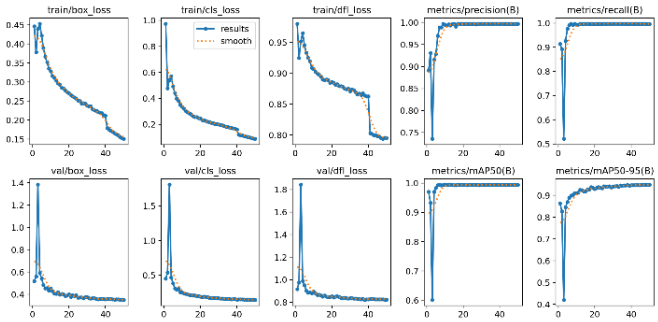


Fig. 6. Result for YOLOv8-m training.



Fig. 7. Sample of YOLOv8-m detection outcome.

Finally, Fig. 7 showcases the resultant images for aluminum can, PET bottle and tetra-pak items after running the inference/testing process with confidence label labelled for each item.

B. User Interface Design Assessment

In this section a snapshot of using the *ARM* system is elaborated and the steps are summarized in Fig. 8. Users can initiate the process by pressing the start button which is represented by the earth icon (see Fig. 8 (a)). This action activates the system and prepares it to receive recyclable materials. Next, recyclable items can be put one by one into the designated slot. As each item is placed inside, the YOLO detection module will scan and detect the material accordingly (Fig. 8(b)). If the system fails to detect an item, the user can remove it and try inserting it again until it is recognized, and if still fails then the material is not acceptable by the system. During this process, the user can monitor each inserted item on the system's screen to ensure proper detection. Once all your recyclable materials have been successfully inserted, the user can press the finish button (Fig. 8(c)) to conclude the recycling process. Eventually, the user can key in a valid phone number into the system to collect any points or rewards associated with the recycling effort. This phone number will typically be used for verification purposes and to credit any recycling points.

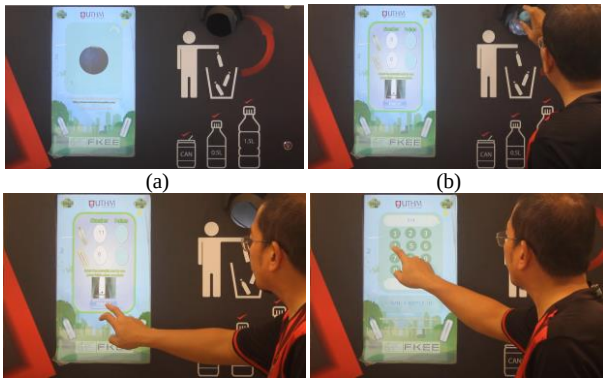


Fig. 8. User interface design consists of four main steps: (a) Start the recycling process. (b) Insert the recyclable items into the recycling system. (c) Finish the recycling process. (d) Enter phone number to claim the points for reward.

C. System Installation and Pilot Testing

Finally, to investigate the feasibility of the *ARM* implementation in actual conditions, the module (Fig. 9(a)) was installed at two different events. In the first case, the module was installed at SWM Pura Kencana (Fig. 9(b)), which is the office of a waste management company in southern Malaysia. In the second case, it was used during an event at a university (see Fig. 10). The installation at the SWM office aimed to study about public acceptance of the module and its effectiveness in encouraging recycling among the community and SWM staff. During the 30 days of implementation, the system managed to collect approximately 21 kilograms of PET bottles and one kilogram of aluminum cans, with snapshots of the collection shown in Fig. 9(c). These results indicate that PET bottles were the dominant item collected and aligning with the goal of reducing plastic waste that ends up in landfills. The system experienced only one instance of downtime due to computer overheating and to address this, a ventilation hole was added to ensure proper airflow within the module.

During implementation of the *ARM* at a showcase event (see Fig. 10), it successfully attracted users to recycle with some of the participants bringing many recyclable items. Primary school students were eager to use the system and deposit most of their plastic bottles into the bin. The one-day event demonstrated that the system could attract first-time users to fully utilize the system. However, to maintain their continuous interest, a different type of reward system should be implemented to sustain their engagement. For instance, integrating point-based reward systems or a lucky draw reward system. Additionally, educational workshops or interactive sessions that explain the importance of recycling and its impact on the environment can further produce a long-term participation.

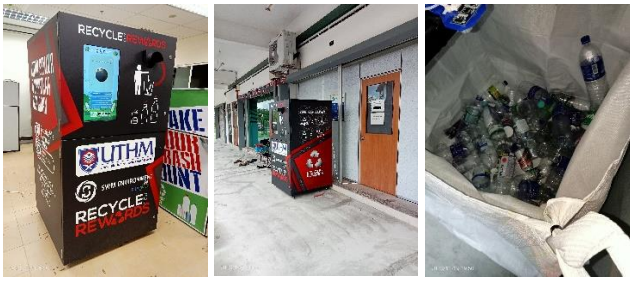


Fig. 9. Automated recycling system module. (a) Complete system module. (b) Installation onsite at SWM Pura Kencana Batu Pahat Johor. (c) Snapshot of collected recycling items.



Fig. 10. Pictures of automated recycle system implementation as a showcase in an event inside university.

IV. Discussion

The key takeaway from the previous analysis of the ARM detection model is that among YOLOv8, YOLOv9 and YOLOv10 architectures, the YOLOv8-m model shows exceptional performance on various object detection metrics for

our dataset and hence make it a top choice for waste classification tasks in the ARM module. YOLOv8-m strength in terms of precision (minimizing false positives) and recall (minimizing false negatives), is critical for robust detection in real-world scenarios. To improve these results even further, advanced data augmentation techniques could be used to diversify the training data and reduce overfitting to ensure even better generalization to new, unseen data. Furthermore, experimenting with fine-tuning the hyperparameters of YOLOv8-m could optimize its performance for specific drop-off types or conditions, leading to even higher precision and recognition rates. Apart from that, integrating ensemble methods combining YOLOv8-m with other models could leverage the strengths of the different architectures, increasing accuracy and robustness. Finally, ongoing monitoring and updating of the model with new data can help to maintain its relevance and effectiveness as waste classification requirements evolve.

Integrating YOLOv8-m model into ARM applications involves addressing software and hardware integration with seamless user interface. To maintain cost competitiveness, a CPU is used as the main framework for detecting the recyclable material types. It will also command the Arduino controller for diverting the waste into the respective bin. To illustrate the practicality of YOLOv8-m model for the ARM implementation, a 30-day continuous usage trial reveals the model capability for effectively detecting the inserted materials and delivering to the respective bin. However, the system encounter challenges in outdoor environments due to overexposure from light entering through the chute, which causes the camera to inaccurately capture the shape of object. Such an issue can be overcome by installing a movable lid over the chute to control the amount of light exposure.

V. Conclusion

In this project, a vision-based YOLO detection module of recyclable items was investigated for the automated recycling machine (*ARM*) module. In the assessment stage, three classes of recycle items which are PET bottle, aluminum can and tetra-pak are used during the training and validation stage with total number of 14,883 and 937 respectively. The optimal YOLO structure that can balance between speed and accuracy requirements is integrated in the *ARM* system.

For the detection model assessments, seven state-of-the art YOLO models namely YOLOv8-n, YOLOv8s, YOLOv8m, YOLOv9-c, YOLOv10-n, YOLOv10-s and YOLOv10-m were used to train the recycle dataset. From series of training and fine tuning, it can be concluded that Yolov8-m metrics show a well balance performance with mAP@0.5 score of 0.994, mAP@0.5:0.95 score of 0.949 and F1 value of 0.997. However, it is also worth mentioning that the performance of other investigated models is not too distinct with this model. Since this model yields the highest accuracy, it was further tested under for the ARM inference module that acquired data straight from life feed camera. For the user interface, the flow of using the ARM is easy to follow with simple operation steps with the option for the user to claim the reward points. As for the ARM pilot testing onsite, it shows a promising outcome to attract users to recycle and can be utilized as an interactive tool to ignite recycling culture among communities.

In future, the research can address the current system limitations which lack redemption (incentive) options. Implementing a redemption model aligned with container deposit legislation would enhance user participation. In the absence of such policies, support from entities through corporate social responsibility initiatives is essential for the success of the redemption stage. Apart from that, the detection model can be tested with various public recyclable datasets such as TACO dataset, ThrashNet dataset and WasteNet dataset to investigate the model reliability and scalability.

Acknowledgment

This research is supported by the Ministry of Higher Education (MOHE) through Prototype Development Research Grant Scheme (PRGS) (PRGS/2/2020WAB02/UTHM/02/2) and The Sumitomo Foundation through Grant for Japan-Related Research Project (FY2022).

References

[1] Malaysia Investment Development Authority, "Waste to Energy for a Sustainable Future", MIDA e-Newsletter, pp. 6-8, 2021.

[2] Jereme, I. A., Siwar, C., & Alam, M. M.,"Waste recycling in Malaysia: Transition from developing to developed country", Indian Journal of Education and Information Management, 4 (1), pp. 1- 14, 2015.

[3] Happonen. A, Santti. U & Auvinen.H., "Digitalization Boosted Recycling: Gamification as an Inspiration for Young Adults to do Enhanced Waste Sorting". AIP Conference Proceedings. 2233. 10.1063/5.0001547.

[4] Xinwen C., Martin S.-P., Mark Y.L. Wang, Markus A. Reuter, “Informal electronic waste recycling: A sector review with special focus on China,” Waste Management 31, 2011, 731–742.

[5] Kurian J. “Electronic Waste Management in India – Issue and Strategies”, Proceedings Sardinia 2007, Eleventh International Waste Management and Landfill Symposium S. Margherita di Pula, Cagliari, Italy; 1 - 5 October 2007.

[6] Greenredeems. (2013). Reward & Recycling: How incentives may have the answer for a ‘zero waste economy’ in UK. Whitepaper.

[7] Schindler, Helen R., Nico S., Vasco S., Jonathan C., Bastian W., and ArneA., "SMART TRASH: Study on RFID tags and the recycling industry". Santa Monica, CA: RAND Corporation, 2012.

[8] Zamali T., Mohd L. A., Abu O. M. T., “An Overview of Municipal Solid Wastes Generation In Malaysia”, Jurnal Teknologi, 51(F) Dis. 2009: 1–15.

[9] Abdouli, S. “RFID Application in Municipal Solid Waste Management System”, Int. J. Environ. Res., 3(3):447-45, 2009.

[10] Belal C. and Morshed U. C., “RFID-based Real-time Smart Waste Management System.” 2007 Australasian Telecommunication Networks and Applications Conference, December 2nd – 5th 2007.

[11] Waheed A. A., Faheem I., Tae M. Y., Chankil L.,” A USN based Automatic Waste Collection System”, ICACT 2012, February 19 – 22, 2012.

[12] Al-Sanjary O. I., Vasuthevan S., Omer H. K., Mohammed M. N. and Abdullah M. I., "An Intelligent Recycling Bin Using Wireless Sensor Network Technology," 2019 IEEE International Conference on Automatic Control and Intelligent Systems (I2CACIS).

[13] Saad M., Ahmad M. B., Asif M., Khan M. K., Mahmood T. and Mahmood M. T., "Blockchain-Enabled VANET for Smart Solid Waste Management," in IEEE Access, vol. 11, pp. 5679-5700, 2023.

[14] R. Tomari, A. A. Kadir, W.N.W Zakaria, M. F. Zakaria. M.H.A Wahab & M.H. Jabbar, "Development of Reverse Vending Machine (RVM) Framework for Implementation to a Standard Recycle Bin", Procedia Computer Science, vol. 105, pp. 75-80, 2017.

[15] R. Tomari, M. F. Zakaria , A. A. Kadir, W.N.W Zakaria, M.H.A Wahab ,"Empirical Framework of Reverse Vending Machine (RVM) with Material Identification Capability to Improve Recycling", Applied Mechanics and Materials, pp. 114-119, 2019.

[16] M. Yang and G. Thung, "Classification of trash for recyclability status", CS229 Project Report 2016, 2016.

[17] A. N. Kokoulin, A. I. Tur and A. A. Yuzhakov, "Convolutional neural networks application in plastic waste recognition and sorting," 2018 IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering (EIConRus), Moscow, 2018, pp. 1094-1098.

[18] A. N. Kokoulin and D. A. Kiryanov, "The Optical Subsystem for the Empty Containers Recognition and Sorting in a Reverse Vending Machine," 2019 4th International Conference on Smart and Sustainable Technologies (SpliTech), Split, Croatia, 2019, pp. 1-6.

[19] E. Kenan, B. Bünyamin & B. Barış. "YOLO -Based Waste Detection". Journal of Smart Systems Research,3. 120-127, 2022.

[20] Wen, S.; Yuan, Y.; Chen, J. A. "Vision Detection Scheme Based on Deep Learning in a Waste Plastics Sorting System".Appl. Sci. 2023, 13, 4634.

[21] Fang B, Yu J, Chen Z, Osman AI, Farghali M, Ihara I, Hamza EH, Rooney DW, Yap PS. "Artificial intelligence for waste management in smart cities: a review". Environ Chem Lett. 2023 May 9:1-31.

[22] E. Oluwatobi, M. Lisa, P. Pramod & S.Paul. "An On-Device Deep Learning Framework to Encourage the Recycling of Waste", Intelligent Systems and Applications, pp. 405-417, 2022.

[23] Munira S, Paul N, Alam MA, et al."Turning Trash into Treasure: Developing an Intelligent Bin for Plastic Bottle Recycling", Journal of Social Computing, 5(1): 1-14. 2024.

[24] Al Rabbani A.M. & Hussai M., "YOLOv1 to YOLOv10: A comprehensive review of YOLO variants and their application in the agricultural domain", arXiv preprint arXiv:2406.10139, 2024.

[25] Glenn J, Ayush C. and Jing Q, "Ultralytics YOLOv8", https://github.com/ultralytics/ultralytics, 2023.

[26] Glenn Jocher, "ultralytics/yolov5: v7.0 - YOLOv5 SOTA Realtime Instance Segmentation". Zenodo, Nov. 22, 2022. doi: 0.5281/zenodo.7347926.

[27] T. -Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan and S. Belongie, "Feature Pyramid Networks for Object Detection," 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 2017, pp. 936-944.

[28] S. Liu, L. Qi, H. Qin, J. Shi and J. Jia, "Path Aggregation Network for Instance Segmentation," 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 2018, pp. 8759-8768, d.

[29] C.-Y. Wang and H.-Y. M. Liao. YOLOv9: Learning what you want to learn using programmable gradient information. arXiv preprint arXiv:2402.13616v2, 2024.

[30] C. -Y. Wang, A. Bochkovskiy and H. -Y. M. Liao, "YOLOv7: Trainable Bag-of-Freebies Sets New State-of-the-Art for Real-Time Object Detectors," 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Vancouver, BC, Canada, 2023, pp. 7464-7475.

[31] A. Wang, H. Chen, L. Liu, Kai Chen, Z. Lin, J. Han, and G. Ding. Yolov10: Real-time end-to-end object detection. arXiv preprint arXiv:2405.14458, 2024.

A Hybrid of Extreme Learning Machine and Cellular Neural Network Segmentation in Mangrove Fruit Classification

Romi Fadillah Rahmat[1]*, Opim Salim Sitompul[2], Maya Silvi Lydia[3],
Fahmi[4], Shifani Adriani Ch[5], Pauzi Ibrahim Nainggolan[6], Riza Sulaiman[7]
Faculty of Computer Science and Information Technology, Universitas Sumatera Utara, Medan, Indonesia[1, 2, 3, 5, 6]
Faculty of Engineering, Universitas Sumatera Utara, Medan, Indonesia[4]
Institute of IR4.0, Universiti Kebangsaan Malaysia, Bangi, Malaysia[7]

***Abstract*—Mangroves are a collection of plants that inhabit the intertidal zone, namely the area between the lowest and highest points reached by the tide. Overall, mangroves provide a range of advantages, including the prevention of coastal erosion, the inhibition of seawater intrusion onto land leading to brackish groundwater, and serving as habitats and food sources for diverse animal species. In addition, many types of mangrove fruit have been used as sustenance for humans and as ingredients in processed food products. Mangrove fruit has a considerable variety of species, each characterized by distinct forms. At now, farmers and the general public rely only on visual observation to identify mangrove fruit species. Consequently, their ability to accurately detect the correct species is not guaranteed. In order to address this issue, this study employs digital image processing using the Extreme Learning Machine technique to facilitate the identification of various kinds and varieties of mangrove fruit by the general public and farmers. The study utilizes gray-scaling and Contrast Enhancement as image processing methods, while segmentation is performed by the use of the Cellular Neural Network approach. Following extensive testing in this study, it was determined that the used methodology effectively identified several species of mangrove fruit. The results yielded an accuracy rate of 94.11% for extracting shape, texture, and color elements, and accuracy rate of 99.63% for extracting texture and color features.**

***Keywords*—Mangroves; mangrove conservation; image processing; ecological informatics; cellular neural network; extreme learning machine**

I. Introduction

Mangroves are highly productive and distinctive ecosystems due to their exclusive ability to grow and endure in the transitional area between the ocean and land, where no other plants can live [1]. Indonesia has over 60% of the whole mangrove population in Southeast Asia, establishing it as the nation with the highest number of mangroves in the region. Indonesia is home to a minimum of 48 out of the 52 known mangrove species. Indonesia has the highest level of mangrove population diversity globally [2]. Mangroves are a kind of arboreal flora that often thrives in aquatic habitats with high salinity, such as marine and brackish waters. Mangroves often thrive in the intertidal zone, including the stretch of shoreline between the lowest ebb tide and the highest water level. Nevertheless, mangroves only thrive in regions characterized by tropical and sub-tropical climates [2].

Mangroves are very productive ecosystems that often serve as an economic resource, particularly for those living in coastal areas. Due to their significant impact on marine ecosystems, particularly fisheries, mangroves play a crucial role [2], [3]. Approximately 77% of all mangroves provide utilitarian value, with the most prevalent uses being medicinal, building, culinary (including vegetables, spices, and fruit), decorative, and as a source of fuel) [4]. Overall, mangroves provide a range of advantages, including the prevention of coastal erosion, the inhibition of seawater intrusion that may lead to the salinization of groundwater, and the provision of habitat and sustenance for several animal species. In addition, some mangrove fruit species have been used as a source of sustenance for humans and as ingredients in processed food products. The desirability of mangrove fruit as an economic resource is high among the community. Mangroves play a crucial role in the economics of coastal settlements due to the market value of goods derived from them and their contribution to sustaining coastal fisheries [2].

Several previous studies related to this research include those studied by Naskar & Bhattacharya [5] who examined the introduction of several types of fruit. There were six types of fruit studied, namely apples, bananas, lychees, oranges, pineapples and pomegranates. This research uses the Artificial Neural Network method, and achieves accuracy above 90%. However, the data used is still small. For the six types of fruit processed, they only used 150 data. The fruit features used are also only texture, color and shape. If other features are added to image processing, accuracy may be higher.

In research conducted by Ji et al. [6] carried out fruit classification by combining biogeography-based optimization (BBO) and feedforward neural network (FNN) methods. The data is processed more efficiently with four stages of preprocessing, then feature extraction (color, texture and shape) is carried out. In the third stage, unnecessary features are removed using component analysis. Finally, fruit classification was carried out using the biogeography-based optimization (BBO) and feedforward neural network (FNN) methods. This research used data from 1653 images from 18 fruit categories, but the accuracy was less high than research

*Corresponding Author.

using other methods, namely 89.11%. Similar with research in [5], if researchers add features and processed data, the accuracy achieved can be even higher.

The study conducted by Lu et al. [7] examines the categorization of fruits for industrial purposes. They used a Convolutional Neural Network consisting of six layers and utilized a dataset of 200 samples for each fruit species. The accuracy achieved by this study was 91.44%. Nevertheless, comprehending the weights of a Convolutional Neural Network using this approach is challenging. Enhancing the classification performance of Convolutional Neural Network requires the use of a more effective approach for parameter configuration. This would enable better handling of bigger datasets including diverse fruit kinds. Rouhi et al. [8] performed research on the categorization of benign and malignant breast cancers using the Cellular Neural Network segmentation approach. The accuracy gained in this study was very high, exceeding 95%. Nevertheless, the dataset used in this work remains quite small, consisting of just 93 photos depicting malignant tumors and 170 images depicting benign tumors. While the suggested approach is really strong, its varying outcomes on the DDSM and MIAS databases are seen as a shortcoming. An in-depth analysis of preprocessing procedures may be conducted to ensure consistent outcomes. The study done by Ghoneim, Muhammad, and Hossain [9] examines the categorization of cervical cancer via the use of Convolutional Neural Network and Extreme Learning Machine classification techniques. The acquired accuracy is quite high, namely 99.7% for 2 class classification and 97.2% for seven class classification. Nevertheless, the dataset used remains quite small, consisting of just 917 data points distributed over seven distinct categorization groups. The use of additional data is very likely to enhance accuracy.

The aforementioned studies conducted Rouhi et al. [8] and Ghoneim, Muhammad, and Hossain [9] demonstrate that Cellular Neural Network is effective for segmentation, while Extreme Learning Machine is effective for classification and image detection. These methods yield higher accuracy when accompanied by preprocessing techniques and a substantial amount of data.

Until now, there has not been much research on the classification of mangrove fruit, we have carried out several studies related to mangroves and machine learning. Among them is research related to the classification of mangrove shoots based on measurements of their morphology. This research uses a Multi-Layer Perceptron to carry out a basic classification of the types of mangroves shoots to be planted. The results obtained were not satisfactory, namely 91.9% using 3000 data [4]. Another research related to mangroves is the classification of maturity levels of mangrove fruit using deep convolutional neural networks. Where in this research the accuracy rate reached 99.1% [3]. From previous research conducted [3], it is deemed necessary to classify fruit types. This is because the level of ripeness and type of fruit are closely related. Apart from that, using the Extreme Learning Machine method with Cellular Neural Network also needs to be tested for its level of accuracy. Apart from that, mangrove fruit has quite a lot of species with different shapes. Currently, farmers or the public generally see and recognize mangrove fruit species only with the naked eye, so the introduction of mangrove fruit species does not necessarily match the species they are looking for. Therefore, an approach was taken to this problem in order to classify mangrove fruit species. Thus, we can form our research gap, which state that mangrove identification with machine learning technique is a must considering the concern of ecological urgency around the world. We also state that the novelty of our data and mangrove fruit classification system is a new field of ecological research in the world.

In this study, we imposed several constraints to restrict the extent of the issue from proliferating. The specific limitations of this problem include the focus on studying mangrove fruit species found in the North Sumatra region, specifically Rhizophora stylosa and Rhizophora mucronata. Additionally, only images with the file extensions .jpg or .jpeg will be processed, and the image size must be 720 × 84 pixels. The research provides several benefits. Firstly, it serves as a valuable reference for future studies on the classification of mangrove fruits. Secondly, it facilitates the identification of different mangrove species based on their fruits, benefiting both the general public and farmers. Thirdly, it aids in the recognition of specific mangrove fruit species, assisting individuals in replanting according to their desired species. Lastly, it simplifies the process for new farmers who wish to propagate mangrove plants by helping them identify different mangrove species, particularly through their fruits.

The content of paper is arranged as follows; in the Section II, we describe the methodology and algorithm used in this study. While in Section III we analyze the experimental results. Section IV is the summarization and future research of this paper.

II. METHODOLOGY

The research in this study included many stages: image acquisition, image preprocessing, image segmentation, feature extraction, and classification using the Extreme Learning Machine. The overall structure illustrating the sequential steps undertaken may be shown in Fig. 1.

A. Image Acquisition

This phase involves gathering visual data of mangrove fruit, which serves as the system's first input. Two sets of data are utilized to classify the mangrove fruit images: training data and testing data. Without utilizing a light, the image was obtained using a DSLR camera equipped with 18 megapixels of resolution. It is imperative that the captured image be situated on a backdrop of black. The information utilized in this investigation was gathered from numerous mangrove forests situated in North Sumatra. Three specific locations in the North Sumatra region provided the information utilized in this study: Percut Sei Tuan, Kampung Nipah Mangrove Beach, and Mangrove Ecotourism, Sicanang Belawan. Seven distinct occasions were utilized to collect the data, spaced at varying intervals.

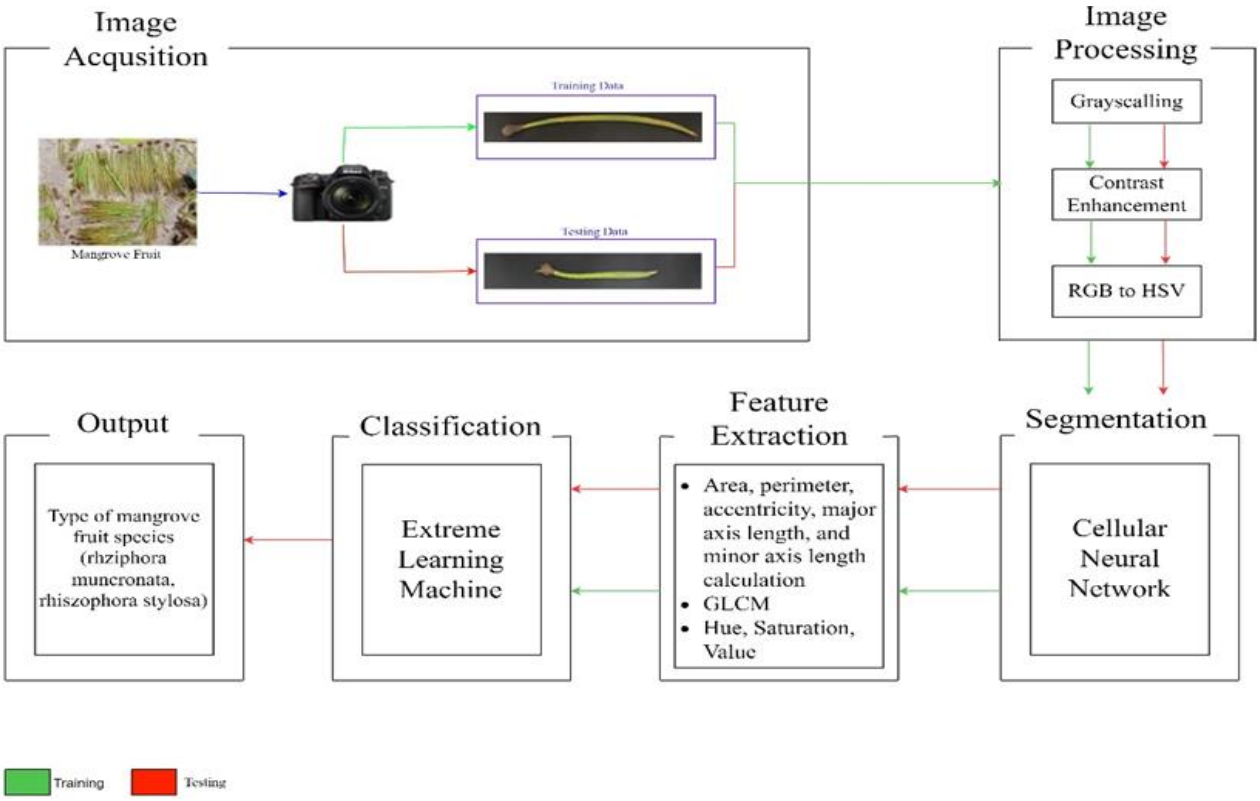


Fig. 1. General architecture.

The study utilizes an image with the file extension .JPG or .JPEG, measuring 720 × 84 pixels in size. Fig. 2 and Fig. 3 provide visual data of mangrove fruit.



Fig. 2. Image of Rhizophora mucronata.



Fig. 3. Image of mangrove fruit Rhizophora stylosa.

B. Image Processing

The phase of preprocessing involves altering of the acquired image to enhance its quality and facilitate further processing. The preprocessing phase includes three steps: grayscaling, contrast enhancement, and conversion of the RGB picture to HSV.

1) Grayscaling: Grayscaling is the first step in the preprocessing phase. At this point, the initial RGB picture is transformed into a grayscale image. The objective of grayscaling is to facilitate the identification of different mangrove fruit species in images. The grayscaling technique used in this study is the luminosity approach. The luminosity technique of grayscaling involves multiplying each value of the red (R), green (G), and blue (B) channels by a certain constant value that has been predetermined. The process of converting RGB photos to grayscale is performed on both training and testing images. The conversion of an RGB picture to a gray image using the luminosity approach can be achieved by using the Eq. (1).

$$GI = 0.2989R + 0.5870G + 0.1140B \quad (1)$$

Images of mangrove fruits after the grayscalling process can be seen in Fig. 4 and Fig. 5.



Fig. 4. Image of Rhizophora mucronate mangrove fruit after the grayscalling process.



Fig. 5. Image of mangrove fruit Rhizophora stylosa after the grayscalling process.

2) Contrast enhancement: Following the completion of the grayscaling procedure, the subsequent step involves enhancing the contrast. Contrast enhancement is performed to augment the contrast value of the captured picture of the mangrove fruit. This process aims to make the image more distinct, increase the visibility of its characteristics, and minimize the presence of noise [10]. The use of contrast enhancement in this study can be achieved utilizing the Eq. (2)

$$f_{0(x,y)} = G\left(f_{i(x,y)} - P\right) + P \quad (2)$$

The image of mangrove fruit that has had contrast enhancement applied can be seen in Fig. 6 and Fig. 7.

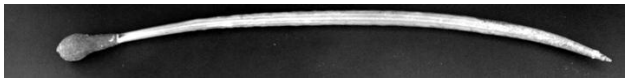


Fig. 6. Image of Rhizophora mucronate mangrove fruit after contrast enhancement.



Fig. 7. Image of mangrove fruit Rhizophora stylosa after contrast enhancement.

3) RGB to HSV conversion: At this stage, the RGB image of the mangrove fruit is converted from the RGB color model to the HSV color model. To get the HSV value, we must first know the RGB value of a pixel. In this research, RGB values were taken and conversion of RGB values to HSV. The results of converting RGB images of mangrove fruit to HSV images can be seen in Fig. 8 and Fig. 9.

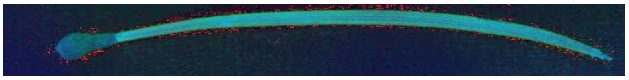


Fig. 8. HSV image of Rhizophora mucronata mangrove fruit.

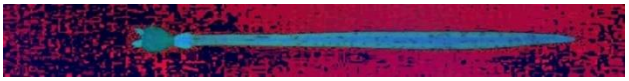


Fig. 9. HSV image of mangrove fruit Rhizophora stylosa.

C. Image Segmentation

Image segmentation follows the preprocessing phase. Image segmentation is used to distinguish the image of mangrove fruit from the surrounding background. This study uses the Cellular Neural Network approach for segmentation.

1) Cellular Neural Network: A Cellular Neural Network (CNN) is a network consisting of linked cells arranged in a MxN matrix, where M represents the number of rows and N represents the number of columns [11]. The architecture of a Cellular Neural Network has resemblance to that of cellular automata, whereby neighboring cells in the network may engage in direct interactions. Cells that lack direct connections may have indirect impact on neighboring cells via the propagation of continuous effects [12].

CNNs have shown efficacy in detecting malignancies and challenging objects, as evidenced by the studies conducted by Döhler et al. [13] and Abdullah et al. [14]. In order to do segmentation, three parameters need to be configured with appropriate values. The inputs consist of the template matrices A and B, as well as a bias value. The approach used in this study to ascertain the suitable values for template matrix A, template matrix B, and bias values was a process of trial and error. Following the completion of many trials, the values for template matrix A, template matrix B, and their respective bias values were acquired, such as:

$$\text{Template A} = \begin{bmatrix} 0 & 0 & 0 \\ 1 & 2 & 1 \\ 0 & 0 & 0 \end{bmatrix} \quad \text{Template B} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

$$\text{Bias} = 0.8$$

Cellular Neural/Nonlinear Networks are complex circuits that exhibit nonlinear dynamical behavior on a vast scale. These networks are comprised of cells, which are analog processing components that are locally coupled. They are characterized by study [12]:

$$\dot{x}_{ij}(t) = -x_{ij}(t) + \sum_{kl \in N_{\bar{r}}} A_{ij,kl} y_{kl}(t) + \sum_{kl \in N_{\bar{r}}} B_{ij,kl} u_{kl}(t) + I_{ij} \quad (3)$$

Where $x_{ij}(t)$ is the state, $y_{ij}(t)$ is the output, $u_{ij}(t)$ is the input, I_{ij} is the cell current, $A_{ij,kl}$ and $B_{ij,kl}$ are the parameters forming the feedback template A and the control template B, respectively, whereas $kl \in \overline{N_r}$ is a grid point in the neighborhood within the radius r of the cell ij [12], [15].

Assign template A as feedback, template B as control, and interpret bias as a threshold value. The input image to this Cellular Neural Network (CNN) is a preprocessed image. The segmentation process utilizing Cellular Neural Networks (CNN) involves the following steps [16]:

- Determine the image that will be used as input.
- Determine the values of template A, template B, and bias.
- Convert the input image to double class.
- If $t < 1$, calculate the output value in vector column form.
- Calculate the state value of the CNN
- Resize the output to its original size

The segmentation result image can be seen in Fig. 10.



Fig. 10. The segmentation result image.

D. Feature Extraction

Following the segmentation process, feature extraction is performed to acquire values from the features present in the picture of the mangrove fruit. The characteristics that are considered include form, texture, and color.

1) Area feature extraction calculation: In order to derive form characteristics from mangrove fruit, a method of feature extraction is conducted, using the object's shape as a basis, resulting in distinct feature values. The feature extraction method involves the computation of several measurements, such as area, perimeter, eccentricity, major axis length, and minor axis length, on the segmented picture. Subsequently, these values will serve as input parameters during the

classification phase. Here we describe area, perimeter and major minor axis length.

$$Area = Number\ of\ White\ Pixel \tag{4}$$

Thus, perimeter calculated by.

$$PR = \sum_{i=1}^{N-1} d_i = \sum_{i=1}^{N-1} |x_i - x_{i+1}| \tag{5}$$

And major axis length.

$$MaAL = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} \tag{6}$$

2) Gray Level Co-Occurrence Matrix (GLCM): Various methods may be used to extract texture characteristics from grayscale photos, including local binary patterns, gray level co-occurrence matrices, and scale invariant feature transform [17]—[19]. Once the form characteristic values of the image have been obtained, the subsequent step involves extracting the texture characteristic value using the Gray Level Co-Occurrence Matrix (GLCM) approach. The input for this GLCM procedure is a grayscale image. This study employs four Gray-Level Co-occurrence Matrix (GLCM) matrices to extract texture features from images of mangrove fruits. Specifically, the GLCM matrices are computed with a spatial distance of 1 and angles of 0°, 45°, 90°, and 135°. The extracted features include entropy, contrast, correlation, energy, and homogeneity. Here we describe the equation of every features.

Entropy is a stochastic metric that quantifies the amount of information related to characterization of image's texture. The entropy value is maximized when all pixel value for grey is equal or exhibits greatest randomness. If the entropy value increase it means the complexity of the image will increase too, which entropy also measure the level of complexity of grey distribution in the image.

$$ENT = \sum_{i,j=0}^{N-1} P_{ij}\,(-In\ P_{ij}) \tag{7}$$

Contrast is used to quantify the variations in intensity between neighboring pixels over the whole image. Contrast is indicative of both the sharpness of the picture and the intricacy of the texture's ridges. The more distinct the texture, the higher the contrast.

$$CNT = \sum_{i,j=0}^{N-1} P_{i,j}\,(i-j)^2 \tag{8}$$

Correlation is used to measure the degree of resemblance between the gray levels of an image in either the horizontal or vertical orientation. The size of the value represents the extent of local gray level correlation. The values of -1 and 1 represent the existence of negative and positive correlation in the image, respectively.

$$Corr = \sum_{i,j=0}^{N-1} \frac{(i-\mu)(j-\sigma)}{\sigma^2} \tag{9}$$

Energy also called as angular second moment, where the homogeneity of gray distribution and the level of texture. A finer texture correspondent to a higher component value in the GLCM matrix which will resulting smaller energy. The energy value will experience a substantial increase when the pixels exhibit a strong correlation, suggesting that the current texture exhibits reasonably consistent and predictable variations. The formula for energy is:

$$EGY = \sum_{i,j=0}^{N-1} (P_{ij})^2 \tag{10}$$

Homogeneity also called deficit moment, is a degree of texture clarity and regularity, refers to a big number indicating a clear image texture with great regularity.

$$HMY = \sum_{i,j=0}^{N-1} \frac{P(i,j)}{1+(i-j)^2} \tag{11}$$

Each value obtained from the process of feature extraction will generate four Grey-Level Co-occurrence Matrices (GLCMs), and each matrix will provide five distinct texture characteristics. Therefore, the total number of features produced is twenty. Nevertheless, the GLCM matrix just captures statistical values for each feature. Consequently, five characteristics will be retrieved and used as input for ELM.

3) Hue Saturation, Value (HSV): During this step, the process of extracting color features is performed, namely by capturing the hue, saturation, and value values. Once the RGB picture is transformed into an HSV image, the HSV value of the image will be extracted.

E. Extreme Learning Machine - Classification

Extreme Learning Machine was the approach that was used in this research project for the purpose of classification. Extreme Learning Machine, often known as ELM, is a learning algorithm that is based on the idea of single hidden-layer feedforward neural networks. (SLFN) [20]. When compared with other algorithms which also have the concept of single hidden-layer feedforward neural networks (SLFN), ELM provides better speed and performance [21], [22]. Because ELM can provide better speed and performance, it is believed that ELM can overcome the learning speed problem that has occurred in other feedforward neural networks algorithms [23], [24]. The learning speed of feedforward neural networks is generally much slower than other algorithms. This is because feedforward neural networks use a slow gradient-based learning algorithm in the training process, and all parameters in the neural networks are determined repeatedly [25].

Let's assume that there are N distinct sets of learning samples. $(x_i, y_i), 1 \le i \le N, X_i = [x_{i1}, x_{i2}, \dots, x_{in}]^T \in R^n, y = [y_{i1}, y_{i2}, \dots, y_{im}]^T \in R^m$, The expression for a neural network with a single hidden layer and L hidden nodes is as follows:

$$\sum_{i=1}^{L} \beta_i g(W_i \cdot X_i + b_i) = O_i, i = 1, \dots, N \quad (12)$$

$W_i = [w_{i1}, w_{i2}, , w_{in}]^T$ is the input weight vector, β_i is the output weight, $g(x)$ is the activation function, b_i is the bias of the ith hidden layer node, $W_i \cdot X_i$ is the inner product of W_i and $X_i, O_i = [o_{i1}, o_{i2}, \dots, o_{in}]^T$ denotes the network output value. The objective of the single hidden layer neural network is to reduce the inaccuracy in the output. Thus, it may be formulated as

$$\sum_{j=1}^{N} \|o_j - t_j\| \quad (13)$$

Exist β_i, W_i and b_i, and make

$$\sum_{i=1}^{L} \beta_i g(W_i \cdot X_j + b_i) = t_j, j = 1, \dots, N \quad (14)$$

It can be matrix expressed as

$$H\beta = Y \quad (15)$$

H is the vector that represents the output of the hidden layer of the neural network is denoted by the symbol β, while the output weight is denoted by the symbol T. The specific way of expressing it is Eq. (15).

$$H(W_1, \dots, W_L, b_1, \dots, b_L, X_i, \dots, X_L) = \begin{bmatrix} g(W_1 X_1 + b_1) & \cdots & g(W_L X_1 + b_L) \\ \vdots & \ddots & \vdots \\ g(W_1 X_N + b_1) & \cdots & g(W_L X_N + b_N) \end{bmatrix}_{NxL} \quad (16)$$

$$\beta = \begin{bmatrix} \beta_1^T \\ \vdots \\ \beta_L^T \end{bmatrix}_{Lxm}, T = \begin{bmatrix} T_1^T \\ \vdots \\ T_N^T \end{bmatrix}_{Nxm}$$

In order to train a single hidden layer neural network, we need to get $\widehat{W}_i$, $\hat{b}_i$, and $\hat{\beta}_i$, such that

$$\|H(\widehat{W}_i, \hat{b}_i)\, \hat{\beta}_i - T\| = min_{W,b,\beta} \|H(W_i, b_i)\beta_i - T\| \quad (17)$$

$i = 1, \dots, L$, that's equivalent to minimizing the loss function

$$E = \sum_{j=1}^{N} \left(\sum_{i=1}^{L} \beta_i g(W_i \cdot X_j + b_i) = t_i \right)^2 \quad (18)$$

In solving such issues, the Extreme Learning Machine (ELM) outperforms other conventional methods that rely on gradient descent algorithms. Unlike classic learning algorithms that require adjusting all parameters throughout each iteration phase, the algorithm employed by the ELM only requires to alter the input weights W_i and hidden bias b_i is determined, the output of the hidden layer matrix H is stochastically determined, resulting in significant savings in human labor and material resources. Consequently, the process of training a neural network with a single hidden layer may be simplified to solving a linear equation. $H\beta = T$ which Simultaneously, the output weight may also be ascertained.

$$\hat{\beta} = \bar{H}T \quad (19)$$

$\bar{H}$ is the generalized inverse of the matrix that Moore and Penrose have developed, and the norm of the solution that has been produced is the lowest and most unique.

The features obtained in the previous stage of feature extraction will be used in this stage as input. This stage is comprised of two distinct phases, namely the training phase and the testing phase. The training phase is conducted to determine the most optimum weights and biases that will subsequently be used during the testing phase. The testing step corresponds to the stage when data validation takes place. The ELM architecture for training has three layers: the input layer, hidden layer, and output layer. The input nodes consist of 13 nodes that include shape, GLCM, and HSV feature extraction. Additionally, there is an input node with 8 nodes that only use GLCM and HSV feature extraction. The following procedures will be executed.

1) Determining the number of nodes in the hidden layer: The values of the nodes in the artificial neural network, particularly in the hidden layer, play a crucial role in processing the training data. The hidden layer is responsible for computing the final outcomes of the artificial neural network. Insufficient hidden layers lead to underfitting, causing suboptimal performance of the accessible nodes in detecting signals from the input layer. An overfitting issue occurs when the hidden layer has an excessive number of node values. This happens because the network's processing capacity is too big to effectively handle the quantity of information acquired from the training data.

2) Determination of the activation function: The choice of the activation function is determined by the process of determining the number of hidden layers. The purpose of selecting this activation function is to use it for neurons throughout both the training and testing stages. The research utilizes the sigmoid function as its activation function. The sigmoid function is a kind of activation function often used in backpropagation algorithms, which aim to decrease computing time.

3) Training process: The first phase conducted by the extreme learning machine involves the training procedure for categorizing mangrove fruit. The training procedure involves configuring an artificial neural network to provide the desired output by using a certain dataset. The ultimate outcome of this procedure is an artificial neural network that has undergone training to provide outcomes that align with the data used in the training phase. The training procedure is conducted in three distinct phases, namely:

a) Randomization of input weight and bias: The first phase of the training procedure included assigning input weight and bias values. The number of neurons in the input layer is adapted to match the number of parameters obtained from the used dataset. The weight and bias values used in this investigation were assigned randomly.

b) Calculation of hidden layer output matrix: Following the completion of the phases that came before it, the calculation of the hidden layer output matrix will be carried out. The hidden layer output matrix is the result of the input processing that neurons in the hidden layer receive from neurons in the input layer. This processing is received by neurons and then processed. The processing has been carried out by using the activation function that was constructed in the phase that came before it, namely the sigmoid function.

c) Calculation of output weight: And last but not least, the computation of the output weight is carried out once the procedure of calculating the hidden layer output matrix has been finished. Following the completion of this procedure, a matrix is produced that is a representation of the weight of each neuron in the output layer.

4) Testing process: The artificial neural network generated after the training phase will undergo testing in the evaluation phase. The testing procedure was conducted to assess the efficacy of the extreme learning machine technique in categorizing mangrove fruit photos according to their species.

5) Output calculation: The calculation process is carried out using an artificial neural network that has been trained in the previous training process. The results of the calculation process will be a classification of mangrove fruit images based on species.

III. Result and Analysis

The tests conducted on mangrove fruit images included two forms of classification: one based on shape, texture, and color features, and another based solely on texture and color features. Here we describe feature extraction result from shape, texture and color feature extraction depicted in Table I, while feature extraction for texture and color depicted in Table II below.

Experimentation was conducted with varying quantities of concealed neurons, commencing with 30, 50, 80, and 100. Conducting experiments with varying numbers of hidden neurons seeks to determine the optimal number of hidden neurons required to accurately distinguish various types of mangrove fruit. The confusion matrix is shown in the Table III.

The accuracy presentation of the system with various hidden neurons is as follows:

$$Accuracy = \frac{Correct\ Testing\ Data}{Sum\ of\ all\ Testing\ Data} \times 100\%$$

TABLE I. Feature Extraction from Shape, Texture, and Color

Area	PR	ECC	MaAL	MiAL	ENT	CNT	Corr	EGY	HMY	H	S	V
609	306.16	0.9998	303.90	6.526	0.559	0.325	0.949	0.271	0.916	-0.1635	-0.0469	-0.1471
673	128.92	0.9998	353.11	6.491	0.558	0.317	0.952	0.274	0.920	-0.1687	-0.0468	-0.1448
697	128.28	0.999	380.65	6.298	0.556	0.321	0.952	0.268	0.919	-0.1685	-0.0464	-0.1461
608	126.69	0.999	387.81	6.132	0.556	0.322	0.951	0.275	0.919	-0.1709	-0.0463	-0.1461
613	127.45	0.999	368.62	6.215	0.556	0.330	0.950	0.270	0.915	-0.1702	-0.0460	-0.1484
…	…	…	…	…	…	…	…	…	…	…	…	…
668	3.556	0.9992	423.26	17.10	0.420	0.207	0.958	0.554	0.955	-0.0948	-0.0363	-0.0590

TABLE II. Feature Extraction from Texture and Color

ENT	CNT	Corr	EGY	HMY	H	S	V
0.5592	0.3254	0.9499	0.2717	0.9163	-0.1635	-0.0469	-0.1471
0.5581	0.3176	0.9520	0.2740	0.9201	-0.1687	-0.0468	-0.1448
0.5565	0.3211	0.9521	0.2684	0.9198	-0.1685	-0.0464	-0.1461
0.5562	0.3220	0.9516	0.2755	0.9195	-0.1709	-0.0463	-0.1461
0.5564	0.3304	0.9502	0.2705	0.9154	-0.1702	-0.0460	-0.1484
…	…	…	…	…	…	…	…
0.4207	0.2071	0.9584	0.5543	0.9555	-0.0948	-0.0363	-0.0590

TABLE III. Confusion Matrix (Shape, Texture, and Color Extraction Feature)

	Shape, Texture, and Color Extraction Features		Texture, and Color Extraction Features		Total
	Rhizopora mucronate	*Rhizopora stylosa*	*Rhizopora mucrotae*	*Rhizopora stylosa*	
Rhizopora mucronate	61	11	72	0	72
Rhizopora stylosa	5	195	1	199	200
Total	66	206	73	199	272

TABLE IV. SYSTEM ACCURACY

Shape, Texture, and Color Extraction Features				Texture, and Color Extraction Features			
Correct amount of test data	*Number of incorrect test data*	*Number of hidden neurons*	*Accuracy*	*Correct amount of test data*	*Number of incorrect test data*	*Number of hidden neurons*	*Accuracy*
253	19	30	93,01%	264	8	30	97,05%
253	19	50	93,01%	268	4	50	98,52%
254	19	80	93,38%	271	1	80	99,63%
256	16	100	94,11%	271	1	100	99,63%

The accuracy presentation of the system with various hidden neurons can be seen in Table IV. From the calculations above, demonstrate that the experiment that achieved a level of accuracy of 94.11% for shape, texture, and color extraction features using 100 hidden neurons. Additionally, the accuracy reached 99.63% for texture and color extraction features with the same number of hidden neurons. While these results indicate a relatively high level of accuracy, they are not perfect. The accuracy achieved in classification using form, texture, and color characteristics is inferior than that achieved using just texture and color features. When observing with the naked eye, one of the most noticeable distinctions between the Rhizophora mucronata and Rhizophora stylosa species is the form characteristics, particularly the size of their mangrove fruit. Rhizophora mucronata often exhibits greater dimensions in comparison to Rhizophora stylosa. The decrease in system accuracy observed when incorporating additional shape features is likely attributed to the limited variation in data from the Rhizophora mucronata species. Consequently, the classification performance of ELM diminishes slightly compared to when only texture and color feature extraction are employed.

The imperfection of this system may be attributed to many variables, primarily the influence of ambient light that results in incorrect image segmentation. Consequently, the accuracy of the Extreme Learning Machine in identifying mangrove fruit species is compromised. Fig. 11 displays an instance of an image that was not successfully categorized. The image in Fig. 11 was not successfully categorized due to the deteriorated state of the mangrove fruit, which was found to be in worse condition compared to the fruit of other Rhizophora mucronata species. Rhizophora mucronata often has a more vibrant green coloration in its fruit and possesses a more uniform appearance, lacking the intricate textural features shown in Fig. 11.



Fig. 11. Example of image that failed to classified.

In our previous study [3], we implemented Deep Convolutional Neural Network in order to classify mangrove fruit ripeness. While it is not relevant to be compared in this study, we see a chance to implement deep-learning based technique in mangrove fruit classification as our foresight research. There are several deep learning techniques that can implement such as deep neural network, Faster-RCNN, AlexNet, RestNet, YOLO or even ensemble learning.

IV. CONCLUSION

The study successfully used the Extreme Learning Machine (ELM) approach to accurately identify mangrove fruit based on species. The system achieved an accuracy of 94.11% when shape, texture, and color features were extracted, and an accuracy of 99.63% when just texture and color features were extracted. Nonetheless, the act of capturing images with a DSLR camera significantly impacts the resulting image. Inadequate suitability of the captured imagine and insufficient brightness of the surrounding light will result in noise and a dark image, so preventing the image features from being prominent. This will interfere with the process of segmenting and classifying. The determination of the value in the segmentation process, particularly the bias value, is very crucial in achieving effective image segmentation. Excessive bias values will result in the segmentation of the image's background as well. Conversely, if the bias value is too low, the segmentation of the mangrove fruit in the image will not be complete.

Several limitations and suggestion have been arising during our study such as, the mobility of our system can be improved using deep learning based mobile application includes its accuracy-concern. While other deep learning methodology can be implemented for our future research in classifying mangrove species. Thus, other limitation including mangrove zone plantation decision system is still open for research which will determine which area is suitable for certain kind of mangrove to increase the life chance of mangrove.

CONFLICTS OF INTEREST

The authors declare no conflict of interest.

AUTHORS' CONTRIBUTION

Conceptualization, RFR and SAC; methodology, RFR; software, SAC; validation, OSL, MSL, and F; formal analysis, RS; investigation, PIN; resources, RS; data curation, F; writing—original draft preparation, RFR; writing—review and editing, OSS, MSL, SAC, RFR; visualization, PIN; supervision, OSS, RS; project administration, OSL, MSL, F, RS; funding acquisition, RFR.

ACKNOWLEDGMENT

This work is supported by Universitas Sumatera Utara WCU Grant 2022 under Research Grant Number 29/UN5.2.3.1/PPM/KP-WCU/2022. Special thanks to Head of Research Program and Rector of Universitas Sumatera Utara for providing us with special grant for this research.

REFERENCES

[1] A. M. Ellison, "Exploring Mangroves Tropical Mangrove Ecosystems A. I. Robertson D. M. Alongi," Bioscience, vol. 44, no. 3, pp. 187–188, Mar. 1994, doi: https://doi.org/10.2307/1312261.

[2] W. Giesen, S. Wulffraat, M. Zieren, and L. Scholten, "Mangrove Guidebook for Southeast Asia," Bangkok, 2006.

[3] S. Faza, R. F. Rahmat, M. Husna, R. Anugrahwaty, R. P. Ahmad, and Onrizal, "Mangrove Fruit Ripeness Classification using Deep Convolutional Neural Network," J Theor Appl Inf Technol, vol. 101, no. 3, pp. 1095–1105, Feb. 2023.

[4] S. Faza, R. F. Rahmat, A. Padly Sembiring, M. Husna, A. S. Chan, and R. Anugrahwaty, "Classification of Mangrove Sprouts Based on Its Morphologhical Measurement," in 2021 International Conference on Data Science, Artificial Intelligence, and Business Analytics (DATABIA), IEEE, Nov. 2021, pp. 97–100. doi: 10.1109/DATABIA53375.2021.9650109.

[5] S. Naskar and T. Bhattacharya, "A Novel Fruit Recognition Technique using Multiple Features and Artificial Neural Network," Int J Comput Appl, vol. 116, no. 20, pp. 23–28, Apr. 2015, doi: 10.5120/20453-2808.

[6] Y. Zhang, P. Phillips, S. Wang, G. Ji, J. Yang, and J. Wu, "Fruit classification by biogeography - based optimization and feedforward neural network," Expert Syst, vol. 33, no. 3, pp. 239‑253, Jun. 2016, doi: 10.1111/exsy.12146.

[7] S. Lu, Z. Lu, S. Aok, and L. Graham, "Fruit Classification Based on Six Layer Convolutional Neural Network," in 2018 IEEE 23rd International Conference on Digital Signal Processing (DSP), IEEE, Nov. 2018, pp. 1–5. doi: 10.1109/ICDSP.2018.8631562.

[8] R. Rouhi, M. Jafari, S. Kasaei, and P. Keshavarzian, "Benign and malignant breast tumors classification based on region growing and CNN segmentation," Expert Syst Appl, vol. 42, no. 3, pp. 990–1002, Feb. 2015, doi: 10.1016/j.eswa.2014.09.020.

[9] A. Ghoneim, G. Muhammad, and M. S. Hossain, "Cervical cancer classification using convolutional neural networks and extreme learning machines," Future Generation Computer Systems, vol. 102, pp. 643–649, Jan. 2020, doi: 10.1016/j.future.2019.09.015.

[10] Z. Gu, C. Chen, and D. Zhang, "A Low-Light Image Enhancement Method Based on Image Degradation Model and Pure Pixel Ratio Prior," Math Probl Eng, vol. 2018, pp. 1–19, Jul. 2018, doi: https://doi.org/10.1155/2018/8178109.

[11] C. Botoca, "Some Aspects of Cellular Neural Networks andTheir Applications," 2003.

[12] L. O. Chua and L. Yang, "Cellular neural networks: theory," IEEE Trans Circuits Syst, vol. 35, no. 10, pp. 1257–1272, Oct. 1988, doi: 10.1109/31.7600.

[13] F. Döhler, F. Mormann, B. Weber, C. E. Elger, and K. Lehnertz, "A cellular neural network based method for classification of magnetic resonance images: Towards an automated detection of hippocampal sclerosis," J Neurosci Methods, vol. 170, no. 2, pp. 324–331, May 2008, doi: 10.1016/j.jneumeth.2008.01.002.

[14] A. Azamimi Abdullah, A. F. Dickson Giong, and N. A. Hanin Zahri, "Cervical cancer detection method using an improved cellular neural network (CNN) algorithm," Indonesian Journal of Electrical Engineering and Computer Science, vol. 14, no. 1, p. 210, Apr. 2019, doi: 10.11591/ijeecs.v14.i1.pp210-218.

[15] P. Vecchio and G. Grassi, "Cellular neural networks: Implementation of a segmentation algorithm on a Bio-inspired hardware processor," in 2012 IEEE 55th International Midwest Symposium on Circuits and Systems (MWSCAS), IEEE, Aug. 2012, pp. 81–84. doi: 10.1109/MWSCAS.2012.6291962.

[16] M. Guo and D. Feng, "Improved Method for Image Segmentation Based on Cellular Neural Network," 2012, pp. 671–678. doi: 10.1007/978-1-4471-2467-2_79.

[17] S. Fekri Ershad, "A Review on Image Texture Analysis Methods," vol. 1, Dec. 2018.

[18] S. Gao, "Gray level co-occurrence matrix and extreme learning machine for Alzheimer's disease diagnosis," International Journal of Cognitive Computing in Engineering, vol. 2, pp. 116–129, Jun. 2021, doi: 10.1016/j.ijcce.2021.08.002.

[19] S. de Roda Husman, J. J. van der Sanden, S. Lhermitte, and M. A. Eleveld, "Integrating intensity and context for improved supervised river ice classification from dual-pol Sentinel-1 SAR data," International Journal of Applied Earth Observation and Geoinformation, vol. 101, p. 102359, Sep. 2021, doi: 10.1016/j.jag.2021.102359.

[20] R. F. Rahmat, A. B. Pangaribuan, E. Suwarno, and T. Z. Lini, "Lake Toba Water Quality Prediction using Extreme Learning Machine," ICIC Express Letters, Part B: Applications, vol. 13, no. 1, pp. 89–97, 2022.

[21] S. Ding, X. Xu, and R. Nie, "Extreme learning machine and its applications," Neural Comput Appl, vol. 25, no. 3–4, pp. 549–556, Sep. 2014, doi: 10.1007/s00521-013-1522-8.

[22] Y. Wei, H. Chen, J. Luo, and Q. Li, "An improved fruit fly optimization enhanced kernel extreme learning machine with application to second major prediction," ICIC Express Letters, Part B: Applications, vol. 8, no. 7, pp. 1015–1021, 2017.

[23] P. Fang, D. Wang, and W. Song, "Identification of fuzzy wavelet neural network by combining extreme learning machine and gradient decent algorithm," ICIC Express Letters, Part B: Applications, vol. 6, pp. 1937–1944, Dec. 2015.

[24] Z. Liu, Y. Song, J. Wang, and K. Li, "Physical Activity Recognition Based On Time Window Selection and Online Sequential ELM," ICIC Express Letters, Part B: Application, vol. 8, no. 1, pp. 1–10, 2017.

[25] Guang-Bin Huang, Qin-Yu Zhu, and Chee-Kheong Siew, "Extreme learning machine: a new learning scheme of feedforward neural networks," in 2004 IEEE International Joint Conference on Neural Networks (IEEE Cat. No.04CH37541), IEEE, pp. 985–990. doi: 10.1109/IJCNN.2004.1380068.

Evaluating the Impact of Yoga Practices to Improve Chronic Venous Insufficiency Symptoms: A Classification by Gaussian Process

Feng Yun Gou
Inner Mongolia University of Technology Hohhot, Inner Mongolia, 010051, China

***Abstract*—Chronic Venous Insufficiency (CVI) is a widespread condition marked by diverse venous system irregularities stemming from occlusion and varicosities. Factors like family history and lifestyle choices amplify CVI's economic consequences, emphasizing the need for proactive measures. The sedentary lifestyle of many individuals can contribute to various diseases, including CVI. Yoga is now endorsed as a multifaceted exercise to alleviate CVI symptoms, offering a holistic approach and complementary therapy for diverse medical conditions. This study developed a method for evaluating and classifying symptoms associated with varicose veins, utilizing the Venous Clinical Score (VCSS) data. A specific emphasis was placed on investigating the impact of yoga on these symptoms, and a comprehensive performance assessment was conducted based on data obtained from a cohort of 100 patients. This paper achieves optimal performance by employing the Gaussian Process Classifier (GPC) along with two optimizers, namely the Crystal Structure Algorithm (CSA) and the Fire Hawk Optimizer (FHO). The results indicate that in predicting VCSS-Pre (reflecting symptoms before engaging in yoga exercises), the GPFH exhibited superior performance with an F1-score of 0.872, surpassing the GPCS, which attained an F1-score of 0.861 by almost 1.26%. Additionally, the prediction for VCSS-1, reflecting symptoms after one month of yoga practices, revealed the GPFH outperforming the GPCS with respective F1-score values of 0.910 and 0.901.**

Keywords*—Chronic venous insufficiency; yoga; Gaussian Process Classifier; Crystal Structure Algorithm; Fire Hawk Optimizer

I. Introduction

Prolonged sitting, prevalent in contemporary office environments that have shifted from active to passive, poses a risk for cardio-metabolic diseases, type 2 diabetes, obesity, coronary artery disease, musculoskeletal conditions, certain cancers, and early death [1], [2]. Sedentary behavior, considered by an energy expenditure of $\leq 1.5\ METs$ while sitting or reclining, is a substantial issue in modern workplaces [3]. Extended periods of sitting have been correlated with elevated risks of obesity and diabetes, with studies indicating a 5% rise in obesity risk and a 7% increase in diabetes risk for every two-hour increase in sitting time [4]. Moreover, prolonged sitting is linked to a heightened possibility of musculoskeletal disorders, particularly low back pain [5]. Other research has demonstrated that occupations requiring significant periods of sedentary behavior are associated with an increased risk of developing certain types of cancers, such as endometrial, prostate, and colorectal cancer [6], [7].

Inactive sitting behavior is closely associated with the risk of cardiovascular disease (CVD), irrespective of one's level of physical activity. This association arises from the impact of sedentary behavior on crucial inflammatory, hemodynamic, and metabolic processes, leading to compromised arterial health. Consequently, these vascular issues directly and indirectly contribute to the development of cardiovascular disease [8], [9]. Some studies have revealed a noteworthy correlation between the duration of individuals' sitting hours in the workplace and the incidence of Chronic Venous Insufficiency (CVI). This association underscores a substantial escalation in the risk of contracting such conditions [10], [11], [12].

CVI is widespread in both developing and developed nations [13]. As outlined in the American Venous Forum's consensus statement, CVI involves various morphological and practical irregularities in the venous system, from telangiectasias to venous ulcers [14]. The recognition of CVI lacks a specific date, but the historical understanding of venous insufficiency can be traced to ancient times when Egyptians, Greeks, and Romans described similar symptoms. In the 17th century, William Harvey, an English physician, significantly advanced the understanding of the circulatory system, though the term "Chronic Venous Insufficiency" is a more recent medical terminology. The precise historical origin of this term is unclear, yet the evolving comprehension and management of venous insufficiency have shaped current diagnostic criteria and treatment approaches [15], [16].

CVI refers to pathological changes in the lower extremities' tissues resulting from anomalies in venous blood flow. These glitches encompass popliteal or iliofemoral vein occlusion, incompetence, and varicosities, mainly in the greater saphenous system associated with valvular leakage or abnormal arteriovenous communications. In some cases, chronic venous insufficiency can arise from large acquired arteriovenous fistulae, certain congenital anomalies, or tumors of the blood vessels [17], [18]. CVI risk factors encompass factors such as family history, aging, long-standing, obesity, an inactive routine, smoking, lower extremity trauma, previous venous thrombosis, the existence of an arteriovenous shunt, high estrogen states, and pregnancy [19], [20], [21]. The economic consequences of CVI are evident through increased health care costs, potential productivity loss due to symptoms such as leg pain and swelling, and the potential threat of disability leading

to unemployment. These challenges have a double effect on individuals and society in general. Implementing effective preventive measures, timely interventions, and creating supportive facilities in the workplace appear as central strategies for reducing the economic burden associated with CVI [22], [23].

Diagnosing CVI relies on a comprehensive assessment encompassing medical history, observed signs and symptoms, and diagnostic tests. A crucial tool for quantifying disease severity is the Venous Severity Scoring system, introduced in 2000 and refined in 2010 [24]. This system comprises three components: The Venous Clinical Severity Score (VCSS), which evaluates clinical symptoms; The Venous Segmental Disease Score (VSDS), which focuses on segment-specific aspects; and the Venous Disability Score (VDS), which provides insights into the functional impact of CVI. These scoring components collectively enhance the accuracy and thoroughness of clinical evaluations for individuals with CVI [25], [26]. Physicians use the VCSS to evaluate the impact of venous disease on a patient's clinical condition and to make informed decisions about appropriate treatment strategies [27].

Presently, yoga is recommended as one of the activities to alleviate symptoms associated with CVI [28]. Yoga, originating in ancient India, is a holistic practice encompassing physical, mental, and spiritual disciplines. It combines physical postures, breath control, meditation, and ethical principles to promote well-being, harmony, flexibility, strength, stress reduction, and mental clarity [29], [30]. Yoga is increasingly integrated into medical care as a complementary therapy, offering benefits for chronic pain, mental health issues, cardiovascular health, cancer care, respiratory conditions, and rehabilitation.

II. Objective

Utilizing a machine learning (ML) approach, this article delves into the effectiveness of yoga exercises in alleviating symptoms among individuals with chronic venous insufficiency. ML, a subset of artificial intelligence (AI), entails creating algorithms and statistical models that empower computers to execute tasks without explicit programming. It constitutes a field of study wherein systems acquire knowledge and enhance performance through experience, enabling pattern recognition, predictions, and adaptation to new data. These algorithms utilize training data to identify inherent patterns, enabling them to make decisions or predictions without explicit programming for each task. The exploration includes predicting and categorizing these symptoms into distinct classifications using the Gaussian Process Classifier (GPC). For the ultimate optimization, the article incorporates two optimizers, namely, the Crystal Structure Algorithm (CSA) and the Fire Hawk Optimizer (FHO), to achieve optimal performance.

III. Methodology and Modeling Approach

A. Gaussian Process Classifier (GPC)

The Gaussian Process (GP) classifier, grounded in Bayesian theory, operates by establishing a Gaussian prior distribution over the estimated function, represented as $p(x) = w^T x + b$. This initial distribution forms the basis for the probabilistic estimation process, and the classifier further incorporates a sigmoid function to construct a probabilistic estimator as outlined in Eq. (1):

$$p(y = 1|x) = sigmoid(f(x)) \tag{1}$$

Determining the distribution of the latent variable y for a test sample involves two key steps. Firstly, it is computed by leveraging the posterior over the latent variables, denoted as $p(f|X, y)$. Following this, the second step entails calculating a posterior using Eq. (2), which builds upon the outcomes from the initial step. This sophisticated method enables the probabilistic estimation of the latent variable y, providing nuanced insights and accurate predictions.

$$p(y = 1|X, y, x) \tag{2}$$

X and Y denote the training samples, with x representing the test sample. While conducting this inference is generally a complex task, there are existing approximations that tend towards an optimal solution with larger datasets. Furthermore, kernel versions of this process offer a more straightforward approach [31].

B. Fire Hawk Optimizer (FHO)

The FHO algorithm replicates the foraging patterns of Fire Hawks, involving the initiation and expansion of fires to capture prey. It commences by generating a set of potential solutions (X) inspired by the position vectors of fire hawks and their prey. The initial positioning of these vectors within the search space is determined randomly, imitating the initial locations of fire hawks and prey, laying the foundation for subsequent optimization stages [32].

$$x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_i \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} x_1^1 & x_1^2 & \cdots & x_1^j & \cdots & x_1^d \\ x_2^1 & x_2^2 & \cdots & x_2^j & \cdots & x_2^d \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ x_i^1 & x_i^2 & \dots & x_i^j & \dots & x_i^d \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ x_N^1 & x_N^2 & \cdots & x_N^j & \dots & x_N^d \end{bmatrix}, \begin{cases} i = 1,2,3,\dots,N. \\ j = 1,2,3,\dots,d. \end{cases} \tag{3}$$

$$x_i^j(0) = x_{i,min}^j + rand.\left(x_{i,max}^j - x_{i,min}^j\right), \begin{cases} i = 1,2,3,\dots,N. \\ j = 1,2,3,\dots,d. \end{cases} \tag{4}$$

X_i represents solution candidates, d is the problem dimension, N is the total number of candidates, x_i^j is a decision variable, $x_i^j(0)$ is the initial position, $x_{i,max}^j$, $x_{i,min}^j$ are variable bounds, and $rand$ is a random number (0, 1). The goal is to identify Fire Hawks in the search space based on higher objective function values. Selected Fire Hawks spread flames around prey, aiding hunting. The primary fire, initially used by Fire Hawks, is assumed to represent the best global solution.

$$PR = \begin{bmatrix} PR_1 \\ PR_2 \\ \vdots \\ PR_i \\ \vdots \\ PR_m \end{bmatrix}, i = 1,2,3,\dots,m \tag{5}$$

$$FH = \begin{bmatrix} FH_1 \\ FH_2 \\ \vdots \\ FH_i \\ \vdots \\ FH_n \end{bmatrix}, i = 1,2,3, \dots, n \quad (6)$$

PR_i signifies the i_{th} fire hawk, and FH_i denotes the i_{th} prey, where *n* is the total number of prey. The subsequent step involves calculating the distance between Fire Hawks and their prey, with D_l^k expressed by the following equation:

$$D_l^k = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}, \begin{cases} k = 1,2,3, \dots, n. \\ l = 1,2,3, \dots, m. \end{cases} \quad (7)$$

The exploration space's total count of prey and fire hawks is denoted by n and m, respectively. The cumulative distance among the i_{th} Fire Hawk (FH) and prey, represented by D_l^k, is determined by the coordinates (x_1, y_1) for FH and (x_2, y_2) for prey. The birds establish territories by classifying the nearest prey and calculating total distances. Gathering hot coals from the main fire to ignite specific spots and engaging in behaviors such as using burning sticks from other territories serve as position updates in the $FHO's$ primary search loop.

$$FH_i^{new} = FH_i + (r_1 \times GB - r_2 \times FH_{near}), i = 1,2,3, \dots, n \quad (8)$$

The primary fire, denoted as GB, signifies the global best solution in the search space. FH_i^{new} represents the new location vector of the i_{th} FH, while r_1 and r_2 are uniformly distributed random numbers in the range $(0, 1)$ representing movements towards the primary fire and other Fire Hawks' territories. FH_{near} designates one of the FH within the exploration space. The algorithm's next stage involves prey movements during fires and guiding position updates. An equation is employed for these actions during place updates to incorporate the importance of territory in animal behavior.

$$PR_s^{new} = PR_s + (r_3 \times FH_i - r_4 \times SP_i), \begin{cases} i = 1,2,3, \dots, n \\ s = 1,2,3, \dots, r \end{cases} \quad (9)$$

The new position vector of the S_{th} prey (PR_s), surrounded by the i_{th} Fire Hawk (FH_i), is represented as PR_s^{new}. SP_i signifies a safe place under the i_{th} Fire Hawk territory. To monitor prey movements toward Fire Hawks and their retreat to safe zones, r_3 and r_4 are uniformly distributed random integers from 0 to 1. Prey may venture into other FH territories or approach those trapped by flames. FH might seek safer areas beyond their territory. The provided equations accommodate these actions in position updates.

$$PR_i^{new} = PR_i + (r_5 \times FH_{Alter} - r_6 \times SP), \begin{cases} i = 1,2,3, \dots, n \\ s = 1,2,3, \dots, r \end{cases} \quad (10)$$

The updated position vector of the i_{th} prey (PR_i), positioned between the i_{th} Fire Hawk (FH_i), is denoted as PR_i^{new}. SP represents a secure area beyond the i_{th} FH territory. FH_{Alter} signifies one of the FH in the search space. The movements of prey toward other FH and the secure region beyond their territories are determined by the uniformly distributed values of r_5 and r_6 in the range (0,1).

The mathematical expression for SP_i and SP is derived with the acknowledgment that, in the natural environment, a secure location is where most animals gather for protection during threats.

$$SP_i = \frac{\sum_{s=1}^{r} PR_s}{r}, \begin{cases} i = 1,2,3, \dots, n \\ s = 1,2,3, \dots, r \end{cases} \quad (11)$$

$$SP = \frac{\sum_{q=1}^{m} PR_q}{m}, q = 1,2,3, \dots, m \quad (12)$$

PR_s represents the prey positioned around the S_{th} fire hawk (FH_i), while PR_q denotes the q_{th} prey within the search space.

C. Crystal Structure Algorithm (CSA)

Solid minerals, comprised of atoms and particles arranged in a crystalline form known as crystals, derive their Grecian meaning from the concept of solidification by cold. Interior particles were initially discovered in 1619 by Kepler, 1665 Hooke, and 1690 Hogens [33]. Crystals display a repeating pattern of atoms in defined spaces, forming a lattice that not only dictates the crystal's shape but also inspires geometric figures derived from infinite natural shapes. The discontinuous crystal structure is crafted by considering an infinite lattice, with each lattice point linked to its position through a vector [34]:

$$r = \sum m_i \, d_i \quad (13)$$

The variables in the model are defined as follows: m_i is an integer, d_i represents the shortest vector along the central crystallographic directions, and i corresponds to the number of crystal corners. The mathematical model of CryStAl is then introduced in this section, incorporating fundamental crystal concepts with notable modifications. The crystal number is initialized as a random number in this model.

$$Cr = \begin{bmatrix} Cr_1 \\ Cr_2 \\ \vdots \\ Cr_i \\ \vdots \\ Cr_n \end{bmatrix} = \begin{bmatrix} x_1^1 & x_1^2 & \dots & x_1^j & \dots & x_1^d \\ x_2^1 & x_2^2 & \dots & x_2^j & \dots & x_2^d \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ x_i^1 & x_i^2 & \dots & x_i^j & \dots & x_i^d \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ x_n^1 & x_n^2 & \dots & x_n^j & \dots & x_n^d \end{bmatrix}, \begin{cases} i = 1,2,3, \dots, n \\ j = 1,2,3, \dots, j \end{cases} \quad (14)$$

$$x_i^j(0) = x_{i,min}^j + \xi \left(x_{i,max}^j - x_{i,min}^j\right), \begin{cases} i = 1,2,3, \dots, n \\ j = 1,2,3, \dots, d \end{cases} \quad (15)$$

In this model, n and d denote the number of crystals and the problem's dimension, respectively. The variables $x_i^j(0)$, $x_{i,max}^j$ and $x_{i,min}^j$ determine the initial positions and permissible values for each decision variable of candidate solutions, while ξ is a random number within [0,1] . Following crystallography principles, primary crystals Cr_m are corner crystals randomly chosen, and the main crystal is selected in each step, excluding the current one. F_c represents the mean merit of randomly chosen crystals and Cr_b is the best-arranged crystal. Based on fundamental lattice cross-section standards, four types of

improvement processes are specified for upgrading candidate arrangements in this crystal model.

Cubicles;

Simple:

$$Cr_n = Cr_o + aCr_m \quad (16)$$

With the finest crystals:

$$Cr_n = Cr_o + a_1 Cr_m + a_2 Cr_b \quad (17)$$

With the mediocre crystals:

$$Cr_n = Cr_o + a_1 Cr_m + a_2 f_c \quad (18)$$

With the finest and mediocre crystals:

$$Cr_n = Cr_o + a_1 Cr_m + a_2 Cr_b + a_3 F_c \quad (19)$$

The set of four equations illustrates the transition between new and old positions, denoted as Cr_n and Cr_o, respectively, incorporating fortuitous numbers a, a_1, a_2 and a_3. The algorithm employs exploration and extraction features, calculated using Eq. (16) to Eq. (19). The optimization process concludes upon reaching the maximum iteration, adhering to a predefined termination criterion with a fixed number of repetitions. A mathematical flag is utilized for solution variables x_i^j, indicating the exterior of factors range and setting a boundary change order.

The CSA pseudo
− code is shown below: Procedure Crystal Structure Algorith
randomly create values for primary positions (x_i^j) of prima
Estimate fitness values for each crystal
while ($t <$ maximum number of iterations)
for $i = 1$: number of initial crystals
Create $\boldsymbol{Cr_m}$
Create new crystals by Eq. (16)
Create $\boldsymbol{Cr_b}$
Create new crystals by Eq. (17)
Create $\boldsymbol{F_c}$
Create new crystals by Eq. (18)
Create new crystals by Eq. (19)
if new crystals violate boundary conditions
Control the position constraints for new crystals and amend
end if
Evaluate the fitness values for new crystals
Update Global Best (GB) if a better solution is found
end for
$t = t + 1$
end while
Return GB
End procedure

IV. MATERIAL

A. Data Description and Analysis

Data mining, acknowledged as a crucial element in the Knowledge Discovery from Databases (KDD) process [35], is progressively gaining significance within the healthcare system. It plays a vital role in precisely predicting medical conditions by automating the extraction of knowledge from extensive datasets [36] and employing a spectrum of techniques, including statistical analysis, machine learning, and database methodologies, to facilitate informed decision-making [37].

The dataset provides an extensive and diverse collection of variables, each with the potential to influence the symptoms associated with varicose veins. This well-rounded dataset includes key variables such as Body Mass Index (BMI), Systolic Blood Pressure Type A (SBPA), Systolic Blood Pressure Type B (SBPB), Ankle-Brachial Pressure Index (ABPI), Diabetes Blood Pressure Type A (DBPA), Diabetes Blood Pressure Type B (DBPB), Pulse Rate (PR), Chronic Fatigue Syndrome (CFS), Hyperhomocysteinemia (HCY), Calf Circumference (CALF-CIR), the Chronic Venous Insufficiency Questionnaire (CVIQ), and the Chalder Fatigue Scale (CFS). In addition to these primary variables, the dataset includes two supplementary variables: VCSS-Pre, which represents symptoms observed before the initiation of yoga practices, and VCSS-1, which captures symptoms observed after one month of participation in yoga sessions. These variables provide a unique opportunity to analyze the potential impact of yoga on varicose vein symptoms over time. To facilitate analysis, the dataset has been categorized into four distinct groups based on the severity of the condition: Absent condition (0-5), Mild (6-10), Moderate (11-20), and Severe (21-30). This classification allows for a more structured examination of the relationship between the various factors and the severity of varicose vein symptoms, offering insights that can inform treatment approaches and patient care strategies.

In this research, Fig. 1 illustrates a correlation matrix that provides a detailed and comprehensive view of the intricate interrelationships among the investigated input and output variables. The matrix reveals how the input data not only significantly influences the output but also affects the relationships between other input variables. For example, the Diabetes Blood Pressure indicators, specifically DBPA (Diabetes Blood Pressure Type A) and DBPB (Diabetes Blood Pressure Type B), show a pronounced effect on both SBPA (Systolic Blood Pressure Type A) and SBPB (Systolic Blood Pressure Type B). This highlights the critical role that blood pressure variations, influenced by diabetes, play in shaping systolic blood pressure outcomes. Additionally, the correlation matrix underscores the significant impact of the Chalder Fatigue Scale's physical and mental components (CFS PHY-Pre and CFS MEN-Pre) on the overall CFS-Pre score. These findings suggest that fatigue, as captured by the Chalder scale, is a crucial predictor of overall chronic fatigue syndrome (CFS) severity. The Chronic Venous Insufficiency Questionnaire (CVIQ) is also identified as a key contributor, influencing both the CFS PHY-Pre and CFS-Pre parameters, further emphasizing the complex interplay between chronic venous insufficiency and fatigue symptoms. Moreover, the analysis identifies other highly influential input parameters, such as the number of standing and sitting hours, the number of working days, CALF-CIR (calf circumference), and HCY (homocysteine levels), which all play significant roles in determining the outcomes. Conversely, the PR (pulse rate) parameter is highlighted as having one of the least impacts on the output, suggesting its relatively minor role in the context of this study. This comprehensive understanding of variable interrelationships provides valuable insights for targeting specific areas in future research and potential interventions.

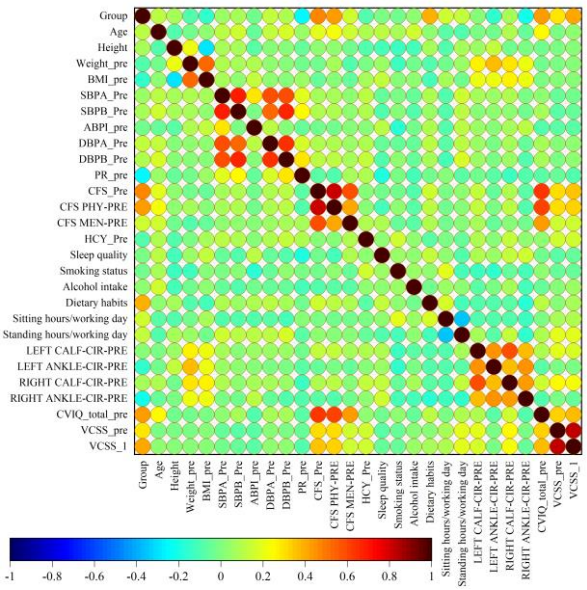


Fig. 1. Correlation matrix to examine how input and output variables are related to one another Evaluation of Model Suitability.

In the realm of classification challenges, Accuracy emerges as a frequently employed metric for evaluating overall model performance, taking into account True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN). Despite its widespread use, Accuracy faces limitations in scenarios with imbalanced data, as it tends to favor the majority class, providing limited insights. The mathematical expression for Accuracy is defined in Eq. (20):

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{20}$$

To overcome this restriction, three more measures are used: $F1 - Score, Precision$, and $Recall$. To reduce False Negatives, recall measures a model's capacity to accurately identify every pertinent instance inside a given class. By quantifying the precision of positive predictions, False Positives are decreased. Combining Precision and Recall, the F1-Score offers a balanced evaluation that is particularly helpful in situations when the data is unbalanced. Together, these metrics which are represented by mathematical formulae (Eq. 21–23) help to provide a more complete picture of the efficacy of a categorization model.

$$Precision = \frac{TP}{TP + FP} \tag{21}$$

$$Recall = TPR = \frac{TP}{P} = \frac{TP}{TP + FN} \tag{21}$$

$$F1_score = \frac{2 \times Recall \times Precision}{Recall + Precision} \tag{23}$$

V. Results

To improve the accuracy of classifying and predicting VCSS before and after one month of yoga exercises, this study utilized two optimization algorithms: Crystal Structure Algorithm (CSA) and Fire Hawk Optimizer (FHO). The data, involving 100 patients, underwent careful evaluation in both training and testing phases following the implementation of these algorithms. The main objective is to refine and optimize model parameters using the mentioned algorithms.

Fig. 2 shows the convergence of developed hybrid models. VCSS-Pre, the GPCS, and GPFH models commenced iterations with nearly identical Precision. Eventually, the GPCS model achieved optimal accuracy in 100 iterations (with 0.860 value), while the GPFH model took 110 iterations (in 0.870). In contrast, in VCSS-1, the GPFH model started with lower accuracy than GPCS, reaching higher accuracy in the 80th iteration (in 0.910 accuracy value), while GPCS reached convergence in the 120th iteration (with 0.900).

Table I showcases extensive metrics, encompassing $Accuracy, Precision, Recall,$ and $F1\ Score$, across all models for the training and test phases. Notably, the GPFH model exhibits excellent performance in both VCSS-Pre and VCSS-1. Specifically, for VCSS-Pre, the model achieves a Precision of 0.870, Accuracy of 0.877, Recall of 0.870, and F1-score of 0.872. Similarly, in the case of VCSS-1, the GPFH model attains $Accuracy, Recall$, and $F1 - score$ values of 0.910, along with a Precision value of 0.916. Fig. 3 presents a bar plot that visually assesses the performance of the advanced models. Furthermore, it provides additional insights into the achieved results. For instance, it is observable that the GPFH, GPCS, and GPC models showcase optimal performance.

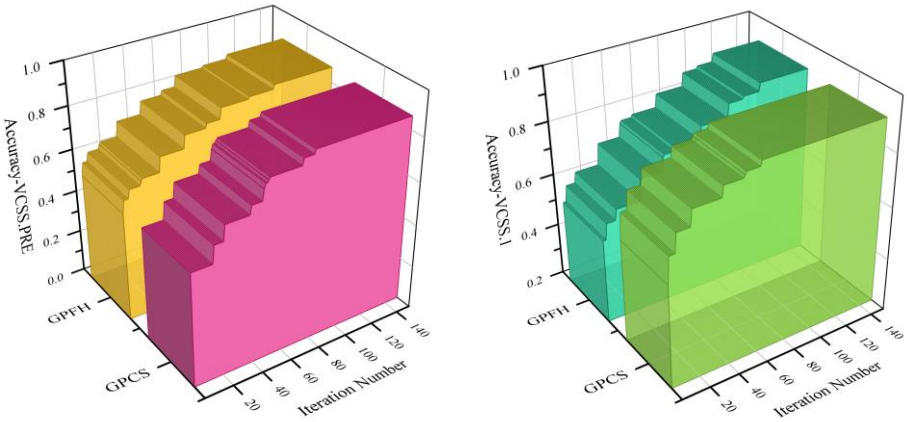


Fig. 2. Convergence curve of hybrid models.

TABLE I. RESULT OF PRESENTED MODELS

	Model	Part	Metric value			
			Accuracy	*Precision*	*Recall*	*F1 _Score*
$VCSS - PRE$	GPC	$Train$	0.900	0.913	0.900	0.901
		$Test$	0.733	0.759	0.733	0.737
		All	0.850	0.868	0.850	0.853
	$GPFH$	$Train$	0.929	0.937	0.929	0.930
		$Test$	0.733	0.739	0.733	0.735
		All	0.870	0.877	0.870	0.872
	$GPCS$	$Train$	0.900	0.915	0.900	0.900
		$Test$	0.767	0.771	0.767	0.767
		All	0.860	0.871	0.860	0.861
$VCSS - 1$	GPC	$Train$	0.957	0.958	0.957	0.957
		$Test$	0.733	0.733	0.733	0.733
		All	0.890	0.890	0.890	0.890
	$GPFH$	$Train$	0.957	0.969	0.957	0.959
		$Test$	0.800	0.800	0.800	0.798
		All	0.910	0.916	0.910	0.910
	$GPCS$	$Train$	0.914	0.926	0.914	0.917
		$Test$	0.867	0.871	0.867	0.867
		All	0.900	0.903	0.900	0.901

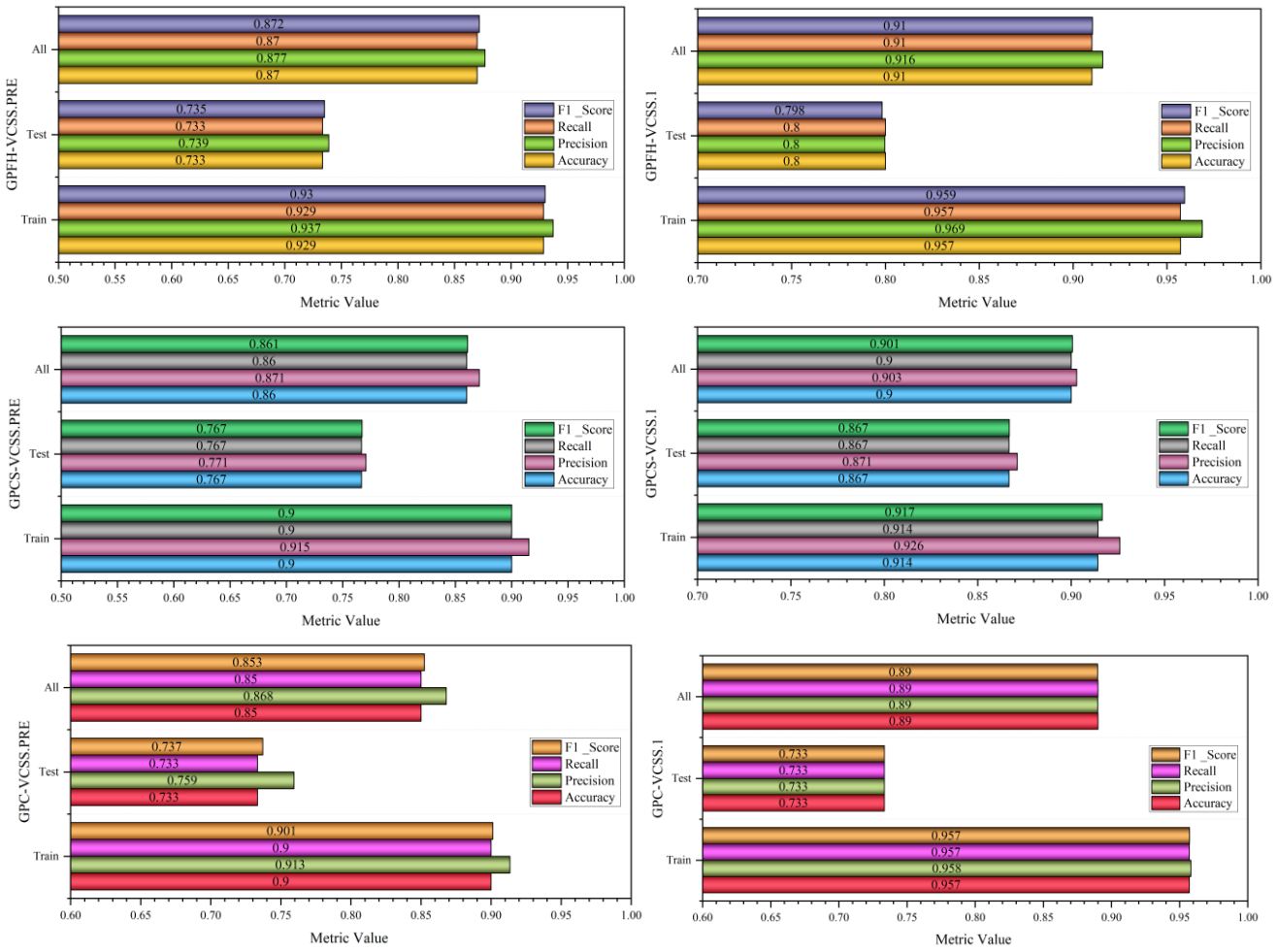


Fig. 3. Bar plot to visually evaluate the created models' performance.

Based on VCSS scores, patients are classified into four groups: Absent (0-5), Mild (6-10), Moderate (11-20), and Severe (21-30) condition. Concerning Precision in VCSS-Pre, as presented in Table II, the GPFH and GPCS models exhibit similar and close performance. Notably, for the Severe condition, both models achieved a value of 1. The GPFH model demonstrates superior performance for Mild and Moderate conditions with values of 0.841 and 0.950, respectively. However, for the Absent condition, the GPCS model outperforms with a value of 0.733. In Table III, considering the Precision values for VCSS-1, both the GPFH and GPCS models exhibit identical values for Severe conditions at 1. However, for Mild (0.912) and Moderate conditions (0.957), the GPFH model outperforms. Notably, akin to VCSS-Pre, in Absent conditions, the GPCS model demonstrates superior performance with a value of 0.813.

TABLE II. PERFORMANCE EVALUATION INDICES FOR THE DEVELOPED MODELS FOR VCSS-PRE

Model	*Condition*	*Metric value*		
		Precision	*Recall*	*F1 − score*
GPC	*Absent*	0.688	0.917	0.786
	Mild	0.804	0.881	0.841
	Moderate	0.972	0.796	0.875
	Severe	1.000	1.000	1.000
GPFH	*Absent*	0.714	0.833	0.769
	Mild	0.841	0.881	0.861
	Moderate	0.950	0.864	0.905
	Severe	1.000	1.000	1.000
GPCS	*Absent*	0.733	0.917	0.815
	Mild	0.826	0.905	0.864
	Moderate	0.946	0.796	0.864
	Severe	1.000	1.000	1.000

TABLE III. PERFORMANCE EVALUATION INDICES FOR THE DEVELOPED MODELS FOR VCSS-1

Model	*Condition*	*Metric value*		
		Precision	*Recall*	*F1 − score*
GPC	*Absent*	0.867	0.929	0.897
	Mild	0.861	0.838	0.849
	Moderate	0.915	0.915	0.915
	Severe	1.000	1.000	1.000
GPFH	*Absent*	0.778	1.000	0.875
	Mild	0.912	0.838	0.873
	Moderate	0.957	0.936	0.946
	Severe	1.000	1.000	1.000
GPCS	*Absent*	0.813	0.929	0.867
	Mild	0.865	0.865	0.865
	Moderate	0.956	0.915	0.935
	Severe	1.000	0.100	1.000

Fig. 4 shows a 3D drop line plot illustrating the difference between measured and forecast values for $VCSS - Pre$ and $VCSS - 1$. Separate graphs are included for each category ($Absent, Mild, Moderate$, and $Severe$), which comprehensively assesses the models' efficacy in classifying.

Upon reviewing the VCSS-Pre diagram, it becomes evident that 12 individuals fall under the Absent category, 42 in Mild, 44 in Moderate, and 2 in Severe. Significantly, the GPC and GPCS models stand out for their remarkable accuracy in classifying the Absent section. In this section, the models demonstrate exceptional Precision in predictions, with a one-unit difference, highlighting the capability to categorize individuals accurately. In Mild and Moderate conditions, the GPCS and GPFH models showcase superior performance with subtle distinctions. In the Severe section, three models demonstrate identical performance.

As illustrated in the figure for VCSS-1, the recorded counts for patients in the $Absent, Mild, Moderate,$ and $Severe$ categories are 14, 37, 47, and 2, individually. Notably, the GPFH model exhibits superior performance in the Absent and Moderate conditions, particularly in the Absent category, where it achieves error-free predictions. The GPCS model performs best in Mild conditions, and for Severe conditions, all utilized models demonstrate superior performance.

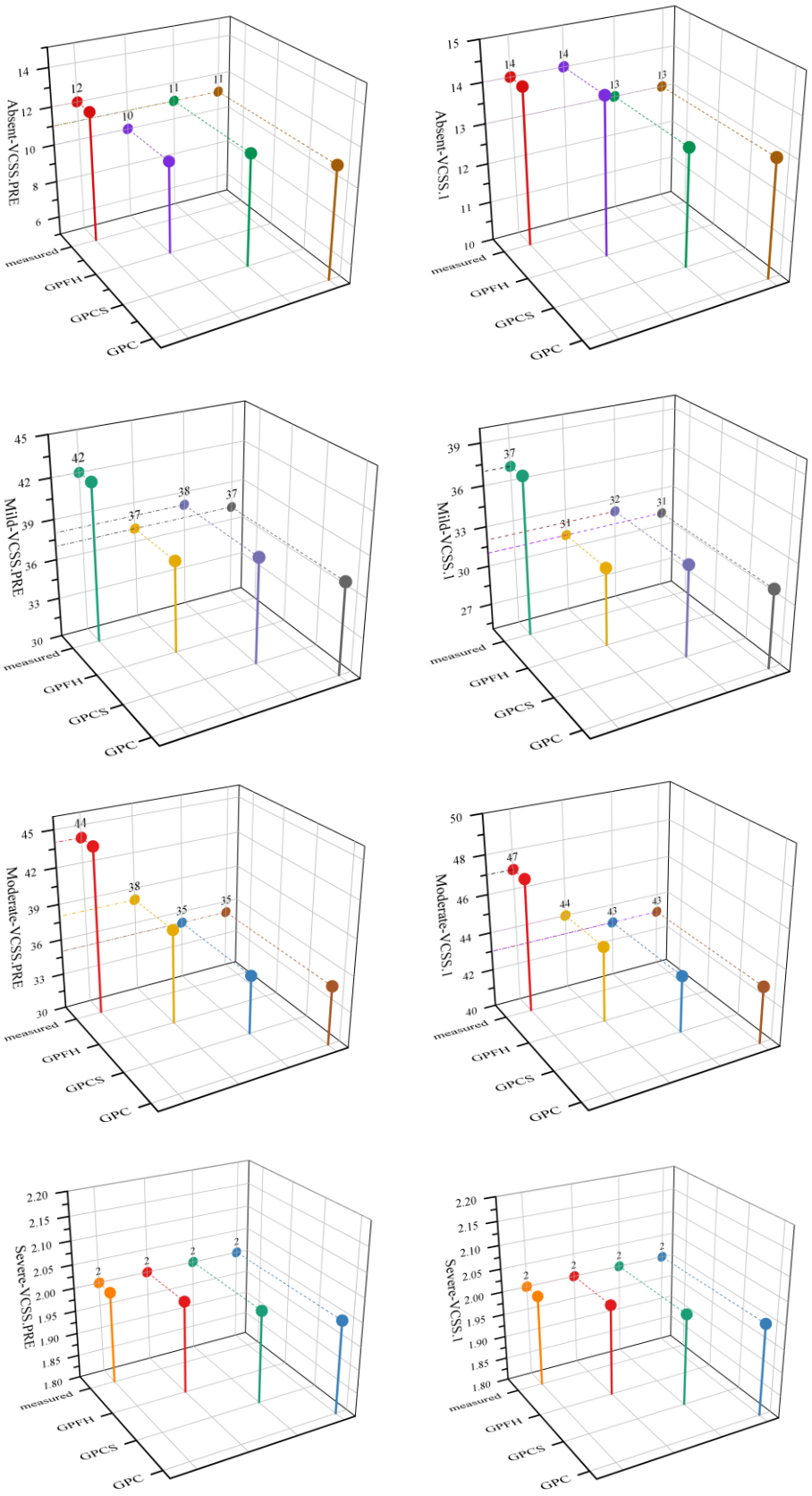


Fig. 4. 3D drop line plot for the difference between measured and forecast values.

Fig. 5 shows the confusion matrix. A confusion matrix summarizes the predictions of the ML model, offering a detailed assessment of its performance in terms of $Accuracy, Precision, Recall,$ and $F1-score$ measures for evaluating effectiveness. Table III displays the confusion matrix corresponding to VCSS-Pre and VCSS-1. In VCSS-Pre, the GPFH model correctly classified 87 patients, breaking down into 10 Absent, 37 Mild, 38 Moderate, and 2 Severe conditions, while 13 patients were misclassified. The GPCS model takes the second position, with 86 patients correctly classified and 14 misclassified. The GPC model ranks third, with 15 patients misclassified; in the context of VCSS-1, the GPFH model successfully classified 91 patients (14 in Absent, 31 in Mild, 44 in Moderate, and 2 in Severe conditions) with only nine misclassifications. The GPCS model secured 90 accurate predictions and 10 incorrect ones, while the GPC model claimed the third rank with 89 correct predictions and 11 incorrect ones.

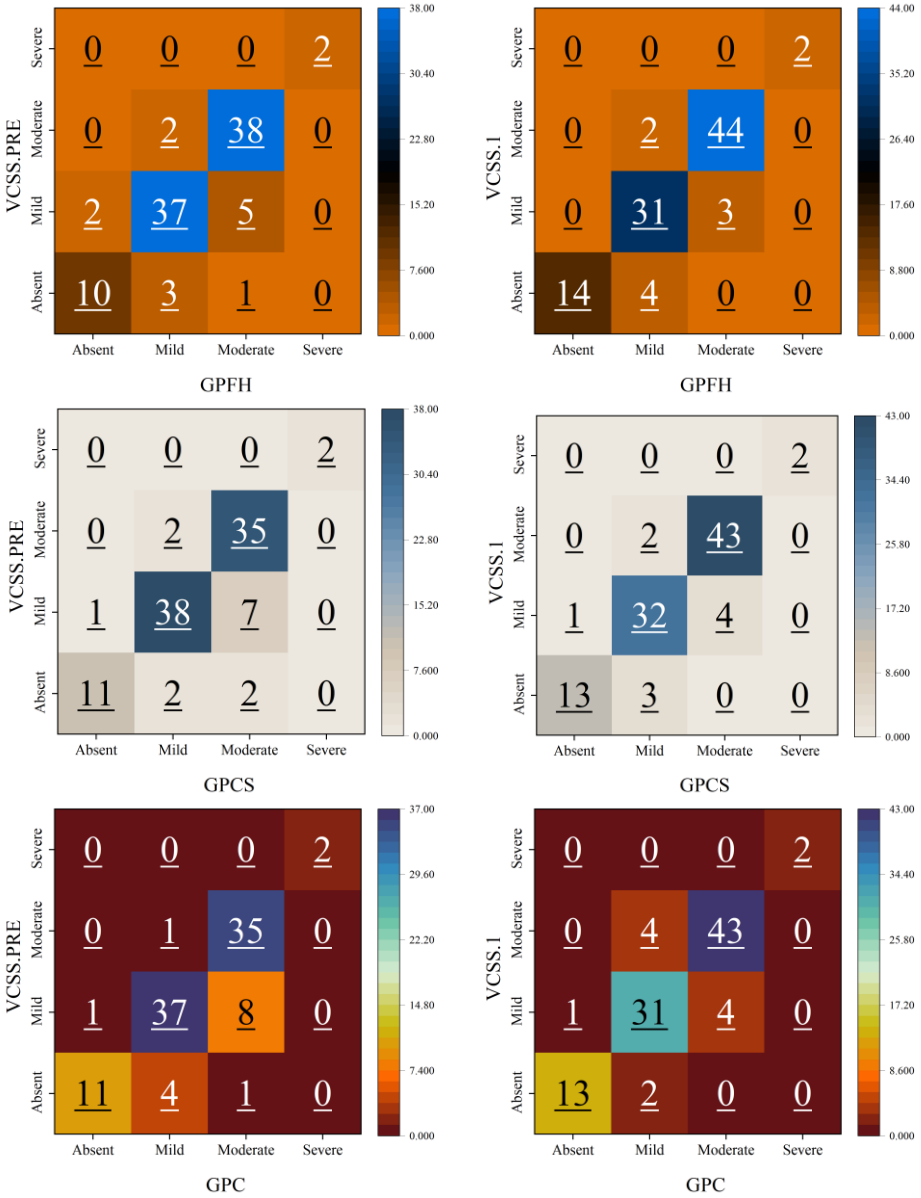


Fig. 5. Confusion matrix for each model's accuracy.

The Receiver Operating Characteristic (ROC) is a visual tool employed in classification to assess a model's performance by mapping the interplay between its false positive rate and true positive rate across diverse thresholds. Illustrating the discriminative capacity of a model, the ROC curve provides a comprehensive overview of its ability to distinguish between classes. Based on the top-performing GPFH model in Fig. 6, the ROC curve analysis indicates that this model serves as a classifier with acceptable performance in predicting VCSS-Pre for Moderate conditions. Within the VCSS-1 framework, the GPC model demonstrates superior performance for Absent and Mild conditions, with curves approaching 1 for each. Finally, it is noteworthy that the Micro Average, drawn for both VCSS-Pre and VCSS-1, supports the best performance in VCSS-1. Consequently, it can be inferred that one month of yoga practice affects CVI.

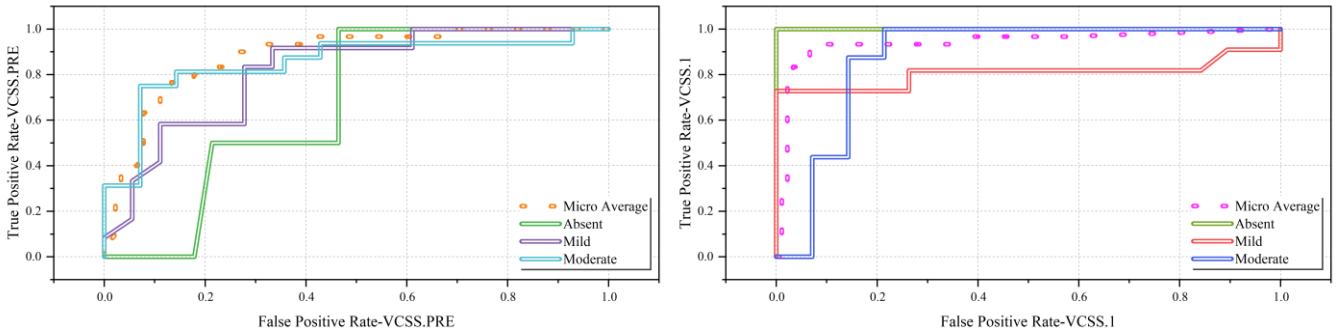
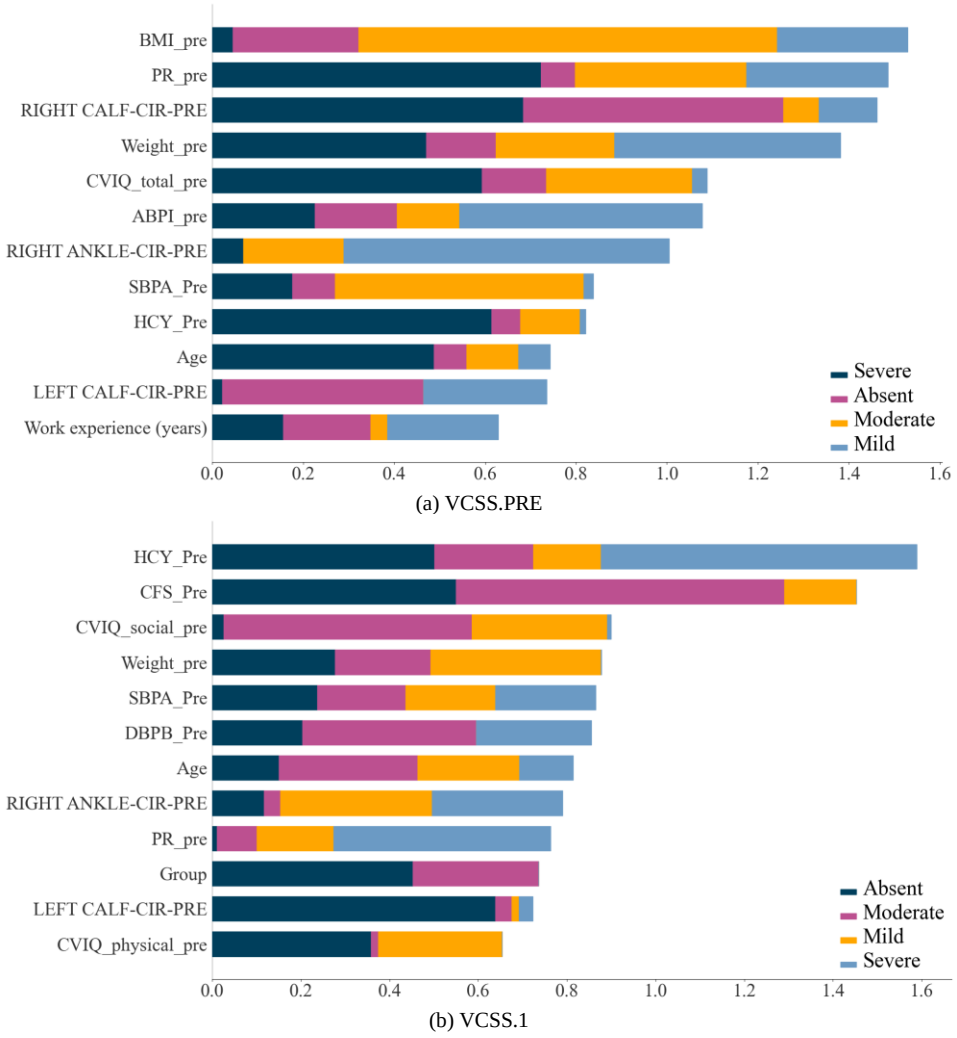


Fig. 6. The outcome is derived from the ROC curve.



(a) VCSS.PRE

(b) VCSS.1

Fig. 7. Impact of input features on model outputs based on SHAP values.

Fig. 7, visually represents the impact of input features on model outputs for predicting symptoms related to CVI based on SHAP (Shapley Additive exPlanations) values. SHAP values are used to interpret the output of machine learning models, providing insights into how much each feature contributes to the prediction. The figure is divided into two plots: (a) VCSS.PRE and (b) VCSS.1, each showing the influence of various input features on the VCSS-Pre and VCSS-1 outcomes, respectively. The VCSS-Pre captures the symptoms observed before the initiation of yoga practices, while VCSS-1 reflects the symptoms after one month of participating in yoga sessions. In plot (a), the features are ranked by their influence on the VCSS-Pre output, with BMI_pre (Body Mass Index before yoga) and PR_pre (Pulse Rate before yoga) showing significant impacts across all severity levels—Severe, Absent, Moderate, and Mild. The RIGHT CALF-CIR-PRE (right calf circumference before yoga) and other physical indicators like the ABPI_pre (Ankle-Brachial Pressure Index before yoga) also contribute

substantially to the prediction outcomes. Plot (b) of the figure highlights the impact of input features on the VCSS-1 outcome. Notably, HCY_Pre (Homocysteine levels before yoga) and CFS_Pre (Chronic Fatigue Syndrome before yoga) emerge as the most influential features, particularly in the Severe category. Other factors, such as CVIQ_social_pre (Chronic Venous Insufficiency Questionnaire-social before yoga) and SBPA_Pre (Systolic Blood Pressure Type A before yoga), also show notable contributions to the model's predictions post-yoga intervention. The colored bars represent different severity levels, with each feature influencing these categories to varying degrees. This detailed analysis helps identify which features are most critical in predicting varicose vein symptoms and how these influences change after yoga intervention. The visual breakdown provides a clear understanding of how input variables contribute to the model's outputs, offering valuable insights for future research and potential therapeutic strategies in managing CVI.

VI. Conclusion

Chronic Venous Insufficiency (CVI) is characterized by impaired blood flow to the heart due to damaged or weakened valves in the leg veins, resulting in symptoms such as swelling, pain, skin alterations, and, more severe instances, persistent ulcers. This article aims to assess the impact of yoga exercises on alleviating symptoms in individuals with CVI. Employing a machine learning technique, the study predicts and categorizes symptoms before and one month after participating in yoga exercises. To ensure optimal performance, the selected model, Gaussian Process Classifier (GPC), for this study went through final optimization utilizing two optimizers, namely the Crystal Structure Algorithm (CSA) and the Fire Hawk Optimizer (FHO). A comprehensive evaluation was performed on 100 patients diagnosed with CVI. This assessment employed key criteria, including $Accuracy, Precision, Recall$, and $F1-score$. The resulting outcomes are presented as follows. In VCSS-Pre, concerning the F1-Score criterion, the GPFH model outperforms the GPCS and GPC models by 1.26% and 2.18%, respectively. Additionally, this superiority in Recall quality translates to 1.15% and 2.3%, respectively. Moreover, In the prediction of VCSS_1, the GPFH model demonstrates superior performance in F1_Score and Recall criteria compared to the GPCS (1% and 1.1%, respectively) and GPC (2.2% for both criteria) models.

Future studies on CVI and its management through yoga and machine learning-driven prediction models could explore several promising avenues. First, expanding the study to include a larger and more diverse patient cohort would enhance the generalizability of the findings, allowing for a more comprehensive understanding of the effectiveness of yoga in various demographic groups and across different stages of CVI. Additionally, longitudinal studies tracking patients over an extended period could provide valuable insights into the long-term effects of yoga on CVI symptoms, as well as any potential cumulative benefits. Another important area for future research is the integration of other complementary therapies with yoga, such as dietary modifications, physical therapy, or mindfulness practices. Investigating the combined impact of these approaches could lead to a more holistic treatment framework, addressing not only the physical but also the psychological aspects of CVI. On the machine learning front, future studies could explore the application of more advanced models, such as deep learning or ensemble methods, to enhance prediction accuracy and robustness. Additionally, incorporating real-time monitoring and data collection through wearable devices could provide more granular data, enabling more precise symptom tracking and personalized intervention strategies.

References

[1] T. S. Church et al., “Trends over 5 decades in US occupation-related physical activity and their associations with obesity,” PLoS One, vol. 6, no. 5, p. e19657, 2011.

[2] S. R. Patel et al., “Real-world experiences with yoga on cancer-related symptoms in women with breast cancer,” Glob Adv Health Med, vol. 10, p. 2164956120984140, 2021.

[3] J. Bames et al., “Standardized use of the terms" sedentary" and" sedentary behaviours",” Applied Physiology Nutrition and Metabolism-Physiologie Appliquee Nutrition Et Metabolisme, vol. 37, pp. 540–542, 2012.

[4] M. T. Hamilton, D. G. Hamilton, and T. W. Zderic, “Sedentary behavior as a mediator of type 2 diabetes,” Diabetes and Physical Activity, vol. 60, pp. 11–26, 2014.

[5] F. Q. S. Dzakpasu et al., “Musculoskeletal pain and sedentary behaviour in occupational and non-occupational settings: a systematic review with meta-analysis,” International Journal of Behavioral Nutrition and Physical Activity, vol. 18, no. 1, pp. 1–56, 2021.

[6] D. Shen et al., “Sedentary behavior and incident cancer: a meta-analysis of prospective studies,” PLoS One, vol. 9, no. 8, p. e105709, 2014.

[7] S. C. Gilchrist et al., “Association of sedentary behavior with cancer mortality in middle-aged and older US adults,” JAMA Oncol, vol. 6, no. 8, pp. 1210–1217, 2020.

[8] S. Carter, Y. Hartman, S. Holder, D. H. Thijssen, and N. D. Hopkins, “Sedentary Behavior and Cardiovascular Disease Risk: Mediating Mechanisms,” Exerc Sport Sci Rev, vol. 45, no. 2, 2017.

[9] A. K. Chomistek et al., “Relationship of sedentary behavior and physical activity to incident cardiovascular disease: results from the Women’s Health Initiative,” J Am Coll Cardiol, vol. 61, no. 23, pp. 2346–2354, 2013.

[10] N. M. Hamburg, “The legs are a pathway to the heart: connections between chronic venous insufficiency and cardiovascular disease,” Eur Heart J, vol. 42, no. 40, pp. 4166–4168, Oct. 2021, doi: 10.1093/eurheartj/ehab589.

[11] J. D. Raffetto and R. A. Khalil, “Mechanisms of lower extremity vein dysfunction in chronic venous disease and implications in management of varicose veins,” Vessel Plus, vol. 5, 2021.

[12] A. Thibert, N. Briche, B. D. Vernizeau, F. Mougin-Guillaume, and S. Béliard, “Systematic review of adapted physical activity and therapeutic education of patients with chronic venous disease,” J Vasc Surg Venous Lymphat Disord, 2022.

[13] H. D. Vlajinac, Đ. J. Radak, J. M. Marinković, and M. Ž. Maksimović, “Risk factors for chronic venous disease,” Phlebology, vol. 27, no. 8, pp. 416–422, 2012.

[14] E. Halliwell, K. Dawson, and S. Burkey, “A randomized experimental evaluation of a yoga-based body image intervention,” Body Image, vol. 28, pp. 119–127, 2019.

[15] M. Anusha, S. Dubey, P. S. Raju, and I. A. Pasha, “Real–time yoga activity with assistance of embedded based smart yoga mat,” in 2019 2nd International Conference on Innovations in Electronics, Signal Processing and Communication (IESC), IEEE, 2019, pp. 1–6.

[16] T. Field, “Yoga research review,” Complement Ther Clin Pract, vol. 24, pp. 145–161, 2016.

[17] D. Neumark-Sztainer, A. W. Watts, and S. Rydell, “Yoga and body image: How do young adults practicing yoga describe its impact on their body image?,” Body Image, vol. 27, pp. 156–168, 2018.

[18] D. Kumar and A. Sinha, Yoga pose detection and classification using deep learning. LAP LAMBERT Academic Publishing London, 2020.

[19] T. Y. Park et al., "Epidemiological trend of pulmonary thromboembolism at a tertiary hospital in Korea," Korean J Intern Med, vol. 32, no. 6, p. 1037, 2017.

[20] D. Morrone and V. Morrone, "Acute pulmonary embolism: focus on the clinical picture," Korean Circ J, vol. 48, no. 5, pp. 365–381, 2018.

[21] A. Pizano et al., "Association between cardiac conditions with venous leg ulcers in patients with chronic venous insufficiency," Phlebology, vol. 38, no. 4, pp. 281–286, 2023.

[22] W. A. Marston et al., "Economic benefit of a novel dual-mode ambulatory compression device for treatment of chronic venous leg ulcers in a randomized clinical trial," J Vasc Surg Venous Lymphat Disord, vol. 8, no. 6, pp. 1031–1040, 2020.

[23] Y. Kim, C. Y. M. Png, B. J. Sumpio, C. S. DeCarlo, and A. Dua, "Defining the human and health care costs of chronic venous insufficiency," in Seminars in Vascular Surgery, Elsevier, 2021, pp. 59–64.

[24] N. Maddukuri and S. R. Ummity, "Yoga Pose prediction using Transfer Learning Based Neural Networks," 2023.

[25] J. Azar, A. Rao, and A. Oropallo, "Chronic venous insufficiency: a comprehensive review of management," J Wound Care, vol. 31, no. 6, pp. 510–519, 2022.

[26] I. Sudoł-Szopińska, A. Bogdan, T. Szopiński, A. K. Panorska, and M. Kołodziejczak, "Prevalence of chronic venous disorders among employees working in prolonged sitting and standing postures," International journal of occupational safety and ergonomics, vol. 17, no. 2, pp. 165–173, 2011.

[27] M. A. Passman et al., "Validation of venous clinical severity score (VCSS) with other venous severity assessment tools from the American venous forum, national venous screening program," J Vasc Surg, vol. 54, no. 6, pp. 2S-9S, 2011.

[28] R. Zulpe, "An Experimental study of yoga therapy on varicose vein," 2023.

[29] M. A. Chaoul and L. Cohen, "Rethinking yoga and the application of yoga in modern medicine," Crosscurrents, vol. 60, no. 2, pp. 144–167, 2010.

[30] T. Field, "Yoga research review," Complement Ther Clin Pract, vol. 24, pp. 145–161, 2016.

[31] M. Kuss, C. E. Rasmussen, and R. Herbrich, "Assessing Approximate Inference for Binary Gaussian Process Classification.," Journal of machine learning research, vol. 6, no. 10, 2005.

[32] M. Azizi, S. Talatahari, and A. H. Gandomi, "Fire Hawk Optimizer: A novel metaheuristic algorithm," Artif Intell Rev, vol. 56, no. 1, pp. 287–363, 2023.

[33] B. A. Averill and P. Eldredge, "Chemistry: principles, patterns, and applications," (No Title), 2007.

[34] S. Talatahari, M. Azizi, M. Tolouei, B. Talatahari, and P. Sareh, "Crystal structure algorithm (CryStAl): a metaheuristic optimization method," IEEE Access, vol. 9, pp. 71244–71261, 2021.

[35] C. Pete et al., "Crisp-Dm 1.0—Step-by-step data mining guide," Cris. Consort, p. 76, 2000.

[36] M. Botlagunta et al., "Classification and diagnostic prediction of breast cancer metastasis on clinical data using machine learning algorithms," Sci Rep, vol. 13, no. 1, p. 485, 2023.

[37] L. Torgo, Data mining with R: learning with case studies. chapman and hall/CRC, 2011.

A Semantic Segmentation Method for Road Scene Images Based on Improved DeeplabV3+ Network

Lihua Bi[1], Xiangfei Zhang[2], Shihao Li[3], Canlin Li[2]
School of Software Engineering, Zhengzhou University of Light Industry, Zhengzhou, China[1]
School of Computer Science and Technology, Zhengzhou University of Light Industry, Zhengzhou, China[2]
School of Information Science and Technology, Beijing Forestry University, Beijing, China[3]

***Abstract*—Semantic segmentation of road scenes plays a crucial role in many fields such as autonomous driving, intelligent transportation systems and urban planning. Through the precise identification and segmentation of elements such as roads, pedestrians, vehicles, and traffic signs, the system can better understand the surrounding environment and make safe and effective decisions. However, the existing semantic segmentation technology still faces many challenges in the face of complex road scenes, such as lighting changes, weather effects, different viewing angles and the existence of occlusions. Combined with the actual road scene image, this paper improves DeeplabV3+ network and applies it to semantic segmentation of road scene image, and proposes a semantic segmentation method of road scene image based on improved DeeplabV3+ network. By adding enhancement strategies for road scene images and hyperparameter adjustment, the method improves the training process of DeeplabV3+ network, and uses SK attention mechanism to improve the feature fusion module in DeeplabV3+, so as to improve the segmentation effect of road scene images. After the validation of Cityscapes and other data sets, the segmentation accuracy index mIoU of the proposed method reaches 79.8%, which can predict better semantic style effect, effectively improve the segmentation performance and accuracy of the model, and achieve better segmentation index results in the comparison network, and the subjective visual effect of the segmentation is also better.**

Keywords*—Image enhancement; attention mechanism; semantic segmentation; road scene images

I. INTRODUCTION

Image semantic segmentation [1] aims to provide richer image semantic information for pixel-level image classification tasks. The need for semantic segmentation is crucial for applications in scenarios that require high-precision target segmentation, such as autonomous driving [2-4], environmental monitoring [5-7], augmented reality [8-10], and security surveillance [11-13]. By accurately segmenting objects in an image, more accurate scene analysis, target recognition and decision making can be achieved, thus enhancing system performance and application experience.

In recent years, the field of automatic driving is constantly developing, and the semantic segmentation technology of road scene plays an important role in the automatic driving system. Image semantic segmentation provides the autonomous driving system with rich road information and high-level understanding of the image, including the classification and accurate positioning of the target, so that the autonomous vehicle can fully understand the complex traffic situation around. However, there are still some specific problems in the existing road scene semantic segmentation technology, such as: insufficient robustness under different lighting conditions, which leads to the deviation of segmentation results. Due to the influence of bad weather (such as rain, snow, fog), the segmentation accuracy decreases significantly. And in crowded and dynamic scenes, it is easy to appear occlusion and confusion. These deficiencies limit the decision-making ability of autonomous vehicles in complex environments, and further research is needed to address these challenges to improve system safety and reliability.

With the continuous development of hardware level and computing power, the rapid development and application of deep learning technology provides new ideas for semantic segmentation research, and deep learning methods can significantly improve the accuracy of semantic segmentation. Full Convolutional Network (FCN) [14] is an important method to deal with image segmentation tasks using deep learning technology, which opens a new era of achieving high-precision semantic segmentation with deep learning as the core technology. However, FCN still has some shortcomings, for example, the relationship between pixels is not fully utilized and the results obtained are still not fine enough. Researchers have gone on to propose many network models with better segmentation results based on different technical features. Badrinarayanan et al. proposed SegNet [15] based on the structure of codec [16]. The innovation of SegNet network is that it reduces the number of fully-connected layers as well as the number of parameters and storage space of the streamlined model, and also outputs the indexing information in the pooling process to improve the image segmentation accuracy and efficiency of the decoding process. Noh et al. proposed DeconvNet [17], which improves the segmentation performance by introducing an inverse convolution layer at the decoder side to recover the resolution of the feature map through the coding-decoding structure. Olaf Ronneberger et al. proposed U-Net [18], whose symmetric coding and decoding network structure captures semantic information at different levels, pinpoints feature map information during up and down sampling, and preserves more information about the original image.

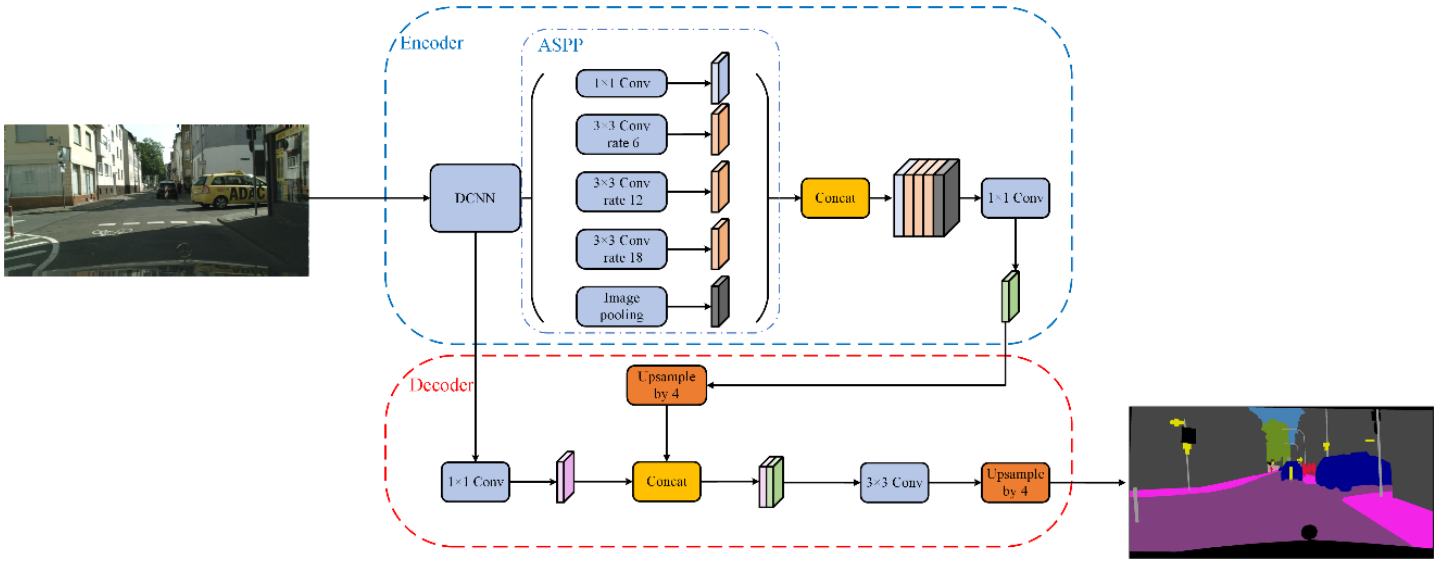


Fig. 1. DeeplabV3+ network architecture.

PSPNet [19] uses the ResNet [20] network as a backbone network, with modifications that use an additional auxiliary loss function, and two losses are assigned different weights, which promotes rapid convergence of the model. Overall, the traditional symmetric coding and decoding structure has more training parameters and complex structure, so it is less effective for the application of scenarios with real-time requirements.

The Deeplab network was proposed by the Google Brain team to solve the problems of category imbalance, voids, and edge details that are difficult to handle in semantic segmentation. Nowadays, Deeplab has developed a series of networks that are constantly employing new techniques and structural optimization algorithms to enhance their performance in various domains. DeepLabV1 [21] utilizes dilated convolutions to expand the receptive field and employs a fully connected conditional random field to enhance detail capturing capability, refining the segmentation object edges. DeepLabV2 [22] introduces the ASPP module on the basis of DeeplabV1, utilizing dilated convolutions with different dilation rates to extract feature information at different scales, enhancing the model's adaptability to objects of different scales, and also incorporating inverse convolutions and batch normalization techniques. DeeplabV3 [23] further expands the depth and width of the network on the basis of V2, adopts cascading or parallel mode to arrange cavity convolution with different cavity rates, optimizes ASPP module, and captures multi-scale features more effectively. DeeplabV3+ [24] extends DeepLabv3 by adopting an encoder-decoder structure to achieve better semantic segmentation performance. This paper proposes a semantic segmentation method for road scene images based on improved DeeplabV3+ network. The backbone network uses lightweight MobileNetV2, combined with the actual application of road scene image, to enhance image data and hyperparameter adjustment, aiming at improving the accuracy and efficiency of semantic segmentation of road scene image. We introduced SK attention module to optimize the feature fusion module in DeeplabV3+ to enhance the model's ability to capture key features. This improvement not only improves segmentation accuracy, especially in complex environments such as bright light, shadows, and dynamic scenes, but also helps reduce computational costs.

The main contributions of this paper are summarized as follows:

- This paper proposes a road scene image semantic segmentation method based on improved DeeplabV3+ network. To reduce the complexity of DeeplabV3+ network, lightweight MobileNetV2 is used as the backbone network.
- Increase the diversity of training data through image enhancement strategies, optimize the generalization of DeeplabV3+ network, and optimize the training process through hyperparameter adjustment.
- By introducing SK attention mechanism to optimize the feature fusion module in DeeplabV3+, adjust the feature weights and improve the accuracy of semantic segmentation.

The rest of this article is structured as follows: Section II describes the network structure of DeeplabV3+. In Section III, the semantic segmentation method of road scene image based on the improved DeeplabV3+ network is introduced in detail. Section IV describes the experimental setup of this paper and the experimental results with other methods. The final conclusion is given in Section V.

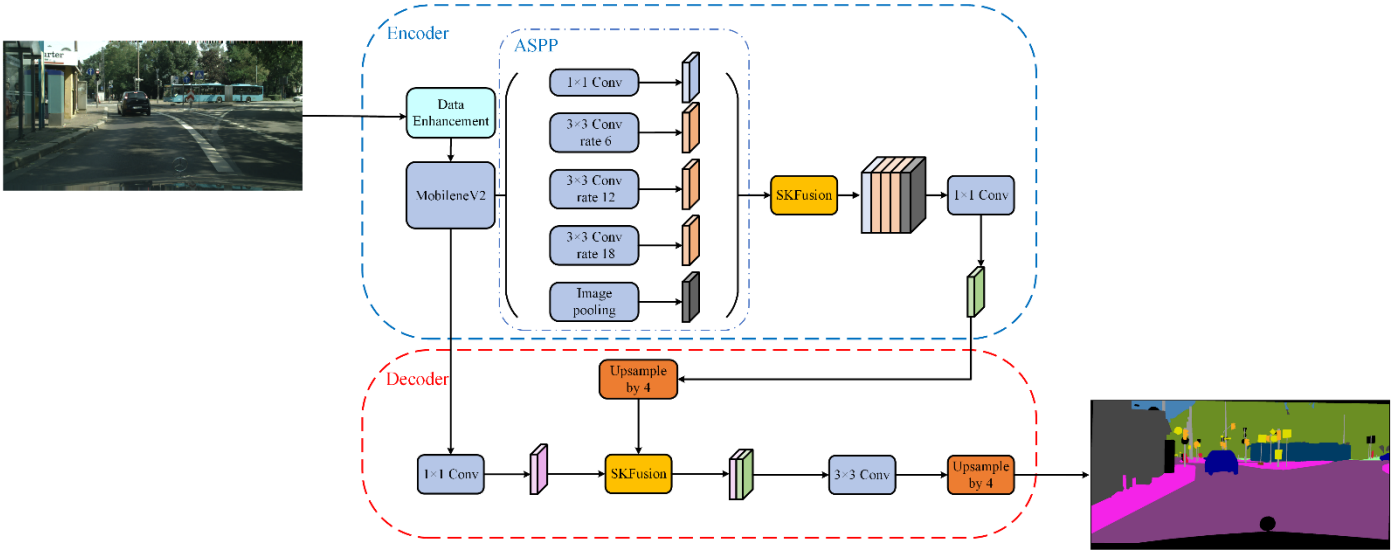


Fig. 2. Overall framework of road scene semantic segmentation based on improved DeeplabV3+.

II. DeepLabV3+ Network

The DeepLabv3+ network architecture is shown in Fig. 1. DeepLabv3+ uses an encoder-decoder structure to improve DeepLabv3. The encoder uses Atrous Spatial Pyramid Pooling (ASPP) to capture context information at different scales, while the decoder refines the target boundary to improve segmentation results. The ASPP module is used to capture semantic information of different scales, splice the feature information after multiple empty convolution operations with different sampling rates, and obtain the feature after 1×1 convolution. Another branch of the Deep Convolutional Neural Network (DCNN) uses 1×1 convolution to process the underlying features of the image to obtain the underlying features of the image. Then the feature is fused with the feature that has been subsampled four times, and the semantic segmentation prediction image is obtained after 3×3 convolution and four times subsampling.

III. A Semantic Segmentation Method for Road Scene Images Based on Improved DeepLabV3+ Network

A. The Overall Framework of the Proposed Methodology

In this paper, DeepLabv3+ model is used to improve image semantic segmentation in road scenes. In the semantic segmentation of road scene image based on DeepLabv3+ network, lightweight MobileNetV2 is adopted as the backbone in this paper. In the training process, the data is enriched by image enhancement. After passing through MobileNetV2 network, ASPP method is used to extract multi-scale feature information from images by different sampling rates. SK attention mechanism [25] is introduced to improve the feature fusion module, and feature fusion is performed on the feature mapping obtained by ASPP module to improve DeepLabv3+ network. The SK feature fusion module is also used in the later feature fusion of encoder and decoder. The improved Deeplabv3+ model is shown in Fig. 2.

B. Image Enhancement

Image enhancement refers to the process of generating new training samples through a series of transformation operations on the original image during image processing or pattern recognition tasks. Through image enhancement, the diversity of training data can be increased, which helps to improve the generalization ability and robustness of the model. The operation process of image enhancement in this paper is shown in Fig. 3.

1) Random left-right flip: In order to increase the diversity of data and make the model have better generalization ability and segmentation effect, the left and right flip of all the training set images is carried out with 50% probability. The same original image was randomly flipped left and right twice, and the resulting image was shown in Fig. 4.

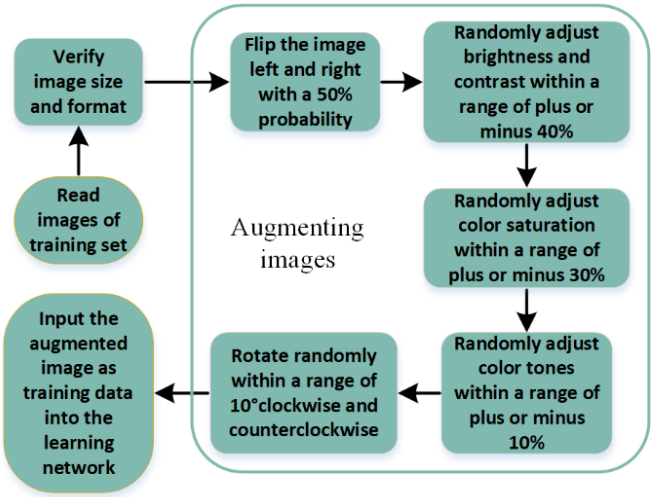


Fig. 3. Improved image augmentation operation process.



(a) Example 1 of result image after randomly flipping



(b) Example 1 of result image after randomly flipping

Fig. 4. Schematic diagram of random left-right flip.

2) Random color adjustment: In the actual road scene image, may encounter a variety of different weather during driving, or due to the impact of camera shooting, resulting in a large difference in the brightness of the input image, so you need to adjust the image brightness, contrast, saturation and tone to increase the generalization ability of color. The pre-processed images are color-adjusted, the brightness and contrast are randomly adjusted within the range of plus or minus 40%, the color saturation is randomly adjusted within the range of plus or minus 30%, and the hue is randomly adjusted within the range of plus or minus 10%. Two random color adjustments were made to the same original training image, and the results were shown in Fig. 5.



(a) Example 1 of result image after randomly adjusting color



(b) Example 2 of result image after randomly adjusting color

Fig. 5. Schematic diagram of random color adjustment.



(a) Example 1 of result image from random rotation



(b) Example 2 of result image from random rotation

Fig. 6. Schematic diagram of random rotation angle.

3) Random rotation angle: In order to increase the generalization ability of oblique images, this paper added the processing of random rotation Angle, set the center of the image as the rotation center, and rotate the image randomly within the range of 10° clockwise and counter clockwise. The same picture was randomly rotated for two times, and the result was shown in Fig. 6.

C. SKFusion

When people observe things, they will selectively pay attention to the more important information, which is called attention. By continuously focusing on this key location to get more information and ignoring other useless information, this visual attention mechanism greatly improves the efficiency and accuracy of our processing on information. The attention mechanism in deep learning is similar to the attention mechanism in human vision, which is to focus attention on the important points with more information, select the key information, and ignore other unimportant information.

SKFusion in the improved DeepLabv3+ network is to adjust the weight of the feature map by using the SK attention mechanism after the concatenated features. The operation process is shown in Fig. 7. We first splice n features together by Concat operation, as shown in formula (1):

$$X = Concat(x_1, ..., x_n) \tag{1}$$

where, $(x_1, ..., x_n)$ a is n eigenvectors.

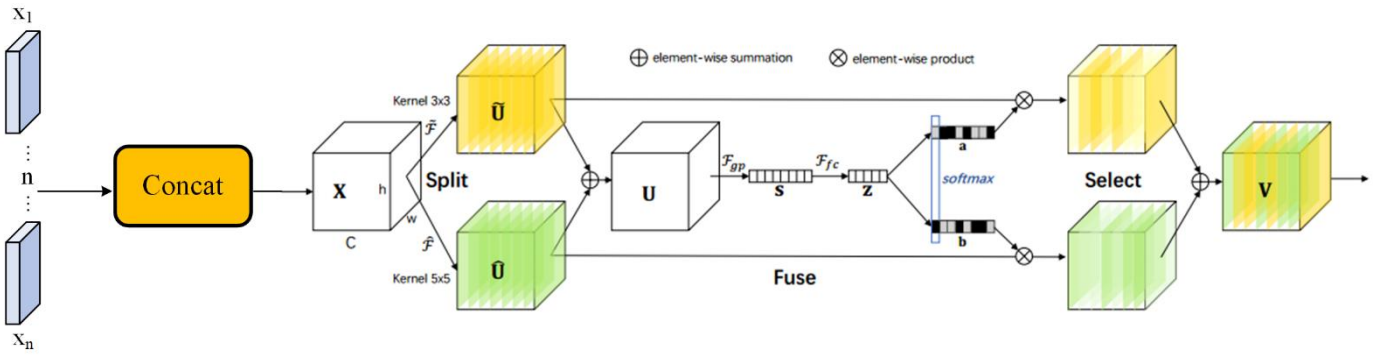


Fig. 7. SKFusion network architecture.

Then the obtained feature X is processed using the SK attention mechanism, and the processed feature is obtained, as shown in formula (2):

$$V = SKAttention(X) \tag{2}$$

SK attention mechanism [25] is mainly divided into three operations: Split, Fuse and Select. The Split operator produces multiple paths of different kernel sizes. The Fuse operators combine and aggregate information from multiple paths to obtain a global and comprehensive representation for selecting weights. The Select operator aggregates feature maps of cores of different sizes based on selection weights.

a) Split: Convolve the input feature graph X through a cavity of different receptive fields. Fig. 7 represents two groups of convolution operations, one with a convolution kernel of 3*3 to obtain the feature graph, and the other with a convolution kernel of 5*5 to obtain the convolution kernel, as shown in formulas (3) and (4):

$$\tilde{U} = Conv_{3*3}(X) \tag{3}$$

$$\hat{U} = Conv_{5*5}(X) \tag{4}$$

b) Fuse: To ensure that the information flow from multiple branches carries information of different sizes into the next layer of neurons, first fuse the results of multiple branches (two branches in Fig. 7) by summing the elements, as shown in formula (5):

$$U = \tilde{U} + \hat{U} \tag{5}$$

Then, the global average pooling of U is performed to obtain s, and the dimensionality is reduced by the fully connected layer to improve the efficiency and z is obtained, as shown in formulas (6) and (7):

$$s = F_{gp}(U) = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} U(i, j) \tag{6}$$

$$z = F_{fc}(s) \tag{7}$$

Where, F_{gp} represents the global average pooling operation, H×W is the spatial dimension size, and F_{fc} represents the fully connected layer.

c) Select: A new feature map computed from convolution kernels with different weights. First do softmax to calculate the weight of each convolution kernel, if there are two convolution cores, then a+b=1. Then, the final feature graph V is obtained by multiplying the weight elements of each convolution kernel, as shown in formula (8):

$$V = a \cdot \tilde{U} + b \cdot \hat{U}, a + b = 1 \tag{8}$$

D. Hyperparameter Adjustment

In the original training hyperparameters of the model, the batch size was set to 16. However, considering that there is still room for expansion in the memory of the graphics card, we have increased the batch size. Increasing the batch size has three following benefits. Firstly, the memory utilization is increased, and the parallelization efficiency of the large matrix multiplication is increased. Secondly, with the same amount of data, the number of iterations required to run through a training round is reduced, further increasing the processing speed. Thirdly, increasing the batch size within a certain range can make the determined descent direction more accurate, thus reducing training oscillations.

When the batch size increases to a certain extent, its determined descent direction may have been largely unchanged. However, since the accuracy of the final convergence is affected by a variety of factors, such as network structure, learning rate, etc., it does not mean that the larger batch size will necessarily get better results. In practice, when the batch size is increased to some extent, to achieve the optimal convergence accuracy, factors such as iteration need to be considered. So the adjustment strategy of our proposed method is to increase the batch size to 32.

For the epoch hyperparameter, the epoch was set to 30 in the initial experiment, which still has room for improvement. The complete process of running the model to complete one forward propagation and back propagation on all the data is called 1 training round, in other words, it means that all the training samples in the dataset have been trained once.

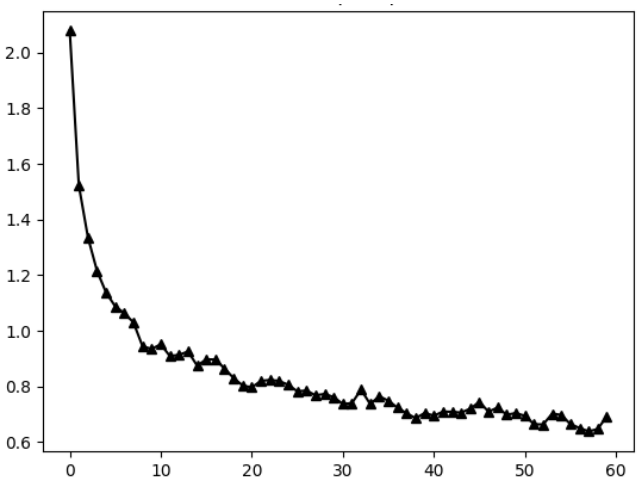


Fig. 8. Loss fluctuation per training set round.

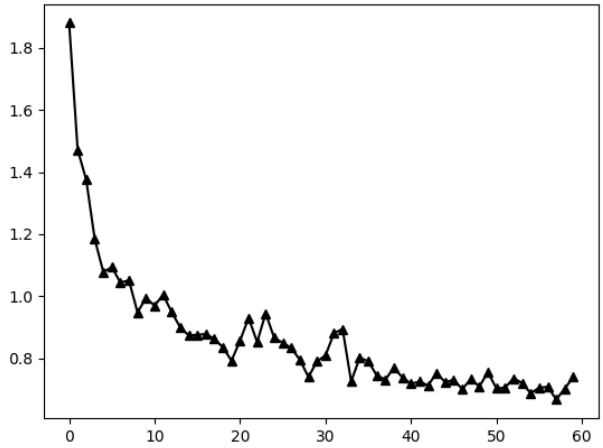


Fig. 9. Loss fluctuation per validation set round.

In the training process of gradient descent model, the neural network gradually transitions from the unfit state to the optimal fit state, then it will enter the overfitting state after reaching the optimal state. However, the number epoch of training rounds is not the larger the better, and it is generally set to between 20 and 200 to achieve good training results. The more diverse the data, the larger the corresponding training rounds. Our proposed method adjusts the epoch to 60.

We visualize the fluctuation of the loss function with the number of training rounds. Both the training and validation sets use the images contained in Cityscapes [26]. The loss fluctuations of training set are shown in Fig. 8, and the loss fluctuations of validation set are shown in Fig. 9.

Through recording the fluctuations of loss function during the training process of model, as can be seen from Fig. 8 and Fig. 9, as the number of training rounds increases, the loss decreases and the segmentation effect gradually improves. Due to the Adam optimization algorithm, after 40 rounds, the learning rate will gradually decrease after being adjusted by the Adam algorithm, so the weight change will also decrease, and the change of corresponding loss function will be minimal. After 60 rounds of training, a model with relatively good segmentation performance can already be obtained.

IV. Experimental Design and Analysis

A. Dataset

The rapid development of deep learning cannot be separated from the development of training data, and the preprocessed dataset can greatly facilitate the training process of image processing without having to put too much effort on the labeling of the dataset. In this paper, we choose the representative dataset Cityscapes [26] as the dataset for semantic segmentation of road scenes. Cityscapes is an image dataset for urban street scenes, and this dataset contains 5000 images that have been annotated at high quality pixel level and covers 19 categories with dense pixel annotations, among which 8 categories have instance-level segmentation. The dataset is divided into three parts: training, validation and testing with 2975, 500 and 1525 images, respectively. In addition, the dataset contains stereoscopic video sequences from street scenes in 50 different cities. The examples on raw and pixel-level labeled images of the Cityscapes dataset are shown in Fig. 10 and Fig. 11, respectively.



Fig. 10. Original image of the road scene.

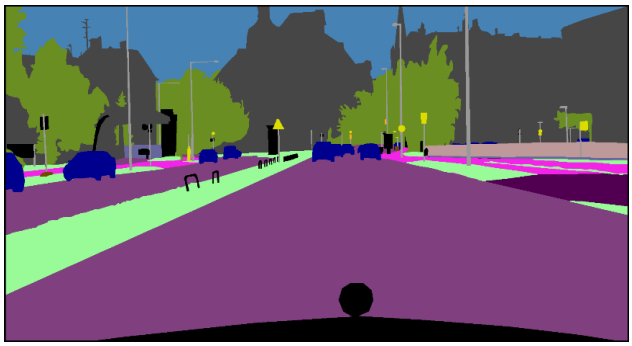


Fig. 11. Pixel-level labeled image.

Cityscapes is very large, with 20,000 weakly annotated frames in addition to 5,000 images with high-quality pixel-level annotations. In addition, the Cityscapes dataset provides fine and coarse metrics. The fine evaluation criterion is based on 5,000 images with fine labels, while the coarse evaluation criterion is based on 5,000 images with fine labels and 20,000 images with coarse labels.

B. Experimental Environment and Parameter Settings

Performing image processing using deep learning involves a large number of floating-point and matrix operations, and it has high requirements for the hardware and software environments, which can affect the effectiveness of deep learning model. The hardware environment for our

experiments involves CPU processor Intel i7-9750H and 128G memory, as well as GPU processor NVIDIA GeForce GTX 1650 with 4G graphics memory. In terms of software environment, Windows 10 64-bit operating system is used, Pytorch deep learning framework is selected, and the classic general-purpose parallel computing architecture CUDA is applied. In addition, dependent libraries such as Numpy, OpenCV, and PIL are also used.

In order to ensure that the segmentation results are only affected by the model itself, the same hardware configuration and software parameter settings are used for different comparison models, so as to ensure the consistency of the experimental environments. For each comparison model, the experimental software and hardware environments are shown in Table I and Table II.

TABLE I. EXPERIMENTAL HARDWARE CONFIGURATION

Hardware	Configuration
CPU	Intel i7-9750H@2.60GHz
GPU	NVIDIA GeForce GTX 1650
Memory	128G
Video memory	4G

TABLE II. EXPERIMENTAL SOFTWARE CONFIGURATION

Software	Configuration
Operating system	Windows10 64-bit
Deep Learning Framework	Pytorch1.12
Programming language	Python3.8
Parallel computing architecture	CUDA 11.6
Main Dependency Libraries	Numpy、Matplotlib、Opencv

In this paper, the uniform parameter settings for model training are as follows. Cross-Entropy Loss (CE Loss) [27] with multiple classes is used. The optimizer adopts Stochastic Gradient Descent (SGD) strategy, where the learning momentum parameter is set to 0.9 and the weight decay is set to 0.00001. The initial learning rate LR is 0.0001, and the learning rate is dynamically adjusted by a poly strategy, which dynamically decreases with the increase of training iterations, and the current learning rate new_lr is updated as shown in formula (9).

$$new_lr = LR * (1 - \frac{epo}{\max_epo})^{power} \quad (9)$$

Where momentum power is 0.9, epo indicates the current number of training iteration, and the maximum number of training iteration max_epo is calculated as shown in formula (10).

$$\max_epo = (\frac{M}{batchsize}) * epoch \quad (10)$$

Where M is the number of training samples 2975, the number epoch of training rounds is the total number of rounds that need to be trained, uniformly set to 60. According to the model size and the graphics memory, batchsize is set to 20. After the above parameter settings, all of comparison models can achieve good convergence results.

C. Comparison of Experimental Effect

To verify the effectiveness of the proposed method, several typical deep learning semantic segmentation networks were selected for experiments, including FCN, SegNet, DeeplabV3 and DeeplabV3+. By comparing with these networks, we aim to comprehensively evaluate the performance of our proposed methods in different scenarios. In the experiment, the segmentation effect is compared on the images of verification set as well as the campus images taken in the field to ensure the universality and reliability of the results. In addition, we will use multiple evaluation metrics such as mean crossover ratio (mIoU) and pixel accuracy (pixAcc) to provide a more comprehensive performance analysis.

1) Comparison of subjective effect: In this paper, five segmentation network models including FCN, SegNet, DeeplabV3, DeeplabV3+ and our improved DeeplabV3+ are applied to test and verify the semantic segmentation of the same image. Fig. 12 shows the segmentation effect diagram of some pictures. Columns (b), (c), (d), (e) and (f) respectively represent the segmentation effect corresponding to FCN, SegNet, DeeplabV3, DeeplabV3+ and our improved DeeplabV3+ network. Images 1 to 5 were selected from part of the validation set in the Cityscapes dataset, including several common road scenes. Images 6 and 7 were two road scenes in the university.

Compared with the segmentation effect of all images, it can be seen from Fig. 12 that SegNet has a slightly higher segmentation accuracy than FCN in general, but the segmentation effect needs to be improved and there are some problems of inaccurate segmentation of details. DeeplabV3 has a better segmentation effect than SegNet and can accurately predict the classification of some detailed images. Our improved DeeplabV3+ works best, with clearer categories. SegNet cannot clearly divide the shape of the category when dividing the categories such as cars and houses. FCN has some deviation when it does not distinguish pedestrian categories well, and the edges are fuzzy, while DeeplabV3 and DeeplabV3+ still have good segmentation effect. Since our improved DeeplabV3+ optimizes the training strategy through image enhancement and hyperparameter adjustment, and uses the SKFusion module DeeplabV3+ network structure, it is more sensitive to small targets and other information, and the segmentation effect is better.

It can be seen that the performance of the improved model at the junction of the motorway and the sidewalk has been significantly improved, and the jagged segmentation phenomenon of the original model has been successfully solved, making the boundary smoother and more natural. This improvement not only improves the visual effect, but also provides a more reliable basis for the division of pedestrians, vehicles and buildings in practical applications. In the segmentation results of images 6 and 7, we can clearly see that the improved DeeplabV3+ model shows higher accuracy and detail in processing various categories. Especially in the

contours of pedestrians, cars and houses, the model can effectively distinguish the boundaries of complex backgrounds and reduce blurring. This further verifies the effectiveness and superiority of the model in multi-class segmentation tasks.

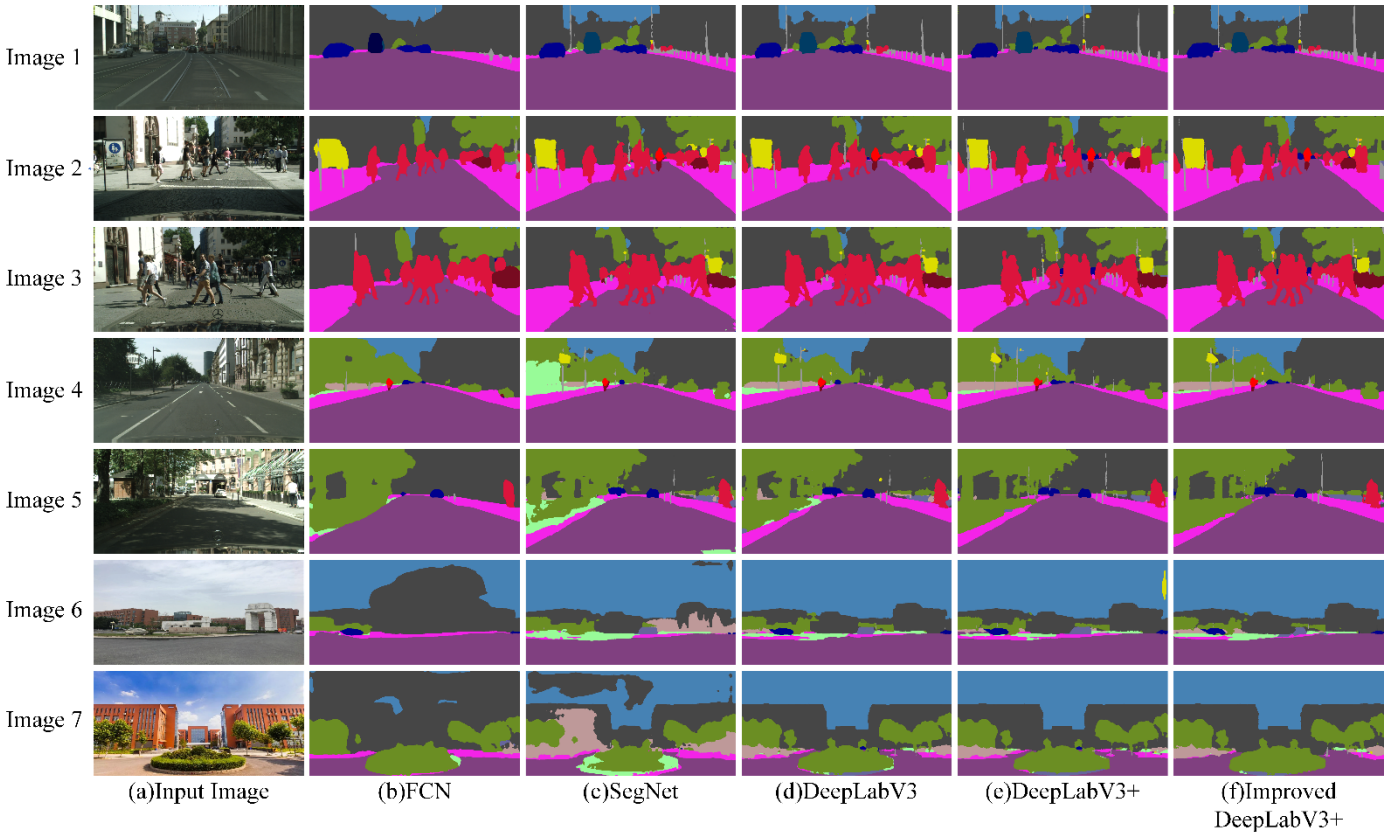


Fig. 12. Comparison of subjective effects. (a)Input Image; (b)FCN; (c)SegNet; (d)DeeplabV3; (e)DeeplabV3+; (f) Improved DeeplabV3+.

2) Comparison of objective effects: Regarding the objective performance evaluation metrics for semantic segmentation on road scene images, this paper adopts the commonly used evaluation metrics for semantic segmentation, pixAcc (PA) and mIoU. PA is the pixel accuracy rate, which is relatively simple to compute, and is the ratio of the number of correctly predicted pixels to the total predicted pixels. mIoU is a commonly used evaluation metric for semantic segmentation tasks, i.e., the average intersection and union ratio which is the ratio of the intersection and union of the true and predicted values. In semantic segmentation, the intersection and union ratio of a single category indicates the ratio of the intersection of the true and predicted values of the category to the union of the category, reflecting the classification accuracy of the model for each category and the overall segmentation effect.

For the test set contained in Cityscapes, the five models FCN, SegNet, DeeplabV3, DeeplabV3+, and our improved DeeplabV3+ are tested and the objective performance evaluation indicators are shown in Table III.

As a whole, combined with Fig. 12, our improved DeeplabV3+ model has better overall segmentation results, with finer segmentation results for some categories such as pedestrians, vehicles, traffic lights, and sign boards, as well as better segmentation results for more detailed categories such as travel lanes. After the validation on the Cityscapes validation set, the segmentation accuracy metric mIoU of our improved DeeplabV3+ reached 79.8%, achieving the best segmentation metric results.

TABLE III. COMPARISON OF OBJECTIVE PERFORMANCE EVALUATION INDICATORS OF FCN, SEGNET, DEEPLABV3, DEEPLABV3+, AND IMPROVED DEEPLABV3+ SEGMENTATION

Model	pixAcc（%）	mIoU（%）
FCN	81.8	32.5
SegNet	84.2	55.4
DeeplabV3	87.6	72.7
DeeplabV3+	89.4	76.5
improved DeeplabV3+	91.2	79.8

V. CONCLUSIONS

For the problem that the semantic segmentation effect of road scene image needs to be improved, this paper proposes a road scene image semantic segmentation method based on the improved DeeplabV3+ network by improving DeeplabV3+ network and applying it to the semantic segmentation of road scene image. The network uses SK attention mechanism to improve the feature fusion module, adjust the feature weights, and optimize the training process by image enhancement and hyperparameter adjustment. Through experiments on cityscape

and other datasets, our method achieves the best segmentation results based on subjective visual effects and objective performance evaluation indicators in the comparison network, among which pixAcc reaches 91.2% and mIoU reaches 79.8% on cityscape dataset. It can be seen that the semantic segmentation effect of our method on road scene image is significantly improved. However, there are two specific limitations to note. Firstly, our method's performance may be compromised in extremely challenging conditions, such as heavy occlusions or severe weather effects, which were not extensively tested in this study. Secondly, while our model achieves high accuracy on urban road scenes, its applicability to rural or less structured environments remains uncertain and may require further adaptation. Future work should focus on developing adaptive models that can learn from real-time data, ensuring robustness in diverse scenarios. Overall, our proposed method not only improves accuracy but also sets the foundation for future innovations in autonomous driving and intelligent transportation systems.

REFERENCES

[1] S Minaee, Y Boykov, F Porikli, et al. “Image segmentation using deep learning: A survey”. IEEE transactions on pattern analysis and machine intelligence, vol.44,2021, pp.3523-3542.

[2] X Li, J Zhang, Y Yang, et al. “Sfnet: Faster and accurate semantic segmentation via semantic flow”. International Journal of Computer Vision, vol.132, 2024, pp.466-489.

[3] D Feng, et al. “Deep multi-modal object detection and semantic segmentation for autonomous driving: Datasets, methods, and challenges”. IEEE Transactions on Intelligent Transportation Systems, vol.22, 2020, pp.1341-1360.

[4] Y Wang, J Zhang, Y Chen, et al. “Automatic learning-based data optimization method for autonomous driving”. Digital Signal Processing, 2024, pp.104428.

[5] A de Silva, R Ranasinghe, A Sounthararajah, et al. “Beyond Conventional Monitoring: A Semantic Segmentation Approach to Quantifying Traffic-Induced Dust on Unsealed Roads”. Sensors, vol.24, 2024, pp.510.

[6] A Alzu'Bi, L Al-Smadi. “Monitoring deforestation in Jordan using deep semantic segmentation with satellite imagery”. Ecol. Informatics, vol.70, 2022, pp.101745.

[7] M Wieland, S Martinis, R Kiefl, et al. “Semantic segmentation of water bodies in very high-resolution satellite and aerial images”. Remote Sensing of Environment, vol.287, 2023, pp.113452.

[8] H Zhang, B Han, et al. “Slimmer: Accelerating 3D Semantic Segmentation for Mobile Augmented Reality”.2020 IEEE 17th International Conference on Mobile Ad Hoc and Sensor Systems (MASS), IEEE, 2020, pp.603-612

[9] S Afzal, IU Khan, I Mehmood, et al. “Leveraging Augmented Reality, Semantic-Segmentation, and VANETs for Enhanced Driver's Safety Assistance”. Computers, Materials & Continua, vol.78, 2024.

[10] D Zhang, L Zhang, J Tang. “Augmented FCN: rethinking context modeling for semantic segmentation”. Science China Information Sciences, vol.66, 2023, pp.142105.

[11] Li L, Dong Z, Yang T,et al. “Deep learning based automatic monitoring method for grain quantity change in warehouse using semantic segmentation”.IEEE Transactions on Instrumentation and Measurement, vlo.70, 2021, pp.1-10.

[12] F Abdullah, A Jalal. “Semantic segmentation based crowd tracking and anomaly detection via neuro-fuzzy classifier in smart surveillance system”. Arabian Journal for Science and Engineering, vol.48, 2023, pp.2173-2190.

[13] Y Wang, Y Shen, B Salahshour, et al. “Urban flood extent segmentation and evaluation from real-world surveillance camera images using deep convolutional neural network”. Environmental Modelling & Software, vol.173, 2024, pp.105939.

[14] J LONG, E SHELHAMER, T DARRELL. “Fully convolutional networks for semantic segmentation”. IEEE Computer Society, 2017, pp.3431-3440.

[15] V BADRINARAYANAN, A KENDALL, R CIPOLLA. “Segnet:a deep convolutional encoder-decoder architecture for image segmentation”. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol.39, 2019, pp.2481-2495.

[16] C Kyunghyun, et al. “Learning Phrase Representations using RNN Encoder–Decoder for Statistical Machine Translation.” Conference on Empirical Methods in Natural Language Processing , 2014.

[17] H Noh, S Hong and B Han, "Learning Deconvolution Network for Semantic Segmentation," 2015 IEEE International Conference on Computer Vision (ICCV), 2015, pp. 1520-1528.

[18] O Ronneberger, P Fischer, and T Brox. “U-net: Convolutional networks for biomedical image segmentation”. In International Conference on Medical image computing and computer-assisted intervention, 2015, pp.234–241.

[19] H Zhao, J Shi, X Qi, et al. “Pyramid scene parsing network”. Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, pp.2881-2890.

[20] K He, et al., "Deep Residual Learning for Image Recognition". 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp.770-778.

[21] L C Chen, G Papandreou, I Kokkinos, et al. “Semantic Image Segmentation with Deep Convolutional Nets and Fully Connected CRFs”. arXiv, 2014.

[22] L C Chen, G Papandreou, I Kokkinos, et al. “Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs”. IEEE transactions on pattern analysis and machine intelligence, 2017, pp.834-848.

[23] L C Chen, G Papandreou, F Schroff, H Adam. “Rethinking atrous convolution for semantic image segmentation.” 2017, arXiv:1706.05587.

[24] L C Chen, Y Zhu, G Papandreou, et al. “Encoder-decoder with atrous separable convolution for semantic image segmentation”. Proceedings of the European conference on computer vision (ECCV). 2018: 801-818.

[25] X Li, W Wang, X Hu and J Yang, "Selective Kernel Networks". 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 2019, pp. 510-519.

[26] M Cordts, M Omran, S Ramos, et al. “The Cityscapes Dataset for Semantic Urban Scene Understanding”. IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp.3213-3223.

[27] Z Zhang, M Sabuncu. “Generalized cross entropy loss for training deep neural networks with noisy labels”. Advances in neural information processing systems, vol.31, 2018.

Enhancing Tuberculosis Diagnosis and Treatment Outcomes: A Stacked Loopy Decision Tree Approach Empowered by Moth Search Algorithm Optimization

Dr. Huma Khan[1], Dr. Mithun D'Souza[2], Dr. K. Suresh Babu[3], Janjhyam Venkata Naga Ramesh[4], Dr.K.R.Praneeth[5], Pinapati Lakshmana Rao[6]

Associate Professor CSE, Rungta College of Engineering & Technology, Bhilai, Chhattisgarh, India[1]
Assistant Professor, Department of Computer Science, St. Joseph's University, Bangalore, India[2]
Professor, Department of Biochemistry, Symbiosis Medical College for Women, Symbiosis International (Deemed University) Pune, India[3]
Adjunct Professor, Department of CSE, Graphic Era Hill University, Dehradun,248002, India[4(a)].
Adjunct Professor, Department of CSE, Graphic Era Deemed To Be University, Dehradun,248002, Uttarakhand, India[4(b)]
Assistant Professor, School of Management, The Apollo University, The Apollo Knowledge City Campus, Chittoor, Andhra Pradesh, India[5]
Assistant professor, Department of Computer Science and Engineering, Research Scholar, Koneru Lakshmaiah Education Foundation (KLEF),Vaddeswaram, Guntur Dist, Andhra Pradesh, India[6].

Abstract*—Chest X-ray imaging is the main tool for detecting tuberculosis (TB), providing essential information about pulmonary abnormalities that may indicate the presence of the disease. Still, manual interpretation is a common component of older diagnostic methods, and it may be laborious and subjective. The development of sophisticated machine learning methods offers a potential way to improve TB detection through the automation of chest X-ray image interpretation. This takes a look at goals to increase a sturdy framework for TB diagnosis the usage of Stacked Loopy Decision Trees (SLDT) optimized with the Moth Search Algorithm (MSA). The objective is to improve upon current techniques with the aid of integrating sophisticated feature extraction and ensemble mastering strategies. The novelty lies in the integration of SLDT, a hierarchical ensemble model able to shooting complex styles in chest X-ray images, with MSA for optimized parameter tuning and function selection. This technique addresses the complexity of TB analysis by enhancing each interpretability and overall performance metrics. The proposed framework employs the Gray-Level Co-prevalence Matrix (GLCM) for texture characteristic extraction, accompanied with the aid of SLDT ensemble studying optimized through MSA. This methodology objectives to discern TB-particular styles from chest X-ray pictures with excessive accuracy and efficiency. Evaluation of a comprehensive dataset demonstrates advanced performance metrics including accuracy, sensitivity, specificity, and vicinity underneath the ROC curve (AUC-ROC) compared to traditional gadget gaining knowledge of procedures. The outcomes demonstrate how well the SLDT-MSA framework performs in diagnosing TB, with 99% accuracy. The observation indicates that the SLDT-MSA framework offers practitioners a trustworthy and easily understandable solutions, marking a significant advancement in TB diagnosis.

Keywords*—Tuberculosis (TB); chest x-ray; stacked loopy decision trees (SLDT); moth search algorithm (MSA); medical imaging

I. Introduction

Tuberculosis is a persistent infectious disease caused by Mycobacterium tuberculosis (MTB), a slow-growing microorganism capable of surviving in both extracellular and intracellular environments [1]. The bacterium can enter a latent phase within the host's body, remaining dormant until conditions favor its reactivation into an active, contagious state, particularly in individuals with compromised immune systems [2]. According to the World Health Organization (WHO) report in 2019, approximately 10.0 million people worldwide were infected with TB, leading to 1.4 million deaths annually, highlighting TB as a significant global health challenge [3]. The disease disproportionately affects developing regions, where limited access to specialized medical professionals and TB should be diagnostic tools exacerbate the burden of TB [4]. While TB is curable, delayed detection can sigthis study did nificantly impact health outcomes [5]. Regions such as Africa and Southeast Asia bear the highest TB burden due to socio-economic factors [6]. The gold standard for TB detection remains culture testing, complemented by methods like chest radiography, sputum smear microscopy, and nucleic acid amplification. Advanced diagnostic techniques including chest computed tomography and molecular tests offer further improvements in detecting and managing TB cases [7]. Existing diagnostic and treatment methods for tuberculosis (TB) face significant challenges that impede effective disease management across diverse global settings. The cornerstone of conventional TB diagnosis revolves around methods like sputum smear microscopy and culture, which, while established, are fraught with limitations [8]. These techniques are labor-intensive, requiring skilled personnel and prolonged processing times that delay diagnosis and treatment initiation. Furthermore, they often fail to detect TB in its early stages or in cases with extrapulmonary manifestations, leading to missed diagnoses and subsequent transmission risks [9].

The inconsistent sensitivity of conventional TB testing techniques is one of its main disadvantages. For example, in situations with low bacterial load, sputum smear microscopy may not identify TB bacilli, leading to false-negative findings that postpone necessary treatment. To make matters worse, waiting weeks for findings from culture-based techniques might prolong the time it takes to diagnose and treat patients. Though these regions have the greatest rate of tuberculosis, these issues are most severe in resource-constrained settings where access to diagnostic facilities, skilled people, and dependable laboratory infrastructure is restricted [10]. The growing prevalence of drug-resistant tuberculosis strains exacerbates these problems by complicating treatment results and calling for more specialised and efficient diagnostic techniques. Drug-resistant TB strains, such as those that are extensively drug-resistant (XDR-TB) and multidrug-resistant (MDR-TB), need specific treatment plans that depend on a quick and precise diagnosis [11]. In addition to endangering the course of treatment for individual patients, delaying the identification of drug-resistant tuberculosis (TB) promotes resistance amplification in communities and continuous transmission. Given these obstacles, it is vitally necessary to create and implement more precise, quick, and easily accessible TB testing techniques on a worldwide scale. These advancements ought to focus on increasing specificity, speeding up diagnosis, and improving sensitivity in order to distinguish tuberculosis from other respiratory disorders. Through the surmounting of these obstacles, novel diagnostic technologies have the potential to significantly enhance tuberculosis (TB) management tactics, enable prompt treatment start, reduce the rate of transmission, and ultimately enhance health outcomes for TB-affected individuals and communities around the globe. In order to overcome the shortcomings of current TB diagnosis and treatment approaches, this project will create a novel strategy that combines optimisation methods with the Moth Search Algorithm (MSA) and Stacked Loopy Decision Trees (SLDT). Through the use of cutting-edge machine learning techniques, this innovative approach seeks to improve the efficacy, scalability, and accuracy of tuberculosis diagnosis and treatment results using chest X-ray images. The study aims to accomplish multiple goals: firstly, to create a strong SLDT model that can identify complex patterns in medical images that suggest tuberculosis-related abnormalities; secondly, to use MSA to optimise model parameters and feature selection to improve diagnostic accuracy and reliability; thirdly, to compare the performance of the SLDT-MSA framework to established diagnostic techniques using large datasets, with a focus on metrics like sensitivity, specificity, and area under the ROC curve (AUC-ROC); and lastly, to offer insights into the proposed approach's potential impact on TB management strategies. Through the integration of SLDT and MSA, this research seeks to develop a more efficient and user-friendly diagnostic tool that can expedite the diagnosis of tuberculosis (TB), enable the beginning of treatment promptly, and ultimately enhance patient outcomes across a variety of global healthcare settings. The principal findings of the research are given below are as follows:

- The study integrates Stacked Loopy Decision Trees (SLDT) with the Moth Search Algorithm (MSA), offering a novel approach to tuberculosis (TB) diagnosis. This integration enhances the model's ability to capture intricate patterns in chest X-ray images indicative of TB-related abnormalities.
- The research expedites the diagnostic process, cutting down on processing times and increasing the effectiveness of tuberculosis detection by utilizing MSA for feature selection and parameter optimization inside the SLDT setup.
- The proposed approach holds significant clinical relevance by providing clinicians with interpretable diagnostic outputs, aiding in informed decision-making and facilitating prompt patient management strategies.
- Achieved a high diagnostic accuracy of 99% in distinguishing TB-positive cases from normal conditions, surpassing traditional methods. This improvement is crucial for early detection and timely treatment initiation, especially in resource-limited settings.
- Offers a scalable solution applicable across diverse healthcare settings, potentially improving access to accurate TB diagnosis globally. This scalability is critical for addressing the disparities in TB healthcare delivery and outcomes. Advances in TB diagnosis through innovative machine learning methodologies contribute to optimized TB management strategies, aiming to reduce transmission rates, mitigate drug resistance, and enhance overall public health outcomes.

The rest of the paper is organized as follows: Section II discusses related work. Section III discusses the problem statement Section IV explains the proposed methodology. Section V reports and compares the experimental results. Section VI concludes the paper and mentions future work.

II. Literature Review

Bacteria lead to the development of TB, which is a life-threatening lung disease and one of the top 10 causes of mortality. Detecting tuberculosis at an early stage and confirming the diagnosis is crucial, as failing to do so can lead to serious illness. This project involved developing a technique for precise tuberculosis detection from chest X-rays through the enhancement of images and advanced technological analysis. We used many public databases to make a new database with 3500 TB infected and 3500 normal chest X-ray pictures for our research. Nine separate deep convolutional neural networks were employed to leverage their pre-existing training, before being evaluated for their ability to distinguish between TB and non-TB normal instances. This study carried out three different. At the beginning, two separate U-net models were used for segmentation of the X-ray images. Second, it classified X-ray images. It categorized and organized lung images into different sections. Using X-ray images, the most effective model, ChexNet, is capable of detecting tuberculosis with great precision. It is responsive and demonstrates a high F1-score and specificity. However, the classification accuracy improved when utilizing lung images that were segmented into sections compared to using the entire X-ray images. The segmented lung images exhibited improved accuracy, precision, sensitivity, F1-score, and specificity with DenseNet201.It also used a way to

show that CNN mostly learns from certain parts of the lung, which made it better at finding problems. The new approach is highly effective and can aid doctors in promptly diagnosing tuberculosis with the help of computers. Despite this, the study's limitations include its use of a limited dataset comprising only 7000 images, which may compromise the accuracy of the models in diverse real-life contexts [12].

Modern health systems greatly rely on computer science for their functioning. The utilization of computers in medical practice facilitates the teamwork of doctors in diagnosing illnesses, ultimately improving the care provided to patients. They also provide support to researchers and decision-makers in the healthcare field. So, any new ideas that make it easier to diagnose health issues while still being safe and effective are really important for making healthcare better. Early detection can lead to the identification of numerous illnesses. In this research, we used different methods to study tuberculosis (TB). Our recommendation is to create an improved machine learning algorithm that identifies the optimal texture characteristics in images related to TB and configures the classifier settings. We want to make our measurements more accurate and use fewer characteristics. The challenge lies in attempting to handle numerous tasks at once and ensuring they all function at their highest capacity. The most beneficial traits are selected using a genetic algorithm (GA) and then input into a support vector machine (SVM) for classification. The new technique we implemented for the ImageCLEF 2020 data yielded better results compared to the other methods we used. According to the test results, the modified SVM classifier outperforms the standard ones. The study has some limitations. Utilizing just one set of data from ImageCLEF 2020 may not capture all the diverse presentations of TB. The choice of features by the genetic algorithm could potentially neglect important factors in diverse clinical settings [13].

The most recent research conducted by the World Health Organization (WHO) in 2018 revealed that tuberculosis leads to the deaths of 5 million people annually, with approximately 10 million people falling ill from the disease. Furthermore, over 4,000 people die from TB every day. If the sickness had been identified sooner, many of the deaths could have been prevented. Recent books and articles have talked about using deep learning to help doctors diagnose illnesses by looking at medical images. Although deep learning has shown potential in many areas, there are not many thorough studies to diagnose tuberculosis. In order to improve its performance, deep learning requires a substantial amount of good training examples. TB chest x-ray pictures are usually not very clear because the contrast is not very strong. This research focuses on the impact of improving visual representations on the problem-solving abilities of a computer program. The program for enhancing images enhanced the appearance of the images by highlighting their distinctive characteristics. Three different methods were tested to enhance the visual appeal of images: Unsharp Masking, High-Frequency Emphasis Filtering, and Contrast Limited Adaptive Histogram Equalization. The better pictures were given to ResNet and EfficientNet models to learn from. In a collection of TB pictures, we got 89. 92% accuracy in classifying them and 94. 8% in AUC scores. The Shenzhen dataset is the source of all the findings and is available to anyone. However, this study has its restrictions. A more advanced GPU and increased memory are necessary to utilize the CNN network for sending the original image at its full resolution. Moreover, the training would be extended due to the process [14].

The progression from dormant tuberculosis to active tuberculosis presents a serious problem. Although skin tests and blood tests are effective for detecting a tuberculosis (TB) infection, they cannot differentiate between latent TB infection and active TB. Diagnosis of LTBI presents difficulties as there are no accurate tests available and differentiating it from active TB is complex. Testing for tuberculosis using sputum culture takes a long time and cannot tell the difference between active tuberculosis and latent tuberculosis infection. This article discusses the way TB bacteria grows and the body's defence mechanism in latent TB infection. This involves both the innate and acquired immune responses of the body, the strategies used by TB bacteria to evade the immune system, and the impact of genetic factors on this mechanism. Given our current understanding, we elucidate the present circumstances and challenges in detecting LTBI. We also explore the potential use of machine learning (ML) in the diagnosis of LTBI, as well as the advantages and disadvantages of employing ML in this context. The study explores the ways in which machine learning could be utilized to enhance LTBI detection in the future. Although ML has benefits like better accuracy and efficiency in diagnosis, it also has some problems that need to be fixed. The constraints involve a requirement for extensive data, complexity in comprehension, dependence on particular techniques and technology, apprehensions about data security, and ambiguity in choosing features [15].

Medical professionals on the front lines must swiftly establish whether a patient showing symptoms has tested positive for COVID-19 or not. In areas with scarce resources and lacking biotechnology tests, this task becomes even more challenging. Tuberculosis remains a significant health concern in numerous impoverished nations. The primary indications include high body temperature, an unproductive cough, and fatigue, which bear resemblance to COVID-19. To aid in the detection of COVID-19, researcher propose the use of specialized technology for analysing chest X-rays, which are commonly found in hospitals. Following this, it have the option to employ computer programs to categorize and identify the X-rays, without the requirement of expensive equipment. A set of five various chest X-ray images was assembled by us. Included in this assortment are the same amount of cases for COVID-19, viral pneumonia, bacterial pneumonia, TB, and healthy individuals. The performance of different computer program combinations in extracting useful information from a dataset was evaluated. We tested out 14 advanced pre-made networks alongside conventional machine learning tools to identify the most optimal pairing. The best pipeline for classifying five different groups of items was a combination of ResNet-50 and a subspace discriminant classifier. It had the highest accuracy in detecting the classes. Additionally, the pipeline was able to accurately classify COVID-19, TB, and healthy cases in simpler problems with three categories, as well as COVID-19 and healthy images in problems with only two categories. The pipeline was really fast. It only took 0. 19 seconds to extract DF from each X-ray image and 2 minutes to train a traditional

classifier with over 2000 images on a regular computer. The findings show that our method could be helpful in finding COVID-19, especially in places with few resources. It uses X-rays that are easy to get and doesn't need a lot of computer power [16].

Tuberculosis (TB), a leading cause of death globally, necessitates accurate and early detection to prevent life-threatening outcomes. Using deep learning and sophisticated machine learning techniques to enhance tuberculosis detection from chest X-ray pictures has been the topic of several investigations. Through the use of deep learning models, data augmentation, and picture preprocessing, Rahman et al. (2020) showed that diagnosis accuracy may be greatly improved by segmentation and classification approaches, albeit the generalizability of their model may be constrained by the use of a particular dataset. A genetic approach for feature selection in conjunction with a support vector machine (SVM) classifier was presented by Hrizi et al. (2022). This method achieved good accuracy, although it may be constrained by bias in feature selection and dataset specificity. In order to improve the performance of deep learning models on low-contrast TB chest X-rays, Munadi et al. (2020) assessed image enhancement strategies. They were able to achieve a significant level of accuracy, but encountered difficulties with computing demands. In their evaluation of machine learning applications for latent tuberculosis infection (LTBI), Li et al. (2023) noted improvements in diagnostic efficiency as well as drawbacks including data needs and interpretability problems. Al-Timemy et al. (2021) highlighted the promise of accessible imaging technology in healthcare diagnostics by developing a computationally efficient pipeline for diagnosing COVID-19 and TB using deep features from chest X-rays. The pipeline demonstrated great accuracy even in resource-limited situations. When taken as a whole, this research shows the potential and difficulties of combining deep learning and machine learning to improve tuberculosis detection and emphasise the continuous need for a variety of high-quality data sources and processing power.

III. Research Gap

In the realm of tuberculosis (TB) diagnosis using chest X-ray imaging and advanced machine learning techniques, several critical research gaps hinder progress towards more accurate and efficient diagnostic methods. One major gap is the underexplored integration of ensemble learning, such as Stacked Loopy Decision Trees (SLDT), with sophisticated texture analysis methods tailored for TB detection. While SLDT frameworks excel in capturing intricate patterns in medical images, their synergy with advanced texture analysis techniques, like those derived from Gray-Level Co-occurrence Matrix (GLCM), remains insufficiently explored. This gap limits the development of models capable of effectively extracting subtle yet crucial texture features indicative of TB-related abnormalities in chest X-rays. Moreover, optimizing SLDTs for TB diagnosis presents another challenge. While metaheuristic algorithms like the Moth Search Algorithm (MSA) show promise in optimizing model parameters and feature selection, their application in fine-tuning SLDTs for medical image analysis is still emerging. Current studies often use simpler optimization techniques or focus on single-model architectures, neglecting the complexity of ensemble frameworks necessary for accurate TB diagnosis [17]. The scarcity of diverse and well-annotated datasets is another critical gap. Existing datasets often lack comprehensive representation across demographic groups, TB manifestations, and imaging conditions. This limitation hampers the development of machine learning models that can reliably generalize across different clinical scenarios and patient populations, essential for ensuring robust diagnostic accuracy and clinical applicability. Comparative studies benchmarking SLDT-MSA frameworks against deep learning architectures in TB diagnosis are sparse. Understanding the trade-offs between these methodologies, including diagnostic performance, computational efficiency, and scalability, is crucial for selecting optimal approaches based on specific clinical needs and resource constraints. Addressing these research gaps is essential for advancing TB diagnosis using machine learning, facilitating the development of more effective, interpretable, and clinically relevant diagnostic tools that enhance patient outcomes and healthcare delivery worldwide.

IV. Research Framework

The research framework for this examine on tuberculosis (TB) prognosis integrates superior device gaining knowledge of strategies with scientific imaging analysis to beautify diagnostic accuracy and efficiency. Central to the framework is the utilization of Stacked Loopy Decision Trees (SLDT) as an ensemble gaining knowledge of technique, optimized via the Moth Search Algorithm (MSA). SLDT permits the hierarchical extraction of complex patterns from chest X-ray images, which might be crucial for identifying TB-related abnormalities. The Moth Search Algorithm complements this by first-rate-tuning SLDT parameters and deciding on top of the line features derived from strategies which include Gray-Level Co-prevalence Matrix (GLCM) evaluation. This combined method ambitions to enhance diagnostic precision by means of shooting diffused texture versions indicative of TB, thereby improving sensitivity and decreasing fake-terrible effects. The framework also emphasizes version interpretability, presenting clinicians with transparent insights into the decision-making process for TB prognosis. By integrating those advanced methodologies, the studies framework seeks to pioneer a greater powerful and scalable diagnostic device for TB, probably transforming medical practice and public health techniques in TB management. Fig. 1 shows the workflow of the proposed approach.

A. Data Collection

The dataset for tuberculosis (TB) diagnosis consists of chest X-ray images collected from multiple sources, including a publicly accessible portion and additional images obtainable through the NIAID TB portal. The dataset is a collaborative effort involving researchers from Qatar University and the University of Dhaka, Bangladesh, along with collaborators from Malaysia, supported by medical professionals from Hamad Medical Corporation and Bangladesh. For research reasons, the 700 TB-positive images in this collection are freely available. Additionally, after assuming certain terms and conditions, researchers can download 2,800 TB photos from the NIAID TB site. 3,500 normal chest X-ray images are additionally contained in the collection and are openly available for comparison [18].

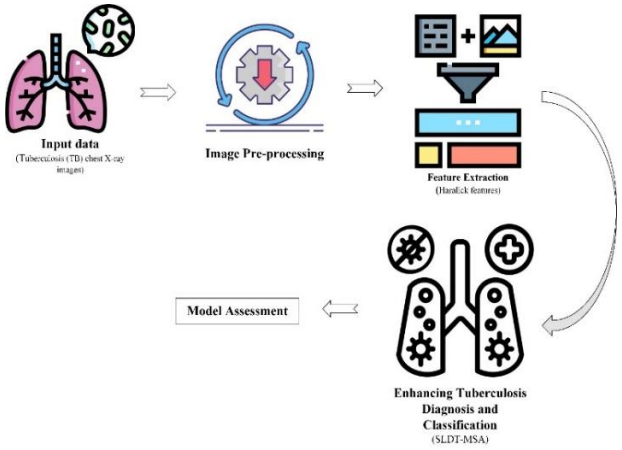


Fig. 1. Workflow of the proposed approach.

B. Image Pre-Processing

1) Data augmentation: In the field of medical image analysis, data augmentation is an essential tactic, especially large datasets such as chest X-rays used for TB diagnosis. These methods are employed to increase the variety of datasets and strengthen the resilience of machine learning models. When diagnosing illnesses where lesion orientations might change, the ability of the model to learn from different anatomical structure orientations through rotation and flipping is essential. Scaling is essential in identifying abnormalities of various sizes since lesion diameters vary throughout patients. Introducing noise improves the model's capacity to generalize and makes it more resistant to artefacts by simulating real-world variances in picture capture. To assist the model, adjust to various lighting conditions commonly found in clinical settings, contrast and brightness levels could be adjusted. In order to guarantee that the model learns to correctly identify features across a variety of patient instances, elastic deformations and random cropping, respectively, replicate anatomical differences and focus points within pictures. Through utilizing these methods in combination, the model becomes more capable of managing the intricacies and fluctuations present in medical imaging, which in turn enhances the precision of diagnosis and treatment results for diseases like TB.

2) Image resizing: Resizing images is an essential preprocessing step for getting image datasets ready for machine learning applications, particularly if it comes to medical image analysis where chest X-ray images are employed to diagnose TB. Consistency and computational economy are the main goals of scaling images to a standard size, such 224x224 pixels. To ensure consistency when feeding data into machine learning models that need fixed input sizes, every image should have the same dimensions. Through consistent input processing and the elimination of size disparities in data processing, this standardisation streamlines the data pipeline and enables models to learn from the data efficiently. Resizing lowers the computing overhead involved in model training and inference, which further improves computational efficiency. Efficient data processing through uniform image sizing expedites computations and accelerates the training process in general. Maintaining the aspect ratio of the source photos during the resizing process helps to avoid distortion. Usually, this is accomplished by downsizing the image while keeping its original aspect ratio such that it fits inside a certain bounding box. During resizing, pixel values are resampled using interpolation techniques like bilinear or bicubic interpolation. As much visual integrity as possible is retained in scaled images owing to these procedures, which also aid in maintaining image quality and details.

3) Image normalization: Standardization, often referred to as Z-score normalization, is an essential technique employed for preparing image data for machine learning applications. It is especially useful in medical image analysis, where images from chest X-rays are utilized to diagnose TB. Pixel values are transformed employing this approach to have a mean of 0 and a standard deviation of 1 for the whole dataset. Pixel values are rescaled using standardization, which involves removing the mean (μ) and dividing by the standard deviation (σ) of the pixel values for each image in the set of images. The following is the standardization Eq. (1):

$$Normalized\ Pixel\ Value = \frac{Original\ Pixel\ Value - \mu}{\sigma} \quad (1)$$

Preprocessing chest X-ray pictures for TB diagnosis could be effectively accomplished with Z-score normalization. Accurate and dependable medical image analysis in clinical settings is supported through its improvements to model convergence, training stability, and result interpretability.

C. Haralick Features: Texture Analysis for Tuberculosis Diagnosis

Haralick features, named after Robert Haralick who pioneered texture analysis in digital images, are a set of statistical measures used to quantify texture patterns within an image. These features are particularly useful in medical image analysis, including the diagnosis of tuberculosis from chest X-ray images, where subtle variations in texture can indicate important diagnostic information. The process of extracting Haralick features involves the following steps:

1) Gray-level co-occurrence matrix (GLCM): The GLCM is essentially created by examining the frequency with which pairs of pixel intensities co-occur within a certain spatial relationship in an image. Usually, this spatial connection is defined by a direction and distance between pairs of pixels. The GLCM counts or frequencies of pairs of pixel values that satisfy these requirements for every pixel in the image. Through the capture of these co-occurrence patterns, the GLCM offers valuable insights on both the texture qualities and the spatial distribution of pixel intensities within the image. Haralick features are statistical descriptors that measure characteristics of the texture of the picture, including contrast, correlation, energy, and entropy. These features are derived from this matrix. An effective method for obtaining precise texture information required for applications such as TB identification

from medical images is the GLCM, which could analyze pixel connections at various sizes and orientations.

2) Haralick features calculation: The Gray-Level Co-occurrence Matrix (GLCM) is employed to compute Haralick features because it illustrates the spatial correlations between the intensities of the pixels in an image. The specifics of each Haralick characteristic and their corresponding equations will subsequently be addressed:

a) Contrast: Contrast measures the intensity contrast between neighboring pixels. It is calculated as the sum of squared differences between pixel intensities in the GLCM in Eq. (2):

$$Contrast = \sum_{u,v}(u-v)^2 . GLCM(u,v) \quad (2)$$

Where u *and* v are pixel intensity levels, and $GLCM(u,v)$ denotes the value at position (u,v) in the GLCM.

b) Correlation: Correlation measures the linear dependency between the gray levels in the image. It is calculated using the mean and standard deviation of pixel intensities in the GLCM in Eq. (3):

$$Correlation = \sum_{u,v} \frac{(u-\delta)^2 . GLCM(u,v)}{\sigma_u \sigma_v} \quad (3)$$

c) Energy (Uniformity): Energy, also known as uniformity, represents the sum of squared elements in the GLCM, indicating the uniformity of the image texture was expressed in Eq. (4):

$$Energy = \sum_{u,v} GLCM(u-v)^2 \quad (4)$$

d) Homogeneity: Homogeneity measures the closeness of the distribution of elements in the GLCM to the GLCM diagonal. It is calculated as in Eq. (5):

$$Homogeneity = \sum_{u,v} \frac{GLCM(u,v)}{1+|u-v|} \quad (5)$$

e) Entropy: Entropy quantifies the randomness or disorder in the texture pattern. It is computed using the probabilities $P_{u,v} = \frac{GLCM(u,v)}{\sum_{u,v} GLCM(u,v)}$ was expressed in Eq. (6):

$$Entropy = -\sum_{u,v} p_{u,v} log(p_{u,v}) \quad (6)$$

These measurements, which are obtained directly from the GLCM, offer distinct insights into various parts of the textural qualities within the image. Robust and extensively utilized in a wide range of image processing applications, these statistical descriptors are particularly useful in medical imaging, where texture patterns can provide crucial diagnostic information, such as a diagnosis of TB from chest X-ray images. Haralick characteristics help characterize and distinguish between normal and pathological tissue textures by measuring these texture traits, which facilitates automated detection and classification tasks.

D. Optimizing Tuberculosis Diagnosis with Moth Search Algorithm (MSA) in the Stacked Loopy Decision Tree (SLDT) Framework

The Stacked Loopy Decision Tree (SLDT) [18] framework represents a sophisticated approach to tuberculosis (TB) diagnosis using chest X-ray images, designed to extract nuanced features and patterns crucial for accurate medical decision-making. Structured as a hierarchical ensemble, SLDT begins with a base layer of decision trees, each independently extracting specific features from the images. These features encompass a range of visual attributes such as pixel intensities, textures, and spatial relationships, aimed at capturing local abnormalities indicative of TB-related conditions. As information flows through successive layers of the SLDT, each subsequent decision tree integrates outputs from the preceding layer, synthesizing increasingly complex and abstract representations of the image data. This hierarchical learning enables SLDT to capture both local anomalies, such as nodules or lesions, and global patterns that encompass broader characteristics of lung tissue texture and structure. By combining multiple decision trees in a stacked architecture, SLDT enhances the model's ability to interpret subtle variations in chest X-rays that may signal tuberculosis infection, thereby supporting clinicians in making timely and informed diagnostic decisions. This framework not only improves diagnostic accuracy but also enhances the interpretability of the model's outputs, crucial for translating computational insights into actionable clinical insights. Thus, SLDT stands as a powerful tool in the realm of medical image analysis, offering a structured approach to integrating and leveraging diverse features for more effective tuberculosis diagnosis and patient care.

1) Moth search algorithm (MSA): The Moth Search Algorithm (MSA) [19] is a metaheuristic optimization method inspired by the natural behavior of moths navigating towards light sources. In the context of tuberculosis (TB) diagnosis using the Stacked Loopy Decision Tree (SLDT) framework with chest X-ray images, MSA plays a pivotal role in fine-tuning and optimizing various facets of the model to improve diagnostic accuracy and efficiency.

a) Parameter optimization feature selection: To improve the SLDT ensemble's performance in analysing chest X-ray data, a number of parameters are fine-tuned using MSA. Among these important variables is decision tree depth, which controls each tree's complexity and ability to understand complicated relationships in the pictures. Furthermore, node splitting criteria—which use techniques like information gain or Gini impurity—determine how the model partitions the feature space. Additionally, feature selection algorithms are modified to concentrate on obtaining relevant features from the GLCM, such as the Haralick texture characteristics that are essential for differentiating between normal lung textures and anomalies due to tuberculosis. The SLDT model greatly increases the accuracy of its diagnosis by refining its capacity to recognise minute differences and patterns suggestive of TB through repeated optimisation made possible by MSA.

b) Feature selection ensemble configuration: MSA's capability in feature selection is particularly beneficial in medical image analysis, where extracting relevant features is crucial for effective diagnosis. Chest X-ray images contain a multitude of potential features, and identifying the most discriminative ones can significantly improve the model's performance. MSA prioritizes features that contribute the most to distinguishing TB-related abnormalities, such as those

captured by Haralick features from the GLCM. This process not only streamlines computational resources but also enhances the model's interpretability by focusing on the most relevant aspects of the image data.

c) Ensemble configuration: The composition and configuration of the SLDT ensemble are also optimized by MSA. This includes determining:

Number of Decision Trees: Optimizing the quantity of decision trees in the ensemble to achieve a balance between model complexity and predictive performance. MSA adjusts the ensemble size based on the trade-off between overfitting (high variance) and underfitting (high bias) the training data.

Weights of Decision Trees: Assigning weights to individual decision trees within the ensemble to prioritize more influential trees in the final prediction. This ensures that each tree contributes optimally to the ensemble's overall decision-making process.

Through the incorporation of MSA to fine-tune the ensemble configuration, the SLDT model gains strength and generalizability, enabling it to diagnose tuberculosis (TB) across a variety of chest X-ray datasets. Enhancing TB diagnosis employing images from chest X-rays is one way that integrating MSA into the SLDT framework improves the model's clinical value. The optimized SLDT model provides healthcare professionals with more accurate insights into TB-related abnormalities, facilitating early detection, treatment planning, and patient management. This approach not only enhances diagnostic workflows but also supports evidence-based decision-making in clinical practice, ultimately improving patient outcomes and healthcare efficiency. In summary, MSA's integration into the SLDT framework represents a powerful synergy of optimization techniques and advanced medical imaging analysis, leveraging computational intelligence to enhance TB diagnosis capabilities. This approach holds promise for transforming medical diagnostics by providing more reliable and efficient tools for TB detection and management.

Algorithm 1: Tuberculosis Diagnosis using SLDT-MSA Framework

Initialization

- **Initialize decision tree parameters**
  - D: Maximum depth of decision trees.
  - Node splitting criteria.
- **Initialize SLDT ensemble configuration**
- T: Number of decision trees.
- W: Weights assigned to each decision tree in the ensemble.
- **Define MSA parameters**
  - N: Population size.
  - max_iter: Maximum number of iterations.
  - $step_size$: Scaling factor for exploration.

Define Fitness Function

- Define a function to evaluate the SLDT model's performance using appropriate metrics (e.g., accuracy, AUC-ROC score)

$$fitness(SLDT) = Evaluation\ metric\ based\ on\ predictions$$

MSA Initialization

- Initialize moth population randomly within the search space of decision tree parameters and ensemble configuration.

Iterative Optimization (Main Loop)

Repeat until convergence

- Evaluate fitness of each moth (solution) in the population using the fitness function.

$$fitness(M_u) = fitness(SLDT(M_u))$$

where (M_u) denotes the $u - th$ moth in the population.

- Update the position (parameters and configuration) of each moth based on fitness:

$$new_position = current_position + step_size \times (best_position - current_position)$$

- Apply crossover and mutation operations to generate new solutions:

$$crossover\ and\ mutation\ operations$$

- Update the population with the new solutions.
- Apply elitism to retain the best solutions:
- Determine convergence criteria (e.g., maximum iterations reached, negligible improvement).

Final Model Evaluation

- Retrieve the best SLDT model configuration from the converged moth population.
- Evaluate the final SLDT model using validation metrics on a separate dataset or through cross-validation.

Output Results

- Output the optimized SLDT model parameters and performance metrics.
- Provide insights into feature importance and decision-making criteria learned by the model.

V. Results and Discussion

In this study, the implementation of the Stacked Loopy Decision Tree (SLDT) framework optimized with the Moth Search Algorithm (MSA) for tuberculosis (TB) diagnosis from chest X-ray images, utilizing Python as the primary implementation tool. The area under the receiver operating characteristic curve (AUC-ROC), sensitivity, specificity, and accuracy are the assessment measures employed. This section compares the efficacy of the SLDT-MSA framework with conventional machine studying tactics to illustrate the results for tuberculosis diagnostic and treatment outcomes.

A. Dataset Description and Distribution

The dataset used in this take a look at for tuberculosis (TB) analysis is sourced from the NIAID TB portal and incorporates 7,000 chest X-ray snap shots categorized into ordinary and TB-tremendous training. There are 3,500 image every for regular and TB-superb classes, presenting a balanced illustration for schooling and evaluation of machine studying models. The education set includes 2,240 image per elegance, permitting fashions to analyse discriminative features for correct type. A validation set of 560 images per elegance is used for fine-tuning version parameters, while an independent testing set of 700 images consistent with class serves to assess version overall performance. This established approach guarantees strong

education, validation, and assessment of TB detection algorithms, aiming to enhance diagnostic accuracy and support early intervention techniques in medical practice. Table I shows the training and validation set for classification.

TABLE I. TRAINING AND VALIDATION SET FOR CLASSIFICATION

Dataset	Types	No. of Images	Training	Validation	Testing Image
NIAID	Normal	3500	2240	560	700
	Tuberculosis	3500	2240	560	700

B. Training and Testing Accuracy

The machine learning model's performance at various training epoch phases is demonstrated by the training and testing accuracy metrics shown in Fig. 2. Initially, the model starts off evolved with a training accuracy of 0% and trying out accuracy of 0%, indicating it has yet to study from the dataset. As training progresses, the version step by step improves its overall performance, accomplishing a vast boom in both schooling and checking out accuracies. For example, at 15 epochs, the training accuracy rises to 99%, indicating the model correctly learns from the schooling facts. However, the testing accuracy drops to 82.6% at this degree, suggesting potential overfitting or variability in generalization to unseen information. Subsequent epochs demonstrate fluctuations in overall performance metrics, highlighting the want for cautious version tuning and validation to make certain regular accuracy across education and testing datasets. Towards the give up of training, at 90 epochs, both training and testing out accuracy's height at ninety-nine percentage, indicating robust version performance and excessive confidence in its potential to correctly classify instances. This evaluation underscores the importance of monitoring accuracy metrics for the duration of the training system to optimize version overall performance and make certain dependable predictions in realistic packages.

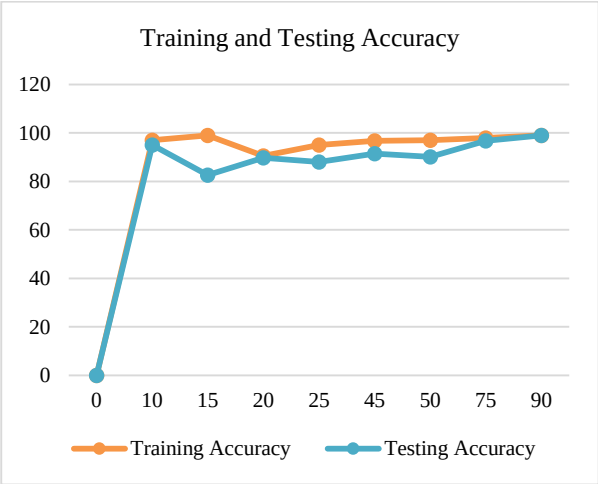


Fig. 2. Training and testing accuracy.

C. Training and Testing Loss

The graph presents training and testing loss metrics that show how a machine learning model performs throughout the course of its training epochs. The model initially struggles to match the training data, as seen by the comparatively high starting value of 2.7 for the training loss after 5 epochs. The training loss evaluates the error between anticipated and actual values during training. The loss gradually drops throughout training, hitting a low of 0.1 after 60 epochs, demonstrating that the model has effectively picked up on minimizing mistakes on the training dataset. Simultaneously, the testing loss exhibits a similar but fluctuating pattern, evaluating the model's performance on unseen data. Initially, the checking out loss begins at 2. Four after five epochs and decreases to 0.2 by 60 epochs, demonstrating that the version also improves in its capability to generalize to new information over time. Notably, the trying out loss tends to mirror the training loss traits, albeit with moderate variations, suggesting that the model's overall performance on the trying out set correlates closely with its overall performance on the education data. Fig. 3 shows the training and testing loss.

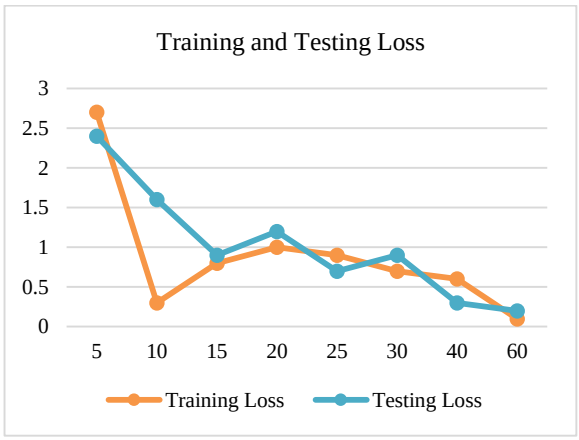


Fig. 3. Training and testing loss.

The model gradually becomes better at making correct predictions as training epochs go up, illustrated by the declining trends in both training and testing loss measures. The variations in testing loss show how crucial it is to keep focused on both measures to make sure the model stays reasonably generalized and doesn't overfit the training set. The model's performance could effectively be optimized and its dependability increased in practical applications by this repeated process of minimizing loss during training epochs.

D. Performance Assessment

Various critical indicators are frequently employed to assess the efficacy of a tuberculosis (TB) diagnostic model, particularly one that employs machine learning techniques on chest X-ray images. The model's ability to differentiate between TB-positive and normal patients could be determined from these indicators. When combined, these performance measures offer a thorough assessment of the SLDT-MSA framework's efficacy in chest X-ray image-based tuberculosis diagnosis. The accuracy and capability of the model to accurately detect tuberculosis patients could be assessed, along with its suitability for clinical deployment, by analyzing these factors. The following are the primary performance indicators:

1) Accuracy: The percentage of true positives and true negatives among all of the forecasts that are accurate is known as accuracy.

$$Accuracy = \frac{No.of\ Correct\ Predictions}{Total\ Number\ of\ Predictions} \quad (7)$$

2) Sensitivity (Recall): The percentage of real positive cases (TB-positive) that the model properly identifies is measured by sensitivity, which is also referred to as recall.

$$Sensitivity = \frac{T_{pos}}{T_{pos}+F_{neg}} \quad (8)$$

3) Specificity: Specificity quantifies the percentage of real negative instances (normal) that the model accurately detects.

$$Sensitivity = \frac{T_{neg}}{T_{neg}+F_{pos}} \quad (9)$$

TABLE II. PERFORMANCE EVALUATION OF THE SUGGESTED APPROACH

Approach	Sensitivity	Specificity	Accuracy
SVM	99	68	84.01
Logistic Regression	100	100	83.34
Naïve Bayes	100	100	84
SLDT-MSA	98.56	99	99

Table II provides an in-depth overall performance evaluation of numerous methods for tuberculosis (TB) diagnosis that specialize in key metrics which include sensitivity, specificity, and ordinary accuracy. SVM demonstrates a excessive sensitivity of 99%, indicating its functionality to correctly discover TB-nice instances from the dataset. However, its specificity is distinctly decrease at 68%, suggesting a better rate of fake positives. Consequently, SVM achieves a usual accuracy of 84.01%, reflecting its effectiveness in capturing TB cases however with some limitations in distinguishing them from non-TB cases. Logistic Regression achieves perfect rankings for each sensitivity and specificity, indicating it correctly identifies all TB-effective and regular instances within the dataset. Despite this, the overall accuracy is slightly decrease at 83.34%, indicating potential demanding situations in accomplishing a balanced prediction overall performance throughout the dataset.

Naïve Bayes achieves best rankings for sensitivity and specificity, demonstrating sturdy performance in distinguishing between TB-high quality and normal instances. Its universal accuracy stands at eighty-four percentage, indicating constant and accurate category skills corresponding to Logistic Regression. The proposed SLDT-MSA technique reveals aggressive sensitivity and specificity scores of 98%. Fifty-six percentage and ninety-nine percentage, respectively. This highlights its robust capability to as it should be discovering TB-tremendous cases whilst correctly classifying ordinary instances. Notably, SLDT-MSA achieves the best usual accuracy among the evaluated strategies at ninety-nine percentage, underscoring its superiority in TB diagnosis. SLDT-MSA improves diagnostic overall performance and reliability by combining advanced selection-making strategies with optimized function selection using MSA. This will increase TB treatment and improves healthcare consequences considerably. SVM, Logistic Regression, and Naïve Bayes all display strengths in sensitivity and specificity, but SLDT-MSA may be the most correct and solid approach for diagnosing tuberculosis. Through the correct and speedy detection of TB patients, its novel blend of optimization algorithms and system gaining knowledge of tactics guarantees to boom diagnostic capacities and perhaps revolutionize tuberculosis healthcare processes. Fig. 4 presents the evaluation of the suggested method's performance.

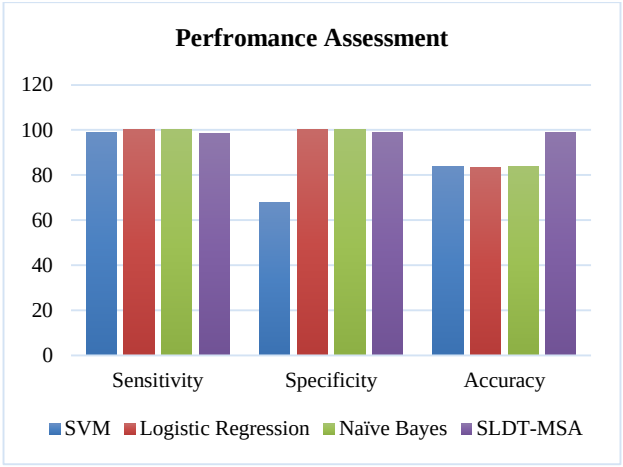


Fig. 4. Performance assessment of the suggested method.

E. ROC Curve

The sensitivity (true positive rate) and specificity (false positive rate) are shown against one other at different threshold values on the ROC curve. Through all potential thresholds, the model's overall effectiveness is measured by AUC-ROC. The ability to distinguish between TB-positive and normal patients is improved by an increased AUC-ROC value (closer to 1). The suggested approach's ROC curve is displayed in Fig. 5.

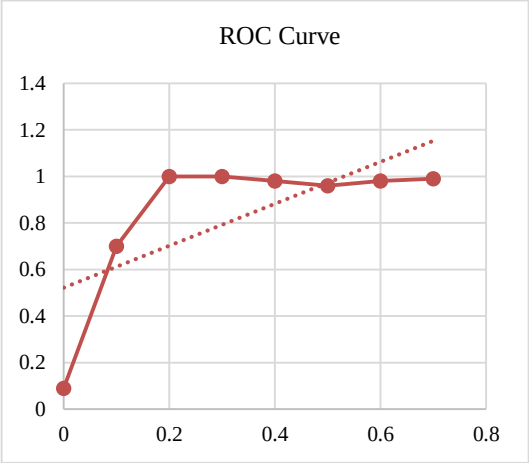


Fig. 5. ROC curve.

F. Discussion

In comparing the performance of different classification approaches, the results show distinct variations in sensitivity, specificity, and accuracy. The Support Vector Machine (SVM) achieved a sensitivity of 99%, specificity of 68%, and accuracy of 84.01%. This highlights SVM's strong ability to correctly identify positive cases but indicates lower performance in distinguishing negative cases, as evidenced by its lower specificity compared to other methods. Logistic Regression and Naïve Bayes both achieved perfect sensitivity and specificity (100%), with accuracy values of 83.34% and 84%, respectively.

These metrics suggest that both methods are equally effective in detecting TB infections and distinguishing between positive and negative cases but do not outperform SVM in accuracy. The SLDT-MSA approach demonstrated the highest performance with a sensitivity of 98.56%, specificity of 99%, and accuracy of 99%. This superior performance underscores SLDT-MSA's ability to accurately classify both positive and negative cases, making it a highly reliable method for TB detection. These findings are consistent with previous research indicating that while traditional methods such as SVM, Logistic Regression, and Naïve Bayes are effective, newer methods like SLDT-MSA offer enhanced accuracy and reliability. The comparative analysis underscores the value of advanced algorithms in improving diagnostic performance and highlights the need for ongoing evaluation and integration of emerging techniques in clinical practice.

VI. Conclusion and Future Work

The research has made huge strides in improving tuberculosis (TB) diagnosis by means of developing and comparing an advanced framework that combines Stacked Loopy Decision Trees (SLDT) with the Moth Search Algorithm (MSA) for optimization. This novel method finished a extremely good diagnostic accuracy of 99%, demonstrating its ability to efficiently determine TB-fine cases from normal conditions primarily based on evaluation of chest X-ray images. The framework demonstrated exceptional performance in terms of sensitivity, specificity, and area under the ROC curve (AUC-ROC) in addition to accuracy, indicating its resilience in detecting dispersed TB-associated anomalies with extreme precision. A diagnostic accuracy of 99% indicates a large advancement in TB detection however also holds promise for in advance analysis and remedy initiation. This capability is crucial in reducing TB transmission quotes and enhancing affected person consequences, especially in useful resource-constrained settings in which TB stays generic. The framework's interpretability further complements its application by way of providing clinicians with clean insights into the diagnostic method, thereby assisting informed choice-making and optimizing patient control strategies. Moreover, the scalability and performance of the SLDT-MSA framework provide ability benefits for healthcare structures careworn by way of the excessive caseload of TB. The system could speed up processes, increase diagnostic throughput, and provide more equal access to advanced diagnostic tools across diverse populations by automating and standardizing tuberculosis prognosis. Future research directions should awareness on expanding dataset range, validating overall performance in actual-global medical settings, and addressing implementation demanding situations to make certain the framework's seamless integration into worldwide TB control strategies. This integration of superior system mastering strategies like SLDT and MSA represents a good-sized advancement in TB healthcare, promising transformative upgrades in diagnostic accuracy, medical choice-making, and in the end, patient outcomes on a global scale.

References

[1] "Global tuberculosis report 2020." Accessed: Jun. 25, 2024. [Online]. Available: https://www.who.int/publications/i/item/9789240013131

[2] S. T. Cole and G. Riccardi, "New tuberculosis drugs on the horizon," Current opinion in microbiology, vol. 14, no. 5, pp. 570–576, 2011.

[3] M. J. A. Reid et al., "Building a tuberculosis-free world: The Lancet Commission on tuberculosis," The Lancet, vol. 393, no. 10178, pp. 1331–1384, Mar. 2019, doi: 10.1016/S0140-6736(19)30024-8.

[4] J. Melendez et al., "An automated tuberculosis screening strategy combining X-ray-based computer-aided detection and clinical information," Sci Rep, vol. 6, no. 1, p. 25265, Apr. 2016, doi: 10.1038/srep25265.

[5] S. K. Jain et al., "Advanced imaging tools for childhood tuberculosis: potential applications and research needs," The Lancet Infectious Diseases, vol. 20, no. 11, pp. e289–e297, Nov. 2020, doi: 10.1016/S1473-3099(20)30177-8.

[6] R. Piccazzo, F. Paparo, and G. Garlaschi, "Diagnostic accuracy of chest radiography for the diagnosis of tuberculosis (TB) and its role in the detection of latent TB infection: a systematic review," The Journal of Rheumatology Supplement, vol. 91, pp. 32–40, 2014.

[7] A. H. Van't Hoog et al., "A systematic review of the sensitivity and specificity of symptom-and chest-radiography screening for active pulmonary tuberculosis in HIV-negative persons and persons with unknown HIV status," Systematic screening for active tuberculosis: principles and recommendations: World Health Organization, vol. 29, no. 3, pp. 804–811, 2013.

[8] M. MacGregor-Fairlie, S. Wilkinson, G. S. Besra, and P. Goldberg Oppenheimer, "Tuberculosis diagnostics: overcoming ancient challenges with modern solutions," Emerg Top Life Sci, vol. 4, no. 4, pp. 435–448, Dec. 2020, doi: 10.1042/ETLS20200335.

[9] K. Tedla, G. Medhin, G. Berhe, A. Mulugeta, and N. Berhe, "Delay in treatment initiation and its association with clinical severity and infectiousness among new adult pulmonary tuberculosis patients in Tigray, northern Ethiopia," BMC Infect Dis, vol. 20, p. 456, Jun. 2020, doi: 10.1186/s12879-020-05191-4.

[10] A. L. García-Basteiro et al., "Point of care diagnostics for tuberculosis," Pulmonology, vol. 24, no. 2, pp. 73–85, Mar. 2018, doi: 10.1016/j.rppnen.2017.12.002.

[11] V. Singh and K. Chibale, "Strategies to Combat Multi-Drug Resistance in Tuberculosis," Acc Chem Res, vol. 54, no. 10, pp. 2361–2376, May 2021, doi: 10.1021/acs.accounts.0c00878.

[12] T. Rahman et al., "Reliable Tuberculosis Detection Using Chest X-Ray With Deep Learning, Segmentation and Visualization," IEEE Access, vol. 8, pp. 191586–191601, 2020, doi: 10.1109/ACCESS.2020.3031384.

[13] O. Hrizi et al., "Tuberculosis Disease Diagnosis Based on an Optimized Machine Learning Model," Journal of Healthcare Engineering, vol. 2022, no. 1, p. 8950243, 2022, doi: 10.1155/2022/8950243.

[14] K. Munadi, K. Muchtar, N. Maulina, and B. Pradhan, "Image Enhancement for Tuberculosis Detection Using Deep Learning," IEEE Access, vol. 8, pp. 217897–217907, 2020, doi: 10.1109/ACCESS.2020.3041867.

[15] L.-S. Li, L. Yang, L. Zhuang, Z.-Y. Ye, W.-G. Zhao, and W.-P. Gong, "From immunology to artificial intelligence: revolutionizing latent tuberculosis infection diagnosis with machine learning," Military Med Res, vol. 10, no. 1, p. 58, Nov. 2023, doi: 10.1186/s40779-023-00490-8.

[16] A. H. Al-Timemy, R. N. Khushaba, Z. M. Mosa, and J. Escudero, "An efficient mixture of deep and machine learning models for covid-19 and tuberculosis detection using x-ray images in resource limited settings," Artificial Intelligence for COVID-19, pp. 77–100, 2021.

[17] X. A. Inbaraj, C. Villavicencio, J. J. Macrohon, J.-H. Jeng, and J.-G. Hsieh, "A novel machine learning approach for tuberculosis segmentation and prediction using chest-x-ray (CXR) images," Applied Sciences, vol. 11, no. 19, p. 9057, 2021.

[18] "Tuberculosis (TB) Chest X-ray Database." Accessed: Jun. 24, 2024. [Online]. Available: https://www.kaggle.com/datasets/tawsifurrahman/tuberculosis-tb-chest-xray-dataset

[19] Y. Feng and G.-G. Wang, "A binary moth search algorithm based on self-learning for multidimensional knapsack problems," Future Generation Computer Systems, vol. 126, pp. 48–64, Jan. 2022, doi: 10.1016/j.future.2021.07.033.

Complex Environmental Localization of Scenic Spots by Integrating LANDMARC Localization System and Traditional Location Fingerprint Localization

Shasha Song[1]*, Cong Li[2]
Tourism College, Yellow River Conservancy Technical Institute, Kaifeng, 475004, China[1]
The Information Engineering Institute, Yellow River Conservancy Technical Institute, Kaifeng, 475004, China[2]

***Abstract*—The scenic spot contains complex and changeable indoor and outdoor environments, some of which may be difficult to work effectively due to signal occlusion, multipath effect and other factors. In response to this problem, this paper proposes a method of Location Identification Based on the Dynamic Active Radio Frequency IDentification Calibration system and fingerprint localization system. It aims to improve positioning accuracy and reliability in the complex environment in the scenic spot. Firstly, the Location Identification Based on Dynamic Active Radio Frequency IDentification Calibration system is analyzed and improved. Then the improved positioning algorithm is applied to the complex environment of the scenic spot. Finally, the positioning results of the improved positioning algorithm in the complex environment of the scenic spot are tested. The experimental results show that when the K value is set to 4, the reader is arranged in the four corners and the center of the area, and the label density is set to 6×6, the average error of the research system in terms of error control is only 0.32, which is 0.28 less than that of the ultrasonic positioning system. All in all, the combination of Location Identification Based on Dynamic Active Radio Frequency IDentification Calibration system and traditional location fingerprint location of the scenic spot complex environment positioning scheme, it has shown great advantages in positioning accuracy, stability and real-time.**

Keywords*—LANDMARC; localization system; fingerprint localization; environmental localization; scenic spot

I. INTRODUCTION

In today's rapidly changing technology, location localization services for scenic spot tourists have become indispensable tools for improving tourist experience and optimizing scenic spot management [1-2]. However, the unique and complex environmental characteristics of scenic spots, such as variable terrain, dense buildings, and pedestrian flow, pose unprecedented challenges to the accuracy and stability of positioning systems. These complex environmental factors may not only hinder the propagation of signals, leading to biased localization results but also cause the localization system to fail in certain areas due to signal interference and occlusion. In recent years, the Location Identification Based on Dynamic Active Radio Frequency Identification Calibration (LANDMARC) localization system has become a focus in the field of indoor localization due to its advantages based on Radio Frequency Identification (RFID) technology, such as high localization precision, excellent stability, and easy deployment. Domestic and foreign scholars have conducted extensive and in-depth research on the performance of the LANDMARC localization system, with a particular focus on improving localization precision and enhancing system stability and environmental adaptability. Duan et al. raised an innovative method to solve the sensitivity of the LANDMARC localization system to environmental noise, which uses Newton interpolation to calculate the distance between the tested label and the reader. This method effectively improved the stability and localization precision of the system, making the LANDMARC localization system more reliable in practical applications [3]. In addition, Duan et al. optimized and improved the original indoor localization algorithm to address the issue of pre-deploying a large number of reference labels in the LANDMARC localization system. Through MATLAB simulation experiments, it verified the significant effect of the improved LANDMARC (I-LANDMARC) localization algorithm in reducing localization errors and improving localization precision [4].

However, the LANDMARC localization system also has some limitations, such as sensitivity to environmental noise and the need to deploy a large number of reference labels in advance. However, traditional location fingerprint localization methods construct a location fingerprint library by collecting environmental signal features, and use machine learning and other algorithms for location estimation. This method does not require additional equipment deployment and exhibits certain robustness to environmental noise. With the vigorous advancement of machine learning, deep learning and other technologies, traditional location fingerprint localization methods have made significant progress in localization precision, real-time performance, and robustness. For example, Lu et al. proposed a fingerprint database matching localization method with homologous multi-channel pseudo satellites, and verified its localization performance under dynamic and static conditions through extensive experiments. In an indoor testing environment, the dynamic average localization precision of this method reached 0.39 meters, with a 95% localization error better than 0.85 meters. In a real airport environment, the dynamic average localization accuracy was 0.75 meters, the maximum localization error was 1.69 meters, and 92% of the localization error was better than 1 meter [5]. At the same time, Han et al. proposed dynamic fusion features as a new fingerprint formation method and tested it in indoor environments. This method improved the system's feature resolution in both fingerprint features and similarity

measurement, had good noise resistance, and effectively reduced localization errors [6]. In addition, to improve the localization effect of GPS in indoor environments, the Uradzinski team proposed a method based on average threshold and effective data domain filtering to optimize the fingerprint database of ZigBee technology. Indoor experiments conducted by Waemmia and Mazuri University denoted that this method extends the localization distance by more than 30 meters without reducing localization precision [7].

Based on this, the study proposed a complex environment localization system for scenic spots that integrates LANDMARC localization system and traditional location fingerprint localization. The research combined the positioning accuracy characteristics of LANDMARC with the environmental adaptability of location fingerprints, filling the gap in the market for high-precision and stable positioning in the complex environment of scenic spots, and opening up a new path for the development of positioning technology. At the same time, by integrating the advantages of the two positioning technologies, the research can effectively overcome the limitations of single technology application and promote the further development of positioning technology and even the entire Internet of Things field.

The research is divided into four sections. Section I is the introduction, which introduces the research method and lays the foundation for the research. Section II is the method of the research, which analyzes the I-LANDMARC localization algorithm and its application in the location of the complex environment of scenic spots. Section III is the result of the research, which is the analysis of the application results of I-LANDMARC localization algorithm in the location of scenic complex environment. Section IV is the conclusion part, that is, the analysis and evaluation of the experimental results of the research model.

II. Methods and Materials

A. LANDMARC Localization System

The LANDMARC localization system is an indoor localization system based on RFID technology. The core algorithm is based on the Received Signal Strength Indication (RSSI) and utilizes the centroid weight algorithm to correct blind spots in traditional localization by real-time obtaining the RSSI value of the reference label, thereby improving the accuracy of object localization [8-9]. In RFID systems, readers are also known as interrogators, readers, or RFID devices. It can read or write data from electronic tags, independently perform data reading and processing, and can also be combined with computers to perform related operations on tags [10-11]. The basic components of the reader are illustrated in Fig. 1.

As shown in Fig. 1, the basic components of the reader mainly include radio frequency (RF) interface module, logic control unit, and antenna. The RF interface module is the core part of the reader, responsible for generating and receiving wireless RF signals [12]. The RF interface module sends energy and information to the electronic tag through an antenna and receives response signals from the electronic tag [13-14]. The logic control unit is the control center of the reader, which receives instructions from the backend application software system and controls the RF interface module to send and receive signals. The logic control unit is also responsible for decoding the response signal of the electronic tag, extracting the data information from it, and transmitting it to the computer's data management system for processing. The antenna is the physical interface for wireless communication between the reader and electronic tags. It is responsible for sending out wireless RF signals generated by the RF interface module and receiving response signals from electronic tags. The layout diagram of the LANDMARC localization system is denoted in Fig. 2.

Assuming that the number of readers in the Spider system is N, the amount of reference tags is Y, and the amount of test tags is P, the expression for the signal strength vector matrix E of the reference tags on each reader is defined, as shown in Eq. (1).

$$E = \begin{bmatrix} E_1^1 & E_2^1 & \cdots & E_n^1 \\ E_1^2 & E_2^2 & \cdots & E_n^2 \\ \vdots & \vdots & \vdots & \vdots \\ E_1^y & E_2^y & \cdots & E_n^y \end{bmatrix} \quad (1)$$

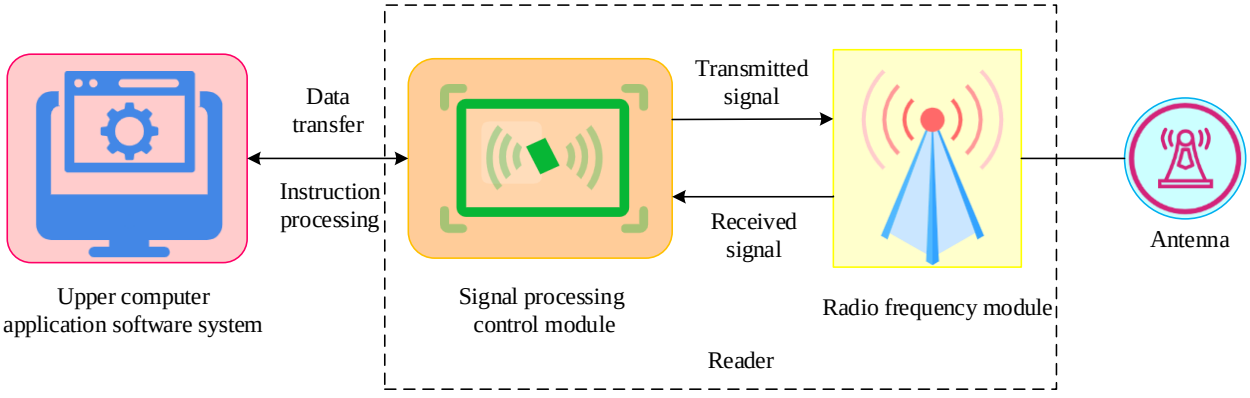


Fig. 1. Schematic diagram of the basic components of the reader.

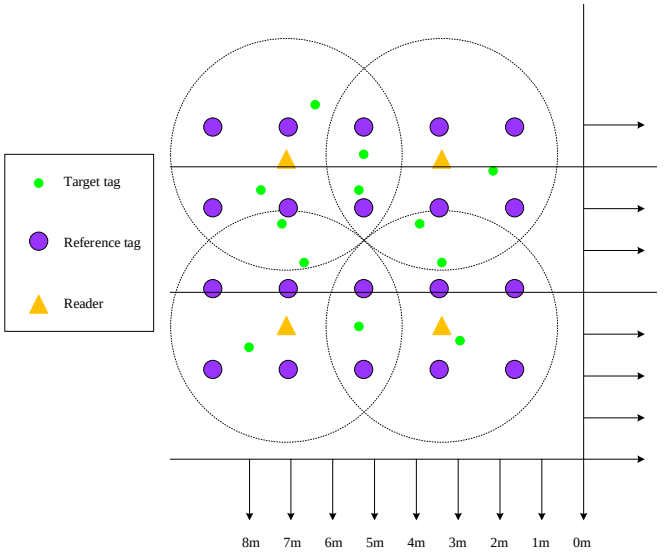


Fig. 2. LANDMARC localization system layout.

In Eq. (1), E_n^m represents the RSSI value when the n th reader reads the y th reference tag. Assuming that the signal strength vector received by the reader when reading an unknown point label is M , the expression for M is shown in Eq. (2).

$$M = \begin{bmatrix} M_1^1 & M_2^1 & \dots & M_n^1 \\ M_1^2 & M_2^2 & \dots & M_n^2 \\ \vdots & \vdots & \vdots & \vdots \\ M_1^x & M_2^x & \dots & M_n^x \end{bmatrix} \quad (2)$$

In Eq. (2), M_n^x represents the RSSI value of the unknown point label x on the n th reader. The LANDMARC localization system utilizes the k-Nearest Neighbor (KNN) algorithm as part of its localization mechanism. The LANDMARC system first collects signal strength or other relevant information about reference labels, and constructs a database containing the location information of these labels [15-16]. Then, when the system receives signals from unknown labels, it calculates the similarity between these signals and the reference label signals in the database, usually measured using metrics such as Euclidean distance. After obtaining the RSSI value of the target label, it is matched with the virtual label in the positioning area. If the difference is less than a specific threshold, the virtual label is marked as valid. If the difference is greater than a specific threshold, it is considered an invalid virtual label and filtered out from the valid neighboring electronic map. The low probability position filtering diagram of adjacent electronic maps is shown in Fig. 3.

Finally, the KNN algorithm finds k reference labels that are most similar to the unknown label signal, and this process is usually achieved by combining the positions of these reference labels through example weighting [17]. The expression for the constructed distance matrix D is shown in Eq. (3).

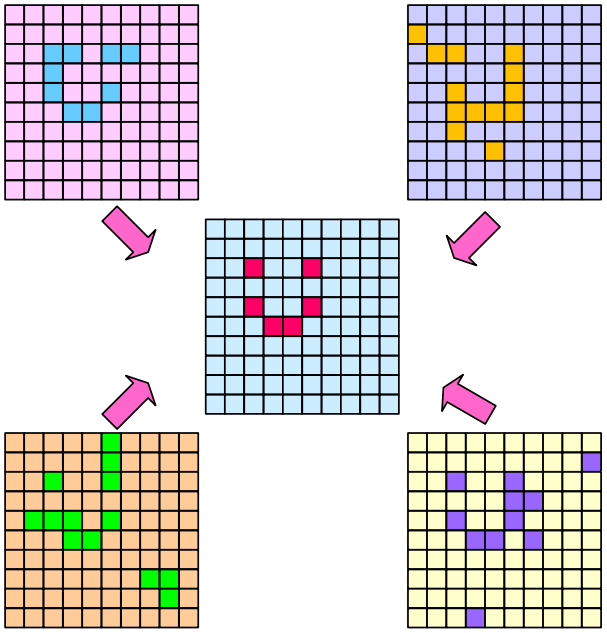


Fig. 3. Small probability location filtering of adjacent electronic map.

$$D = \begin{bmatrix} D_1^1 & D_2^1 & \dots & D_y^1 \\ D_1^2 & D_2^2 & \dots & D_y^2 \\ \vdots & \vdots & \vdots & \vdots \\ D_1^x & D_2^x & \dots & D_y^x \end{bmatrix} \quad (3)$$

In Eq. (3), D_y^x represents the Euclidean distance between the unknown point label D_y^x and the y th reference label, as expressed in Eq. (4).

$$D_y^x = \sqrt{\sum_{k=1}^{n} (E_k^y - S_k^x)} \quad (4)$$

In Eq. (4), k represents the nearest neighbor reference label (NNRL), S_k^x represents the RSSI value of unknown point tag x on reader n . In the LANDMARC indoor localization system, the smaller the D_y^x , the higher the similarity or proximity between the unknown point label and the reference label. For the unknown point label x , the known coordinate information and corresponding weights of k nearest neighbor labels can be used for calculation. The coordinate calculation expression for the unknown point label x is shown in Eq. (5).

$$(e, f) = \sum_{i=1}^{k} w_i (e_i, f_i) \quad (5)$$

In Eq. (5), i means the number of NNRLs, w_i represents the weight, and the calculation method for w_i is shown in Eq. (6).

$$w_i = \frac{\frac{1}{(D_i^x)^2}}{\sum_{i=1}^{k} \frac{1}{(D_i^x)^2}} \quad (6)$$

The RSSI value label correlation mainly refers to the relationship between RSSI values between different reference points in wireless communication or localization systems. Due to the influence of various factors during the propagation of wireless signals, there may be some correlation between RSSI value labels between different reference points. Assuming two variables are H, Z, the expression for the correlation coefficient is shown in Eq. (7).

$$r_{H,Z} = \frac{Cov(H,Z)}{\sigma_H, \sigma_Y} = \frac{\frac{1}{j}\sum_{i=1}^{j}(h_i - u_H)(z_i - u_Z)}{\sigma_H, \sigma_Y} \quad (7)$$

In Eq. (7), $Cov(H,Z)$ represents the covariance difference of H,Z, σ_H, σ_Y represent the variance of H,Z, u_H, u_Z represent the mean of H,Z, and j represents the number of samples.

B. LANDMARC Localization System based on Position Fingerprint Localization

LANDMARC localization technology has shown its unique advantages in many application scenarios, however, its technology precision is still limited. The main drawback is that the precision of LANDMARC localization is highly dependent on the precision of reference label sampling values. In complex environments, due to the presence of various interference factors, the real-time RSSI values obtained fluctuate greatly, and the degree of interference received by adjacent labels varies. This inconsistency brings errors to the position calculation of unknown points. To further raise the precision and stability of the LANDMARC localization algorithm, the study combines the stable fingerprint library of the position fingerprint method with the real-time signal strength information of the LANDMARC system. The position fingerprint localization method is an advanced technology based on wireless signal features for position estimation. This method achieves precise localization by linking different positions in the actual environment with their unique "fingerprints". Among them, fingerprint data is usually established by collecting RSSI values received at various locations, which represent the unique signal characteristics of each location. The fingerprint database localization process is shown in Fig. 4.

In the process of building a fingerprint information database, the study first built a LANDMARC localization system, which reads and collects the RSSI values of all reference labels through multiple readers [18-19]. These collected data form vectors, each containing RSSI measurements from different readers for the same reference label. Then, for each reference label y, the study extract the corresponding $RSSI$ value from each set of vectors. In order to evaluate the stability and distribution of these $RSSI$ values, statistical analysis was conducted on these signal samples, and the mean u and variance σ of the $RSSI$ samples for each label were calculated. The expression for calculating the sample mean u is indicated in Eq. (8).

$$u = \frac{1}{l}\sum_{y=1}^{l} RSSI_y \quad (8)$$

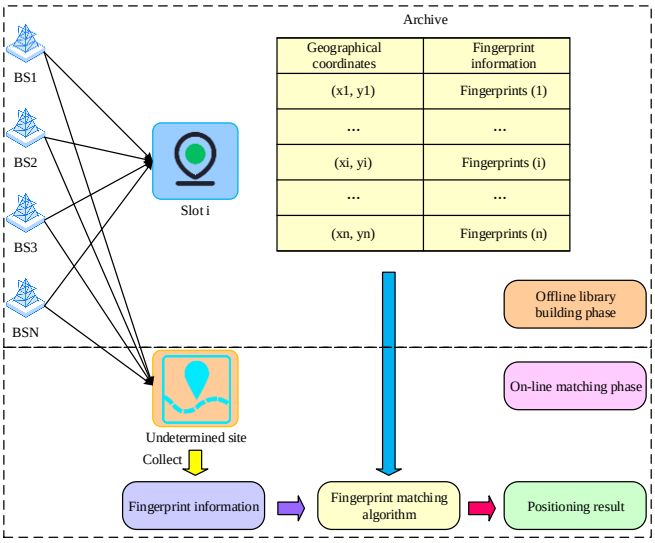


Fig. 4. Fingerprint database localization process.

In Eq. (6), l means the total amount of samples, and $RSSI_y$ represents the $RSSI$ value of the reference label y in the table. In order to fully utilize the collected data and improve positioning precision, study needs to ensure that fingerprint information contains as much and effective evidence as possible. To this end, the variance information of the reference point is included in the fingerprint for calculation. The variance information reflects the degree of dispersion of the RSSI value at the reference point, providing important information about the stability of the signal at that point [20-21]. The calculation formula for variance σ is shown in Eq. (9).

$$\sigma = \frac{1}{l}\sum_{y=1}^{l}(RSSI_y - u)^2 \quad (9)$$

In the process of constructing a fingerprint information database, to raise the stability and reliability of the data, a limited amplitude sliding filter algorithm was studied to preprocess the sequence composed of every seven consecutive sample data. The processed data is statistically calculated to obtain the mean u of each group of data and the variance σ of the original data, which together constitute the fingerprint information F_y of the reference label y. The expression for the fingerprint library value F_y of the reference tag y is shown in Eq. (10).

$$F_y = (a_y, b_y, u_1, ..., u_k, \sigma_1, ..., \sigma_k) \quad (10)$$

In Eq. (10), (a_y, b_y) represents the position coordinates

of reference label y, and u_1 represents the sample mean. In the final stage of building a fingerprint database, the study will summarize the fingerprint information of all reference labels processed on all readers, in order to establish a complete and comprehensive fingerprint database. The flowchart of the LANDMARC localization system based on location fingerprint localization is shown in Fig. 5.

C. Application of I-LANDMARC Localization Algorithm in Complex Environment Localization of Scenic Spots

The research will apply the LANDMARC localization system based on location fingerprint localization to a certain scenic spot, where n readers and y reference tags will be deployed to achieve precise location estimation of tourists or other moving targets. However, due to the unique geographical environment and limitations of the scenic spot, the layout of reference labels did not follow the traditional regular layout. The schematic diagram of the location environment of a certain scenic spot is shown in Fig. 6.

Due to the irregular layout of the scenic spot, it may lead to the misselection of neighboring or problematic labels, resulting in a decrease in localization precision. Therefore, the study introduces a quadratic weighted localization method to precisely calculate the coordinates of the labels to be located [22-23]. Firstly, by calculating the Euclidean distance between the reference label and the label to be located, k reference labels with the smallest distance are selected as the set of NNRLs, denoted as k1, k2, k3, and k4. Then, based on the NNRL, the weighted position coordinates of the label to be located are preliminarily calculated, denoted as O (x ', y'), as

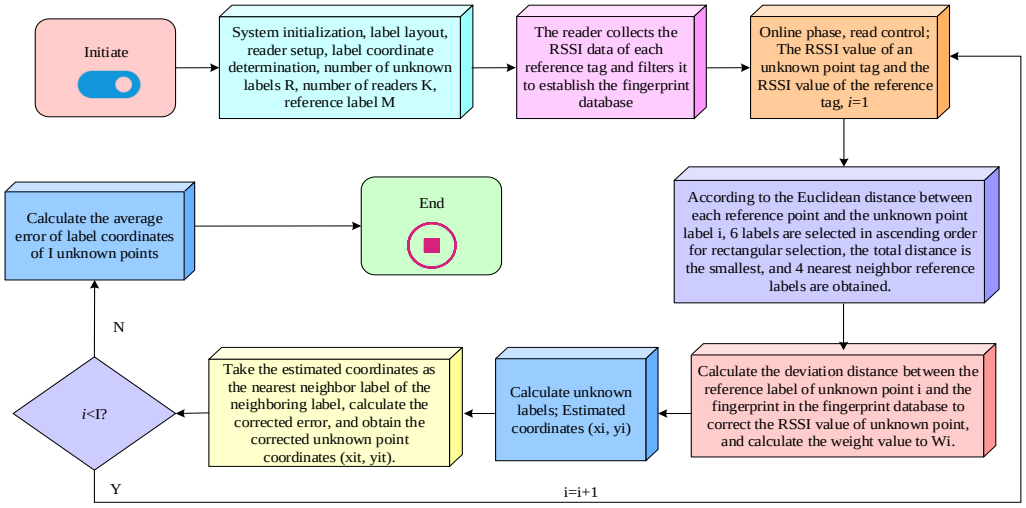


Fig. 5. Flowchart of LANDMARC localization system based on location fingerprint localization.

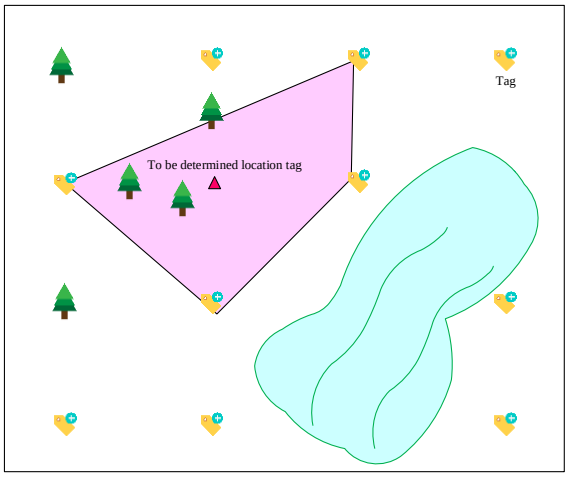


Fig. 6. Location environment diagram of a scenic spot.

In Fig. 7, in the irregular quadrilateral region, the line segment formed by connecting point O (x ', y') with four known points k1, k2, k3, and k4 is divided into four triangular sub regions. To determine the center coordinates of the inscribed circles in each triangle subregion, the study labeled these centers as O1, O2, O3, and O4. Since the process of solving the center of the inscribed circle in each triangle is the same, taking the solution of the center of the inscribed circle O3 in a triangle as an example, assuming that the coordinates of points A, B, and C are A (xa, ya), B (xb, yb), and C (xc, yc), respectively. Based on these coordinate points, the expression for the slope k_{AB}, k_{BC}, k_{AC} of the equation of the line segment AB, AC, and BC can be derived, as shown in Eq. (11).

$$k_{AB} = \frac{y_a - y_b}{x_a - x_b}, k_{BC} = \frac{y_c - y_b}{x_c - x_b}, k_{AC} = \frac{y_c - y_b}{x_c - x_b} \quad (11)$$

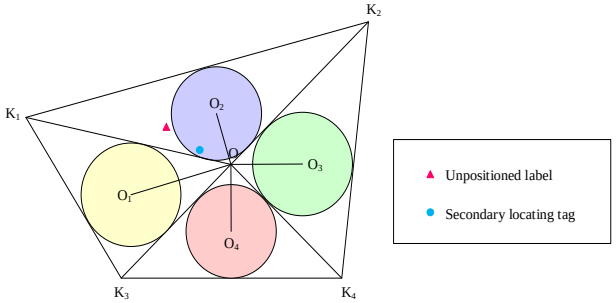


Fig. 7. The irregular layout of labels.

From Eq. (11), the linear equations of AB, AC, and BC can be obtained, and the calculation method for the linear equations of AB, AC, and BC are shown in Eq. (12).

$$\begin{cases} L_{AB}: k_{AB}x - y + b_1 = 0 \\ L_{BC}: k_{BC}x - y + b_2 = 0 \\ L_{AC}: k_{AC}x - y + b_3 = 0 \end{cases} \tag{12}$$

The calculation method for b_1, b_2, b_3 in Eq. (12) is shown in Eq. (13).

$$\begin{cases} b_1 = \dfrac{y_b x_a - y_a x_b}{x_a - x_b} \\ b_2 = \dfrac{y_b x_c - y_a x_c}{x_c - x_b} \\ b_3 = \dfrac{y_b x_c - y_a x_c}{x_a - x_c} \end{cases} \tag{13}$$

In triangle ABC, it assumes that the coordinates of the center O3 of its inscribed circle are (x3, y3). Due to O3 being the center of an inscribed circle, according to the properties of the inscribed circle, the distance between O3 and the three sides AB, AC, and BC of triangle ABC must be equal [24-25]. This property can be formalized through a mathematical expression, where the vertical distance between O3 and edges AB, AC, and BC have the same value, as shown in Eq. (14).

$$\frac{k_{AB}x_3 - y_3 + b_1}{\sqrt{k_{AB}^2 + (-1)^2}} = \frac{k_{BC}x_3 - y_3 + b_2}{\sqrt{k_{BC}^2 + (-1)^2}} = \frac{k_{AC}x_3 - y_3 + b_3}{\sqrt{k_{BC}^2 + (-1)^2}} \tag{14}$$

By using Eq. (14), the coordinates of the three inscribed circle centers O3 (x3, y3) of triangle ABC can be determined, and the same method can be applied to obtain the coordinates of the other three inscribed circle centers O1 (x1, y1), O2 (x2, y2), and O4 (x4, y4). To evaluate the relationship between the centers of these four inscribed circles and the preliminary estimated position of the target label, the study calculated the distance between the centers of these four inscribed circles and the first weighted position coordinates (x ', y') of the target label, denoted as d_1, d_2, d_3, d_4 . The calculation expression for d_1, d_2, d_3, d_4 is shown in Eq. (15).

$$\begin{cases} d_1 = \sqrt{(x_1 - x') + (y_1 - y')^2} \\ d_2 = \sqrt{(x_2 - x') + (y_2 - y')^2} \\ d_3 = \sqrt{(x_3 - x') + (y_3 - y')^2} \\ d_4 = \sqrt{(x_4 - x') + (y_4 - y')^2} \end{cases} \tag{15}$$

The study uses d_1, d_2, d_3, d_4 as weight factors for quadratic weighted localization, which reflect the proximity between the target label and the center of each inscribed circle, and can be used to optimize the accuracy of localization results. The expression for calculating the weight of quadratic weighting is shown in Eq. (16).

$$\omega_j^{'} = \frac{\frac{1}{d_j^2}}{\sum_{j=1}^{k} \frac{1}{d_j^2}}, j = 1, 2, ..., k \tag{16}$$

The coordinates obtained from the second weighted calculation are used as the final position coordinates of the label to be located, denoted as (x ", y "). This coordinate needs to comprehensively consider the geometric relationship between the center of each inscribed circle and the label to be located, in order to ensure the accuracy and reliability of the final position coordinates. The calculation method for coordinates (x ", y ") is shown in Eq. (17).

$$(x'', y'') = \sum_{i=1}^{k} \omega_j^{'}(x_i, y_i) \tag{17}$$

In order to evaluate the accuracy of positioning, the study introduced positioning error e'. The localization error e' represents the degree of difference between the actual position and the final position calculated by the algorithm. The expression for localization error e' is shown in Eq. (18).

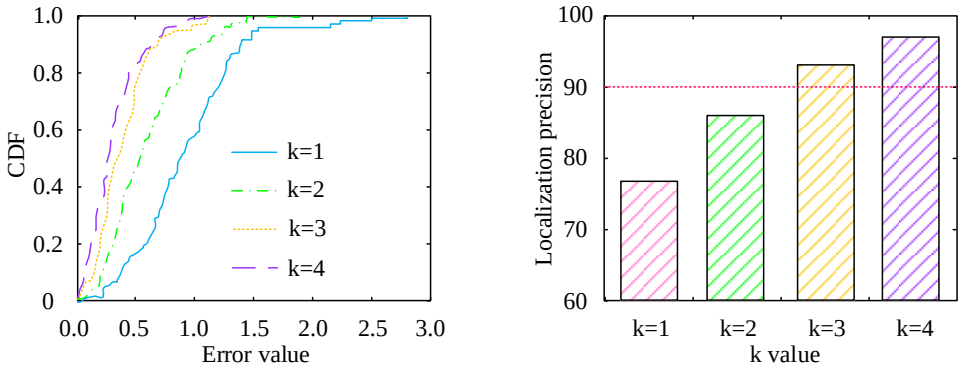
$$e' = \sqrt{(x'' - x_0)^2 - (y'' - y_0)^2} \tag{18}$$

III. Results

A. Simulation Analysis of I-LANDMARC System based on Fingerprint Library

To ensure the stability of the performance of the I-LANDMARC system based on fingerprint library, it is first necessary to determine the k value of the KNN algorithm. In practical applications, it needs to select the appropriate k value with specific scenarios and requirements. This usually requires experimentation and simulation of the system to find the optimal k value setting. Set different k values for experiments and compare the error and localization precision of the system with different k values, as indicated in Fig. 8.

As shown in Fig. 8 (a), as the amount of NNRLs k gradually increased from 1 to 4, the probability of positioning error less than 2m significantly increased from 50% to 70%. This trend indicated that increasing the value of k helps to improve the precision of the localization system. In Fig. 8 (b), when the k value was set to 4, the localization precision was highest at 98.27%, while when the k value was set to 1, the localization precision was lowest at 78.61%. In summary, under the condition of k=4, studying the positioning system can obtain the best localization results, which not only ensures the precision of localization but also takes into account the performance of the system. Therefore, the study chooses to set the k value of the system to 4. In order to verify the impact of reader placement and its correctness on the performance of the localization system, a study randomly placed 100 test points and tested them through different layout schemes of the system reader. The cumulative distribution function (CDF) of errors corresponding to different reader placement methods is shown in Fig. 9.



(a) Cumulative function curves for different values of k

(b)Localization precision different K values

Fig. 8. Comparison chart of error and localization precision of different k values.

Fig. 9 (a) shows the layout 1 of the reader in the original LANDMARC system, where the reader was located at four corner positions (0,0), (0,20), (20,0), and (20,20), with a localization error value of 36. Fig. 9 (b) shows the improved position layout 2 of the reader, as shown in Fig. 9 (b). The position coordinates of the reader were (5,5), (5,15), (15,5), and (15,15). At this point, the localization error value of layout 2 was 32. Fig. 9 (c) shows a layout 3 where a reader was added to the center of the region based on the improved layout 2, with its center coordinates located at (10,10). At this point, the localization error value of layout 3 was 34. Therefore, the study chose layout 2 as the localization scheme. Node density refers to the average connectivity of nodes in a network. A high node density indicates good network connectivity and more frequent communication between nodes, which can improve localization precision. Under other unchanged conditions, the value of k was set to 4, and the reader adopted the optimal layout 2. Simulation experiments were conducted on different placement densities of reference labels, and the simulation results of reference label layout with different densities are shown in Fig. 10.

Fig. 10 (a) is a dense layout 1 of 21 × 21, and 441 reference labels were required for this layout. In this layout, the localization error value of the system was 33. Fig. 10 (b) is layout 2 of 11 × 11, and 121 reference labels were required for this layout. Under this layout, the localization error value of the fixed system was 31. Fig. 10 (c) is a layout 3 of 6 x 6, and 36 reference labels were required. Under this layout, the localization error value of the system was 28. Through comparative analysis, it was found that the localization error gradually decreased as the density of the reference labels increased. Taking into account localization precision, system complexity, and cost, the optimal choice for the research was the 6 × 6 layout 3. This layout maintained high localization precision while also controlling the complexity and cost of the system, providing feasible solutions for practical applications. To assess the localization effect of the research system, a network consisting of four readers and 36 reference labels was constructed. These reference labels were evenly distributed at intervals of 4 meters, and 100 test labels were randomly placed. The localization outcomes of the research system are denoted in Fig. 11.

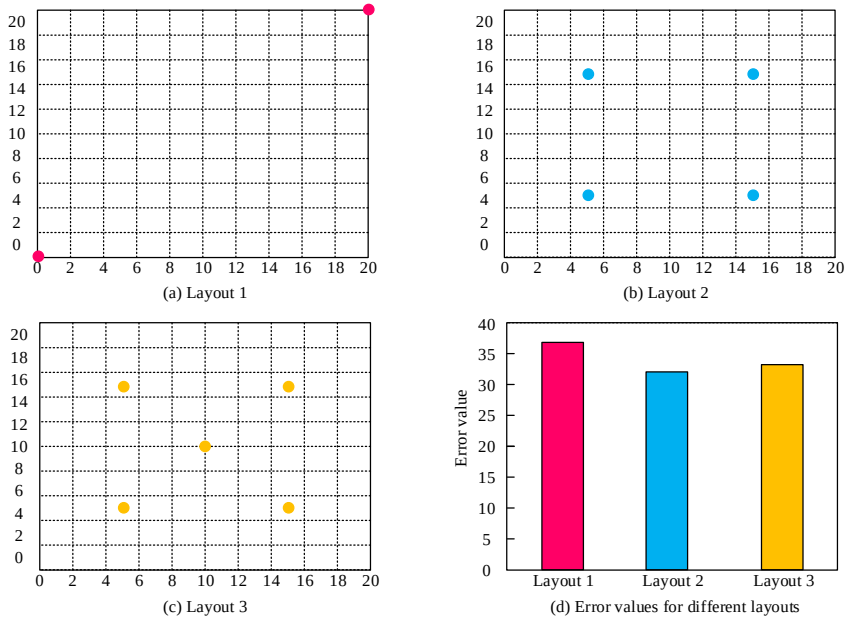


Fig. 9. The cumulative error distribution function graph of different reader placement modes.

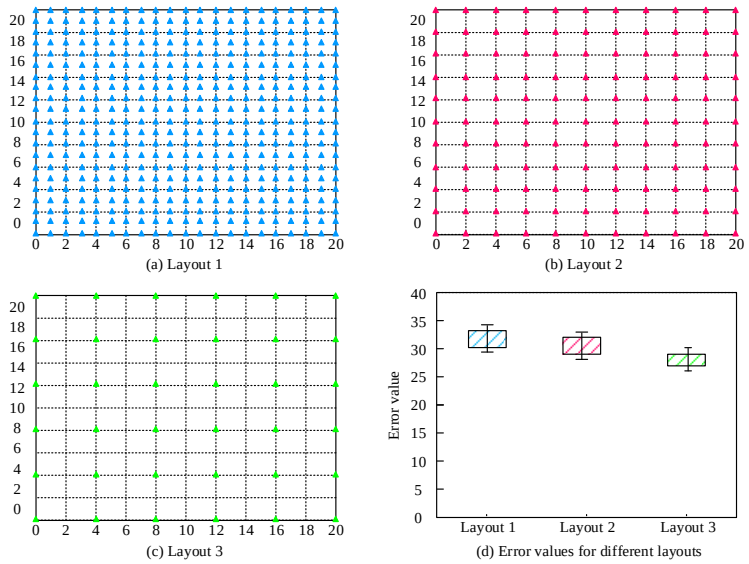


Fig. 10. Reference label layouts of different densities.

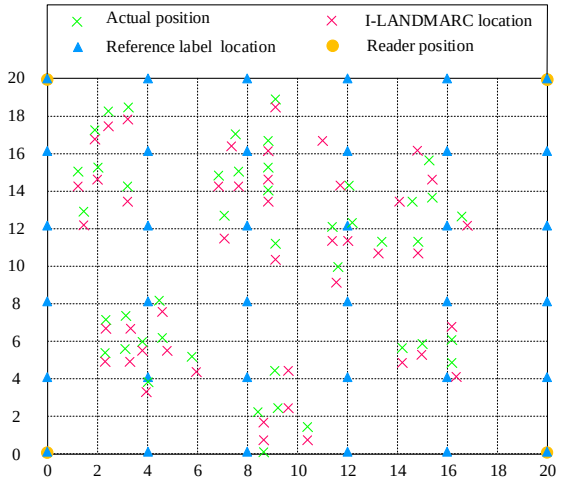


Fig. 11. Research system localization results in schematic diagram.

In Fig. 11, the research system also showed significant advantages in locating label locations. The research system was able to precisely locate the position of labels, thanks to its unique algorithm design and optimization, as well as the full utilization of reference label data. Specifically, the research system utilized advanced signal processing techniques, machine learning algorithms, or optimization algorithms to precisely calculate the distance or angle relationship between the tested label and the reference label, thereby determining the precise position of the label. In addition, the research system also considered the influence of environmental factors on the localization signal. Through appropriate compensation and correction, the localization precision was further improved, making the research system have greater potential and value in application scenarios that require high-precision localization.

B. *Analysis of Application Results of I-LANDMARC Localization Algorithm in Complex Environmental Localization of Scenic Spots*

To prove the practical application effect of the research localization system in scenic spots, a simulation testing environment was designed, in which four readers, 20 reference tags, and 8 labels to be located were deployed. The reference labels were arranged with a regular spacing of 2m to ensure that the localization algorithm was evaluated under unified and standard conditions. The simulation diagram of the scenic spot is denoted in Fig. 12.

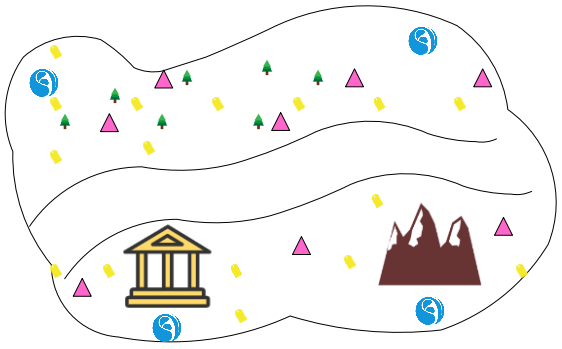


Fig. 12. Scenic spot simulation diagram.

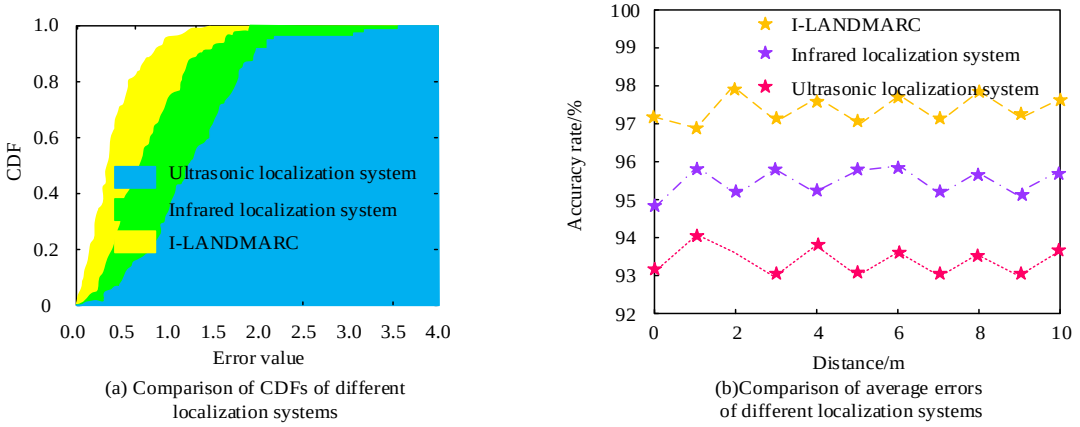
The research system's development environment and operation environment are indicated in Table I.

TABLE I. SYSTEM DEVELOPMENT AND OPERATION ENVIRONMENT

Operating system	Microsoft Windows7 Service Packl
Monitoring and development platform	LABVIEW
Development language	C, C++, VB, G
Database	Microsoft SOLServer2005 database
Other software and platform	Chengdu Wireless Long CC2431 positioning system, MATLAB
Operating system	Microsof Window system: CPU frequency :200 MHZ or higher, 500MHz or higher recommended, minimum memory: 2GRAM. Resolution: pixel cannot be less than 800*600;

To prove the effectiveness of the I-LANDMARC system based on fingerprint database, a comparative experiment was conducted on the error of traditional localization systems such as ultrasonic positioning system, infrared localization system, and LANDMARC localization system based on fingerprint database. Among them, the ultrasonic positioning system mainly determines the position of the object by measuring the time or phase difference of the ultrasonic signal propagating in space, and the infrared positioning system mainly uses the infrared propagation characteristics for positioning. The CDF and precision comparison of different localization systems are shown in Fig. 13.

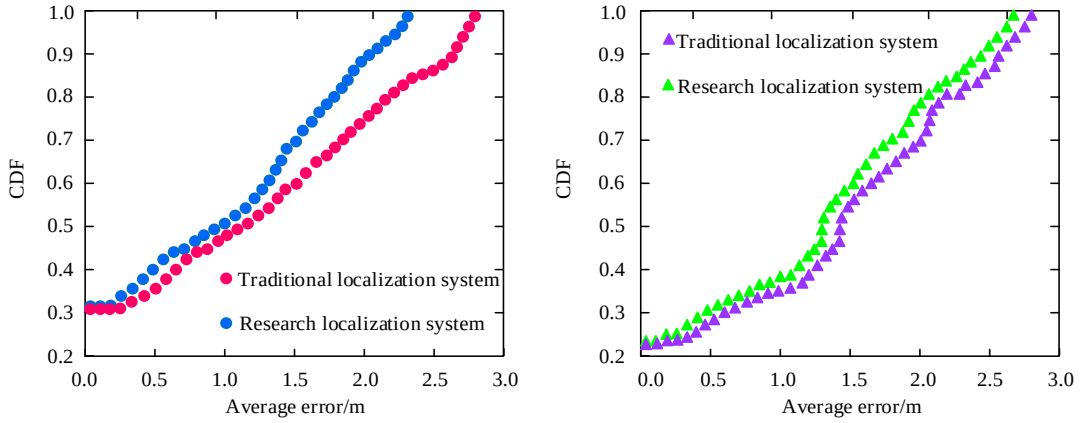
Fig. 13 (a) shows a comparison of the CDFs of different localization systems. From Fig. 13 (a) when the error value of the research system reached 1.86, its CDF curve gradually tended to stabilize, indicating the stability and efficiency of the research system in error control. Fig. 13 (b) shows a comparison of the accuracy of different localization systems. It can be seen from Fig. 13 (b) that the accuracy curve of the research system fluctuated around 97.6%, the accuracy curve of the ultrasonic localization system fluctuated around 95.8%, and the accuracy curve of the infrared localization system fluctuated around 93.4%. In summary, the average error of the research system was significantly lower than the other two traditional localization systems, which effectively reduced the generation of large errors. At the same time, the accuracy of scenic spot localization was extremely high. In practical applications, localization systems may face various complex environments and conditions. To prove the effectiveness of traditional localization systems and research localization systems in practical applications, a total of 500 sets of localization experiments were conducted to compare the effectiveness of traditional localization and research localization systems. The effect diagram of unknown point label distance and deviation localization is shown in Fig. 14.



(a) Comparison of CDFs of different localization systems

(b)Comparison of average errors of different localization systems

Fig. 13. CDF and accuracy of different positioning systems.



(a) Unknown point label distance localization effect

(b) Unknown point deviation localization effect diagram

Fig. 14. Unknown point label distance and deviation from the localization effect.

Fig. 14 (a) shows the effect of unknown point label distance localization. As shown in Fig. 14 (a), when the average error value of the research system reached 2.37, the value of the CDF rapidly increased to 1. However, traditional localization systems required a higher average error value of 2.83 to achieve a CDF value of 1. Fig. 14 (b) shows the localization effect of unknown point deviation. From Fig. 14 (b), when an unknown point deviated, traditional positioning systems needed to reach an average error value of 2.76 in order to achieve a CDF value of 1. Under the same deviation conditions, the research system only needed an average error value of 2.53, and the CDF value reached 1, indicating that the research system can still maintain

high localization precision and stability when facing unknown point deviations. In summary, the research system has shown superior performance in both unknown point label distance localization and unknown point deviation localization in scenic areas compared to traditional localization systems, especially in terms of localization precision and stability, with significant improvements.

C. Discussion

In the research of complex environment positioning in scenic spots, high-precision and efficient localization services are crucial for tourist safety, scenic spot management, and personalized services. The LANDMARC localization system has been widely used in the field of wireless localization due to its unique working mechanism and advantages. However, a single technology often fails to meet the localization requirements in complex environments. Therefore, combining LANDMARC with traditional location fingerprint localization technology can integrate the advantages of both and achieve high-precision and high-efficiency localization. This fusion technology not only improves localization precision, but also reduces computational complexity and time cost through optimization algorithms. This is similar to the results obtained in the study of progressive target localization for underground tunnels based on compressed sensing grids by Tian et al. [26].

The study first explored the key parameters that affect the localization system. The value of k, as an important parameter for the amount of NNRLs, has a significant impact on localization precision. Research has found that increasing the k value appropriately can effectively improve the accuracy of the localization system. However, excessively high k values could also increase computational complexity and time costs. Therefore, in practical applications, it is necessary to balance localization precision and calculation speed, and choose an appropriate value of k. In the study, k=4 has been proven to be the best compromise solution, providing valuable reference for similar research in the future. In the study, k=4 proved to be the best compromise, which is consistent with the results of Ashenafi's team, Q's team, and K's team [27-29]. The placement of the reader also has a significant impact on localization precision. Research has found that moving the reader from the boundary to the middle can significantly improve localization precision. However, when the amount of readers increased to a certain extent, the improvement of localization precision by further increasing the number became limited. Therefore, in actual deployment, it is necessary to choose the appropriate number and placement of readers based on specific circumstances and cost factors. The layout of reference labels also had a significant impact on positioning accuracy. Research has found that localization error gradually decreased with the increase of reference label density, indicating that increasing the number of reference labels could improve localization precision. Therefore, the selected 6 × 6 layout scheme in the study maintained high localization precision while also controlling the complexity and cost of the system. This is consistent with the results of Rahmatillah et al. in the study on time difference detection of reference signals from in orbit cubic satellites based on atomic clocks [30].

To comprehensively evaluate the performance of the research system, the study compared it with ultrasonic and infrared localization systems. The localization accuracy of three systems were tested under the same testing environment and conditions. The data showed that the accuracy curve of the research system fluctuated around 97.6%, the accuracy curve of the ultrasonic localization system fluctuated around 95.8%, and the accuracy curve of the infrared localization system fluctuated around 93.4%, indicating the superiority of the research system in the field of scenic spot localization. Furthermore, in-depth research has been conducted on two aspects: distance localization of unknown point labels and deviation localization of unknown points. In terms of distance localization of unknown point labels, 500 sets of experiments were designed to simulate the random movement of tourists within the scenic area. When the average error value of the research system was 2.37, the value of the CDF quickly rose to close to 1, indicating that the system can accurately complete localization within this error range. This result was similar to the results obtained by Zhang et al. in their study on corner detection using point to the center of mass distance technology [31]. In terms of unknown point deviation positioning, the study simulated the possible deviation path of tourists in the scenic spot. Traditional localization systems needed to achieve an average error value of 2.76 to achieve a CDF value of 1. However, under the same deviation conditions, the research system only needed an average error value of 2.53 to achieve a CDF value of 1, indicating that the research system can still maintain high localization accuracy and stability when facing unknown point deviation. Hassan et al. also obtained similar results in the review of system integration and current integrity monitoring methods for localization in intelligent transportation systems [32].

IV. Conclusion

A complex environment positioning system for scenic spots that integrates LANDMARC positioning system and traditional location fingerprint positioning was proposed to address the issue of low localization effectiveness. Its effectiveness was verified through simulation experiments and actual deployment tests. This system showed significant advantages in localization precision, stability, and practicality, which was significantly improved compared to traditional ultrasonic and infrared localization systems. Especially in terms of distance localization of unknown point labels and deviation localization of unknown points, the system showed better performance than traditional localization systems, providing effective solutions for scenic spot localization problems. Although the research system showed superior performance improvement, there were still some potential shortcomings. The system has a high dependence on hardware devices, including the layout and density of readers and labels, which may require further optimization and adjustment in practical applications. Future research can further explore how to reduce the system's dependence on hardware devices, improve the system's robustness and scalability, and better adapt to different scenic environments and localization needs. At the same time, it is possible to conduct in-depth research on various challenges and problems that the system may face in practical applications, in order to propose more effective solutions and improvement measures.

ACKNOWLEDGMENT

The research is supported by: Henan Provincial Department of Education, "Henan Province Higher Vocational School Youth Backbone Teacher Training Plan" Exploration and Practice of Teaching Reform in Higher Vocational Colleges under the Background of the "Three Education" Reform - Taking the Course "Cocktail Practice" as an Example, (No. 2020GZGG055).

REFERENCES

[1] Y. Junjie and Y. Yuan. "Evaluation of Development Model of Community Residential Areas in a World Heritage Site: A Case Study of Wulingyuan Scenic Spot in Zhangjiajie," c. Res., vol. 14, no. 1, pp. 104-106, Jan, 2022. DOI: 15.04/landsc7106676694.

[2] Y. Gao, Y.Y. Chiang and X. Zhang. "Traffic volume prediction for scenic spots based on multi-source and heterogeneous data," Trans. GIS., vol. 26, no. 5, pp. 2415-2439, Nov, 2022. DOI: 10.1111/tgis.12975.

[3] R. Duan, Z. Li and Y. Yin. "Improvement of LANDMARC Indoor Positioning Algorithm," Int. J. Perform. Eng., vol. 16, no. 3, pp. 446-453, Mar, 2020, DOI: 10.23940/ijpe.20.03.p14.446453.

[4] J. Xu, Z. Li and K. Zhang. "The principle, methods and recent progress in RFID positioning techniques: A review," IEEE J. Radio Freq. Identif., vol. 7, pp. 50-63, Mar, 2023, DOI: 10.1109/JRFID.2022.3233855.

[5] H.U. Lu, B.G. Yu and H.S. LI. "Pseudolite Fingerprint Positioning Method under GNSS Rejection Environment," Acta Electron. Sinica., vol. 50, no. 4, pp. 811-822, July, 2022. DOI: 10.12263/DZXB.20211167.

[6] K. Han, Y. Xu and Z. Deng. "DFF-EDR: An Indoor Fingerprint Location Technology Using Dynamic Fusion Features of Channel State Information and Improved Edit Distance on Real Sequence," China Commun., vol. 18, no. 4, pp. 40-63, May, 2021.

[7] M. Uradzinski, H. Guo and M. Yu. "Improved indoor positioning based on range-free RSSI fingerprint method," J. Geodetic Sci., vol. 10, no. 1, pp. 23-28, Nov, 2020. DOI: 10.1515/jogs-2020-0004.

[8] H. Ai, X. Sun and J. Tao. "DRVAT: Exploring RSSI series representation and attention model for indoor positioning," Int. J. Intell. Syst., vol. 37, no. 7, pp. 4065-4091, July, 2021. DOI: 10.1002/int.22712.

[9] W Fan, L. Luo and H. Song. "Fault Early Warning in Air-insulated Substations by RSSI-Based Angle of Arrival Estimation and Monopole UHF Wireless Sensor Array," IET Gener. Transm. Distrib., vol. 14, no. 12, pp. 2345-2351, Dec, 2020. DOI: 10.1049/iet-gtd.2019.0813.

[10] D. Zhong, J. Zhou and G. Liu. "Missing Unknown Tag Identification Protocol Based on Priority Strategy in Battery-Less RFID System," IEEE Sens. J., vol. 23, no. 18, pp. 20845-20855, Sept, 2023. DOI: 10.1109/JSEN.2023.3239610

[11] W. Shi, J. Gao and Y. Cao. "Gain characteristics estimation of heteromorphic RFID antennas using neuro-space mapping," IET Microw. Antennas Propag., vol. 14, no. 1, pp. 1555-1565, Jan, 2020. DOI: 10.1049/iet-map.2020.0105.

[12] B. Meher and R. Amin. "A location-based multi-factor authentication scheme for mobile devices," Int. J. Ad Hoc Ubiquit. Comput., vol. 41, no. 3, pp. 181-190, Mar, 2022. DOI: 1504/IJAHUC.2022.126113

[13] C.Y. Cheng, J.C. Jhuang and P.Y. Wul. "Surface characteristics of polycarbonate by radio-frequency linear dielectric barrier plasma activation," Surf. Interface Anal., vol. 54, no. 1, pp. 3-12, April, 2022. DOI: 10.1002/sia.7009.

[14] M. R. Masinter. "Court ruling suggests electronic databases inaccessible to screen readers are not forbidden," Disabil. Complianc. High. Educ., vol. 27, no. 5, pp. 3-12, Oct, 2021. DOI: 10.1002/dhe.31182.

[15] K. Bhosle and V. Musande. "Evaluation of Deep Learning CNN Model for Recognition of Devanagari Digit," Artif. Intell. Appl., vol. 1, no. 2, pp. 114-118, Feb, 2023. DOI: 10.47852/bonviewAIA3202441.

[16] Z. Liu, Q.L. Lu and J. Gao. "A similarity-based data-driven car-following model considering driver heterogeneity," Transp. Res. Procedia, vol. 78, pp. 611-618, May, 2024. DOI: 10.1016/j.trpro.2024.02.076.

[17] N.N.Y. Liu, N.N.Z. Chen and A.W.C. Fu. "Optimal location query based on k nearest neighbours," Front. Comput. Sci. China, vol. 15, no. 2, pp. 105-117, Sep, 2021. DOI: 10.1007/s11704-020-9279-6.

[18] M. Kim, S.P. Hong and M. Kang. "Fingerprint-Based Millimeter-Wave Beam Selection for Interference Mitigation in Beamspace Multi-User MIMO Communications," Comput. Mater. Continua, vol. 2021, no. 1, pp. 59-70, February, 2021. DOI: 10.32604/cmc.2020.013132.

[19] C. Wu, S.N. Qi and C. Zhao. "Fingerprint location algorithm based on K-means for spatial farthest access point in Wi-Fi environment," J. Eng., vol. 2020, no. 4, pp. 115-119, April, 2020. DOI: 10.1049/joe.2019.0995.

[20] L Huang, Bao-guo Yu and Hong-sheng Li. "Pseudolite Fingerprint Positioning Method under GNSS Rejection Environment," Acta Electronica Sinica, vol. 50, no. 04, pp. 811-822, Sep, 2022. DOI: 10.12263/DZXB.20211167.

[21] X. Lv, L. Ding and G. Zhang. "Research on fingerprint feature recognition of access control based on deep learning," Int. J. Biometrics, vol. 13, no. 1, pp. 80-95, Dec, 2021. DOI: 10.1504/IJBM.2021.10034247.

[22] L. Dumitrescu, W. Qian and J.N.K. Rao. "Inference for longitudinal data from complex sampling surveys: An approach based on quadratic inference functions," Scand. J. Stat., vol. 48, no. 1, pp. 246-274, July, 2020. DOI: 10.1111/sjos.12448.

[23] B. Bai, Z. Xiao and Q. Wang. "Multi-Objective Trajectory Optimization for Freight Trains Based on Quadratic Programming," Transp. Res. Rec., vol. 2674, no. 11, pp. 466-477, October, 2020. DOI: 10.1177/0361198120937307.

[24] V. Arya, R. Goyal and M. Majji. "Linear Quadratic Regulator Weighting Matrices for Output Covariance Assignment in Nonlinear Systems," J. Guid. Control Dyn., vol. 46, no. 2, pp. 264-276, January, 2023. DOI: 10.2514/1.G006584

[25] X. Zeng, J. Zhang, and H. Li. "Application of the hybrid genetic particle swarm algorithm to design the linear quadratic regulator controller for the accelerator power supply," Radiat. Detect. Technol. Methods, vol. 5, no. 1, pp. 128-135, April, 2021.

[26] Z.J. Tian, X.W. Gong and F.Y. He. "Compressed sensing grid-based target stepwise location method in underground tunnel," Sensor Rev., vol. 40, no. 4, pp. 397-405, June, 2020. DOI: 10.1108/SR-12-2019-0303.

[27] M.K. Ashenafi and V.G. Paolo. "An Innovative Location Value Determination: Domain Disaggregation Additive Regression Approach," Int. J. Tomogr. Simul., vol. 34, no. 2, pp. 1-28, March, 2021. DOI: 10.15866/356134429.

[28] Q. D. Vo and P. De, "A Survey of Fingerprint-Based Outdoor Localization," in IEEE Communications Surveys & Tutorials, vol. 18, no. 1, pp. 491-506, Firstquarter 2016, doi: 10.1109/COMST.2015.2448632.

[29] K. Han and S. H. Cho, "Advanced LANDMARC with adaptive k-nearest algorithm for RFID location system," 2010 2nd IEEE International Conference on Network Infrastructure and Digital Content, Beijing, China, 2010, pp. 595-598, doi: 10.1109/ICNIDC.2010.5657852.

[30] R. Rahmatillah,

[31] R. Ninagawa and K. Aheieva. "Time Difference Detection of Atomic Clock-Based Reference Signal from a CubeSat in Orbit," Int. Rev. Aerosp. Eng. IREASE, vol. 1, no. 15, pp. 36-49, Novembe, 2022. DOI: 10.15866/irease.v15i1.21136.

[32] S. Zhang, L. Huangfu and Z. Zhang. "Corner detection using the point-to-centroid distance technique," IET Image Process., vol. 14, no. 14, pp. 3385-3392, August, 2020. DOI: 10.1049/iet-ipr.2020.0164.

[33] T. Hassan, El cm owafy Ahmed and K. Wang. "A review of system integration and current integrity monitoring methods for positioning in intelligent transport systems," IET Intell. Transport Syst., vol. 15, no. 1, pp. 43-60, February, 2020. DOI: 10.1049/itr2.12003.

Development of a 5G-Optimized MIMO Antenna with Enhanced Isolation Using Neutralization Line and SRRs Metamaterials

Chaker Essid[1], Linda Chouikhi[2], Alsharef Mohammad[3], Bassem Ben Salah[4], Hedi Sakli[5]*
SERCOM Laboratory-Tunisia Polytechnic School, University of Carthage, La Marsa 2078, Tunisia[1, 2, 4]
Department of Electrical Engineering-College of Engineering, Taif University, Taif, Saudi Arabia[3]
EITA Consulting, 7 Rue du Chant des oiseaux, 78360 Montesson, France[3, 5]
MACS Research Laboratory RL16ES22-National Engineering School of Gabes, Gabes University, Gabes, 6029, Tunisia[4, 5]

***Abstract*—This paper presents the design of a Multiple Input Multiple Output (MIMO) antenna intended for 5G wireless applications operating in the 3.5 GHz frequency range. The MIMO system consists of two adjacent antennas, measuring 100 mm × 80 mm × 1.6 mm, with spacing between radiating elements equal to one-eighth of the wavelength (λ/8). The antenna is constructed on an FR4 substrate with a permittivity of 4.3, and a microstrip line is employed for feeding the patch. Several techniques are employed to enhance the isolation between the antennas. Specifically, two decoupling methods are explored: the use of a neutralization line (NL) and the incorporation of metamaterial split-ring resonators (SRRs). Simulation results demonstrate substantial isolation, exceeding 20 dB with SRR implementation and more than 23 dB with the NL approach. Both individual antennas and the MIMO configuration are simulated, analyzed, and then physically fabricated for measurement, exhibiting good agreement between measured and simulated results. The study investigates the impact of each technique on antenna to determine the optimal configuration for the applications of 5G and IOT in different fields such as health (wireless medical telemetry systems (WMTS)). Remarkably, the introduction of metamaterial (MTM) with SRRs achieves a noteworthy reduction of mutual coupling by 23 dB while minimizing the mutual coupling to about 23 dB with NL insertion.**

***Keywords*—5G; antenna; MIMO; SRRs metamaterials; isolation; IOT; neutralization line (NL)**

I. INTRODUCTION

The increasing requirement for significantly improved data rates and extensive capacity has fueled the continuous evolution of technology and the progression of mobile and wireless communication networks. This necessity has led to the emergence of a new phase in mobile communication, known as the fifth generation or 5G [1]. 5G cellular networks are considered a critical component in enabling the next generation of the Internet, commonly known as the Internet of Things (IoT) [2, 3]. IoT essentially involves connecting a wide range of devices and people to a central IoT platform [4]. The core concept of IoT revolves around using sensors to collect data, which is then sent to a server for analysis. This analysis generates insights that are transformed into actions, leading to the creation of intelligent environments such as smart homes and vehicles [5]. The evolution of IoT enables the connection of numerous devices to the internet, with predictions suggesting that over 28 billion smart devices will be connected worldwide by 2021 [6]. Moreover, machine-to-machine (M2M) communication is expected to be utilized in over 15 billion devices, emphasizing the magnitude of IoT's impact. However, it is crucial to acknowledge that most of these connected devices depend on batteries for power, presenting environmental challenges due to their polluting nature [7].

To address this concern, researchers at the Georgia Institute of Technology in the United States have developed a rectifying antenna, often called a Rectenna [8]. The Rectenna integrates a rectifier and an antenna, acting as a vital element in wireless telecommunications systems. Its distinctive ability is to convert 5G waves into electric current, which could eventually enable mobile operators to serve as power providers for IoT devices [9, 10]. With the introduction of 5G, mobile operators are tasked with deploying 5G base stations in the 3.4-3.6 GHz band, using a specific operational mode called Time Division Multiplexing (TDD) [11]. TDD involves dividing the frequency band into separate transmit and receive segments, allowing the same frequency band to be used for both transmission directions. This feature is particularly beneficial for Massive Multiple Input Multiple Output (MIMO) technology [12], which has gained traction in wireless communication systems, especially in the context of 5G applications.

The antennas operating at 3.5 GHz are used in wireless medical telemetry systems (WMTS) to monitor patients' vital signs and transmit this data wirelessly to medical personnel [13]. These antennas enable the transmission of medical data over short to medium distances in various healthcare settings, providing reliable and high-speed communication for real-time monitoring and prompt responses to emergency situations [14,15].

However, the increasing number of antennas on a limited-size electrical component can lead to a phenomenon called mutual coupling (MC) [15, 16]. MC occurs when current densities generated by nearby antennas alter the radiation characteristics of neighboring antennas, potentially degrading MIMO system performance [17,18]. Although various techniques, including metamaterials (MTM) [19-21],

neutralization lines (NL) [22, 23], Defected Ground Structure (DGS) [24], parasitic or slot elements [25], have been explored to mitigate MC, there is ongoing research to enhance the isolation characteristics of MIMO antennas.

A variety of research has been done to increase the isolation characteristics of MIMO antennas by inserting MTM cells. The authors in [19] propose a flag-shaped MIMO antenna using MTM technique in the frequency of Sub-6 GHz 5G applications. It consists of two antennas printed on the FR-4 substrate and separated by a distance of λ/2. Results show a low mutual coupling reduction of 16dB, and the gain obtained was 3.28 dB at 3.1 GHz. Research in [20] applies the same technique to increase the decoupling over the two MIMO antennas spaced 2 mm. This method provides a coupling of minus of −45 dB in the 4.8 - 5 GHz band to be a suitable candidate for 5G applications.

Another idea developed by the researcher in [21], it is based on placing a MTM isolating structure between two circular antenna patches. After insertion of this cell MTM, it provides -23 dB of mutual coupling reduction between the two radiator elements. Other types of metamaterials are used to increase the decoupling between antennas, such as Electromagnetic Band Gap (EBG) [22], and Frequency Selective Surface (FSS) [23].

An alternative approach to enhance the performance of MIMO systems is introduced in [22], which involves the incorporation of a neutralization line between four L-monopole antennas with a semi-elliptical radiating patch. This method achieves a coupling reduction of approximately 20 dB between the elements of the antenna array within the frequency range of 3.36 - 3.68 GHz. Essentially, this line is modeled as an inductance to impede mutual coupling, functioning as a rejection filter. Similarly, a neutralization line is employed between two elliptical multi-antennas around the resonant frequency of 3.5 GHz for 5G mobile applications [24]. This configuration results in a transmission magnitude of less than −24 dB at the operating frequency, thereby improving gain, diversity, and radiation efficiency. Other researchers have adopted this isolation technique in [25], which consists of two tri-band antenna elements separated by 0.03 λ0. These two antennas are connected by U-shaped and inverted U-shaped NLs, suppressing the MC about −15 dB, −30 dB, and −20 dB at 2.3 GHz, 3.5 GHz, and 5.7 GHz, respectively.

The band 3.5 GHz (3.4 - 3.8) GHz has been exclusively allocated to the 5G mobile network, serving as the "heart" of 5G and providing subscribers with a high quality of service. Among the various frequencies utilized by 5G (700 MHz, 3.5 GHz, and 26 GHz), this band is considered sufficient as it offers adequate coverage and capacity [25]. The primary objective of this research is to evaluate the performance of MIMO antennas employing various techniques, with a focus on minimizing mutual coupling. Given that placing antennas in close proximity can lead to MC, the literature has explored methods for reducing MC, with an emphasis on isolation techniques involving NL and metamaterials. This paper also provides insights into the magnetic phenomena underpinning these methods and emphasizes the importance of parametric studies to optimize simulation results.

However, the deployment of 5G networks and IoT devices faces several challenges, including mutual coupling in Multiple Input Multiple Output (MIMO) antennas, which can degrade system performance. Addressing these challenges is crucial for realizing the full potential of 5G and IoT. This research is motivated by the need to enhance the isolation characteristics of MIMO antennas, thereby improving the overall performance and reliability of 5G networks and IoT applications. By developing novel techniques to reduce mutual coupling, we aim to contribute to the advancement of wireless communication systems and enable more efficient and reliable connectivity for a wide range of applications.

The main contributions of this paper are the development and evaluation of two innovative techniques for enhancing the isolation characteristics of Multiple Input Multiple Output (MIMO) antennas operating in the 3.5 GHz frequency band, which is crucial for 5G applications. Firstly, we propose the use of a neutralization line (NL) to significantly reduce mutual coupling between adjacent antennas, achieving an isolation of over 23 dB. Secondly, we introduce the integration of metamaterial split-ring resonators (SRRs) to further minimize mutual coupling, resulting in an isolation improvement of approximately 23 dB. These techniques are thoroughly analyzed through simulations and validated through physical measurements, demonstrating their effectiveness in real-world scenarios. The implications of these contributions are substantial for 5G and IoT applications. By improving the isolation characteristics of MIMO antennas, we can enhance the overall performance and reliability of 5G networks, enabling higher data rates, reduced interference, and better coverage. This is particularly beneficial for IoT applications, where reliable and high-speed communication is essential for real-time monitoring and control of various devices and systems. Our findings provide valuable insights and practical solutions for the design of advanced MIMO antennas, contributing to the advancement of wireless communication technologies and the realization of the full potential of 5G and IoT.

The structure of the paper is as follows: Section II elaborates on the design of the proposed antenna. Subsequently, Section III presents the simulation results of the MIMO antenna system. Section IV delves into the characteristics of the MIMO system incorporating decoupling techniques. Lastly, the conclusions are summarized in the final section.

II. Design of New Patch Antenna

To ensure the optimal performance of our advanced multi-antenna system, it is imperative to begin with a comprehensive understanding of the fundamental principles of antenna design. In our proposed project, our focus has been centered on three critical aspects: the antenna's geometric configuration, the designated frequency range for operation, and the implementation of techniques to minimize MC within the multi-antenna system.

Our initial step entailed the design of a rectangular patch antenna, which was powered through a microstrip transmission line with a specific width denoted as wf, adhering to the requisite 50-ohm characteristic impedance for the transmission line. The radiating element, characterized by dimensions Lp × Wp, was physically fabricated on an FR-4 substrate, with dimensions L × W, and a specified thickness (h). The substrate was chosen for its cost-effectiveness and the expeditious prototyping capabilities it offers. The resonant frequency was set at 3.5 GHz, a choice determined through calculations employing equations outlined in [26].

A pivotal step in enhancing the antenna's performance was the strategic alteration of its structure, which involved the incorporation of cuts and slots in both the radiating element and the ground plane (show Fig. 1). Fig. 2 illustrates the results of simulation of these structural adjustments in terms of the reflection coefficient (S11).

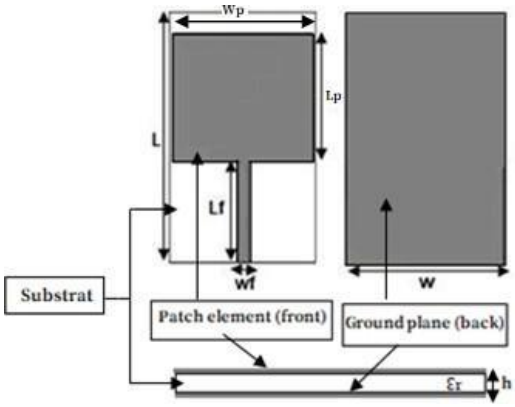


Fig. 1. Structure of the basic antenna in front and back views.

Remarkably, the implementation of partial modifications to the ground plane, featuring two small square patches on the left and right, yielded a maximum S11 value of approximately -20 dB within the 3.4 - 3.8 GHz operational band. To further broaden the bandwidth and ensure that the signal reflection remained below -10 dB, we introduced two triangular cuts at the top of the radiating element. This modification extended the frequency band from 2.2 GHz to 4.5 GHz. The incorporation of two additional triangular cuts at the bottom of the patch resulted in a notable enhancement in the antenna's adaptation to the 3.5 GHz frequency, with an S11 value now falling below -59 dB.

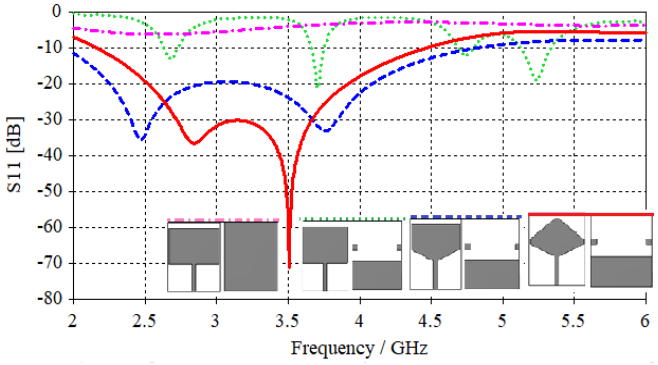


Fig. 2. Reflection coefficient S11 of the proposed antenna with and without modifications and optimized final antenna.

In order to enhance the performance of our antenna, the first crucial step involves modifying the structure by incorporating cuts and slots in the radiating element and the ground plane. The reflection coefficient (S11) results for all configurations are illustrated in Fig. 2. By making the ground plane partial, a maximum S11 value of approximately -20 dB in the 3.4-3.8 GHz operating band is achieved. To further extend the bandwidth (ensure a signal below -10 dB), two triangular cuts are applied at the top of the radiating element, resulting in a band spanning from 2.2 GHz to 4.5 GHz (as shown in Fig. 3).

Fig. 3 indicates that the optimum position is L4 = 8 mm, which provides the highest S11 at 3.5 GHz. By introducing two additional triangular cuts at the bottom of the patch, we fixed the optimal value of L4 at 8 mm and varied the values of L3 (ranging from 10 to 16 mm).

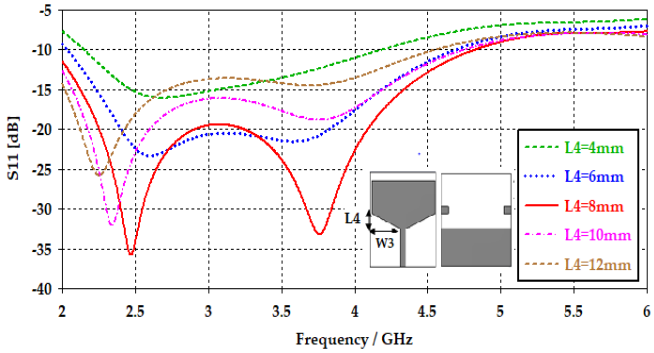


Fig. 3. Reflexion coefficient of the proposed antenna with different values of L4.

As illustrated in Fig. 4, the adaptation to the 3.5 GHz frequency shows improvement for L3=14 mm compared to the previous results, with S11 below -59 dB.

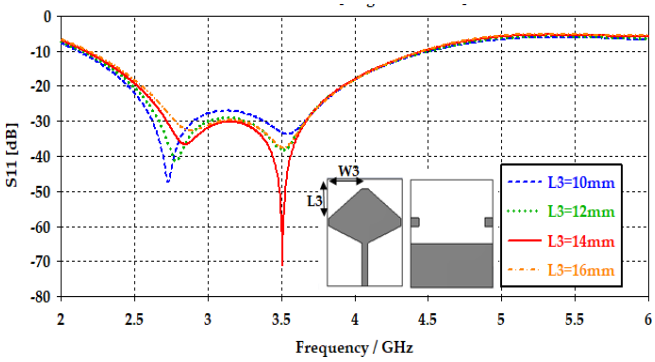


Fig. 4. Reflexion coefficient of the proposed antenna with different values L3.

The selections made for the dimensions presented in Table I were made with precision, considering both impedance matching and the constraints related to system size, all while considering the specific applications for which the antenna is intended. The incorporation of a slotted ground within the proposed antenna serves the purpose of enhancing the antenna's performance by amplifying inter-element isolation and reducing overall reflection (see Fig. 5).

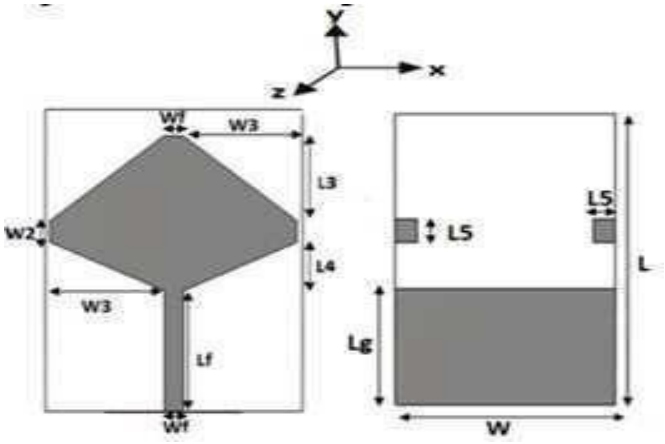


Fig. 5. Proposed antenna in front and bottom views.

TABLE I. OPTIMIZED DIMENSIONS OF THE PROPOSED ANTENNA

Components	Parameters	Values (mm)
Patch	Lf	20
	Wf	3.1
	W2	3.6
	W3	18.5
	L3	14
	L4	8
Dielectric substrate	W	40
	L	50
	h	1.6
	εr	4.4
Ground Plan	Lg	20
	L5	5

Fig. 6 presents the real and imaginary parts of the antenna impedance, revealing that at 3.5 GHz, the real part is approximately 50 Ω and the imaginary part is nearly 0 Ω. Consequently, the impedance demonstrates good adaptability. The antenna gain as a function of frequency is displayed in Fig. 7 within the [3.4-3.8] GHz band. It is observed that the maximum gain value obtained is around 2.8 dB at 3.5 GHz.

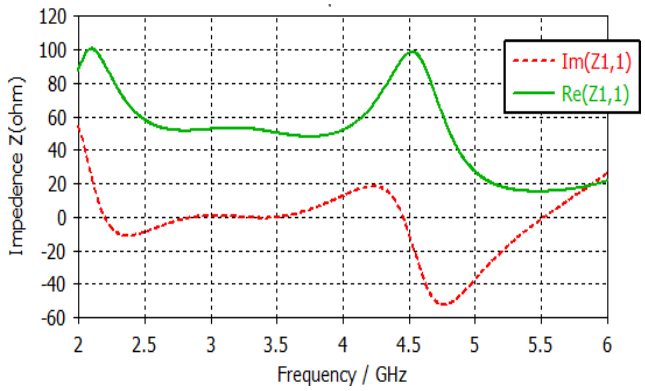


Fig. 6. Input impedance of the studied antenna.

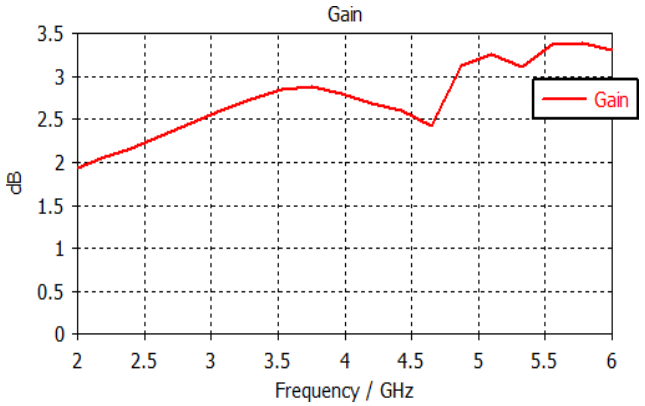


Fig. 7. Variation of gain as a function of frequency.

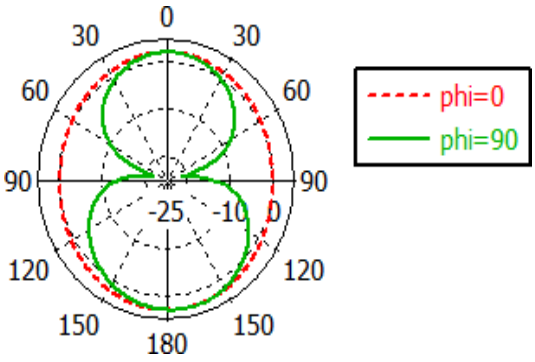


Fig. 8. Antenna radiation pattern at 3.5 GHz in E and H planes.

In Fig. 8, the radiation patterns of the proposed antenna are illustrated in the E and H planes (YZ plane with φ= 90° and XZ plane with φ= 0°, respectively) at 3.5 GHz. The antenna exhibits a unidirectional radiation pattern, which remains relatively consistent throughout the entire band.

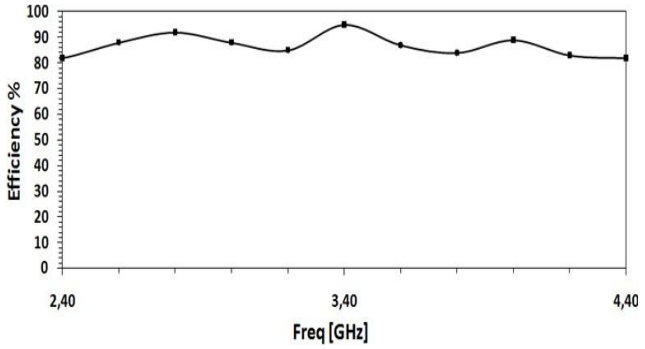


Fig. 9. Efficiency of the studied antenna versus frequency.

Fig. 9 provides compelling evidence of the exceptional efficiency of the antenna design we have put forth. Within the operational band, the efficiency of this antenna consistently surpasses the remarkable threshold of 80%. This is a noteworthy achievement, signifying the antenna's ability to effectively convert a substantial portion of the input power into radiated energy, thus enhancing its performance and applicability across a wide range of scenarios and applications. Fig. 10 shows the realized prototype of our antenna which is dedicated for many applications with an isolation necessarily exceeding 15 dB.



Fig. 10. Prototype of the proposed antenna.

The comparison between the simulation and the measured simulation results of S11 is thoughtfully displayed in Fig. 11. This comprehensive illustration underscores the integrity of our research, revealing a compelling alignment between the simulated data and the actual measurements across various frequencies within the operating band. This robust congruence between simulation results and real-world performance, particularly in the context of S11, stands as a testament to the accuracy and reliability of our antenna analysis. It signifies that our findings and the practical implementation of our antenna consistently correspond, instilling confidence in the veracity of our results and reinforcing the practical utility of the studied antenna design across a broad spectrum of frequencies within the operational range.

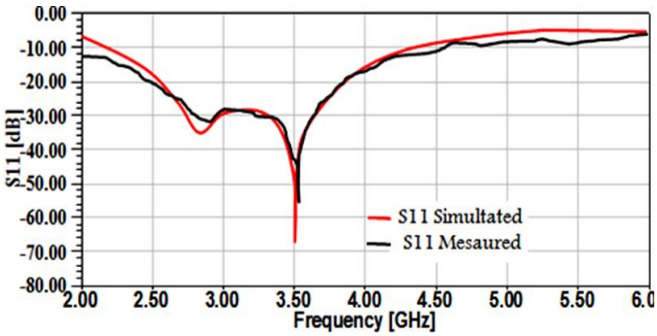


Fig. 11. Measured and simulated of S11 of the studied antenna versus frequency.

III. Study of Characteristics of the System MIMO

In this part, two techniques will be applied to two different antenna systems and each system consists of two antennas of the same configuration and size with left and right edges symmetrically. The first technique is based on the insertion of a neutralization line (NL) to reduce mutual coupling. The second technique is based on the integration of metamaterial, composed of a periodic structure consisting of three spring ring resonator (SRR) cells which are well detailed in the following section.

All measurements were conducted in an anechoic chamber, as shown in Fig. 12. The structure of the MIMO system before the application of the isolation methods is depicted in Fig. 13. The gap distance between the two-end edges of the two antennas is represented by 'e' and is optimized to 51 mm. An additional substrate of length 'a' equal to 11 mm is also introduced.

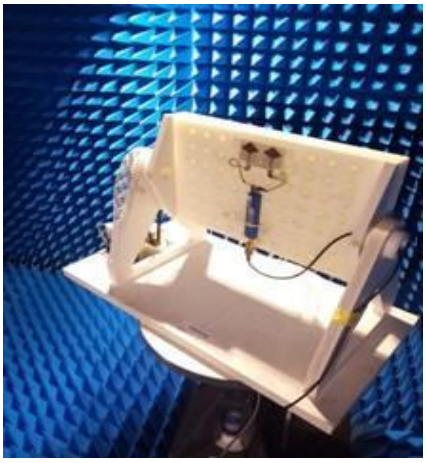


Fig. 12. Apparatus used to measure the antenna in the anechoic chamber.

The separation distance between the two edges of the antennas, represented by 'e', is set to e = λ/8 (where λ is the free space wavelength at the center frequency of 3.5 GHz). The overall size of the studied MIMO antenna system, comprising both substrates, is 100 mm × 80 mm × 1.6 mm, resulting in enhanced performance at 3.5 GHz.

The mutual coupling coefficients S21 and the simulated reflection coefficients S11 for various values of e (from 12.6 to 22.6 mm) are obtained and presented in Fig. 14 and 15, respectively. As shown in Fig. 12, we note that the impact on the S11 is relatively limited with unchanged bandwidth. From Fig. 13, we can see that the weakest coupling S21 is achieved with the smallest value of d (from 12.6 to 22.6 mm). The lowest coupling is obtained with the smallest value of e (e is 12.6 mm) and to maintain the antenna quality and the overall size of the MIMO antenna. MIMO antennas must be independent with minimal coupling to work well. So, to reduce the space between them, an efficient approach is to work on the space between them.

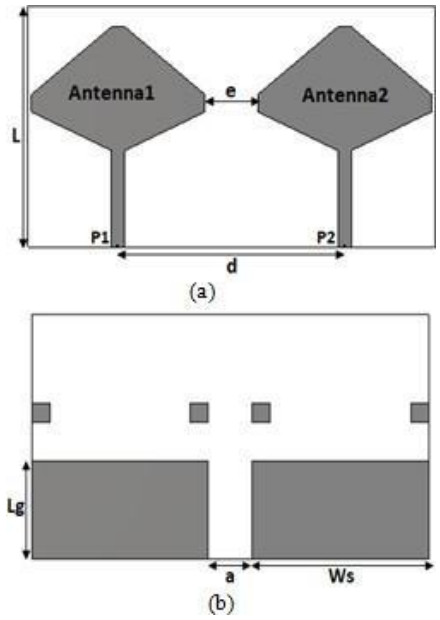


Fig. 13. Structure of the system MIMO antenna: (a) front view, (b) back view.

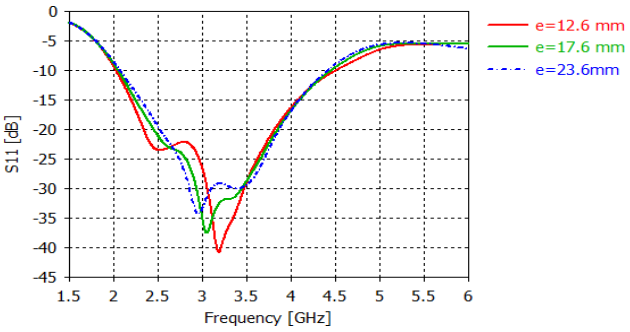


Fig. 14. Reflection Coefficient S_{11} of the suggested antenna at various e values.

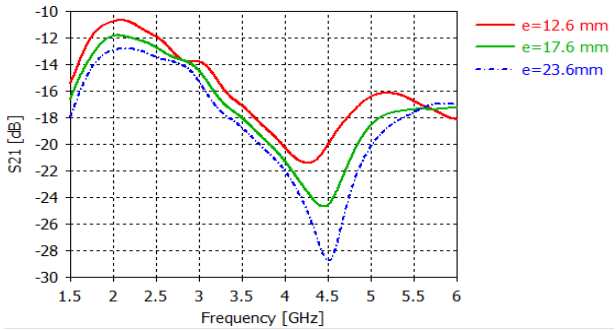
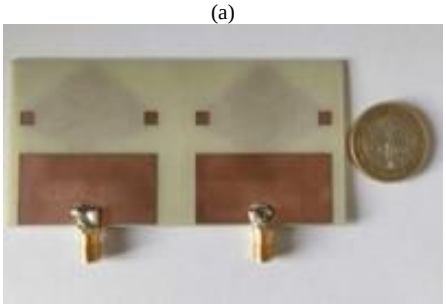


Fig. 15. Transmission coefficient S21 of the proposed antenna in different values of e.

Fig. 16 presents the realized prototype of the two symmetric antennas. The measured reflection (S11) and transmission (S21) parameters are displayed in Fig. 17 and 18, respectively. These figures provide a comparison of the reflection and transmission parameters between the measured and simulated results for the two antennas without any isolation elements. The discrepancies between the curves can be primarily attributed to the circuit construction and the antenna connectors.



(a)



(b)

Fig. 16. Photos of the prototype antenna system: (a) in front view, (b) in back view.

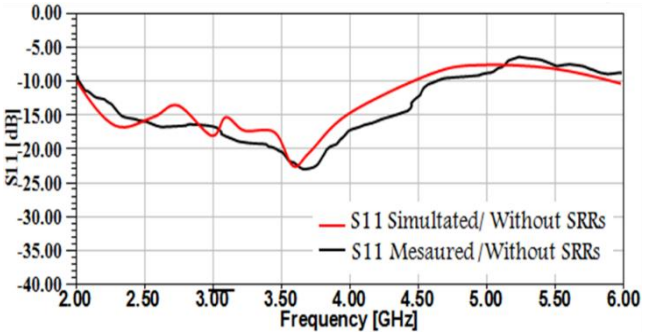


Fig. 17. Measured and simulated and S11 of the MIMO antenna.

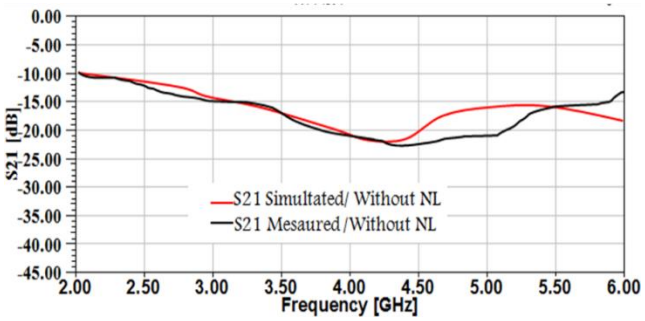


Fig. 18. Measured and simulated and S21 of the MIMO antenna.

IV. Reduction of Mutual Coupling Based on Neutralization Line and MTM

A. With NL

By positioning the antenna, you may keep some distance between them. We suggest a method to address the isolation issue, based on the placement of a neutralization line between the two antennas that were printed on the top surfaces of the FR-4 substrate (see Fig. 19).

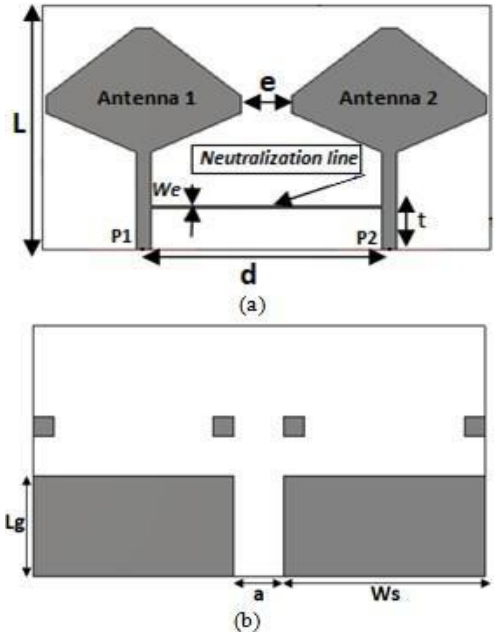


Fig. 19. Layout of a two-antenna MIMO system with NL: (a) front view, (b) bottom view.

To address the challenge of isolation while positioning the antennas with a certain degree of separation, we introduce an initial method. This method entails the insertion of a neutralization line positioned between the two antennas, printed onto the top surfaces of the FR-4 substrate, as depicted in Fig. 19. This technique is based on connecting the radiating elements to better decouple their power ports. In effect to make sure that the energy transmitted by one antenna is not wasted and radiated to the second antenna. Hence, it is important to reduce the parameter characterizing S21 which is taken as a parameter of the isolation between antenna ports. Indeed, the neutralization line was thin, short and will be taken as an inductance (see Table II).

TABLE I. PARAMETERS OF POSITION OF THE NEUTRALIZATION LINE

Parameters	Values (mm)
e	12.6
a	3.1
L	3.6
We	18.5
t	14

A parametric study on the position of the neutralization line in relation to the two supply ports is done to specify the efficiency of this line (NL). This study has been done with NL dimensions maintained at 45.5mm × 0.5mm. The reflection and transmission coefficients are illustrated in Fig. 20 and 21, respectively. In these two figures, the optimum position is t=9 mm, this value provides the high isolation at 3.5 GHz. Fig. 22 presents the transmission of radiation from antenna 1 to antenna 2 and the adaptation of antenna 1, in both situations of a MIMO system with and without NL. In simulation with a neutralization line, mutual coupling (S21) can be enhanced from -16 dB to more than -36 dB at 3.5 GHz. When the neutralization line was inserted, reflection coefficient was varied slightly. Fig. 23 shows that when antenna 1 is exited, there is less current in antenna 2 because all the current is concentrated on the neutralization line. So, the isolation in this case is enhanced.

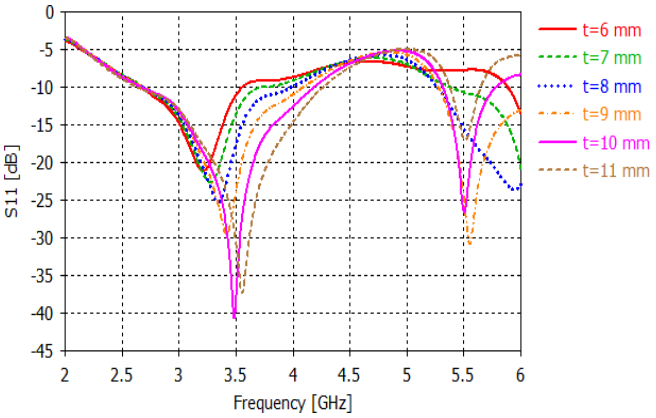


Fig. 20. Reflexion coefficient for different values of *t*.

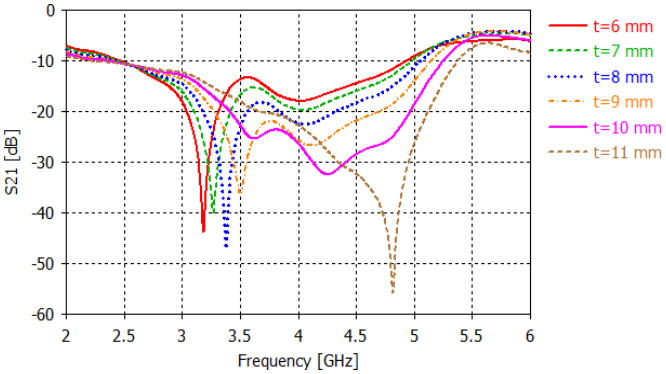


Fig. 21. Transmission coefficient for different values of *t*.

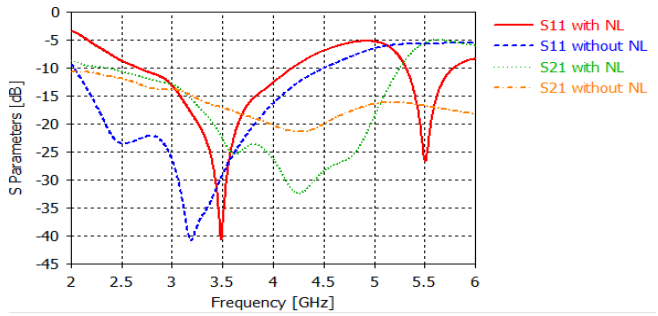


Fig. 22. Reflexion and transmission coefficients without and with NL.

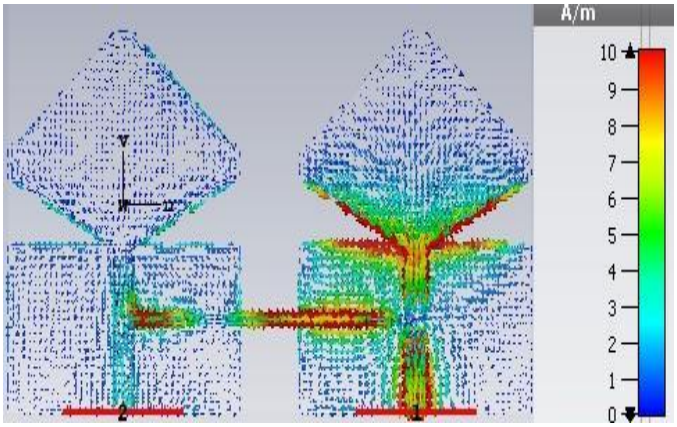


Fig. 23. Current distributions of the two antennas with NL at 3.5 GHz.

Theoretically, the NLs move some of the existing current from one antenna to the second antenna to remove the existing coupling. The envelope correlation coefficient (ECC) is an important factor to evaluate the capabilities of the MIMO/diversity antenna. ECC of the MIMO antenna system can be calculated using the radiation pattern or the S-parameters of the antenna. The envelope correlation can be expressed using the following expression:

$$ECC = \frac{|S_{11}^{*}S_{12}^{*} + S_{21}^{*}S_{22}^{*}|^2}{(1-|S_{11}|^2-|S_{21}|^2)(1-|S_{12}|^2-|S_{22}|^2)} \quad (1)$$

Another necessary parameter is its diversity gain (DG). It can be calculated with the following relation (2):

$$DG = 10\sqrt{1-(ECC)^2} \quad (2)$$

It is well that in desired band, the correlation envelope is very low (less than 0.2) as seen in Fig. 24. Also, we see that the gain is between 7 and 10 dB because the DG depends to a great part on the ECC and when ECC is minimal, DG is maximal as indicated in Fig. 25.

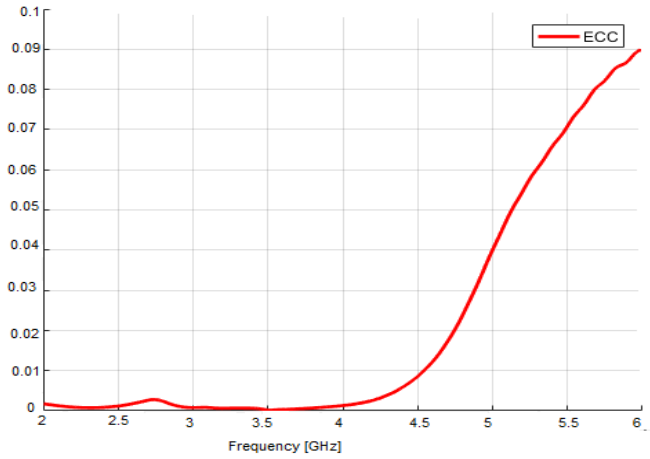


Fig. 24. ECC simulated for antenna 1 with NL.

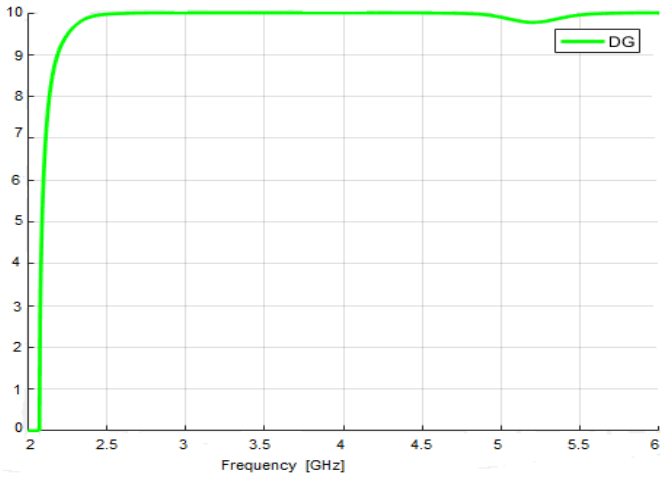


Fig. 25. DG simulated for antenna 1 with NL.

In Fig. 26, the radiation patterns of our investigated antenna are thoughtfully presented in both the horizontal (H) and vertical (E) planes. In this analysis, port 1 is excited while port 2 is loaded with 50 Ω impedance. This data showcases the antenna's remarkable unidirectional radiation pattern across both the E and H planes, consistently observed within the 3.5 GHz band. The tangible prototype of the two antennas integrated with the neutralization line (NL) is visually portrayed in Fig. 27, while Fig. 28 and 29 provide a direct comparison between the simulated and measured results of our antenna system. Notably, these figures illustrate the remarkable concordance between the two sets of data, further underscoring the efficacy of our antenna design. Furthermore, it's worth highlighting that the impedance bandwidth, defined by S11 < -10 dB, remains intact within the [3.4 - 3.8] GHz range, while S21 consistently maintains a level of less than -30 dB, affirming the antenna's outstanding performance in maintaining minimal mutual coupling.

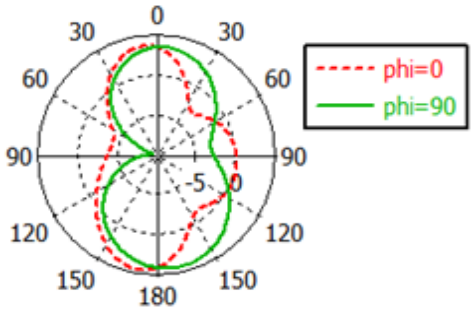


Fig. 26. Radiation patterns of antennas with NL in E and H planes at 3.5 GHz.



Fig. 27. Photos of the manufactured antenna system with NL technique.

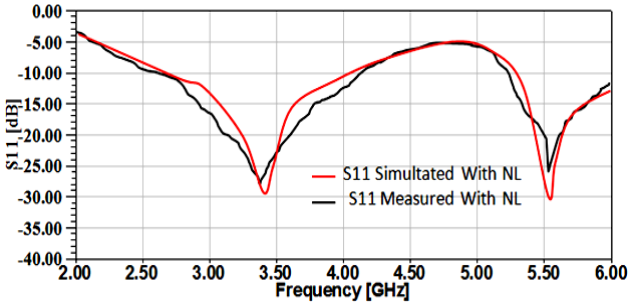


Fig. 28. Measured and simulated S11 of the MIMO antenna NL.

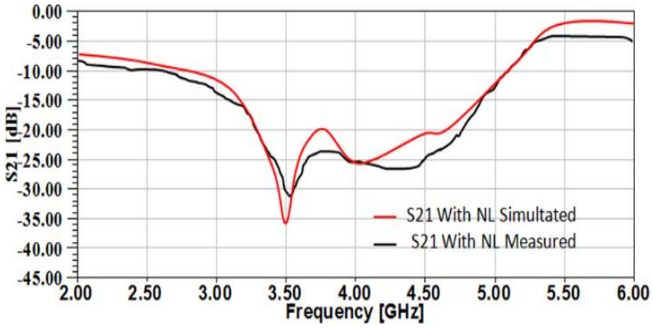


Fig. 29. Measured and simulated S21 of the MIMO antenna NL.

B. With MTM

Another technique to reduce the mutual coupling, we propose the insertion of a periodic array of SRR metamaterial composed of 3 cells. This technique will be implemented between two antenna elements with an edge-to-edge separation λ/8 (12.6 mm) at 3.5 GHz (Fig. 13). Fig. 30 presents the structure of the studied SRR, which is designed on a FR-4 dielectric substrate (thickness h = 1.6 mm, relative permittivity = 4.3).

SRRs possess the unique capability to manipulate electromagnetic fields, allowing us to create an environment that attenuates interference between antennas. By implementing SRRs, we establish an isolating structure that minimizes mutual coupling, which is essential for optimizing the performance and capacity of our MIMO system. In essence, SRRs act as a shield against interference, ensuring our antennas can function independently and efficiently, ultimately improving their overall performance.

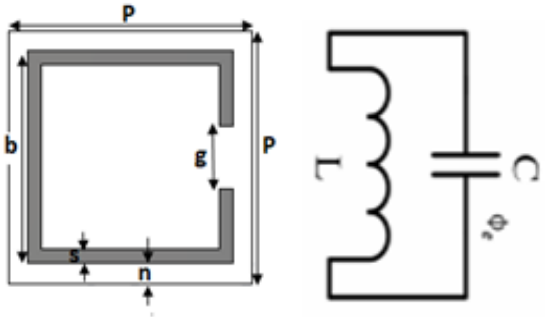


Fig. 30. Geometry of the SRR and his circuit equivalent model.

The resonance of the metamaterial can be characterized as an LC resonator and the resonant frequency is given by w=1/LC. The square-shaped SRR consists of a ring which represents the inductor effect and the gap is represented as the capacitor effect.

Fig. 31 shows the result of the simulation of the S parameters at the resonant frequency of 3.5 GHz. A reflection S11 around 0 dB with a transmission S21 tends towards -35 dB which confirms a phenomenon of stop band at 3.5 GHz of the metamaterial cell to ensure a better mutual decoupling in a cell network.

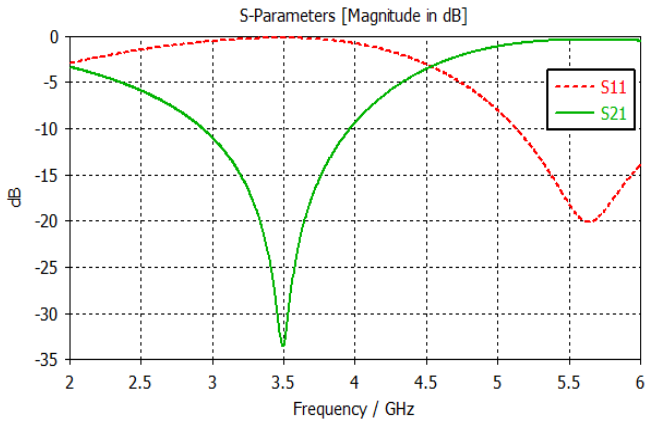


Fig. 31. Reflection and transmission coefficients of the SRR.

The square SRR under investigation has a side length P of 11 mm, with a 0.8 mm space (n) between the inner ring and the cell extremity, and a ring width (s) of 0.5 mm. When a perpendicular magnetic field is applied to the ring's plane, it begins to conduct, creating a current flow. This current enables the ring to function as an inductor, while the dielectric gap (n) generates a mutual capacitance. The resonant frequency of the SRR is determined by its dimensions, which are provided in Table III.

TABLE II. PARAMETERS OF SRR

Parameters	Values (mm)
p	12.6
b	8.9
s	0.5
n	0.8
g	3

The results of the simulation of S11 and S21 shown in Fig. 32 and 33, these results are in good accord in the whole [3.4 - 3.8] GHz band and illustrate that the cells of SRR act as a stop band (S11 = 0 dB and S21 <-10 dB).

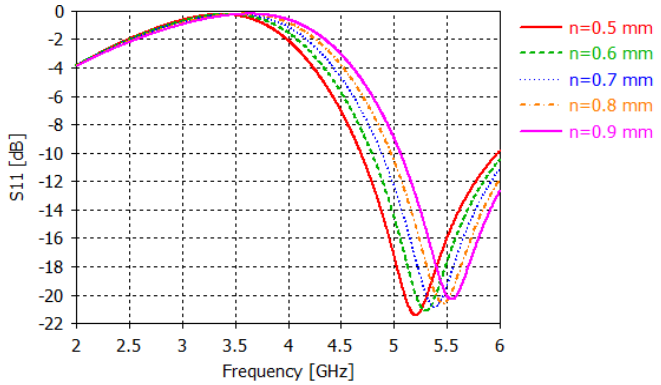


Fig. 32. S11 for a ran10ge of n-gap values.

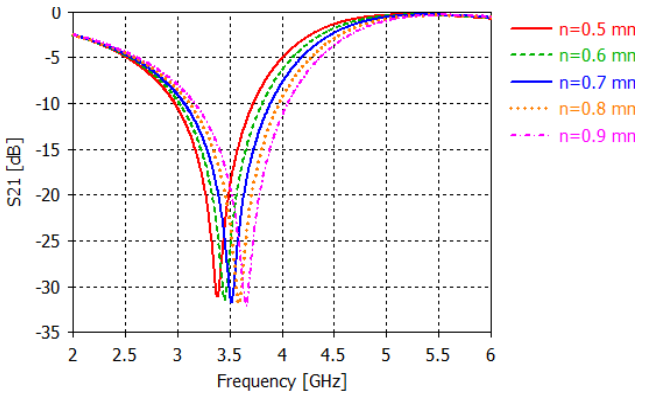


Fig. 33. S21 for a range of n-gap values.

In summary, the study demonstrates that the improvement in the n value is directly proportional to the increase in resonant frequency. To achieve the minimum value of S21, the chosen value for n is 0.8. After investigating the SRR, a new metamaterial structure consisting of a linear array of five identical SRR-type unit cells is placed between the two antennas to enhance the isolation at 3.5 GHz. The spacing between the two radiating elements' leading edges is set at 12.6 mm (Fig. 34). The dimensions and geometry of the studied SRR are provided earlier.

Fig. 35 shows that S11 remains below -10 dB throughout the 5G band, indicating good performance for the antenna both with and without the SRR. Moreover, the separation between the two antenna ports, P1 and P2, can be observed

from the S21 plot. By utilizing the SRR, the energy radiated from antenna 1 to antenna 2 is coupled at the SRR cell, resulting in improved isolation between the two antennas (S21 <-34 dB), as illustrated in Fig. 35. The parameter S21 is reduced by nearly 17 dB at 3.5 GHz, achieving good impedance matching for both antennas in the MIMO antenna system. Fig. 36 presents the current distribution of the MIMO antenna elements for both cases, with and without SRRs, when the P1 port is fed, and the P2 port is connected to a 50 Ω load. It can be observed that a portion of the current matches the SRR region, and the magnetic field is confined within the SRR region, leading to a decrease in mutual coupling between the antennas.

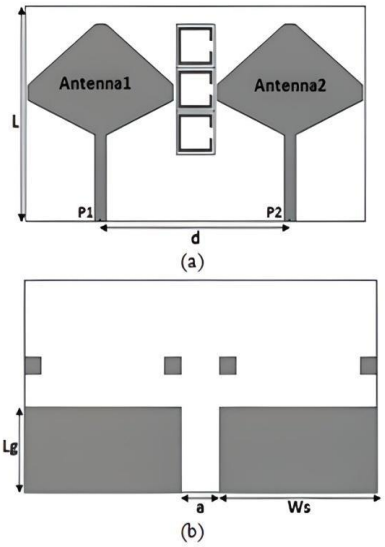


Fig. 34. Design of a two-antenna SRR MIMO system with SRRs: (a) in front view, in (b) bottom view.

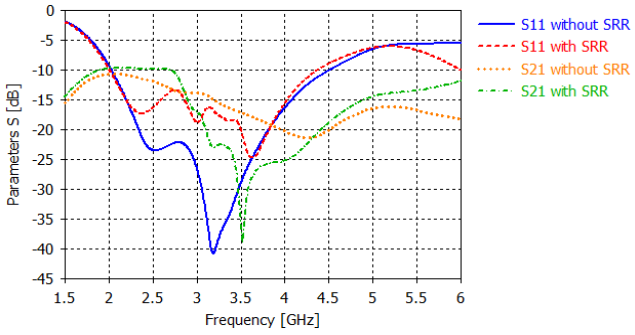


Fig. 35. Reflection and transmission coefficients without and with SRRs.

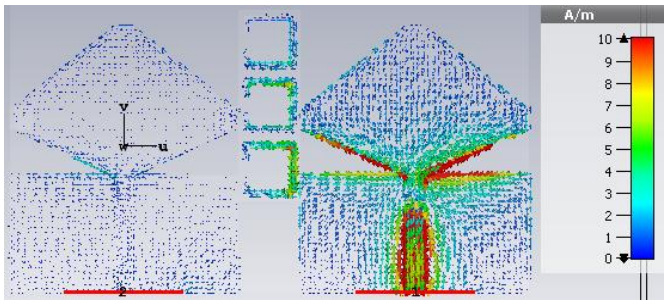


Fig. 36. Current distributions of the two antennas with SRRs at 3.5 GHz.

Fig. 37 and 38 provide crucial insights into the diversity gain (DG) of our antenna system. Notably, the diversity gain surpasses 9 dB, and its highest point aligns with the region where the correlation envelope is at its lowest. This observation underscores the significant impact of the envelope correlation coefficient (ECC) on diversity gain. A minimal correlation envelope signifies that the antennas are functioning independently and in an uncorrelated manner, which is instrumental for achieving higher diversity gain. In essence, the data in these figures emphasizes the pivotal role of ECC in influencing the diversity gain of our antenna system, highlighting the importance of ensuring low correlation between antennas for optimal performance.

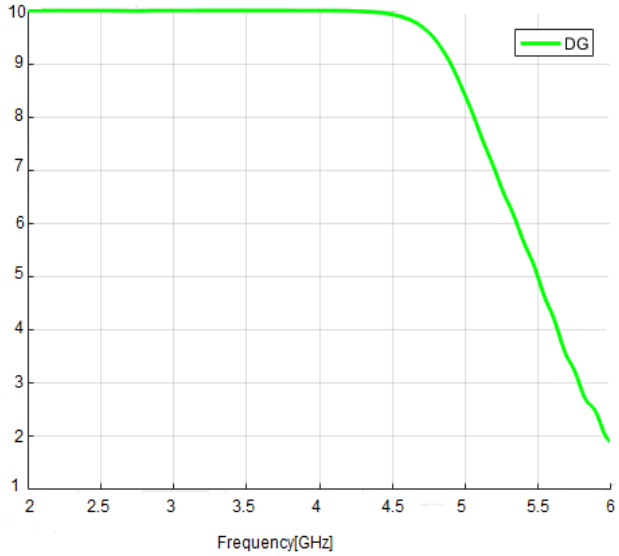


Fig. 37. DG simulated antenna 1 with SRRs.

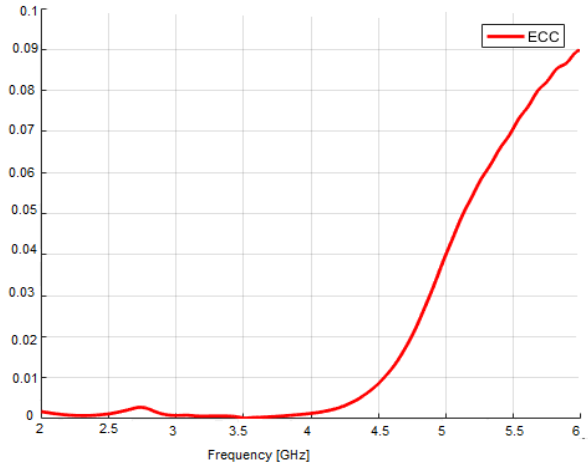


Fig. 38. ECC simulated for antenna 1 with SRRs.

Fig. 39 shows the radiation patterns of our MIMO antenna system in both cases with and without SRR. It is clearly seen that the radiation pattern is only affected by the SRR structure in the E and H planes. A prototype of the MIMO antenna described above has been manufactured and tested. Photos of the manufactured MIMO antenna are shown in Fig. 40.

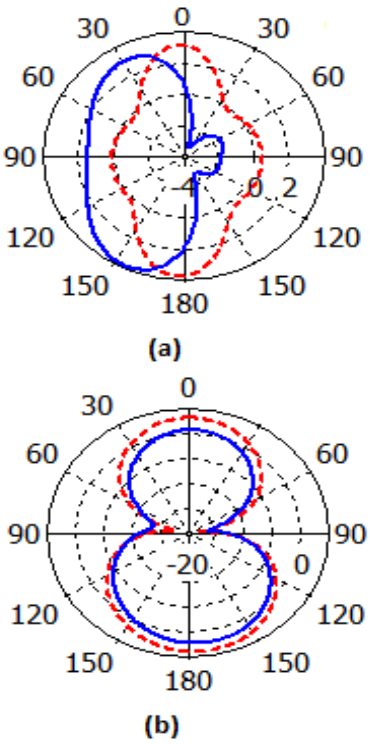
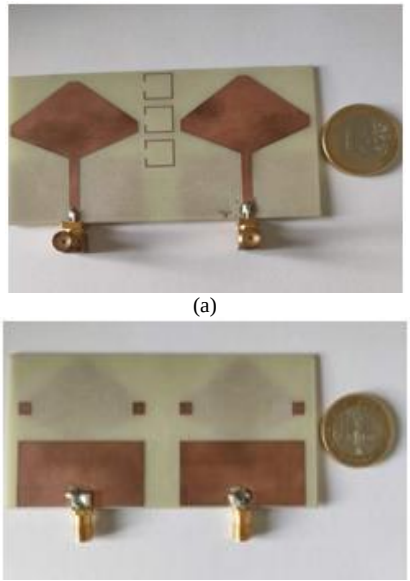


Fig. 39. Radiation patterns of antennas in both cases with and without SRR at frequency 3.5 GHz: (a) in E plane, (b) in H plane: dashed-line without SSR; solid-line with SRR.



(a)

(b)

Fig. 40. Photos of the prototype antenna system with SRR technique: (a) in front view, (b) in back view.

The depicted Fig. 41 and 42 illustrate a comparison between the measured and simulated S parameters, specifically S11 and S21, for the MIMO antenna equipped with SRRs. These visual representations highlight a notable level of agreement, essentially validating our work. It's worth noting that any slight disparities observed between the two datasets can be largely attributed to factors like losses incurred by the measuring cable. Nonetheless, the overall congruence between the measured and simulated results reinforces the effectiveness of our antenna design and the reliability of our simulation, making a strong case for the practical viability of our approach.

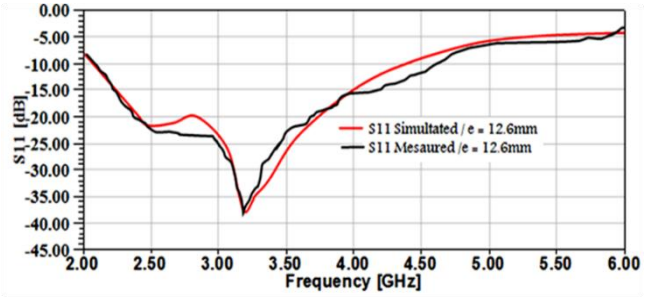


Fig. 41. Measured and simulated S11 of the MIMO antenna with SRRs.

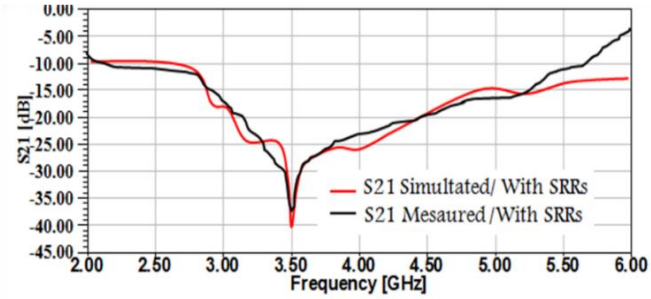


Fig. 42. Measured and simulated S21 of the MIMO antenna with SRRs.

V. Discussion and Results

The results presented in this study demonstrate the significant potential of the proposed techniques for enhancing the isolation characteristics of MIMO antennas operating in the 3.5 GHz frequency band, which is crucial for 5G and IoT applications. The use of a neutralization line (NL) and the integration of metamaterial split-ring resonators (SRRs) have shown remarkable improvements in reducing mutual coupling between adjacent antennas. The NL approach achieved an isolation of over 23 dB, while the SRR technique resulted in an isolation improvement of approximately 23 dB. These findings are particularly noteworthy when compared to existing methods in the literature. For instance, previous studies have reported lower isolation improvements, such as the 16 dB reduction achieved by a flag-shaped MIMO antenna using MTM technique and the 23 dB reduction with a circular antenna patch and MTM isolating structure. In contrast, our proposed techniques not only surpass these isolation levels but also maintain high gain and efficiency, as evidenced by the measured and simulated results. The gain of our antenna remained above 3.5 dB, and the efficiency reached above 80% in the working band with a frequency bandwidth of about 2.3 GHz. These results highlight the superior performance of our MIMO antenna design, making it a promising candidate for 5G and IoT applications. Furthermore, the practical implications of our findings are substantial. By improving the isolation characteristics of MIMO antennas, we can enhance the overall performance and reliability of 5G networks, enabling higher data rates, reduced interference, and better coverage. This is particularly beneficial for IoT applications,

where reliable and high-speed communication is essential for real-time monitoring and control of various devices and systems. Our findings provide valuable insights and practical solutions for the design of advanced MIMO antennas, contributing to the advancement of wireless communication technologies and the realization of the full potential of 5G and IoT. The effectiveness of the NL and SRR techniques is clearly demonstrated in Fig. 43 and 44, which show the reflection and transmission coefficients without and with the NL and SRRs, respectively. These figures illustrate the significant reduction in mutual coupling achieved by our proposed methods.

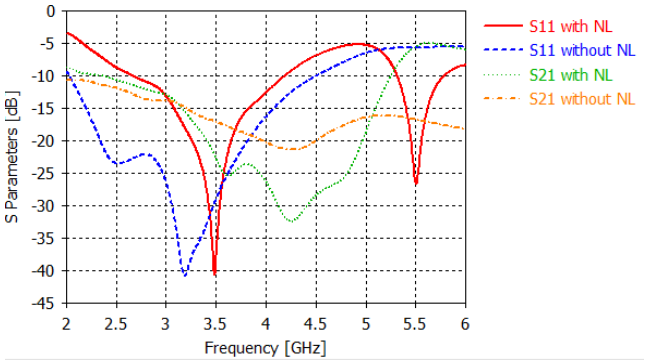


Fig. 43. Reflexion and transmission coefficients without and with NL.

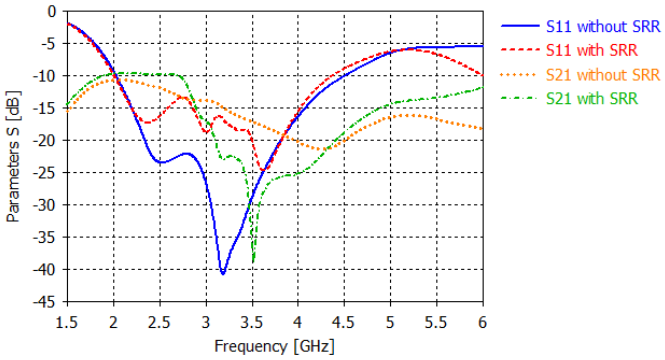


Fig. 44. Reflection and transmission coefficients without and with SRRs.

Table III presents a comparative analysis with other works in the field, highlighting the potential of our proposed techniques. This comparison underscores the advantages of our approach in terms of isolation enhancement, gain, and efficiency, demonstrating the superior performance of our MIMO antenna design for 5G and IoT applications.

TABLE III. TABLE OF COMPARISON WITH OTHER WORKS

Refs.	TECHNIQUE OF ISOLATION	EDGE-TO-EDGE GAP	MAX. ISOLATION IMPROVEMENT (dB)	GAIN (dB)	EFFICIENCY (%)
[27]	Metamaterial (MTM)	-	40	5.5	63
[28]	Metasurface (MTS)	0.26 ʎ	32	-	-
[29]	Slots	0.37 ʎ	26	4.4	-
THIS WORK	MTM/NL	0.12 ʎ	39/42	4.5	80

VI. CONCLUSION

In our paper, we have proposed a novel design of a system MIMO antenna array for 5G and IoT applications in the [3.4 - 3.8] GHz band of interest. We have described two techniques to present techniques for decreasing mutual coupling in antenna arrays applied to MIMO systems and a simulation of a MIMO antenna has been fully presented. These techniques are used to increase the isolation of our initially presented two-element system. The enhancement of the performance of our antenna system is obtained by using the neutralization line method to have isolation higher than 20 dB. On the other hand, with the method of insertion of the SRR periodic structures, the mutual coupling is minimized by about 23 dB. Simulation results show that the gain remains above 3.5 dB and the efficiency reaches above 80% in the working band with a frequency bandwidth about 2.3 GHz. Detailed parametric studies are done to find the right structure which is capable of responding to the demands of modern communication systems dedicated to 5G and IoT. Also, it is possible to apply the proposed neutralization line and the split ring resonator decoupling methods to design MIMO antennas with more than two radiation elements.

ACKNOWLEDGMENT

This research was funded by Taif University, Saudi Arabia, Project N° (TU- DSPP-2024-70).

REFERENCES

[1] Devi, Delshi Howsalya, Duraisamy, Kumutha, Armghan, Ammar, et al. 5g technology in healthcare and wearable devices: A review. Sensors, 2023, vol. 23, no 5, p. 2519.

[2] Khanh, Quy Vu, Hoai, Nam Vi, Manh, Linh Dao, et al. Wireless communication technologies for IoT in 5G: Vision, applications, and challenges. Wireless Communications and Mobile Computing, 2022, vol. 2022, no 1, p. 3229294.

[3] Huseien, Ghasan Fahim et Shah, Kwok Wei. A review on 5G technology for smart energy management and smart buildings in Singapore. Energy and AI, 2022, vol. 7, p. 100116.

[4] V. Avula, R. Nanditha, S. Dhuli, and P. Ranjan, "The Internet of Everything: A Survey," *2021 13th International Conference on Computational Intelligence and Communication Networks (CICN)*. IEEE, 2021. p. 72-79.

[5] Al-Malah, Duha Khalid Abdul-Rahman, Majeed, Ban Hassan, et Alrikabi, Haider Th Salim. Enhancement the educational technology by using 5G networks. International Journal of Emerging Technologies in Learning (Online), 2023, vol. 18, no 1, p. 137.

[6] Al-Gburi, Ahmed Jamal Abdullah, Zakaria, Zahriladha, Ibrahim, Imran Mohd, et al. Microstrip patch antenna arrays design for 5G wireless backhaul application at 3.5 GHz. In : Recent Advances in Electrical and Electronic Engineering and Computer Science: Selected articles from ICCEE 2021, Malaysia. Singapore : Springer Singapore, 2022. p. 77-88.

[7] Z. Davoody-Beni, N. Sheini-Shahvand, H. Shahinzadeh, M. Moazzami, M. Shaneh, and Gharehpetian, G. B, "Application of IoT in Smart Grid: Challenges and Solutions," In : *2019 5th Iranian Conference on Signal Processing and Intelligent Systems (ICSPIS)*. IEEE, 2019. p. 1-8.

[8] Anbarasu, Muthumanickam et Nithiyantham, Janakiraman. Performance analysis of highly efficient two-port MIMO antenna for 5G wearable applications. IETE Journal of Research, 2023, vol. 69, no 6, p. 3594-3603.

[9] Rafique Umair Khan, Suleman, Ahmed, Muhammad Mansoor, et al. Uni-planar MIMO antenna for sub-6 GHz 5G mobile phone applications. Applied Sciences, 2022, vol. 12, no 8, p. 3746.

[10] Surender, D., Khan, T., Talukdar, F. A., De, A., Antar, Y. M., and Freundorfer, A. P., "Key components of rectenna system: a comprehensive survey," *IETE Journal of Research*, 2020, p. 1-27.

[11] Lee, Young Seung, Jeon, Sang Bong, Pack, Jeong-Ki, et al. An In Vitro Experimental System for 5G 3.5 GHz Exposures. IEEE Access, 2022, vol. 10, p. 94832-94840.

[12] Z. Ren, A. Zhao, and S. Wu, "MIMO antenna with compact decoupled antenna pairs for 5G mobile terminals," *IEEE Antennas and Wireless Propagation Letters*, 2019, vol. 18, no 7, p. 1367-1371.

[13] Du, K., Wang, Y., & Hu, Y. (2022). Design and analysis on decoupling techniques for MIMO wireless systems in 5G applications. Applied Sciences, 12(8), 3816.

[14] Yang, C., Lu, K., & Leung, K. W. (2022). Dielectric decoupler for compact MIMO antenna systems. IEEE Transactions on Antennas and Propagation, 70(8), 6444-6454.

[15] Nadeem, I., and Choi, D. Y., "Study on mutual coupling reduction technique for MIMO antennas." *IEEE Access*, 2018, vol. 7, p. 563-586.

[16] Jiang, W., Liu, B., Cui, Y., and Hu, W., "High-isolation eight-element MIMO array for 5G smartphone applications," *IEEE Access*, 2019, vol. 7, p. 34104-34112.

[17] Zhou, W., Yue, C., and Li, Y., "Metamaterial Promoting 5G MIMO Antenna with Isolation Enhancement." In : *2021 International Applied Computational Electromagnetics Society (ACES-China) Symposium*. IEEE, 2021. p. 1-2.

[18] Daghari, M., Abdelhamid, C., Sakli, H., and Nafkha, K., "High Isolation with Neutralization Technique for 5G-MIMO Elliptical Multi-Antennas." In : *International conference on the Sciences of Electronics, Technologies of Information and Telecommunications*. Springer, Cham, 2018. p. 124-133.

[19] He, D., Yu, Y., and Mao, S., "Characteristic Mode Analysis of a MIMO Antenna with DGS." *2021 IEEE International Symposium on Antennas and Propagation and USNC-URSI Radio Science Meeting (APS/URSI)*. IEEE, 2021. p. 1143-1144.

[20] Alibakhshikenari, M., Vittori, M., Colangeli, S., Virdee, B. S., Andújar, A., Anguera, J., and Limiti, E. "EM isolation enhancement based on metamaterial concept in antenna array system to support full-duplex application." *2017 IEEE Asia Pacific Microwave Conference (APMC)*. IEEE, 2017. p. 740-742.

[21] Wan, L., Guo, Z., and Chen, X. (2019). "Enabling efficient 5G NR and 4G LTE coexistence." *IEEE Wireless Communications*, 2019, vol. 26, no 1, p. 6-8.

[22] Vaughan, R. G., and Andersen, J. B. "Antenna diversity in mobile communications." *IEEE Transactions on vehicular technology*, 1987, vol. 36, no 4, p. 149-172.

[23] Chen, J. H., Ye, L. H., Liu, T., & Wu, D. L. (2023). A low-profile dual-polarized patch antenna with simple feed and multiple decoupling techniques. IEEE Antennas and Wireless Propagation Letters, 22(8), 1883-1887.

[24] Alibakhshikenari, M., Virdee, B. S., See, C. H., Abd-Alhameed, R. A., Falcone, F., and Limiti, E. "Surface wave reduction in antenna arrays using metasurface inclusion for MIMO and SAR systems." *Radio Science*, 2019, vol. 54, no 11, p. 1067-1075.

[25] Megahed, A. A., Abdelazim, M., Abdelhay, E. H., and Soliman, H. Y. "Sub-6 GHz highly isolated wideband MIMO antenna arrays." *IEEE Access*, 2022, vol. 10, p. 19875-19889.

[26] Abd-Alhameed, R. A., and Limiti, E. "Mutual-coupling isolation using embedded metamaterial EM bandgap decoupling slab for densely packed array antennas." IEEE Access, 2019, vol. 7, p. 51827-51840.

[27] Supreeyatitikul, Nathapat, Phungasem, Anupan, et Aelmopas, Pongphol. Design of wideband sub-6 GHz 5G MIMO antenna with isolation enhancement using an MTM-inspired resonators. In : 2021 Joint International Conference on Digital Arts, Media and Technology with ECTI Northern Section Conference on Electrical, Electronics, Computer and Telecommunication Engineering. IEEE, 2021. p. 206-209.

[28] Sultan, Kamel S., Abdullah, Haythem H., Abdallah, Esmat A., et al. Metasurface-based dual polarized MIMO antenna for 5G smartphones using CMA. Ieee Access, 2020, vol. 8, p. 37250-37264.

[29] Sufian, Md Abu, Hussain, Niamat, Askari, Hussain, et al. Isolation enhancement of a metasurface-based MIMO antenna using slots and shorting pins. IEEE Access, 2021, vol. 9, p. 73533-73543.

Deep Learning Fusion for Intracranial Hemorrhage Classification in Brain CT Imaging

Padma Priya S. Babu[1], Dr. T. Brindha[2]
Department of Computer Science and Engineering, Noorul Islam Centre for Higher Education, Tamil Nadu, India[1]
Department of Information Technology, Noorul Islam Centre for Higher Education, Tamil Nadu, India[2]

***Abstract*—Brain hemorrhages are characterized by the rupture in the arteries of brain due blood clotting or high blood pressure (BP), presents a significant risk of traumatic injury or even death. This bleeding results in the damage in brain cells, with common causes including brain tumors, aneurysm, blood vessel abnormalities, amyloid angiopathy, trauma, high BP, and bleeding disorders. When a hemorrhage happens, oxygen can no longer reach the brain tissues and brain cells begin to die if they are depleted of oxygen and nutrients for longer than three or four minutes. The affected nerve cells and the related functions they control are damaged as well. Early detection of brain hemorrhages is crucial. In this paper an efficient hybrid deep learning (DL) model is proposed for the intracranial hemorrhage detection (ICH) from brain CT images. The proposed method integrates DenseNet 121 and Long Short-Term Memory (LSTM) models for the accurate classification of ICH. The DenseNet 121 model act as the feature extraction model. The experimental results demonstrated that the model attained 97.50% accuracy, 97.00% precision, 95.99% recall and 96.33% F1 score, demonstrating its effectiveness in accurately identifying and classifying ICH.**

Keywords*—Intracranial hemorrhage; deep learning; DenseNet 121; LSTM; brain CT images

I. INTRODUCTION

Hemorrhage describes the occurrence of bleeding either internally or externally from the body. A sudden blood clot in arteries can cause brain hemorrhage, which can lead to symptoms such as tingling, palsy, weakness, and numbness. Recognizing these symptoms is crucial for initiating immediate treatment. Brain hemorrhages occur when a blood vessel in the brain leaks or bursts, results in hemorrhagic stroke. ICH is a life-threatening neurological condition that can occur due to various causes, such as increased BP, hemorrhage secondary to infarct, trauma, tumor hemorrhage, and more [1]. One of the most common causes of ICH is traumatic brain injury. When blood pools within the brain parenchyma, it forms a hematoma, which increases pressure on the surrounding brain tissues. This pressure leads to reduced blood flow and ultimately kills brain cells.

There are various forms of hemorrhages, each occurring in different areas of the skull. Epidural hemorrhage (EPD) occurs when there is damage to both the skull and the dura mater, leading to bleeding. An accumulation of blood within the brain's tissues causes an intraparenchymal hemorrhage (ITP), also, the bleeding in the brain's ventricular system is known as intraventricular hemorrhage (ITV). Subdural hemorrhage refers to a collection of blood within the subdural spaces, which are potential spaces between the dura and arachnoid of the meninges surrounding the brain. Subarachnoid hemorrhage is an extra-axial intracranial hemorrhage located within the subarachnoid spaces [2].

Diagnosing a brain hemorrhage is challenging because some individuals do not exhibit any physical signs. Computed tomography (CT) is the prime method used for the diagnosis of ICH. During a CT scan, a set of images is generated using X-ray beams, capturing the various intensities of brain cells from their X-ray absorbency levels [3]. The regions of ICH are depicted as hyperdense areas without a defined architecture. Radiologists analyze these scanned images and confirm the presence, type and location of ICH. However, the accuracy of the diagnosis depends on the accessibility and experience of the radiologist, which can lead to ineffective and impressive results [4].

In recent years, artificial intelligence (AI), particularly DL, has significantly transformed image analysis, becoming an essential tool in medical diagnostics. This study employs a new hybrid DL approach to detect and classify ICH from brain CT images. This approach combines the strengths of Advanced Neural Network (ANN) such as DenseNet 121 and LSTM to increase the accuracy and reliability of hemorrhage detection. The performances of the model are assessed using the metrics of precision, accuracy, recall, and F1 score, demonstrating its effectiveness in identifying brain hemorrhages. The proposed work offers some key contributions as follows.

- To develop a hybrid DL-based computer aided diagnosis system for early detection and classification of brain hemorrhage using brain CT images.
- To reduce the error rate of brain hemorrhage detection system.
- To train the Neural Network using more images to avoid over fitting problem.
- To improve the performance evaluation parameters precision, accuracy, F1 score and recall of the system.
- To develop a new model based on deep learning for effective segmentation from brain CT images.

The next sections of the study as follows: The current approaches are discussed in Section II. The methodology proposed is illustrated in Section III. The results are presented in Section IV. Section V concludes the paper.

II. Literature Review

The efficiency of a DL model for the identification of ICH and its subtypes on non-contrast CT scans was evaluated by Yeo et al. (2023) [5]. The algorithm was tested and trained on an open-source retrospective data. The training dataset was obtained from four different countries and test data were obtained from India. The performance of the convolutional neural network (CNN) was compared to that of other models that were similar. The performances of the model were evaluated using the area under curve characteristics (AUC-ROC) and micro-averaged precision (mAP) score. The performance of mAP increased from 0.77 to 0.93, and AUC-ROC increased from 0.854 to 0.966. The study highlighted its limitations, that the performance of the model was not tested on images that were subjected to different CT image reconstruction methods. Additionally, the datasets used in the study contained class imbalances.

Rajagopal et al. (2023) [6] provided an ICH classification with six separate types of hemorrhages in circumstances where patients experienced several hemorrhages continuously. The different types of ICH present were detected and classified using the CT scan of patient's skull. This study presents a hybrid approach combining CNN and LSTM for enhanced performance. The RSNA dataset was utilized to assess the model's performance. The model achieved 93.87% sensitivity, 96.45% specificity, 95.21% precision and 95.14% accuracy.

Luis et al. (2023) [7] proposed an Efficient DL method for the diagnosis of hemorrhages in patients. The technique classified the slices of CT scans for hemorrhage. The model was evaluated to check whether the patient was positive and achieved 0.978 ROC AUC and 92.7% accuracy. The study proves that the framework can be used as an assistant for CT-based diagnosis.

Tharek et al. (2022) [8] created an algorithm for identifying ICH in head CT scan. This algorithm module was developed by CNN using deep learning. In this study 200 data were collected from a public dataset. The algorithm was trained using Jupyter Notebook platform. A confusion matrix was used to evaluate the model's performance which results in 96.94% sensitivity, 93.14% specificity, 93.14% precision and 95.00% of accuracy with 95% F1 score. The study proves that DL using CNN created an accurate classifier.

An AI-based method for ICH on non-contrast CT images was presented by Seyam et al. (2022) [9]. The entire ICH detection attained 93% diagnostic accuracy, 97.8% negative predictive value and 87.2% sensitivity. The study highlighted its limitation such as, the proposed work did not contain complete metrics data for all patients.

Ganeshkumar et al. (2022) [10] proposed an unsupervised DL framework for CT image-based ICH identification. Principal Component Analysis (PCA-Net) was utilized by the model to extract features from CT scans. The models were tested and trained using 752 and 1750 CT slices. The proposed framework achieved 67% accuracy, 80% weighted average precision and 67% weighted average recall additionally with an F1 score of 72% indicates that the method act as an unsupervised framework for identifying ICH from CT images.

Ganeshkumar et al. (2022) [11] studied the segmentation and identification of ICH regions in CT images and proposed a one-stop model. The framework used ResNet for identification and an adversarial network SegAN for segmentation. Therefore, the approach achieved an F1 score of 0.91 on a macro average, 0.80 on sensitivity, and 0.99 on specificity for ICH identification. Additionally, the model received a dice score of 0.32 for the ICH segmentation. Thus, the segmentation and identification method helped doctors in making accurate diagnoses. The study noted limitations, such as the method not being capable of classifying the identified ICH and the dataset used being of small size.

A DL-based ICH diagnosis model called GC-SDL was created by Anupama et al. (2022) [12] utilizing GrabCut-based segmentation with synergic deep learning (SDL). The framework used a Gabor filter for removing the noise. The segmentation method was applied to identify the portions affected. The feature extraction process was performed using SDL and softmax was applied as a classifier. The results showed that the model achieved 97.78% specificity, 95.73% accuracy, 94.01% sensitivity, and 95.79% precision.

Wu et al. (2021) [13] introduced an ensemble deep neural network (DNN) for the identification and categorization of ICH in their study. There are two parallel network pathways in the method. Both pathways used the EfficientNet-B0 as their core architecture, which was then assembled to produce predictions. The models were trained and evaluated using the ICH detection dataset comprising of 674,259 head noncontract CT images from 19,531 patients. In addition, the generalizability of the trained model was tested using another dataset, CQ500. The performance of ICH detection dataset resulted in 95.7% accuracy, 85.9% sensitivity and 86.7% F1 score. Similarly, it resulted in 92.4% accuracy, 93.4% F1 score and 92.6% sensitivity when utilizing the CQ500 dataset. A limitation of the study highlighted that the performance of the framework was not compared with the radiologist diagnostic performance.

In order to accomplish accurate ICH identification, Wang et al. (2021) [14] developed a DL technique that replicates the radiologist's interpretation process by merging a 2D CNN model with a two-sequence model. The 2019-RSNA Brain CT Hemorrhage Challenge dataset, which included over 25,000 CT images that correctly identified ICH, was used to create the technique. Using 491 and 75 CT images, respectively, the system was tested on two separate external validation sets, maintaining 0.949 AUC and 0.964 ICH detection. The results demonstrate the suggested model's excellent performance and capacity for generalization. The study found that the model's training required greater time and complexity.

To save the time needed to identify hemorrhages, Rohit (2021) [15] looked at the ICH detection problem in his study and created a DL model and a transfer learning (TL) model. To ensure the accuracy of the model, DenseNet 121, CNN, and Xception were compared using a variety of evaluation criteria. The accuracy of 91% was achieved in the ICH detection and classification using the suggested CNN and TL

models. The Xception model was capable of identifying the ICH with lower risk of failure.

Using Inception Network for effective image segmentation, Mansour and Aljehane (2021) [16] created a DL-based ICH model (DL-ICH). The model proposed involves several processing steps. A multilayer perceptron (MLP) was utilized for classification, while an inception v4 network was employed for feature extraction. The model's output is compared with a series of simulations showed that its accuracy, precision, and sensitivity were 95.06%, 97.56%, 95.25%, and 93.56%, respectively.

In their study Bhadauria et al. (2021) [17] presented a segmentation method for the extraction of hemorrhage from brain CT images using fuzzy clustering features. Fuzzy clustering was utilized to estimate the parameters that controls the propagation of the level set function. The model utilized a dataset consisting of 300 CT images with various shapes of hemorrhages and sizes. The performance of the proposed model resulted in 85.40% accuracy, 98.79% specificity and 79.91% sensitivity. The study noted a limitation in that using a denoising algorithm on the system, resulted in increased complexity and longer processing times.

A densely connected CNN (DenseNet) with extreme machine learning (EML) was developed by Santhoshkumar et al. (2021) [18] for the purpose of diagnosing and classifying ICH. The model uses DenseNet for feature extraction and Tsallis entropy in conjunction with a grasshopper optimization approach for image segmentation. The results of the simulation verified that the model has reached its highest accuracy level of 96.34%.

Lee et al. (2020) [19] in their study introduced a DL algorithm for artificial neural network (ANN). This study evaluated the feasibility of using the algorithm for detecting and classifying ICH without employing CNNs. The model achieved 0.859 AUC, 78% sensitivity and 80% specificity for the detection of ICH. For the localization of ICH, the CT images attained 0.903 AUC, 82.5% sensitivity and 84.1% specificity. The accuracy rate is 91.7% for the classification of ICH. The study reduced the ICH diagnosis time. The limitation suggest that the size of the dataset used was small.

There are several notable gaps in the current methods for analyzing brain CT images. The magnification factor affects brain CT images, leading to misclassification and reduced performance. Image segmentation cannot be used in certain works, showing the difficulty in determining the correct location of internal bleedings from CT scans. Another disadvantage is the dependence on a manual reference provided by a single human expert. Hemorrhages and microaneurysms have same characteristics and are only distinguishable by their color and size on color fundus images, making them easily misunderstood. False positive rates are high for classification and segmentation. Lastly, existing classification methods suffer from lower accuracy and sensitivity indicating a clear need for improvement in these areas.

III. Materials and Methods

The block diagram depicted in Fig. 1, illustrates the process of ICH classification using a hybrid deep learning approach from brain CT images. The model takes CT images from the dataset as input. The images undergo data preprocessing and augmentation. The DenseNet 121 model uses the preprocessed images as input to extract features. The extracted features are reshaped into sequential patterns and given as an input to the LSTM. The model grasps the temporal connections and sequential dependencies within the data. The LSTM output is directed to a dense layer to group the learned features followed by a softmax layer for final classification. The classification outputs determine whether the CT image indicates a normal or hemorrhagic condition, thus ensuring accurate diagnosis.

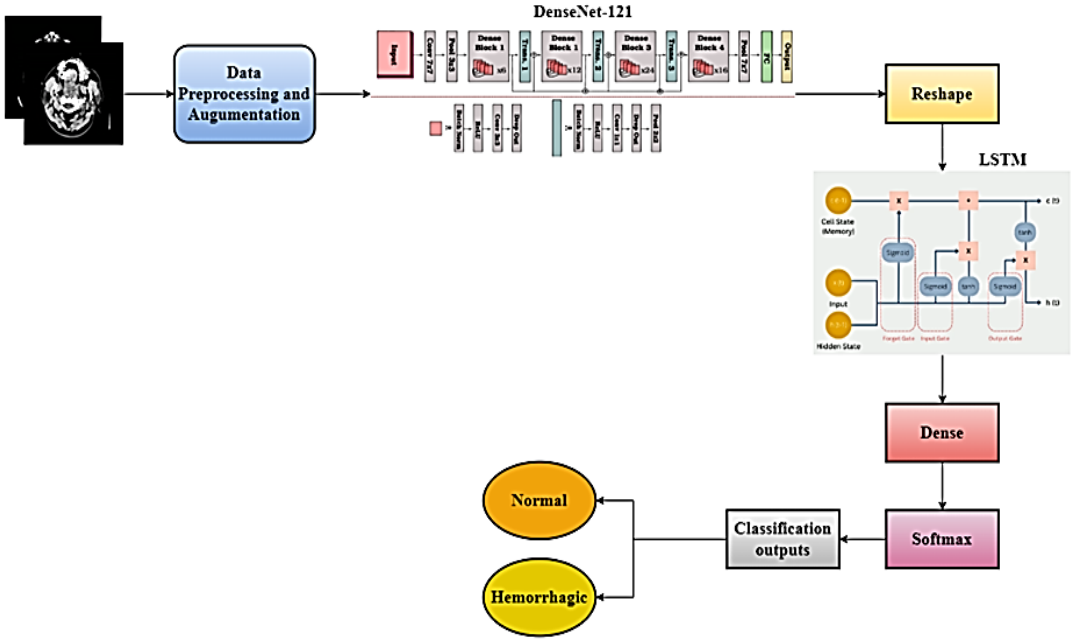


Fig. 1. Block diagram of proposed model.

A. Dataset Description

The dataset required for the classifying ICH from CT images is obtained from the Kaggle repository [20]. It contains images of normal and hemorrhagic CT scans collected from the Near East Hospital, Cyprus. It is having a total of 2600 images. The dataset consists of two folders each for testing and training. Again, each folder is further divided into two groups as "normal" and "hemorrhagic". A total of 1600 images were used for training, 600 images for testing, and 400 images for validation. Fig. 2 depicts the sample images from the dataset.

B. Data Preprocessing and Augmentation

In the classification of ICH from brain CT images, preprocessing and augmentation plays an essential role in increasing the performance of the DL models. These processes standardize the input data through techniques such as resizing, normalization, and contrast adjustment to standardize the input data, ensures that the suggested model can effectively learn and generalize from the images. In this study the 'rescale' parameter normalize the pixel values to a range between 0 and 1. Data augmentation enhances the training dataset's variability by applying random transformations like shearing, zooming, horizontal and vertical flips, and random rotations up to 30 degrees. These augmentations are applied to the input images during the training process, enhancing the diversity of the dataset and promoting better generalization of the deep learning model improving its ability to classify normal and hemorrhagic CT scans accurately. To separate the dataset into training sets and testing sets, an 80:20 ratio is employed.

C. Proposed Methodology

1) DenseNet 121: DenseNet is a CNN architecture where two layers are interconnected to all other layers deeper in a network. This is designed to allow the greatest flow of data among network layers. Firstly, it adopts a dense block structure, ensuring each layer connects to every other layer in a feedforward manner. Secondly, it incorporates bottleneck layers, which effectively decrease the parameter count while preserving the number of learned features within the network. As of Fig. 3 the DenseNet 121 consist of 121 layers and is characterized by its three main building blocks such as transition layers, dense blocks, and global average pooling layer [21].



(a) Hemorrhagic CT Images

(b) Normal CT Images

Fig. 2. Sample images from the dataset.

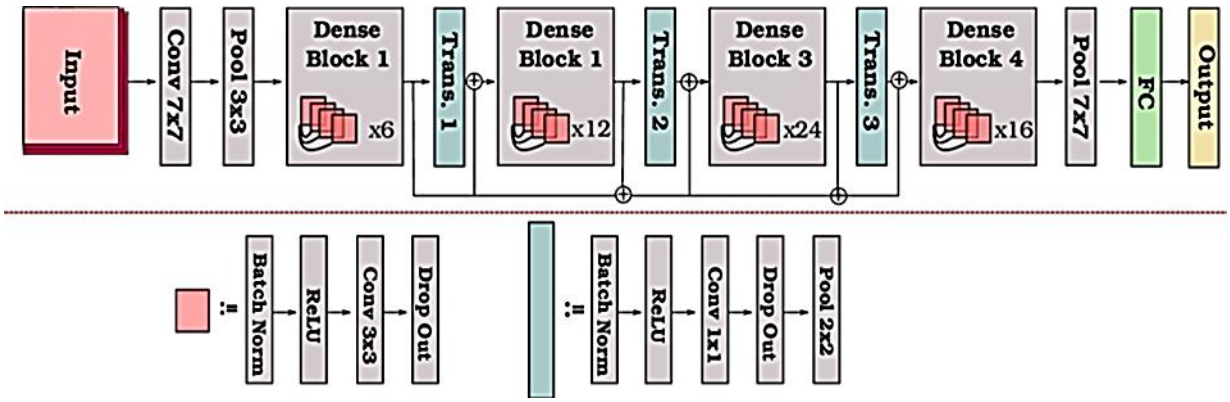


Fig. 3. Basic architecture of DenseNet 121.

In this architecture, the layers are connected through dense blocks, allowing each layer to utilize input from all preceding layers. This enables the creation of feature maps that propagates data to all subsequent layers. Each Dense blocks consist of several layers, composed of batch normalization, ReLU activation function and convolution. Let's denote the output of the t^{th} layer as m_t where all previous feature maps are received as inputs. It is expressed as:

$$m_t = H_t([m_1, m_2, \dots\dots\dots\dots\dots\dots\dots\dots m_{t-1}]) \quad (1)$$

Where, $[m_1, m_2, \dots\dots\dots\dots\dots\dots\dots\dots m_{t-1}]$ represents the concatenation of feature maps from the t^{th} layer and H_t

represents the t^{th} layer, it's a sequence of three successive operations forming a composite function.

DenseNet 121 incorporates four dense blocks, with transition layers scattered between each block. These transition layers facilitate down-sampling of the feature maps by implementing a 1 x 1 convolution followed by a 2 x 2 average pooling layer. If the input to a transition layer is m, the output is given as,

$$m_{trans} = AvgPool(Conv(m, \theta, 1 \times 1) \quad (2)$$

Here, θ represents the reduction factor utilized to decrease the feature maps size.

The dense block comprises multiple convolutional layers interconnected in series, facilitating cross-layer connections between distant layers. DenseNet 121 utilizes the ReLU activation function to enhance the non-linearity of the frame work. The ReLU function defined as follows,

$$ReLU(m) = \max(0, m) \quad (3)$$

The final layer contains a fully connected layer followed by a softmax function, which is utilized for predicting the probability of the CT image class. The softmax function specified as follows:

$$sm(v)_i = \frac{e^{v_i}}{\sum_{j=1}^{D} e^{v_j}} \; for \; i = 1 \; to \; D \quad (4)$$

Here, $v = (v_1, \ldots\ldots\ldots\ldots, v_D) \in \mathbb{R}^D$. The input vector, v, undergoes exponential computation for each value of v_i. The output vector's sum $sm(v)$ is equal to 1.

2) Long short term memory (LSTM): LSTM is used to solve the vanishing gradient problem. The model integrates memory cells capable of storing and retrieving information from long sequences. The architecture consists of three gates shown in Fig. 4, namely input gate, output gate and forget gate. The information that needs to be deleted and that needs to be kept from the previous stages and current input are determined by the forget gate. These values are then passed into a sigmoid function, that provide output values between 0 and 1, shows that if all previous information is lost, then the value is 0, and if all previous information is preserved, then the value is 1. The input gate determines the significance of the current input necessary for solving the task. Finally at the output gate, the model computes its output based on the hidden state. To perform this, sigmoid function determines which information should be allowed to pass through the output gate. After the decision, the cell state is activated with the tanh function and then undergoes multiplication [22].

The LSTM input at a particular time stamp *t*, is said to be (x_t) the data, (c_{t-1}) the cell state and $(h_{t-1}) \cdot c_{t-1}$ is the hidden state where h_{t-1} is the output in the previous time stamp. The three gates of the LSTM cell to control the flow of information are initialized as input gate (i_t), output gate (o_t) *and* forget gate (f_t). All gates utilize h_{t-1} and x_t as the inputs, employing the sigmoid function as activation function. The information stored at c_{t-1} is removed by the forget gate.

The forget gate is expressed as,

$$f_t = sigm(W_{fx}x_t + W_{fh}h_{t-1} + b) \quad (5)$$

For each current time stamp the learned representation is denoted by $\widetilde{c_t}$. Subsequently a point wise multiplication is carried out to preserve essential information. Computing both the cell state (c_t) and input gate (i_t), the equations are termed to be.

$$\widetilde{c_t} = tanh(W_{gx}x_t + W_{ih}h_{t-1} + b) \quad (6)$$

$$i_t = sigm(W_{ix}x_t + W_{ih}h_{t-1} + b) \quad (7)$$

$$c_t = f_t \odot c_{t-1} + i_t \odot \quad (8)$$

At the current time stamp, the LSTM cell output is denoted by the hidden state (h_t), which is utilized for making decisions. The hidden state determines by eliminating non-essential information that does not contribute to the decision at time stamp *t*. This can be achieved by utilizing the output gate (o_t). The expression for obtaining o_t and h_t are

$$o_t = sigm(W_{ox}x_t + W_{oh}h_{t-1} + b) \quad (9)$$

$$h_t = o_t \odot c_t \quad (10)$$

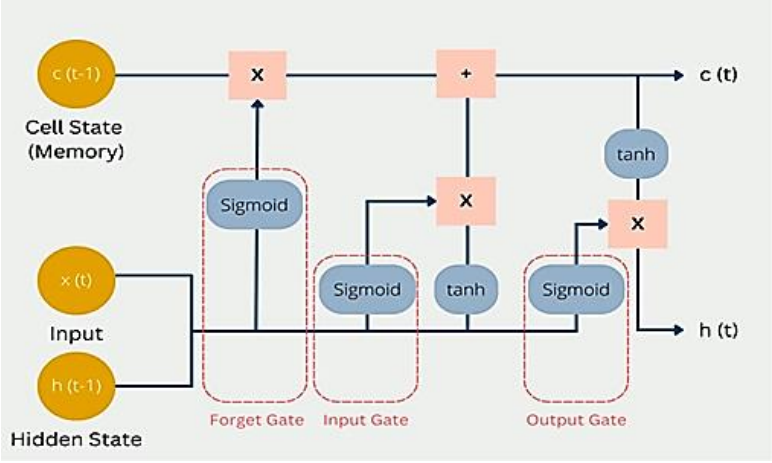


Fig. 4. Basic architecture of LSTM.

The proposed hybrid model combines the two-network architecture DenseNet 121 and LSTM. This combination utilizes the advantages of each architecture in capturing both spatial and temporal information's. The DenseNet 121 model serve as a feature extractor, capturing detailed spatial features from the CT images. These spatial features represent information about the structure and content of the images. In this case the LSTM model is employed to obtain temporal connections within the data, which could be important for analyzing how features change over time or sequence within the images. After extracting the spatial and temporal features, the model uses a dense layer with a softmax function as the output layer. The softmax function converts the output of the previous layers into probability scores corresponding to each class. The model performs binary classification to the brain CT images and further classified as "Normal" or "Hemorrhagic". Finally, the model's performance is assessed using a range of metrics such accuracy, F1 score, recall, and precision. Fig. 5 illustrates the architecture of the proposed work, while Table I provides a summary of the proposed model.

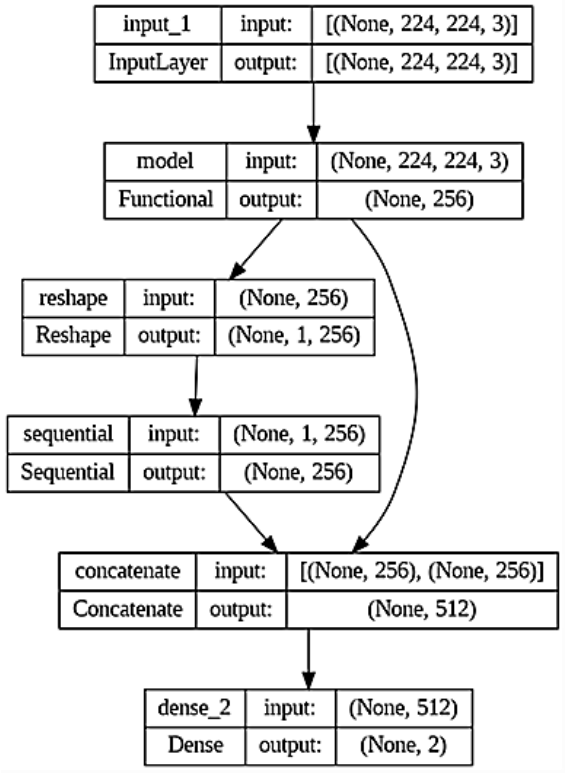


Fig. 5. Proposed model architecture.

TABLE I. SUMMARY OF THE PROPOSED MODEL.

Total Parameters	7531074
Trainable Parameters	7447426
Non-Trainable Parameters	83648

The proposed model algorithm is outlined below.

Algorithm

Input: Brain CT image dataset, labels determine Hemorrhage or Normal

Output: Predictions of whether the input image contains hemorrhage or not

Begin:

Load and preprocess data:

1. Collect dataset: $C= \{(A_i, b_i)$, where A_i is a brain CT image and $b_i \in \{0,1\}$ b_i $i \in \{0,1\}$ (1: Normal, 0: Hemorrhage).
2. Preprocess:
   - Resize: $A_i \rightarrow A_i' \in R^{224\times224}$
   - Normalize: $A_i' \rightarrow \frac{A_i'-\mu}{\sigma}$
   - Data Augmentation: $A_i' \rightarrow \{A_i''\}$(Shear, Zoom, Flipp (horizontal and vertical), Rotation)

Define Base Models:

1. Load Model: DenseNet 121
2. Input: $224 \times 224 \times 3$
   Global Average Pooling 2D ()
   Dense (256, activation='relu')
   LSTM (128)
   Dense (256, activation='relu')
   Concatenate ()
   Dense (2, activation='softmax')

Model Compilation and Training:

1. Compile each model M:
   Optimizer=Adam ()

Loss=binary _crossentropy
Metrics=[accuracy]

2. Train: M.fit (X_{train} , y_{train} ,validation_data= (X_{val}, y_{val}))

Model Evaluation and Comparison:

1. Evaluate:
   metrics=M.evaluate(X_{test} , y_{test}), where metrics include accuracy, precision, recall.

Save the Model:

End

IV. Results and Discussion

A. Software and Hardware Setup

The model development and training were carried out using Google Colaboratory, employing Python and the Keras framework throughout the entire process. The Colab notebooks were equipped with TensorFlow, a GPU, 12.75 GB of RAM, 68.50 GB of disk space, and a 64-bit version of Windows 10. The efficiency of the model proposed was examined using its predictions on the test dataset. The hyperparameters of DNN are determined through empirical methods and significantly influence the learning process, as outlined in Table II. To identify the model that delivers the best classification performance, a wide range of variables are tested and evaluated.

B. Experimental Results

Accuracy and loss plots are utilized to visualize the effectiveness of a machine learning (ML) model throughout its training and validation phases. In the case of ICH classification from brain CT images, these plots offer valuable information on how well the model can differentiate between hemorrhagic and normal images. The accuracy plots illustrate the fluctuations in model accuracy across epochs during both training and validation. However, the loss plot shows the alterations in the loss function throughout the training and validation phases.

Fig. 6 and Fig. 7 show the accuracy plot and loss plot of the model. At the initial epochs the accuracy of the model is low suggest that the model is struggling to learn the patterns. However, as epoch progresses there is an improvement in the accuracy indicates the model is making correct predictions. This means that the model is correctly identifying images that contain signs of hemorrhage and those that are not. In some cases, there is a slight decrease in accuracy can occur due to fluctuations in training process or adjustments made to the learning rate. Similarly, the loss of the system is high at the initial epoch. As the epoch progresses the loss decreases steadily and towards the final epoch the loss become smaller indicates that the model is improving in its ability to differentiate between hemorrhagic and normal images. It is observed that when the learning rate decreases after a certain epoch it can affect the convergence of the model and lead to fluctuations in loss values.

TABLE II. Hyperparameter Specification

Hyperparameters	Values
Batch Size	64
Loss Function	Binary Crossentropy
Optimizer	Adam
Activation function	Softmax
Number of epochs	20

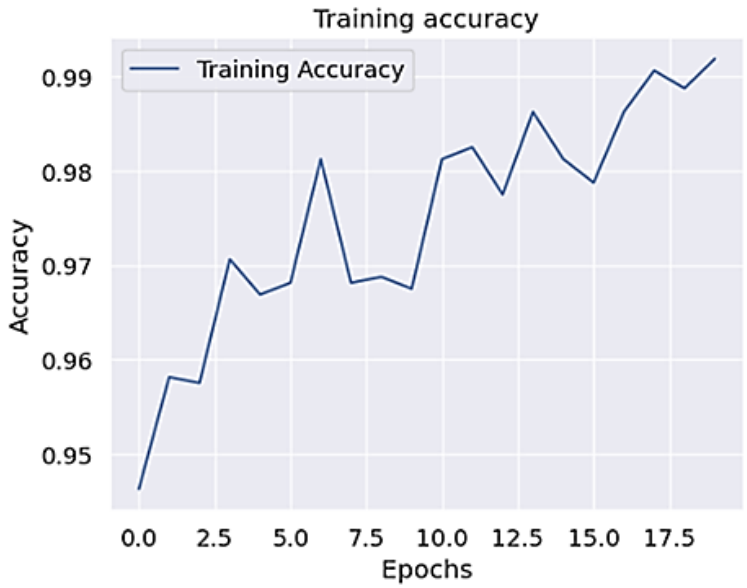


Fig. 6. Accuracy plot of proposed model.



Fig. 7. Loss plot of proposed model.

Evaluation metrics are essential tools used in measuring the performance of ML models, providing details about accuracy, reliability and generalization ability. Some of the common evaluation metrics used in the classification tasks.

$$Accuracy = \frac{T_p+T_n}{T_p+T_n+F_p+F_n} \quad (11)$$

$$Recall = \frac{T_p}{T_p+F_n} \quad (12)$$

$$Precision = \frac{T_p}{T_p+F_p} \quad (13)$$

$$F1\ Score = 2 * \frac{(Recall*Precision)}{(Recall+Precision} \quad (14)$$

Where T_p is True Positive, T_n is True Negative, F_n is False Negative and F_p is False Positive.

The performance of the model is impressive with an accuracy of 97.50%, indicating overall correctness in its predictions. The precision signifies a low false positive rate at 97% highlighting the potential of the model to accurately determine positive instances. Additionally, 95.99% of recall suggest that the model accurately captures a significant portion of actual positive instances. The balanced F1-score of 96.33%, reflects a harmonious combination of recall and precision, reinforcing the model's robustness in handling both false positives and false negatives. Table III shows the proposed model's classification report.

TABLE III. CLASSIFICATION REPORT OF PROPOSED MODEL

Performance Metrics	Obtained Results
Accuracy	97.50 %
Precision	97.00 %
Recall	95.99 %
F1-Score	96.33 %

The efficacy of a classification algorithm is evaluated using a confusion matrix. Fig. 8 visualizes the classification of the ICH from brain CT images. It helps the model to correctly identify hemorrhagic and normal images. It analyzes the types of errors made such as missed hemorrhages and refine the model to improve the diagnostic accuracy. Ensuring reliable and effective medical diagnosis. In this confusion matrix, 295 images were correctly predicted as Hemorrhage, while 277 images were correctly predicted as Normal. Additionally, 25 images were misclassified as Hemorrhage, and three images were misclassified as Normal.

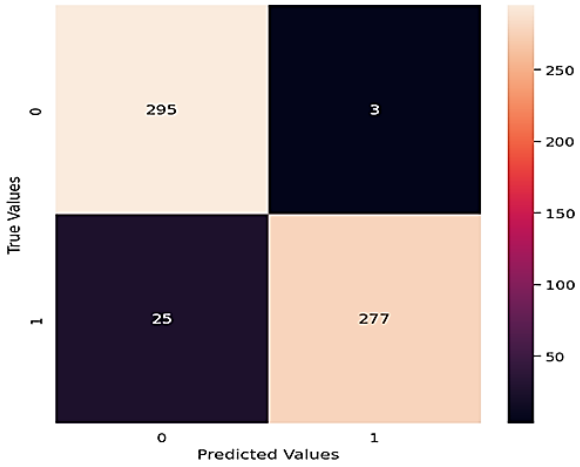


Fig. 8. Classification report of proposed model.

A randomly selected image from the dataset is classified using the suggested model, correctly classifying it as either "Normal" or "Hemorrhage.". As seen in Fig. 9, this successful classification highlights the efficiency of models and reliability in precisely identifying and classifying images within the dataset. Table IV presents a comparison of the accuracy between the proposed and current techniques.

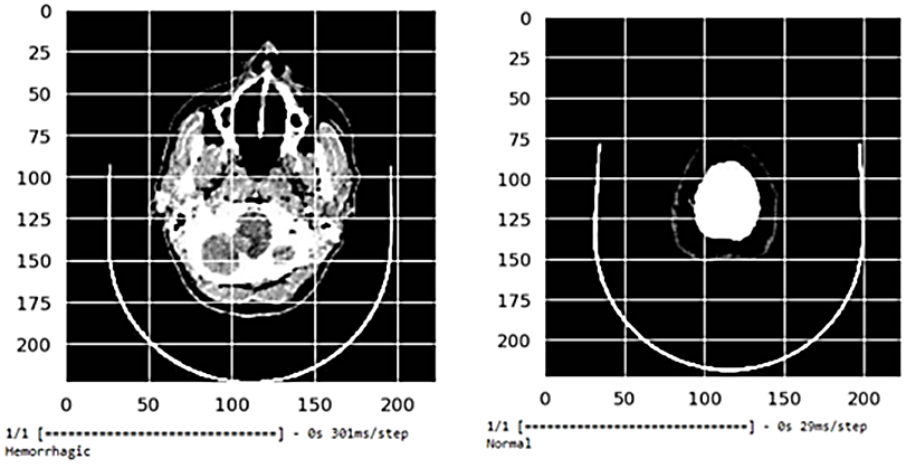


Fig. 9. Classification outputs.

TABLE IV. COMPARISON OF PROPOSED METHOD AND CURRENT APPROACHES

Sl. No:	Author	Methodology	Accuracy	Precision	Recall	F1 Score
1.	Rajagopal et al. [2]	CNN and LSTM	95.14	95.21	93.87	-
2.	Tharek et al [4]	CNN	95.00	93.14	92.94	95.00
3.	Anupama et al. [8]	Deep Learning	95.73	95.79	94.01	-
4.	Mansour and Aljehane [12]	Inception V4	95.06	95.25	93.56	-
5.	Venugopal et al. [23]	DL	96.56	96.43	95.65	-
6.	Qui et al. [24]	U-Net	94.1	93.5	95	-
7.	**Proposed Hybrid DenseNet 121- LSTM**		**97.50**	**97.00**	**95.99**	**96.33**

C. Discussion

Fig. 10 compares the accuracy of various models used for intracranial hemorrhage classification from brain CT images. The proposed hybrid model, which combines DenseNet121 and LSTM, achieves the highest accuracy at 97.50%. In comparison, a hybrid approach combining CNN and LSTM shows an accuracy of 95.14%, while a model using only CNN attains 95.00%. A deep learning model reaches an accuracy of 95.73%, and one using the Inception V4 architecture achieves 95.06%. Another deep learning model (DL) records an accuracy of 96.56%, and a model utilizing U-Net achieves 94.1%.

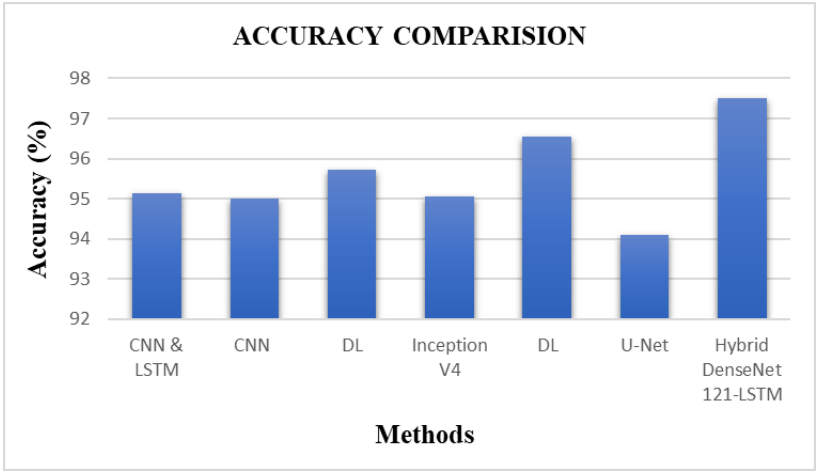


Fig. 10. Accuracy analysis of DenseNet 121-LSTM method with existing approaches.

Fig. 11 compares the precision of different models used for ICH classification from brain CT images. The proposed hybrid model, combining DenseNet121 and LSTM, achieves the highest precision at 97.00%, indicating its superior ability to accurately identify positive instances with minimal false positives. In comparison, the hybrid CNN and LSTM approach shows a precision of 95.21%, while a model using only CNN records a precision of 93.14%. A deep learning model achieves a precision of 95.79%, and the Inception V4 architecture reaches 95.25%. Another deep learning model (DL) shows a precision of 96.43%, and a U-Net model achieves 93.5%.

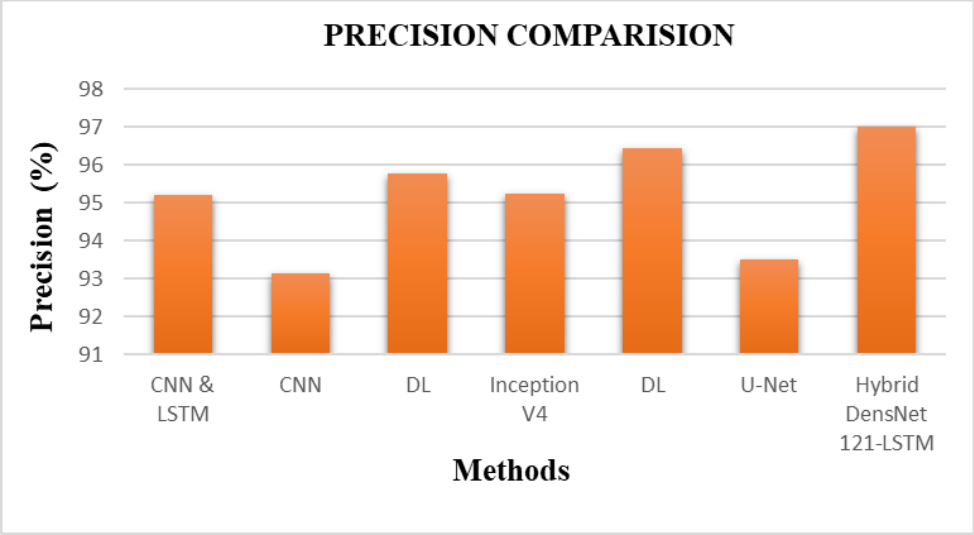


Fig. 11. Precision analysis of DenseNet 121-LSTM method with existing approaches.

Fig. 12 compares the recall of various models used for intracranial hemorrhage classification from brain CT images. The proposed hybrid model, combining DenseNet121 and LSTM, achieves a recall of 95.99%, indicating its effectiveness in correctly identifying a high proportion of actual positive cases. In comparison, the hybrid CNN and LSTM approach shows a recall of 93.87%, while a model using only CNN achieves 92.94%. A deep learning model records a recall of 94.01%, and the Inception V4 architecture reaches 93.56%. Another deep learning model (DL) shows a recall of 95.65%, and a U-Net model achieves recall at 95%.

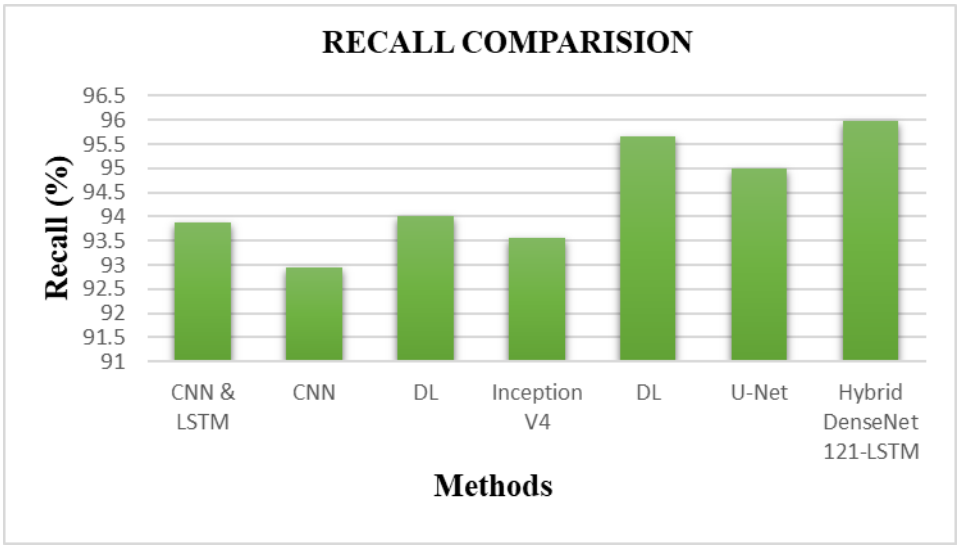


Fig. 12. Recall analysis of DenseNet 121-LSTM method with existing approaches.

V. Conclusion

ICH refers to acute bleeding inside the brain or skull. It is a serious condition that can results in severe disability or death. It is occurred due to a variety of factors such as trauma, vascular disease or congenital development. As a result of the severe compression caused by excessive bleeding, oxygen-rich blood is unable to flow to the brain tissue. AI-driven disease prediction can be used to identify individuals who are at higher risk of developing a certain disease, which can help inform decisions about preventative care and early intervention. This method uses a new hybrid DL based model for the classification of ICH from brain CT images. The model

utilized DenseNet 121 and LSTM for accurate classification of ICH. The model demonstrates better performance exhibiting 97.50% accuracy, 97% precision, 95.99% recall and 96.33% F1 score highlighting its effectiveness and potential to improve diagnostic accuracy and patient outcomes in the detection and classification of ICH. A real time ICH detection and classification framework will be implemented in future.

ACKNOWLEDGMENT

I would like to express my sincere gratitude to all those who contributed to the completion of this research paper. I extend my heartfelt thanks to my supervisor, my family, my colleagues and fellow researchers for their encouragement and understanding during the demanding phases of this work.

REFERENCES

[1] Rocha, E., Rouanet, C., Reges, D., Gagliardi, V., Singhal, A. B., & Silva, G. S. (2020). Intracerebral hemorrhage: update and future directions. *Arquivos de Neuro-Psiquiatria*, *78*, 651-659.

[2] Nizarudeen, S., & Shunmugavel, G. R. (2023). Multi-Layer ResNet-DenseNet architecture in consort with the XgBoost classifier for intracranial hemorrhage (ICH) subtype detection and classification. *Journal of Intelligent & Fuzzy Systems*, *44*(2), 2351-2366.

[3] Rava, R. A., Seymour, S. E., LaQue, M. E., Peterson, B. A., Snyder, K. V., Mokin, M., ... & Ionita, C. N. (2021). Assessment of an artificial intelligence algorithm for detection of intracranial hemorrhage. *World Neurosurgery*, *150*, e209-e217.

[4] Rao, B., Zohrabian, V., Cedeno, P., Saha, A., Pahade, J., & Davis, M. A. (2021). Utility of artificial intelligence tool as a prospective radiology peer reviewer—detection of unreported intracranial hemorrhage. *Academic radiology*, *28*(1), 85-93.

[5] Yeo, M., Tahayori, B., Kok, H. K., Maingard, J., Kutaiba, N., Russell, J., ... & Asadi, H. (2023). Evaluation of techniques to improve a deep learning algorithm for the automatic detection of intracranial haemorrhage on CT head imaging. European Radiology Experimental, 7(1), 17.

[6] Rajagopal, M., Buradagunta, S., Almeshari, M., Alzamil, Y., Ramalingam, R., & Ravi, V. (2023). An efficient framework to detect intracranial hemorrhage using hybrid deep neural networks. *Brain Sciences*, *13*(3), 400.

[7] Cortés-Ferre, L., Gutiérrez-Naranjo, M. A., Egea-Guerrero, J. J., Pérez-Sánchez, S., & Balcerzyk, M. (2023). Deep learning applied to intracranial hemorrhage detection. *Journal of Imaging*, *9*(2), 37.

[8] Tharek, A., Muda, A. S., Hudi, A. B., & Hudin, A. B. (2022). Intracranial hemorrhage detection in CT scan using deep learning. *Asian Journal Of Medical Technology*, *2*(1), 1-18.

[9] Seyam, M., Weikert, T., Sauter, A., Brehm, A., Psychogios, M. N., & Blackham, K. A. (2022). Utilization of artificial intelligence–based intracranial hemorrhage detection on emergent noncontrast CT images in clinical workflow. *Radiology: Artificial Intelligence*, *4*(2), e210168.

[10] Ganeshkumar, M., Sowmya, V., Gopalakrishnan, E. A., & Soman, K. P. (2022). Unsupervised deep learning approach for the identification of intracranial haemorrhage in CT images using PCA-Net and K-Means algorithm. In *Intelligent vision in healthcare* (pp. 23-31). Singapore: Springer Nature Singapore.

[11] Ganeshkumar, M., Ravi, V., Sowmya, V., Gopalakrishnan, E. A., Soman, K. P., & Chakraborty, C. (2022). Identification of intracranial haemorrhage (ICH) using ResNet with data augmentation using CycleGAN and ICH segmentation using SegAN. *Multimedia Tools and Applications*, *81*(25), 36257-36273.

[12] Anupama, C. S. S., Sivaram, M., Lydia, E. L., Gupta, D., & Shankar, K. (2022). Synergic deep learning model–based automated detection and classification of brain intracranial hemorrhage images in wearable networks. *Personal and Ubiquitous Computing*, *26*(1), 1-10.

[13] Wu, Y., Supanich, M. P., & Jie, D. (2021). Ensembled deep neural network for intracranial hemorrhage detection and subtype classification on noncontrast CT images. *Journal of Artificial Intelligence for Medical Sciences*, *2*(1-2), 12-20.

[14] Wang, X., Shen, T., Yang, S., Lan, J., Xu, Y., Wang, M., ... & Han, X. (2021). A deep learning algorithm for automatic detection and classification of acute intracranial hemorrhages in head CT scans. *NeuroImage: Clinical*, *32*, 102785.

[15] Kumar, R. (2021). *Intracranial Hemorrhage Detection Using Deep Learning and Transfer Learning* (Doctoral dissertation, Dublin, National College of Ireland).

[16] Mansour, R. F., & Aljehane, N. O. (2021). An optimal segmentation with deep learning-based inception network model for intracranial hemorrhage diagnosis. *Neural Computing and Applications*, *33*(20), 13831-13843.

[17] Bhadauria, N. S., Kumar, I., Bhadauria, H. S., & Patel, R. B. (2021). Hemorrhage detection using edge-based contour with fuzzy clustering from brain computed tomography images. *International Journal of System Assurance Engineering and Management*, *12*(6), 1296-1307.

[18] Santhoshkumar, S., Varadarajan, V., Gavaskar, S., Amalraj, J. J., & Sumathi, A. (2021). Machine learning model for intracranial hemorrhage diagnosis and classification. *Electronics*, *10*(21), 2574.

[19] Lee, J. Y., Kim, J. S., Kim, T. Y., & Kim, Y. S. (2020). Detection and classification of intracranial haemorrhage on CT images using a novel deep-learning algorithm. *Scientific reports*, *10*(1), 20546.

[20] *brain-ct-hemorrhage-AMINE-dataset*. (2022, April 27). Kaggle. https://www.kaggle.com/datasets/mahjoubimohamedamine/braincthemorrhageaminedataset

[21] Albelwi, S. A. (2022). Deep architecture based on DenseNet-121 model for weather image recognition. *International Journal of Advanced Computer Science and Applications*, *13*(10).

[22] Jyotishi, D., & Dandapat, S. (2023). An Attention Based Hierarchical LSTM Architecture for ECG Biometric System. *Authorea Preprints*.

[23] Venugopal, D., Jayasankar, T., Sikkandar, M. Y., Waly, M. I., Pustokhina, I. V., Pustokhin, D. A., & Shankar, K. (2021). A Novel Deep Neural Network for Intracranial Haemorrhage Detection and Classification. *Computers, Materials & Continua*, *68*(3).

[24] Qiu, Y., Chang, C. S., Yan, J. L., Ko, L., & Chang, T. S. (2019, October). Semantic segmentation of intracranial hemorrhages in head CT scans. In *2019 IEEE 10th International Conference on Software Engineering and Service Science (ICSESS)* (pp. 112-115). IEEE.

Romanian Sign Language and Mime-Gesture Recognition

Enachi Andrei[1], Turcu Cornel[2], George Culea[3], Sghera Bogdan Constantin[4], Ungureanu Andrei Gabriel[5]
Faculty of Electrical Engineering and Computer Science, Ștefan cel Mare University of Suceava, Suceava, Romania[1, 2, 4, 5]
Department of Energetics and Computer Science, Vasile Alecsandri University of Bacău, Bacău, Romania[1, 3, 4, 5]

***Abstract*—This paper presents a comprehensive approach to Romanian Sign Language (RSL) recognition using machine learning techniques. The primary focus is on developing and evaluating a robust model capable of accurately classifying hand and mime gestures representative of RSL and converting it into speech through an application. Utilizing a dataset of hand landmarks captured and stored in CSV format, the study outlines the preprocessing steps, model training, and performance evaluation. Key components of the methodology include data preparation, model training, performance evaluation and model optimization. The results demonstrate the feasibility of using machine learning for RSL recognition, achieving promising accuracy rates. The study concludes with a discussion on potential applications and future enhancements, including real-time gesture recognition and expanding the dataset for improved generalization. This work contributes to the broader effort of making sign language more accessible through technology, particularly for the Romanian-speaking deaf and hard-of-hearing community.**

***Keywords*—RSL; sign language; machine learning; model; mime gestures**

I. Introduction

Romanian Sign Language (RSL) serves as a vital means of communication for the deaf and hard-of-hearing community in Romania. Despite its significance, the accessibility and recognition of sign language pose substantial social and technological challenges. Recent advancements in machine learning and gesture recognition offer promising opportunities for developing innovative solutions to enhance interaction and integration for individuals who rely on sign language. This paper focuses on the development and evaluation of a machine learning model for recognizing hand and mime gestures specific to RSL [1-6]. By training a model on hand landmark data captured and stored in CSV format, the goal is to create a system capable of accurately and efficiently recognizing gestures used in RSL communication. The process involves several essential stages: data collection and preprocessing, model training and optimization, performance evaluation, and model conversion for deployment on mobile and embedded devices. Each of these stages is detailed, highlighting the methodologies and technologies employed to ensure high accuracy and efficient model implementation. In the data collection phase, a diverse dataset of RSL hand gestures (30 gestures) was compiled, ensuring representation across various gestures to improve the model's robustness. Preprocessing steps, such as normalization and augmentation, were applied to enhance data quality and model generalization. During the model training phase, different neural network architectures were explored, and hyperparameter tuning was conducted to optimize the model's performance. The evaluation phase included extensive testing using a confusion matrix to identify areas of improvement and validate the model's accuracy. Finally, the trained model was converted to TensorFlow Lite format, enabling its use in resource-constrained environments such as mobile and embedded devices. This conversion is crucial for practical applications, allowing the model to be deployed in real-world scenarios where computational resources are limited. Through this research, we aim to improve e and pave the way for practical applications that support the communication and integration of deaf and hard-of-hearing individuals. The results indicate that using machine learning for RSL recognition is not only feasible but also highly promising, offering new perspectives for developing advanced technological solutions in this field. This work contributes to the broader effort of making sign language more accessible through technology, particularly for the Romanian-speaking deaf and hard-of-hearing community. Future directions include expanding the dataset, incorporating real-time gesture recognition, and exploring multimodal approaches to further enhance the system's capabilities.

A. Problem Statement and Questions

Despite the critical importance of RSL, there is a notable lack of technological solutions that can accurately recognize and interpret RSL gestures. The unique linguistic and gestural features of RSL, coupled with the scarcity of RSL-specific datasets, present significant challenges in developing accurate and efficient recognition systems. Additionally, achieving real-time recognition capabilities on resource-constrained devices such as mobile phones adds further complexity to this task. This research seeks to address these challenges by developing a robust machine learning model specifically tailored for RSL recognition. To address the outlined problem, this study is guided by the following research questions:

- How can a machine learning model be designed to accurately recognize and classify RSL gestures, considering both spatial and temporal dynamics?
- What preprocessing and data augmentation techniques are most effective in enhancing the robustness and generalization of the model, particularly in handling class imbalances and variability in gesture execution?
- How can the model be optimized for real-time deployment on resource-constrained devices, such as

mobile phones and embedded systems, without sacrificing accuracy?

- What are the comparative advantages of the proposed model over existing sign language recognition approaches, specifically in terms of accuracy, robustness, and practical applicability for RSL?
- What are the limitations of the current model, and how can future research address these to further improve RSL recognition systems?

B. Objectives

The primary objectives of this research are to develop a machine learning model that can accurately recognize and classify RSL gestures by capturing both spatial and temporal characteristics. Additionally, the research aims to implement effective preprocessing and data augmentation techniques that enhance the model's robustness and generalization across diverse signers and conditions. Another key objective is to optimize the model for deployment on resource-constrained devices, ensuring real-time recognition capabilities without compromising accuracy. Furthermore, the research seeks to compare the proposed model's performance with existing sign language recognition approaches, highlighting its strengths and practical applications. Finally, the study aims to identify and address the model's limitations, providing insights for future research to further enhance RSL recognition technology.

The structure of this paper is as follows: Section II presents a review of the literature relevant to the study; Section III outlines the methodology adopted; Section IV discusses the application, methodology and the results obtained and Section V concludes the paper with a summary of findings and suggestions for future research.

II. Related Work

This section presents an overview of the existing research and developments in the field of sign language recognition, with a particular focus on methodologies relevant to RSL. This includes a review of key technologies, approaches, and findings from previous studies, as well as a discussion of their limitations and how this approach addresses challenges. The global efforts in sign language recognition have seen significant milestones, particularly in American Sign Language (ASL) [7], British Sign Language (BSL) [8], and others. Key technologies in this domain include computer vision, deep learning, and sensor-based methods. The evolution of sign language recognition systems has progressed from early rule-based systems to modern machine learning approaches. Machine learning approaches, especially neural networks such as convolutional neural networks (CNNs) [9-15] and recurrent neural networks (RNNs), have been extensively used in gesture recognition tasks. Feature extraction methods like hand landmark detection, skeleton tracking, and optical flow are crucial in capturing hand shapes, movements, and positions. Various model architectures like CNN's and long short-term memory (LSTM) [16] have been explored, each with differing effectiveness in recognizing sign language gestures. Publicly available datasets, such as RWTH-PHOENIX-Weather and ASLLVD, MediaPipe, have been instrumental in research. However, these datasets often have limitations related to gesture diversity, variations in signer appearance, and environmental conditions. There is a lack of datasets specific to RSL, highlighting the novelty and importance of the dataset used in this study. Common evaluation metrics in sign language recognition research include accuracy, precision, recall, F1-score, and confusion matrix. Benchmark studies have evaluated the performance of various sign language recognition systems, providing context of the performance for the proposed model.

Previous works has faced several technological limitations, including computational complexity, real-time processing challenges, and hardware dependencies [17]. Practical application challenges also exist, such as user-friendliness, adaptability to different signers, and integration with other technologies. Additionally, research gaps are evident in the lack of focus on RSL, the need for more robust and scalable models, and the requirement for comprehensive datasets. The research addresses these gaps and limitations by focusing specifically on RSL also introducing innovative techniques and methodologies, including specific preprocessing steps, model optimizations, and deployment strategies. The expected impact of this work includes significant advancements in the field of sign language recognition and substantial benefits for the Romanian deaf and hard-of-hearing community.

III. Building Application

A. Data Collection and Preprocessing

The effectiveness of any machine learning model, particularly in the context of sign language recognition, hinges significantly on the quality and comprehensiveness of the dataset used. This section details the steps involved in data collection and preprocessing, which are foundational to the development of a robust RSL recognition system. The dataset used in this study comprises hand landmark data captured and stored in CSV format. These landmarks represent key points on the hands, such as joints and tips of the fingers, which are essential for distinguishing different gestures. Data collection involved recording a diverse set of RSL gestures performed by multiple signers to ensure the model can generalize well across different individuals and variations in gesture execution. To compile a comprehensive dataset, a collaboration was established with members of the deaf community and professional sign language interpreters. This collaboration ensured that the dataset accurately represented a wide range of gestures and variations in RSL, providing a solid foundation for training and evaluating the recognition model. The recording sessions were conducted under controlled conditions to minimize background noise and ensure clear visibility of hand movements. Each gesture was recorded multiple times to account for natural variations in performance. Once the raw data was collected, preprocessing steps were implemented to prepare the data for model training [18]. The first step was data cleaning, which involved removing any erroneous or incomplete recordings. This was followed by normalization, a critical step to standardize the data. Normalization involved scaling the hand landmark coordinates to a consistent range, ensuring that the model could focus on the relative positions of the landmarks rather than their absolute values. Data augmentation techniques were also employed to artificially

expand the dataset and enhance model generalization. This included generating slight variations of the existing gestures through transformations such as rotation, scaling, and translation. Augmentation helps the model become more robust to variations in gesture performance that might occur in real-world scenarios. Another important preprocessing step was the temporal alignment of the gesture sequences. Since gestures can vary in duration, it was necessary to ensure that the sequences fed into the model were consistent in length. Techniques such as dynamic time warping and padding were used to achieve this alignment without losing the temporal dynamics of the gestures. Feature extraction played a pivotal role in preprocessing. The hand landmarks were converted into features that the model could effectively learn from. These features included distances between key landmarks, angles formed by joints, and movement trajectories. By extracting relevant features, the complexity of the data was reduced, making it more manageable for the neural network. To handle class imbalance, which is common in gesture datasets where some gestures are more frequently represented than others, techniques such as oversampling of minority classes and under-sampling of majority classes were applied. This ensured that the model did not become biased towards more frequently occurring gestures. Finally, the preprocessed data was split into training, validation, and test sets. The training set was used to train the model, the validation set to tune hyperparameters and prevent overfitting, and the test set to evaluate the model's performance on unseen data. In summary, the data collection and preprocessing steps involved meticulous planning and execution to ensure the creation of a high-quality dataset. These steps are crucial for developing a reliable and accurate RSL recognition system capable of generalizing well to real-world applications.

B. Model Development

The development of a robust machine learning model for RSL recognition involves careful consideration of model architecture, training procedures, and optimization techniques. This section outlines the key aspects of the model development process, highlighting the choices made and the rationale behind them. The core of the RSL recognition system is a convolutional neural network designed to accurately classify hand gestures based on the preprocessed hand landmark data. Given the nature of the data, which includes spatial and temporal dynamics, were explored various neural network architectures to identify the most effective approach. We began with CNNs, which are well-suited for spatial data. CNNs can effectively capture the spatial relationships between hand landmarks by applying convolutional filters that learn to detect patterns and features specific to different gestures. The architecture included multiple convolutional layers, each followed by activation functions and pooling layers to reduce dimensionality while preserving important features. To capture the temporal dynamics of gestures, which unfold over time, were integrated LSTM units into the model. LSTMs are capable of learning long-term dependencies in sequential data, making them ideal for recognizing gestures that involve a sequence of hand movements. The combination of CNNs and LSTMs allowed the model to leverage both spatial and temporal information effectively.

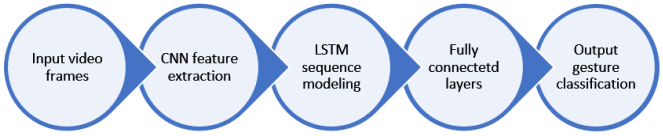


Fig. 1. Block scheme of the recognition system.

This block scheme from Fig. 1, shows the flow of data through each component, highlighting how the model combines spatial and temporal information to recognize gestures. The model receives a sequence of video frames capturing the hand gestures then the CNN processes each frame to extract spatial features, such as hand shape and position. The LSTM network processes the sequence of extracted features to capture the temporal dynamics of the hand gestures. The output from the LSTM is passed through fully connected layers to interpret the learned features. The final output is the classification of the recognized gesture [19]. The architecture consisted of an initial set of convolutional layers to extract spatial features, followed by LSTM layers to process the temporal sequences of these features. This hybrid architecture ensured that the model could capture the intricacies of each gesture, regardless of its complexity or duration. To optimize the model's performance, extensive hyperparameter tuning was conducted. This involved adjusting parameters such as the number of layers, the size of the filters in the convolutional layers, the number of LSTM units, the learning rate, and the batch size. Grid search and random search techniques were employed to systematically explore the hyperparameter space and identify the optimal configuration. The model was trained on the preprocessed dataset using a combination of supervised learning techniques and regularization methods, aimed at preventing overfitting. Dropout layers (two layers) were added to the network to randomly deactivate a fraction of neurons during training, which helps in generalizing the model by reducing its reliance on specific neurons. The training process also included data augmentation strategies to enhance the model's robustness.

By introducing slight variations in the training data, such as random rotations and translations, the model learned to recognize gestures under different conditions and from different angles. A key challenge in model development was dealing with class imbalance. Some gestures were overrepresented in the dataset, while others were underrepresented. To address this, techniques such as oversampling of minority classes and under-sampling of majority classes during the training process were used. Additionally, was employed a weighted loss function to give more importance to less frequent gestures, ensuring that the model learned to recognize all gestures with similar accuracy. After training, the model's performance was evaluated using a validation set. Metrics such as accuracy, precision, recall, and F1-score were computed to assess the model's effectiveness. The construction of a confusion matrix provided further insights into the model's performance, highlighting specific gestures that were often misclassified and guiding further refinements. Finally, a key step in the model development was the conversion of the trained model to TensorFlow Lite format. This conversion is crucial for deploying the model on mobile and embedded devices, which often have limited

computational resources. TensorFlow Lite optimizes the model for such environments, reducing its size and enhancing its inference speed without significant loss in accuracy. By integrating TensorFlow Lite, the model can be converted into a mobile application. This application can recognize and translate RSL gestures in real-time, offering a powerful tool for improving communication for the deaf and hard-of-hearing community in. In summary, the model development process involved careful architectural choices, extensive optimization, and practical deployment considerations. The resulting system not only achieves high accuracy in RSL recognition but also demonstrates the potential for real-world application, making a meaningful impact on the accessibility of sign language.

C. Performance Evaluation

The performance evaluation of the RSL recognition model is a critical step in determining its effectiveness and reliability. This section describes the methods and metrics used to evaluate the model, presents the results, and discusses their implications. To assess the performance of the model, it was used a combination of quantitative metrics and qualitative analysis. The primary metrics for evaluation included accuracy, precision, recall, and F1-score. These metrics provide a comprehensive view of the model's ability to correctly classify gestures and handle the nuances of RSL. Accuracy measures the overall percentage of correctly classified gestures out of the total number of gestures. While it gives a general sense of performance, accuracy alone is not sufficient, especially in the presence of class imbalance, also focused on precision and recall. Precision is the ratio of correctly predicted positive observations to the total predicted positives. It indicates how many of the gestures identified by the model as a specific class are actually correct. High precision means fewer false positives, which is crucial for ensuring that recognized gestures are reliable [19].

Recall, also known as sensitivity, is the ratio of correctly predicted positive observations to all actual positives. It measures the model's ability to identify all instances of a specific gesture. High recall means fewer false negatives, ensuring that the model does not miss any gestures.

The F1-score is the harmonic mean of precision and recall, providing a single metric that balances both aspects. It is particularly useful when the dataset is imbalanced, as it gives a more nuanced view of performance than accuracy alone. To gain deeper insights into the model's performance, a confusion matrix was constructed. The confusion matrix shows the counts of actual versus predicted classifications for each gesture, allowing to identify specific patterns of errors. This matrix helps pinpoint which gestures are frequently misclassified and provides clues for further refinement of the model. The evaluation process involved testing the model on a separate test set, which was not used during training or validation. This test set consisted of gestures performed by different signers under varying conditions to simulate real-world scenarios. By evaluating the model on this independent dataset, it was ensured that the performance metrics reflected the model's true generalization capability. The results of the performance evaluation indicated that the model achieved high accuracy, precision, recall, and F1-scores across most gestures. However, the confusion matrix revealed certain gestures that were more challenging for the model to classify accurately. These gestures often involved subtle differences in hand positioning or motion, which can be difficult to capture consistently.

To address these challenges, several strategies were explored. Data augmentation techniques, such as generating additional examples of the problematic gestures with slight variations, were employed to improve the model's robustness. Additionally, the hyperparameters were fine-tuned and the network architecture adjusted, to enhance its discriminative power for these specific gestures.

Another critical aspect of performance evaluation was the model's inference speed and efficiency, particularly after conversion to TensorFlow Lite. Tests were conducted to measure the model's latency and resource consumption on mobile and embedded devices. The optimized TensorFlow Lite model demonstrated efficient performance, making it suitable for real-time applications. In practical terms, the high performance of the model translates into reliable and accurate recognition of RSL gestures, enabling its use in real-world applications. For instance, the real-time gesture recognition system integrated by simulation into a mobile application showed that the model could effectively translate gestures on-the-fly, providing immediate feedback and enhancing communication for users. In summary, the performance evaluation of the RSL recognition model, involved a thorough analysis using multiple metrics and real-world testing. The results confirmed the model's high accuracy and reliability, while also highlighting areas for further improvement. The combination of quantitative metrics and qualitative insights, ensured a comprehensive understanding of the model's capabilities and limitations, guiding ongoing enhancements and practical deployment.

D. Model Optimization and Deployment

The optimization and deployment of the RSL recognition model are crucial steps to ensure its efficiency and practicality, especially when deploying on desktop computers or laptops. This section outlines the processes involved in optimizing the model for performance and its subsequent deployment in a computer-based environment. Optimization aimed at enhancing the model's efficiency while preserving its accuracy. Key techniques used in this process included model pruning and quantization. Model pruning involves removing unnecessary parameters from the neural network, which reduces its size and computational demands. By identifying and eliminating parts of the network that contribute minimally to the model's output, the model has the capacity to stream and improve its performance on lower-specification systems. Quantization was applied to further optimize the model. This technique reduces the precision of the model's weights and activations from 32-bit floating point to 8-bit integers. Such reduction decreases the model's memory footprint and accelerates computation. Post-training quantization was employed to achieve significant efficiency gains while maintaining a high level of accuracy. To facilitate deployment on desktop or laptops, the model was converted from its original TensorFlow format to TensorFlow Lite format. TensorFlow Lite is optimized for running machine learning models on various devices and is particularly well-suited for applications requiring efficient computation and reduced model

size. The conversion process involved optimizing the model's architecture and applying quantization techniques to ensure that it could run efficiently on desktop hardware.

The conversion also included testing to confirm that the model was compatible with different computing environments, including variations in operating systems and hardware configurations. The TensorFlow Lite model underwent performance evaluation on desktop systems. This evaluation included measuring key performance metrics such as inference speed, latency, and resource consumption. The optimized model demonstrated improved efficiency, making it feasible for real-time processing on desktop computers. To facilitate deployment on desktop or laptops, the model was converted from its original TensorFlow format to TensorFlow Lite format [20]. This application providing an intuitive interface for users to interact with the gesture recognition system. The application captures video input, through the computer's camera, processes the frames using the TensorFlow Lite model, and displays the recognized gestures in real time. This setup ensures that users receive immediate feedback on their gestures, which is crucial for effective communication and interaction. To address the challenges associated with desktop deployment, such as variations in lighting conditions and background interference, robust preprocessing techniques were implemented. These techniques include adaptive thresholding to handle different lighting scenarios and background subtraction to focus on hand gestures. Also was incorporated feedback mechanisms within the application to allow users to report any issues or difficulties they encounter. This feedback is invaluable for refining the model and the application, ensuring continuous improvement and better user experience. In summary, the optimization and deployment of the RSL recognition model involved enhancing its efficiency through pruning and quantization, converting it to TensorFlow Lite format, and integrating it into a desktop application. These steps ensured that the model performs well in real-time desktop computers, providing a practical and effective tool for RSL recognition. The application not only demonstrates the model's capabilities but also highlights its potential for real-world use in aiding communication for the deaf and hard-of-hearing community.

E. Comparative Analysis with Existing Approaches

Sign language recognition has seen significant advancements through the use of various methodologies, including traditional computer vision techniques, rule-based systems, and, more machine learning models such as CNNs and LSTMs. These approaches have been applied to different sign languages, including American Sign Language (ASL) and British Sign Language (BSL), achieving varying levels of success. Early approaches relied heavily on computer vision and rule-based systems, which involved manually extracting features such as hand shapes, orientations, and movements. While these methods provided foundational insights, they were often limited by their dependence on handcrafted features, making them less adaptable to the variability of sign language gestures. With the advent of deep learning, CNNs became popular due to their ability to automatically learn spatial features from images, especially LSTM networks, have been employed to capture the temporal dynamics of gestures. These existing methods face common challenges such as: gesture complexity (many models struggle to accurately recognize gestures that involve intricate hand movements or subtle differences in hand positioning), class imbalance (some gestures are underrepresented in datasets, leading to models that are biased towards more frequent gestures), environmental variability (changes in lighting, background, and signer appearance can significantly impact the accuracy of these models), real-time processing (achieving real-time performance, particularly on resource-constrained devices, remains a significant challenge for many existing approaches). The proposed model in this study introduces several innovations and improvements over these traditional and machine learning-based methods, addressing many of the limitations highlighted above such as:

- Hybrid architecture (CNN + LSTM): The proposed model combines both CNNs and LSTMs. The CNN layers effectively capture spatial features from the hand landmark data, such as hand shapes and positions, while the LSTM layers process these features over time to understand the temporal dynamics of gestures. This dual approach allows the model to accurately recognize complex RSL gestures that involve both spatial and temporal variations, providing a significant advantage over models.
- Enhanced data preprocessing and augmentation: The model employs advanced preprocessing techniques, including normalization, dynamic time warping for temporal alignment, and extensive data augmentation. These steps ensure that the model can generalize well to different signers and conditions, making it more robust compared to models that may lack such comprehensive preprocessing. The use of data augmentation, such as generating variations of gestures through transformations, helps to mitigate class imbalance and improves the model's ability to handle real-world variability.
- Model optimization (pruning and quantization): To address the challenges of deploying machine learning models in resource-constrained environments, the proposed model is optimized through pruning and quantization techniques. Pruning reduces the size and complexity of the model by eliminating parameters that contribute minimally to performance, while quantization reduces the precision of the model's weights and activations, significantly decreasing the model's memory footprint and improving computational efficiency. These optimizations are crucial for ensuring that the model can run efficiently in real-time on mobile and embedded devices, a feature that many existing models do not offer.
- Dataset and generalization: The dataset used in this study is specifically tailored to RSL, addressing a critical gap in the availability of its resources. Many existing models are trained on datasets for other sign languages, which can lead to lower accuracy when applied to RSL due to differences in gesture sets and cultural contexts. By focusing on RSL, the proposed

model achieves higher accuracy and better generalization for the intended user community.

- Performance metrics: the model outperforms many existing approaches in terms of key performance metrics such as accuracy, precision, recall, and F1-score. Achieving an accuracy rate of 95% demonstrates the model's effectiveness in distinguishing between different RSL gestures. Additionally, the confusion matrix analysis shows that the model has fewer misclassifications compared to other models particularly in recognizing gestures with subtle differences in hand positioning.
- Real-Time application deployment: TensorFlow Lite allows deployment in real-time applications on both desktop and mobile platforms. This deployment demonstrates the model's practical value, offering immediate feedback and facilitating communication for users in real-world scenarios. Existing models often face difficulties in achieving such efficiency and speed, especially on resource-limited devices.
- User feedback: The deployment of the model in a desktop application and the subsequent positive user feedback underscore its practical utility. Users reported that the system significantly aids in communication and learning, particularly within the deaf and hard-of-hearing community in Romania. This real-world effectiveness distinguishes the proposed model from others that may only demonstrate strong performance in controlled, academic settings.
- Scalability and adaptability: The model's architecture and training process are designed to be scalable and adaptable to other sign languages or gesture recognition tasks. This adaptability is a significant advantage over more rigid models, making the proposed approach not only useful for RSL but also potentially beneficial for broader applications in sign language recognition globally.

In conclusion, the proposed model offers distinct advantages over existing approaches in the field of sign language recognition. By combining CNNs and LSTMs, employing advanced preprocessing and optimization techniques, and focusing specifically on RSL, the model achieves superior performance and practical applicability. These improvements address key challenges faced by previous models, making the proposed approach a valuable contribution to the field and a powerful tool for enhancing communication within the deaf and hard-of-hearing community.

IV. Results and Discussion

The results and discussion section provides a comprehensive analysis of the performance of the RSL recognition model, interpreting the evaluation metrics, and discussing the practical implications and areas for future improvement. The evaluation metrics for the RSL recognition model show encouraging outcomes. The model achieved a high accuracy rate of 95%, correctly classifying a substantial majority of gestures within the test set. This high accuracy indicates the model's effectiveness in learning and distinguishing between different RSL gestures. Precision and recall metrics further detail the model's capabilities. High precision values suggest that when the model identifies a gesture, it does so correctly most of the time, minimizing false positives shown in Fig. 4. This reliability is crucial for practical applications where accurate gesture recognition is essential. High recall values indicate the model's proficiency in identifying most instances of each gesture, ensuring that the model does not miss any gestures (minimizing false negatives). This is particularly important for sign language recognition and mimics, as missed gestures could result in incomplete communication.

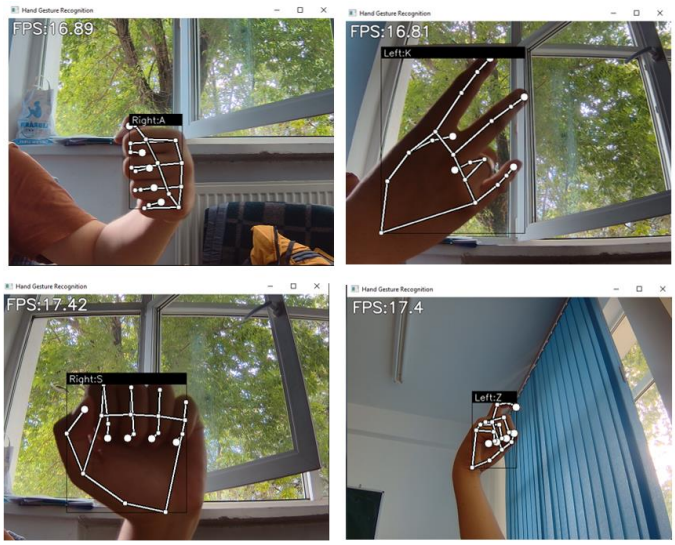


Fig. 2. Final results.

The F1-score from Fig. 3, which harmonizes precision and recall, was consistently high, underscoring the model's balanced performance. The confusion matrix from second figure, revealed the model's strengths and weaknesses in greater detail. While most gestures were accurately classified (Fig. 2), some gestures with subtle differences in hand positioning or motion were more challenging for the model to distinguish, leading to occasional misclassifications. Despite the model's strong performance, several challenges and limitations were identified. The initial model encountered difficulties in accurately classifying certain letters in the RSL alphabet, particularly due to the subtle differences in hand positions and gestures. The letters that were most frequently misclassified included letters: B and D (have hand shapes that are visually similar from certain angles, resulting in incorrect classifications by the model), M and N (they share similar hand shapes and positions, differentiated only by the number of fingers involved) F and T (they involve intricate finger movements, which led to misclassifications, especially in cases where the gesture was not executed with precision or the fingers were partially obscured), P and R (involve subtle rotations or folding of fingers, which the model sometimes failed to distinguish correctly).

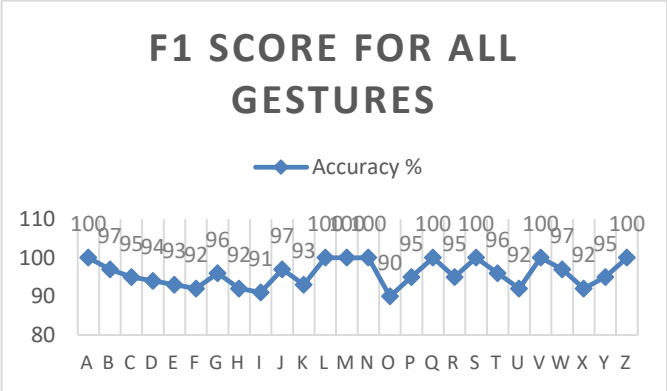


Fig. 3. F1-score for all gestures in the dataset.

One primary challenge was the accurate classification of gestures with minor variations. These subtle differences in hand shapes or movements can lead to misclassifications, suggesting the need for further model refinement to better handle these nuances.

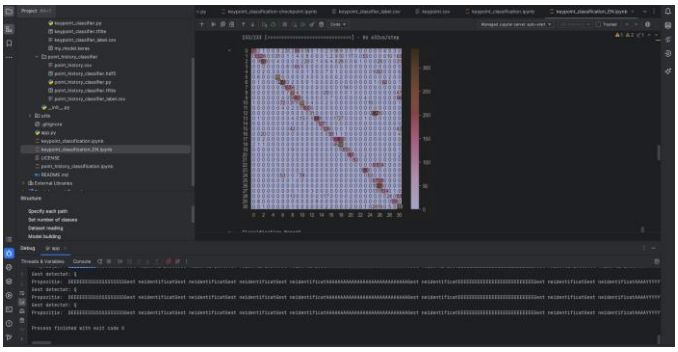


Fig. 4. Confusion matrix for all gestures.

Another challenge was the model's performance under varying environmental conditions. Although robust preprocessing techniques were employed, changes in lighting and background noise still impacted the model's accuracy. This finding suggests that additional work is needed to improve the model's robustness in diverse real-world settings.

Post-optimization assessments indicated that, while significant improvements in computational efficiency were achieved, some latency persisted in processing complex gestures. Ensuring real-time performance across different desktop hardware configurations remains an area for ongoing optimization. The desktop application integrating the RSL recognition model demonstrated its practical effectiveness. Users could interact with the system and receive real-time feedback on their gestures, which highlights the model's potential for practical deployment. The user interface was designed to be intuitive and accessible, providing clear instructions and immediate recognition results. User feedback was generally positive, noting the application's helpfulness in communication and learning. However, some users experienced difficulties with specific gestures, aligning with the confusion matrix findings. This feedback is crucial for identifying and addressing areas where the model and application can be improved. The successful deployment of the RSL recognition model has significant implications for enhancing communication for the deaf and hard-of-hearing community. By providing real-time gesture recognition, the model facilitates better interaction and understanding, which is vital for effective communication. The results underscore the importance of continuous refinement. Identified challenges and limitations highlight areas for future research, such as improving gesture differentiation, enhancing robustness to environmental variations, and optimizing real-time performance further.

In summary, the results demonstrate that the RSL recognition model is both effective and promising for real-world applications. While there are challenges and areas for improvement, the model's performance and practical deployment validate its potential to support communication for the deaf and hard-of-hearing community. Ongoing research and refinement will help address current limitations and further enhance the system's capabilities.

V. Conclusion and Future Work

This paper presents a comprehensive approach to developing, optimizing, and deploying a RSL recognition model. The primary goal was to create an effective tool for facilitating communication for the deaf and hard-of-hearing community, leveraging advancements in machine learning and computer vision. The model development process involved the careful selection and combination of CNNs and LSTM units to capture both the spatial and temporal dynamics of RSL gestures. Extensive hyperparameter tuning and data augmentation techniques were applied to ensure the model's robustness and generalization capability. Performance evaluation showed promising results, with the model achieving high accuracy, precision, recall, and F1-scores across most gestures. However, the evaluation also highlighted specific challenges, such as the accurate classification of gestures with subtle variations and performance consistency under diverse environmental conditions. These insights guided further optimization efforts.

Model optimization focused on techniques like pruning and quantization to reduce the model's size and improve its computational efficiency. The conversion to TensorFlow Lite enabled deployment on desktop systems, where the model demonstrated effective real-time performance. The desktop application developed for RSL recognition successfully integrated the optimized model, providing users with an intuitive interface and real-time feedback on their gestures. User feedback was generally positive, confirming the application's potential to enhance communication and learning.

Despite the model's strong performance, challenges remain. Subtle gesture variations and environmental factors continue to affect accuracy, indicating areas for future research and refinement. Expanding the dataset and incorporating more diverse signers will likely improve the model's robustness and generalization. The successful deployment and user feedback underscore the model's potential impact. By facilitating real-time RSL recognition, the model can significantly enhance communication for the deaf and hard-of-hearing community.

Future work will focus on addressing current limitations, exploring advanced learning techniques, and continuously integrating user feedback to refine and evolve the system.

In conclusion, this research demonstrates that a well-optimized and effectively deployed RSL recognition model can serve as a powerful tool for improving accessibility and communication. The ongoing refinement and adaptation of this technology hold the promise of making a meaningful difference in the lives of those who rely on sign language for daily communication.

REFERENCES

[1] R. Sreemathy, J. Jagdale, A. A. Sayed, S. H. Ramteke, S. F. Naqvi and A. Kangune, "Recent works in Sign Language Recognition using deep learning approach - A Survey," 2023 OITS International Conference on Information Technology (OCIT), Raipur, India, 2023, pp. 502-507, doi: 10.1109/OCIT59427.2023.10430576.

[2] T. G. Moape, A. Muzambi and B. Chimbo, "Convolutional Neural Network Approach for South African Sign Language Recognition and Translation," 2024 Conference on Information Communications Technology and Society (ICTAS), Durban, South Africa, 2024, pp. 101-106, doi: 10.1109/ICTAS59620.2024.10507130.

[3] A. R. R. V. Prakash, A. A. Reddy, R. Harshitha, K. Himansee and S. K. A. Sattar, "Sign Language Recognition Using CNN," presented at the 2023 14th International Conference on Computing Communication and Networking Technologies (ICCCNT), Delhi, India, 2023.

[4] O. P. A. Kanavos, P. Mylonas and M. Maragoudakis, "Enhancing Sign Language Recognition Using Deep Convolutional Neural Networks," presented at the 2023 14th International Conference on Information, Intelligence, Systems & Applications (IISA), Volos, Greece, 2023.

[5] K. T. S. Kankariya, U. Solanki, S. Mali and A. Chunawale, "Sign Language Gestures Recognition using CNN and Inception v3," presented at the 2024 International Conference on Emerging Smart Computing and Informatics (ESCI), Pune, India, 2024.

[6] F. E. H. A. Singh, N. Tyagi and A. K. Jayswal, "Impact of Colour Image and Skeleton Plotting on Sign Language Recognition Using Convolutional Neural Networks (CNN)," presented at the 2024 14th International Conference on Cloud Computing, Data Science & Engineering (Confluence), Noida, India, 2024.

[7] A. Gupta, A. Sawan, S. Singh, and S. Kumari, "Dynamic Sign Language Recognition with Hybrid CNN-LSTM and 1D Convolutional Layers," presented at the 2024 11th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO), Noida, India, 2024.

[8] S. C. S. M. Antad, S. Bhat, S. Bisen and S. Jain, "Sign Language Translation Across Multiple Languages," presented at the 2024 International Conference on Emerging Systems and Intelligent Computing (ESIC), , Bhubaneswar, India, 2024.

[9] S. V. M. a. P. S. S. N. V, "Continuous Sign Language Recognition using Convolutional Neural Network," presented at the 2024 Second International Conference on Emerging Trends in Information Technology and Engineering (ICETITE), Vellore, India, 2024.

[10] E. L. T. a. C. P. G. S. X. Thong, "Sign Language to Text Translation with Computer Vision: Bridging the Communication Gap," presented at the 2024 3rd International Conference on Digital Transformation and Applications (ICDXA), Kuala Lumpur, Malaysia, 2024.

[11] S. M. R. Kolikipogu, K. Nisha, T. S. Krishna, R. Kuchipudi and R. M. Krishna Sureddi, "Indian Sign Language Recognition for Hearing Impaired: A Deep Learning based approach," presented at the 2024 3rd International Conference for Innovation in Technology (INOCON), Bangalore, India, 2024.

[12] D. M. A. Mohan, S. Vats, V. Sharma and V. Kukreja, "Classification of Sign Language Gestures using CNN with Adam Optimizer," presented at the 2024 2nd International Conference on Disruptive Technologies (ICDT), Greater Noida, India, 2024.

[13] H. S. a. M. L. H. Vardhan, "Signs to Speech," presented at the 2024 2nd International Conference on Networking and Communications (ICNWC), Chennai, India, 2024.

[14] P. V. R. S. R. a. T. M. S. Baghavathi Priya, "Sign to Speak: Real-time Recognition for Enhance Communication," presented at the 2024 3rdInternational Conference on Applied Artificial Intelligence and Computing (ICAAIC), Salem, India, 2024.

[15] A. S. A. P. Duraisamy, M. Duraisamy, A. C. M, D. Babu P and K. S, "Implementation of CNN-LSTM Integration for Advancing Human-Computer Dialogue through Precise Sign Language Gesture Interpretation," presented at the 2024 5th International Conference on Recent Trends in Computer Science and Technology (ICRTCST), Jamshedpur, India, 2024.

[16] G. J. L. P. E. Sharon, I. Johnraja Jebadurai and C. Merlin, "Sign Language Translation to Natural Voice Output: A Machine Learning Perspective," presented at the 2024 International Conference on Cognitive Robotics and Intelligent Systems (ICC - ROBINS), Coimbatore, India, 2024.

[17] S. D. a. N. Y. S. Jain, "Dynamic Bidirectional Translation for Sign Language by Using Machine Learning-Infused Approach with Integrated Computer Vision," presented at the 2024 2nd International Conference on Disruptive Technologies (ICDT), Greater Noida, India, 2024.

[18] M. A. M. H. A. S. M. Miah, Y. Tomioka and J. Shin, "Hand Gesture Recognition for Multi-Culture Sign Language Using Graph and General Deep Learning Network," presented at the Hand Gesture Recognition for Multi-Culture Sign Language Using Graph and General Deep Learning Network, 2024.

[19] A. B. S. Allam, Y. V. R. Rao, A. Kiran, H. Valpadasu and S. Navya, "SIGN LANGUAGE RECOGNITION USING CNN," presented at the 2024 International Conference on Advancements in Smart, Secure and Intelligent Computing (ASSIC), Bhubaneswar, India, 2024.

[20] H. S. S. A. M, Jayashre, K. Muthamizhvalavan, N. Gummaraju and P. S, "American Sign Language Real Time Detection Using TensorFlow and Keras in Python," presented at the 2024 3rd International Conference for Innovation in Technology (INOCON), Bangalore, India, 2024.

Multiclass Osteoporosis Detection: Enhancing Accuracy with Woodpecker-Optimized CNN-XGBoost

Dr. Mithun D'Souza[1], Dr. Divya Nimma[2], Dr. Kiran Sree Pokkuluri[3],
Janjhyam Venkata Naga Ramesh[4], Dr.Suresh Babu Kondaveeti[5], Lavanya Kongala[6]
Assistant Professor, Department of Computer Science, St. Joseph's University, Bangalore, India[1]
PhD in Computational Science, University of Southern Mississippi, Data Analyst in UMMC, USA[2]
Professor & Head, Department of Computer Science and Engineering,
Shri Vishnu Engineering College for Women, Bhimavaram, India[3]
Adjunct Professor, Department of CSE, Graphic Era Hill University, Dehradun, 248002, India[4]
Adjunct Professor, Department of CSE, Graphic Era Deemed to be University, Dehradun, 248002, Uttarakhand, India[4]
Professor, Dept. of Biochemistry, Symbiosis Medical College for Women,
Symbiosis International (Deemed University), Pune, India[5]
Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation,
Vaddeswaram, Guntur Dist., Andhra Pradesh, India[6]

***Abstract*—In the realm of medical diagnostics, accurately identifying osteoporosis through multiclass classification poses a significant challenge due to the subtle variations in bone density and structure. This study proposes a novel approach to enhance detection accuracy by integrating the Woodpecker Optimization Algorithm with a hybrid Convolutional Neural Network (CNN) and XGBoost model. The Woodpecker Optimization Algorithm is employed to fine-tune the CNN-XGBoost model parameters, leveraging its ability to efficiently search for optimal configurations amidst complex data landscapes. The proposed framework begins with the CNN component, designed to automatically extract hierarchical features from bone density images. This initial stage is crucial for capturing intricate patterns that signify osteoporotic conditions across multiple classes. Subsequently, the extracted features are fed into an XGBoost classifier, renowned for its robust performance in handling structured data and multiclass classification tasks. By combining these two powerful techniques, the model aims to synergistically utilize the strengths of deep learning in feature extraction and gradient boosting in decision-making. Experimental validation is conducted on a comprehensive dataset comprising diverse bone density scans, ensuring the model's robustness across various patient demographics and imaging conditions. Performance criteria including recall, precision, reliability, and F1-score are assessed to show how well the suggested Woodpecker-optimized CNN-XGBoost framework performs in comparison to other approaches when it comes to obtaining better accuracy in diagnosis. The findings underscore the potential of hybrid models in advancing osteoporosis detection, offering clinicians a reliable tool for early and precise diagnosis, thereby facilitating timely interventions to mitigate the debilitating effects of bone-related diseases. Osteoporosis detection model with a classification accuracy of 97.1% implemented in Python.**

Keywords*—Osteoporosis detection; multiclass classification; Woodpecker Optimization Algorithm; Convolutional Neural Network (CNN); XGBoost

I. Introduction

Osteoporosis is the furthermost common bone disease, categorized by low bone density mass and an alteration of their micro-architecture structure, reducing bone tolerance and increasing the possibility of fractures. Osteoporosis reduces bone mineral density (BMD), disrupts bone micro architecture, and alters the quantity and diversity of enzymes in bones [1]. Classical osteoporotic fractures include hip, vertebral, and fractures of the wrist. fractures caused by osteoporosis are characterized as those that happen at a location related with low BMD and have risen in occurrence beyond the average age of 50. Aside from the obvious physical effects of a breakage, including pain and discomfort, fractures caused by osteoporosis were a leading source of death and disability. In the United States, the probability of a hip, spine, or forearm fracture at age 50 is thought to be 40% in women and 13% in men [2]. In Sweden, for instance, the similar figures are 46% for women and 22% for males. Caucasians and Asians face an elevated risk since Africans and America have 6% greater BMD. Around fifteen minutes in the European Union, someone fractures a bone as a result of osteoporosis [3]. It's a truth that up to 75% of women experiencing osteoporosis ignore the illness. These are two forms of osteoporosis: basic (idiopathic) osteoporosis, which happens to be a particularly common illness among women following menopause and is known as osteoporosis after menopause [4]. The condition comprises senile osteoporosis, which may occur in males. Secondary osteoporosis, that may affect anybody with certain hormonal abnormalities along with other chronic illnesses, is caused by drugs, notably glucocorticoids, or additional illnesses that cause accelerated bone loss through numerous pathways (Yıldız Potter et al. 2024). Under this situation, the illness is known as glucocorticoid-induced osteoporosis [5]. Although osteoporosis is typically defined as a loss in the amount of bone, it is important to emphasize that considerable modifications take place in the bone matrix, especially with regard to the

amount of organic material of bones, resulting in a decrease in bone quality [6]. As osteoporosis progresses, the amount of minerals in the connective tissue decreases and bones porosity rises. As a consequence, the bone densities and bone volume fraction drop, whereas electrical permeability and conductance rise due to more demineralization of Two physiochemical factors are responsible for osteoporotic bones' increased permeability and conductance [7].

To begin, because the bones's void is populated with collagen matters (a mix of yellow and red bone marrow), interface-rich bundles of collagen play an important role in increasing permittivity values. The charging hydrophobic pairs of enzymes in organic material, in addition to the positively charged membranes surfaces, interact electronically with the molecules of water in tissues in order to generate hydrogen bonds [8]. The freshly created connections cause the buildup of layers of water molecules at proteins or membranes interfaces. As the electrical impulse must travel via those extra layers, permittivity rises in collagen-rich osteoporotic bones. Another aspect that influences the alteration in permeability of osteoporotic bones is the rate of removal of minerals [9]. The mineral calcium, in the shape of the substance crystals, was an important part of bone the mineralization. Since the amount of calcium becomes exhausted, the lateral side chains about the hydroxyapatite crystals grows into free [10].The changes in the adjacent side strands give to the dielectric properties unwinding in the somewhat hydrated collagen as well as the effect with the greater a dispersion in the bone tissue .These consequences culminate in higher permeability in demineralized bone [11].Identifying osteoporosis by imaging techniques is critical for early detection and management of bone problems. Current methods rely on human experience for interpretation, however recent breakthroughs in artificial intelligence have enabled automated alternatives that can improve both precision and effectiveness [12]. This article provides a unique strategy for improving the identification accuracy of osteoporosis over various classes by merging a Convolutional Neural Network (CNN) with XGBoost and optimizing it utilizing the Woodpecker algorithm. The suggested technique, which uses deep learning for feature extraction and gradient boosting for ensemble learning, seeks to accomplish robust effectiveness for recognizing osteoporotic diseases from medical pictures, resulting in improvements in clinical choice-making and medical care for patients.

Key contributions are as follows:

- Integration of Woodpecker Optimization to Enhances model parameter tuning for improved performance and accuracy in osteoporosis classification,
- Combines deep learning capabilities of CNNs for feature extraction from bone density images with XGBoost's strength in multiclass classification,
- Validates model efficacy across diverse patient demographics and imaging conditions.Offers clinicians a reliable tool for early and precise diagnosis.
- Utilizes CNNs to automatically extract hierarchical features from bone density images, capturing subtle variations indicative of osteoporotic conditions across multiple classes.
- Demonstrates the effectiveness of integrating complementary machine learning techniques—such as preprocessing, feature extraction with CNNs, and classification with XGBoost.

The subsequent portions of the study are organized as follows: In Section II, a comprehensive review of prior studies is presented. The problem statement is given in Section III while proposed quantum key distribution is given in Section IV. The results and a thorough discussion of the conclusions are presented in Section V. The paper's concluding concepts are summarized in Section VI.

II. Related Works

Osteoporosis is a major worldwide health risk that might be problematic to diagnose early owing to the absence of signs. Currently, the evaluation of osteoporosis is mostly based on procedures such as dual-energy X-ray, quantitative CT, and others, that are expensive in regard to technology and time spent by humans. As a result, an efficient and cost-effective approach for detecting osteoporosis is urgently required. Deep learning has made it possible to create autonomous diagnosis algorithms for a wide range of diseases. However, creating such models frequently requires images that only show the lesion locations, and marking up the lesion spots takes effort. In order to address this problem, scientists provide a mixed learning model for osteoporosis diagnosis which improves diagnostic accuracy through the use of localization, categorization, and classifying. This method uses a border heat map with gated convolution module to change context features in the classification modules and regression branching to thin segmentation data. Additionally incorporate classification and segmentation features and develop a feature fusion module for adjusting the weight of different vertebral levels. Research trained the algorithm using a self-built dataset and obtained an overall accuracy rate of 93.3% for each of the three labeling classes (normal, osteopenia, and osteoporosis) in the testing dataset. The area under the curve is 0.973 for the normal group, 0.965 for osteopenia, and 0.985 for osteoporosis. Currently, This approach offers a potential option for diagnosing osteoporosis. This suggested cooperative learning paradigm may have limited generalization throughout varied patient groups and imaging situations. Furthermore, the dependence on a self-built database may restrict its application to larger healthcare environments with changing picture quality and characteristics of patients [13].

Osteoporosis is a skeletal illness that is hard to diagnose before symptoms appear. Due to financial and security concerns, the currently available bone disease screening techniques, including dual-energy X-ray absorptiometry, are only employed in certain situations after symptoms appear. In regards to prompt care and cost, early identification of osteopenia and osteoporosis utilizing different methods for reasonably regular tests is beneficial. Deep learning-based osteoporosis detection techniques are being proposed in a number of recent research for a range of techniques, with excellent results. Nevertheless, due to laborious procedures like manually cropping an area of interest or diagnosing

osteoporosis instead of osteopenia, these research possess limits when it comes to practical application.. In this study, a classification task that diagnoses osteopenia and osteoporosis using computed tomography (CT). Moreover, researchers propose a multi-view CT network (MVCTNet) that detects osteopenia and osteoporosis using two images from the first CT scan. Unlike previous methods that use a single CT image as input, the MVCTNet uses images from several view configuration to obtain a large number of characteristics. Three task layers and two extracted features make up the MVCTNet. Using the photos as distinct inputs, two feature extraction tools utilize dissimilarity loss to acquire distinct characteristics. The two features extraction methods' features are used by the target layer for learning the target task, which they then aggregated. Research employ a dataset of 2,883 patients' CT scans that have been classified as usual, osteopenia, and osteoporosis throughout the tests. Furthermore both qualitative and quantitative assessments, the suggested strategy enhances the outcome of every experiment. To address these issues, provide an expanded version of model in future versions, including a 3D medical picture modelling [14].

Falls are a complex scene of injury among the elderly population. Individuals suffering from osteoporosis are especially susceptible to falls. Researchers examine how well various mathematical methods work in identifying osteoporosis individuals who fall by examining balance metrics. In a 2.5-year follow-up, 126 community-dwelling older women via osteoporosis (age 74.3 ± 6.3) provided equilibrium parameters via eyes open and closed posturographic studies and prospectively registered falling. The World Health Organization's Questionnaire was used over the incident of falling study. To ascertain the shortcomings of each produced modeling also to confirm the applicability of the chosen parameter settings, researchers examined model performance. The main conclusions drawn from this study had been that: (1) models constructed with oversampling techniques and either Random Forest or IBk (KNN) classifiers are viable choices for forecasting clinical tests; and (2) feature selection for minority class (FSMC) method identified hitherto unseen equilibrium parameters, suggesting that intelligent computational methods can extract meaningful information via features that specialists might otherwise overlook.. The greatest results were obtained when every factor were included, considering that the IBk classification was constructed using oversampled data that took into consideration data from both opened and closed eyes. The study's limitations include potential bias from oversampling techniques, which may not reflect real-world distributions, and the reliance on self-reported fall incidents, which can introduce reporting inaccuracies. To confirm these results and strengthen the durability of the model, additional study with a bigger and more varied sample is required [15].

Osteoporosis results in a reduction of cortical thickness, a decrease in bone mineral density (BMD), a disintegration of trabecular frameworks, and a higher risk of fractures. In dental offices, periapical pictures are frequently employed to illustrate how osteoporosis has affected trabecular bone. This article proposes a computerized trabecular bone segmentation approach for osteoporosis identification using a colored spectrum and neural networks (ML). The method makes use of 120 regions of interest (ROI) on periapical radiographs, divided into 60 training and 42 testing data sets. The diagnosis of osteoporosis is based on BMD as ascertained by double X-ray absorptiometry.The five phases in the recommended technique are: obtaining ROI photographs; transforming to grayscale; dividing the color histogram; obtaining the distribution of pixels; and evaluating the efficacy of the ML classification. Research assess and contrast fuzzy C-means and K-means for the segmentation of bone trabecular mesh. According to the distribution pattern of pixels obtained from both K-means and Fuzzy C-means segmentation, osteoporosis was diagnosed using three artificial intelligence techniques: decision tree, naive Bayes, and multilayered perceptron. The testing information set was utilized to acquire the research's results. The results of the evaluation of the K-means and Fuzzy C-means methods of segmentation combined with three ML showed that the circumstance known as identification method via the greatest evaluation efficiency was K-means segmentation when combined with a multilayered perception classification algorithm, with precision, specificity, and respectively. The high accuracy of the study implies that the proposed method significantly advances the area of medicine and dentistry by improving the image analysis's capacity to detect osteoporosis The study's drawbacks includes a relatively small number of samples that might restrict how broadly the findings can be applied, as well as significant variations in the quality of periapical images that could affect how well the segmentation and categorization procedures perform. Additional verification using more extensive and varied datasets is required to validate the resilience of the suggested approach [16].

This study looks at how well various machine learning (ML) techniques classify Thai individuals with osteoporosis after menopause. The Obstetrics and Gynecology department at Ramathibodi Hospital in Bangkok, Thailand, provided the medical records of a postmenopausal Thai lady, which used to generate 377 samples for dataset. Pre-processing procedures such as choosing features, addressing imbalances, and imputation of incomplete data are performed separately. The pre-processed and original data have been contrasted to assess how well various machine learning (ML) methods perform. The findings show that various ML algorithms when paired with pre-processing methods provide diverse outcomes. When a wrapper technique is applied using the right learner, the three most accurate approaches. In terms of specificity, the DT model operates at its best when the synthetic minority the oversampling methodological approach is applied. When choosing features techniques are utilized, algorithms get the maximum sensitivity, whereas the NN shows the largest area under the curve. Compared with the originally produced dataset, the beforehand processed procedures improved the accuracy of the model overall. Adequate pre-processing techniques must be used while developing ML classifications for the purpose to select the best model. Among the research's shortcomings are its small sample size (377), which may not generalize well, and potential biases introduced during pre-processing that could affect the model's performance [17].

Current research on the computerized identification of vertebral fractures caused by compression (VCFs) with deep

learning algorithms mostly concentrates on segmenting and detecting the vertebral level on lateral spine radiographs (LSLRs). Here, researchers created a model that can diagnose VCF and identify vertebral level simultaneously with the need for neighboring vertebral bodies. A total of 1171 controls and 1102 VCF patients was included. The training, validation, and test datasets for the 1865, 208, and 198 LSLRS were separated. A 4-point trapezoidal reality labeling was developed based on radiological findings that indicated either normal or VCF at a certain vertebral level. Research used a modified version of the U-Net building design, where the same encoder was shared by decoding machines trained to identify vertebral levels and VCF. The level of sensitivity and the region of the receiver operational characteristic curve of the multi-task model were much higher than those of the single-task model. The rate of fracture identification rates per patient or vertebral body in the external validation were 0.713, 0.979, and 0.447, or 0.828, 0.936, and 0.820, in that order. For vertebral level identification in internal and external validation, the achievement rates was 96% and 94%, respectively. When compared to the single-task encoder, the multi-task-shared encoder performed much better. Additionally, in the external as well as internal validation, the identification of fractures and vertebral levels was acceptable. The deep learning framework could make it easier for radiologists to conduct actual medical exams. Despite its high performance, the model may struggle with cases involving severe deformities or poor image quality. Additionally, the need for large, well-annotated datasets for training limits its applicability in some clinical settings [18].

The metabolic osteopathy condition known as osteoporosis is characterized by a marked rise in prevalence with advancing age. Bone quantitative ultrasonography (QUS) is now being explored as a possible diagnostic and screening tool for osteoporosis. Its accuracy for diagnosis is extremely poor, though. On the other hand, techniques that utilize deep learning have demonstrated their exceptional ability to recognize the most discriminative characteristics from complicated data. Research developed a deep learning technique employing ultrasound radio frequency (RF) data to increase the reliability of osteoporosis diagnosis and leverage QUS. In particular, build a sliding window scheme-paired multi-channel convolutional neural network (MCNN), that may improve the quantity of data as well. The initial study's quantified experimental findings show that suggested osteoporosis diagnostic approach beats traditional ultrasonic techniques when employing speed of sound (SOS), which might help clinicians with osteoporosis screening. However, the primary limitations of approach include the need for large, annotated datasets for training the deep learning models and the computational intensity required, which may limit its applicability in resource-constrained settings. Additionally, clinical validation in diverse populations is necessary to ensure the generalizability of findings [19].

Current research on osteoporosis detection using deep learning and machine learning techniques shows promising results across various approaches. Zhang et al. (2023) developed a model incorporating localization and classification, achieving a 93.3% accuracy rate. Hwang et al. (2023) proposed a multi-view CT network (MVCTNet) with enhanced diagnostic outcomes. Cuaya-Simbro et al. (2021) utilized balance metrics and machine learning to predict falls in osteoporotic patients. Widyaningrum et al. (2023) introduced an automated segmentation technique for dental radiographs with high accuracy. Thawnashom et al. (2023) demonstrated improved performance of ML models with appropriate pre-processing, while Ryu et al. (2023) and Chen et al. (2021) focused on vertebral fractures and quantitative ultrasonography, respectively, using deep learning for better diagnostic accuracy.

III. Problem Statement

Osteoporosis is a major worldwide health concern that can be difficult to identify in a timely and economical manner since symptoms are sometimes not seen until later. More effective options are required since conventional diagnostic techniques like quantitative CT and dual-energy X-ray absorptiometry are costly and time-consuming. Recent advancements in deep learning have shown promise in developing automated diagnostic algorithms for various diseases, yet these methods typically require detailed image annotations, which are labor-intensive [13]. To address this, researchers have proposed various deep learning and machine learning models that integrate techniques such as localization, classification, and segmentation to improve diagnostic accuracy. Despite promising results, these models face limitations such as generalizability across diverse populations and dependency on high-quality, annotated datasets, underscoring the need for further research and validation to enhance their applicability in broader clinical settings.

IV. Proposed Quantum Key Distribution (QKD) Integration for Secure Data Transmission in Cloud Computing Environments

Fig. 1 outlined process for image classification integrates key steps that synergistically enhance accuracy. Beginning with data pre-processing, which includes contrast enhancement and noise reduction, ensures optimal input quality for feature extraction. Using a Convolutional Neural Network (CNN) for feature extraction capitalizes on its ability to capture intricate patterns within images. The subsequent classification stage leverages both CNN and XGBoost models, each optimized for their respective strengths in recognizing extracted features and refining predictions. Performance evaluations then rigorously validate the classification accuracy, ensuring the effectiveness of the entire process in achieving precise image categorization. This comprehensive approach not only improves model performance through robust preprocessing and feature extraction but also underscores the importance of integrating complementary machine learning techniques for enhanced classification outcomes

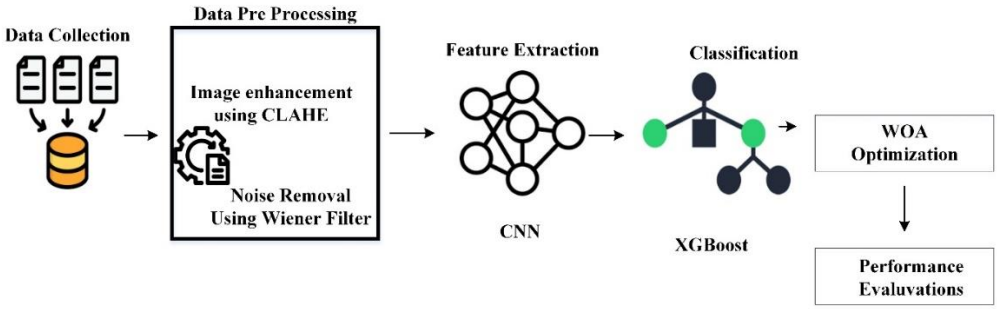


Fig. 1. The conceptual diagram of the proposed model.

A. Dataset Collection

Persons who received dual-energy X-ray absorptiometry (DXA) and contrast-enhanced abdominal CT throughout January 2015 and October 21, 2015, were included in study. 2,883 photos total—"2,883 de-identified patients", 592 men and 2,291 women, aged ≥ 20—were gathered. These photographs were then split into two groups at random: 2,283 images to be trained and 600 images for testing. Additionally, we split the experimental datasets in an 8:2 ratio across the validationu and training datasets. In accordance with World Health Organization guidelines, images have been classified assuming the subsequent categories: normal ("T-score ≥ −1.0"), osteopenia (−2.5 < T-score < −1.0), and osteoporosis (T-score ≤ −2.5) ..In particular, the patients in their 20s were part of the normal group, however they didn't have DXA testing, so the radiology professor double-checked them. Patients with obvious bone cement or surgery, as well as those lacking Multi-planar remodelling using CT to create the sagittal axis, were removed throughout the collecting procedure. Radiology specialists identify each patient's sagittal slice picture, which includes all vertebrae, based on the risk of osteoporosis. Next, we employ every slice that the specialists have identified. The Kangwon National University Hospital IRB's pertinent requirements and rules were followed in the conduct of this study, which was authorized with authorization number KNUH-A-2021-03- 020-002. Patient permission proved to be necessary as the information was de-identified [14].

B. Image Pre-Processing

Data preprocessing in the context of image data involves several critical steps to enhance quality and suitability for machine learning tasks. Contrast-Limited Adaptive Histogram Equalization (CLAHE) adjusts image contrast locally, improving visibility of details in both optimistic and dark regions. Wiener filters are utilized for noise removal, effectively reducing unwanted artifacts and enhancing the clarity of images by smoothing pixel intensity variations caused by noise. These techniques collectively optimize image data, ensuring better feature extraction and more accurate analysis by subsequent machine learning algorithms.

1) Image enhancement using contrast-limited adaptive histogram equalization (CLAHE): The pixel dispersion is shown by the photo histogram. The contrasting qualities of the image can be improved by altering the pixel distribution. A map-based modification of the initial image's grey level called equalization of the histogram can improve the fluctuation in the quantity of grey within every pixel. As a result, the image has greater brightness. An adaptive equalization of histograms (AHE) technique tends to overamplify noises in usually uniform areas of the image. The solution to this issue was suggested to be the CLAHE approach. Divide the image into portions that don't overlap. Typically, an area measure of 8 by 8 is used. Use the value of the threshold to trim your histogram once you have the histogram for each section.

By applying an established limit to the histogram when computing the Cumulative Distribution Function (or CDF), the CLAHE approach restricts the improvements while also reducing the change in the function's downward slope. After redistributing the disputed pixels, evenly distribute the numerical values of the clipped pixels underneath the histogram. Fig. 3 shows the regional equalization of the histogram for each region [20].

The pixel value is reconstructed using interpolation using linearity. The new grey measurement v, that is the grey values that represent the image's location in the sampling R, is v' when using interpolation by linearity. Let R'1, R'2, R'3, and R'4 be the collection sites for surrounding areas.The grey-level mappings for u is gr(u).

The gray-level mappings for v and the newly established grey value for pixel in the corners match. In Eq. (1), the altered gray value is stated.

$$u' = gr_1(u) \tag{1}$$

Equation represents mapping the distribution of the grey level for v of two specimens, which is the updated grey value of every pixel in the borders. Eq. (2),

$$u' = (1-\alpha)gr_1(u) + \alpha gr_2(u) \tag{2}$$

Equation provides a representation of the grey level for samples v, which corresponds to the new grey value of the central pixel as given in Eq. (3),

$$u' = (1-\beta)\big((1-\alpha)gr_1(u) + \alpha gr_2(u)\big) + \beta((1-\alpha)gr_3(u) + \alpha gr_4(u)) \tag{3}$$

Hence, with relation to point R1, the standardized lengths are α and β. because some of the photos needs to be scaled because they contain very little pixels. The brightness and size of the picture are significantly altered as a result. For different acquisition equipment, there are many sets of variables. All pixel densities were adjusted to fall between [-1, 1] in order to

guarantee accurate and noise-free data. The normalizing calculations of Eq. (4) reduced the sensitivity of the model to small weight variations. Eq. (7) provides the Normalization of image INorm as,

$$I_{Norm} = (I - min_i)\left(\frac{2}{max_i - min_i}\right) - 1 \quad (4)$$

Where, min_i and max_i are minimum image and maximum image.

2) Noise removal by wiener filters: Excessive data is removed from the image using a statistical approach. It achieves the optimal trade-off among noise flattening and reversed filtration, whatever decreases noise and blur in the picture the longest [21].

Filter function is given in Eq. (5) as,

$$f(x,y) = \left[\frac{H(x,y)^*}{H(x,y)^2 + \left[\frac{S_n(x,y)}{S_i(x,y)}\right]}\right] G(x,y) \quad (5)$$

Here G(x,y) denotes the deteriorated picture, $H(x,y)$ is the degrading function, $S_n(x,y)$ is the noise power radio spectrum $S_i(x,y)$, and displays the brightness spectrum of the initial image.

C. CNN for Feature Extraction

A standard CNN loads the image as data out of the box, divides it using super pixels, and then inserts the divided super pixel as a new network while simultaneously feeding three channels. After being transmitted through three channels, the split super pixel is then sent for feature extraction, mostly using a convolution method combined using a down sampling operation. Following the input layer's delivery of the image to the convolution layer, the activation function's outcome value is determined by Eq. (6)

$$y^l = \text{f}\,(W^l y^l + b^l) \quad (6)$$

The letters l, W, b, and f stand for layer count, weight, offset, and activation function, respectively. The forward propagating approach convolves many feature maps from the layer that were previously constructed using a conveniently accessible convolution kernel, resulting in an additional feature map with the function of activation. Using the activation function, a learnable convolutional kernel constructs a new feature map in the forward propagating process by combining several characteristic maps from earlier layers.as stated in Eq. (7)

$$y_j^l = f(\sum_{i \in N_j} Y_i^{l-1} * k_{ij}^l + b_j^l) \quad (7)$$

The down sampling algorithm is represented by down in this instance. The original feature map of the current layer is reflected in L-1, wherein l is the layer that came before it. The offset number b_j^l is corresponds to the first feature map of the j previous layer, the first feature map of the current layer, and the y_j^l convolution kernel, in that order. When the down sampling layer is included after the convolution layer, the relative positioning modifications of the goal's tilt and rotation can be disregarded. This improves the method's efficacy and adaptability, shrinks the map of features, and partially prevents over-fitting. The down sampling layer in Eq. (8) is found using the method outlined below.

$$y_j' = \text{f}\,(\beta_j' down(y_j^{l-l}) + b_j') \quad (8)$$

where, down is the representation of the down sampling function. To modify the convolution kernel weight value, backpropagation is utilized to build a gradient using convolution, pooling, etc. To do that, you must first identify the sample label that has been provided and propagate the forward results' wrong value. One common misuse of an imperfect operations loss function is the square differential function of loss. Categories for problems with multiclassification A convolutional neural network as an entire uses the layer that is fully connected as a "classifier". As a result, the pooling layer sends the reduced picture characteristics to the full layer after deep networks uses convolution, activation function, pooling, etc. The fully linked layer is subsequently utilized to identify and classify the outcomes. Eq. (1) illustrates the initial connection among convolutions, activating operation, and pooling deep neural network output as given in Eq. (9)

$$E^N = \frac{1}{2}\sum_{n=1}^{N}\sum_{k=1}^{c}\sqrt{(t_k^n - x_k^n)} \quad (9)$$

This indicates the n measurement mistake of the k sample & f the total error of the N samples, wherein E^N is the k sample's n dimensional output. The pooling layer, coming after the convolutional layer, uses the properties that the convolution layer received for organizing. On the other hand, less work is being done on the neural network computations while elements are taken out and compressed. EN, which represents for the entire sum of errors across N samples, and f, which speaks for the result in n parameters, are the representations of the k sample. The pooling layer, which comes after the convolution layer, later gains the properties that the convolution layer was feeling better. The key trait lies in the typical decrease in size. Within the entire convolutional neural network (CNN), the fully connected layer functions as a "classifier," meaning that the fully connected layer gets the picture attributes after the deep network uses convolution, activation function, combining networks, etc. to minimize them [22]. Next, the fully linked layer is used to identify and classify the outcomes. combining the results of deep networks. Convolution and the activation functional are originally related.

D. XGBoost Model for Classification

Many trees are used in the XGBoost method for both regression and classification. Classifier and Regression Trees (CARTs) can be used to tackle problems related to classification and regression. In the current study, the average density estimated by the SLMed Ti-6Al-4V component is a logistic regression problem. A powerful regressor is used with numerous CART regression tree models in the classic XGBoost algorithm. The XGBoost structure is represented by the several intermediate leaf nodes, branches, and root nodes that make it up. For arriving at the first judgments, this framework provides the input, x_i the i-th parameters, via each root node of the CARTs.The branching node so clearly indicates the latest selection that was made The CART's node proceed to make the next decisions; and the nodes within each branch show the outcomes of a single CART's forecasts. Ultimately, the

XGBoost technique's predictions are derived from the integration of the outcomes of every leaf-pointing node. In the i-th set (a_i, b_i): b_i for example, the XGBoost tree of regression model may be expressed mathematically in the following manner a_i is the data being used that contains several characteristics, and is the actual outcome of the trial.

$$b_i=\alpha \sum_{k\prime=1}^{k\prime} f'_k(a_i) \quad (10)$$

wherein α is the expected learning rate of each element in the algorithm's regression tree, K is the anticipated quantity correlating to input a_i,, and f'_k is the result of the k-th predictions trees.

The anticipated score f'_k is the total of all standards, as demonstrated by Eq. (10)

b_i is the expected value for input a_i, α is the coefficient that represents the regression tree's algorithm's estimated rates of learning for each and every component, and $f'_k$is the result of the k-th regression tree model. The total amount of CARTs used is denoted by K'. The predicted score $f'_k$is the total of each of the requirements, as Eq. (1) illustrates.

The average of all requirements, as shown by Eq.(7), in which is the estimated rate of learning of each and each components in the algorithm's regression tree, b_(i) is the expected result matching to inputa_i,, and f'_k is the result of the k-th regress trees. The total quantity of CARTs used is denoted by K'. The predicted f'_k score is the total of all the standards.

Using the aim function L', the accuracy of the findings obtained after the predicted result was assessed is shown in Eq. (11)

$$L'=\sum_i^n l(b_i, b'_i)+\sum_{k\prime=1}^{k\prime} \Omega\ (f'_a) \quad (11)$$

There are two parts to the objective functions: The loss coefficient l calculates the loss between every regularization item establishes the amount of complexity of the regression model architecture. In reference to a CART, Ω was described in Eq. (12)

$$\Omega(f')=\Upsilon\ T+\frac{1}{2}\lambda\sum_{J=1}^{T} w'^2_J \quad (12)$$

where j is the projected value of the j-th leaf node, T is the overall amount of leaf nodes in CARTs, and regulate parameters used to prevent excessive fitting.

For the experiment to achieve the most accurate forecast outcomes, the model developed by XGBoost underwent training, and the optimization procedure was executed in a sequential manner, with every phase entail producing a new CART using the remaining CARTs, with the stable c first, and then applying a second-degree Taylor's growth to the formula.

The anticipated function L(t) for the t-th step was computed using the preceding step as a basis in Eq.(13)

$$L^{(t)}=\sum_i^n(l(b_i, {b'_i}^{(t-1)} +g'_i f_t(a_i)+\frac{1}{2} h_i {f_t}^2\ (a_i)))+ b_i(f_t)+c \quad (13)$$

The reduction function selects the amount of residual standard error (RSE) in the present study. It translated each of the input variables, a_i Because each input variable, a_i was allocated to a CART's leaf nodes $f_k(a_i)$was defined as follows in this study, where the loss functions selects the standard error of the residual (RSE) is shown in Eq.(14)

$$f_k(a_i) = \omega_q(a_i),\ \omega\epsilon R'^T, q(a_i): R'^T \rightarrow \{1,2,\ldots\ldots,T\} \quad (14)$$

wherein d is the value of the input, a_i is the significance for this particular leaf node, and q(a_i): is the position of a particular leaf node. A T-dimensional vectors is represented by R'^T, whereas a d-dimensional vectors by Eq.(15) and it was written as Eq. (16)

$$G'_J=\sum_{i\prime\epsilon I'_j} g'_i \text{ and } H_J==\sum_{i\epsilon I'_j} h'_i, \text{when } \omega_J=-\frac{G'_J}{H_J+\lambda},\ L'_{min} \quad (15)$$

$$L'_{min}=\frac{1}{2}\sum_{J=1}^{T}\frac{{G_J}^2}{H_J+\lambda}+\Upsilon\ T+c \quad (16)$$

Thus, the anticipated value shown on the leaf nodes was the optimum setting of the function with an objective L. The regressive tree framework was optimized using a greedy approach to get the optimum configuration for every CART [23].

Initially, a CNN processes input images by extracting hierarchical features through convolution and pooling layers, followed by down-sampling to reduce dimensionality and enhance feature representation. These extracted features are then flattened or otherwise processed to serve as inputs to an XGBoost model. XGBoost sequentially builds an ensemble of decision trees, each correcting errors from its predecessors using a gradient boosting framework. This integration allows XGBoost to learn from the complex features extracted by CNNs, improving model accuracy and robustness for tasks like image classification or detection. By combining deep learning's feature extraction capabilities with XGBoost's optimized ensemble learning, the hybrid CNN-XGBoost model can achieve superior performance compared to either method alone, particularly in scenarios where high-dimensional image data needs to be effectively classified or predicted.

E. The Woodpecker Optimization Algorithm

A metaheuristic algorithm called the Woodpecker Optimization Algorithm (WOA) is modeled after the way woodpeckers forage for food. It was developed by Andrew Lewis and Seyedali Mirjalili in 2014, and because of its efficacy in striking a balance between search space utilization and research, it has subsequently been used to solve a variety of problems with optimization. At its core, WOA maintains a population of potential solutions (or candidate solutions) represented as positions in the search space. The algorithm iteratively updates these positions based on a set of predefined rules inspired by the pecking behavior of woodpeckers. Here's a succinct explanation of the algorithm:

1) Initialization: In this phase, the algorithm begins with an initial population of candidate solutions. These solutions are randomly generated within the feasible region of the problem space. This step ensures that the algorithm starts with a diverse set of potential solutions to explore.

2) Objective function evaluation: Once the initial solutions are defined, the algorithm evaluates each solution's fitness or quality by computing the objective function value associated with each solution. The performance of every solution in

relation to the objectives of the optimization problem is quantified by an objective functions.

3) Update positions: During each iteration, two main operations are performed to update the positions of the solutions:

a) Exploration phase: Randomly select one solution (e.g., the leader) and adjust the positions of all solutions towards it to encourage exploration of the search space. This is done using the following equation for updating the position $X_{i(t+1)}=X_{rand} - A.D$

Where X_{rand} is a randomly selected solution, A is a coefficient that controls the step size, and D is a vector representing the distance between the current solution X_i and X_{rand}

Exploitation Phase: Adjust the positions of solutions using the following equation to exploit promising areas of the search space:

$$X_{i(t+1)}=X_{best} - A.|C.\ X_{best} - X_I\ |$$

where, X_{best} is the best solution found so far, C is a random coefficient, and |.| denotes element-wise absolute difference.

4) Boundary constraints handling: Throughout the algorithm's execution, it's crucial to ensure that the solutions generated and updated adhere to any constraints defined by the problem. If a solution violates these constraints after an update, adjustments are made to bring it back within the feasible region of the search space.

5) Update parameters: Parameters such as the step size coefficient AAA are dynamically adjusted over iterations. This adjustment helps in striking a balance between exploration (discovering new solutions) and exploitation (refining existing solutions), thereby enhancing the algorithm's effectiveness in finding optimal or near-optimal solutions.

6) Termination: Up until the termination requirement is satisfied, an algorithm repeatedly proceeds through the update stages. A certain amount of cycles, a suitable degree of solution quality, or a minimal progress over subsequent rounds constitute standard ending conditions.

The Woodpecker Optimization Algorithm (WOA) significantly enhances the tuning process for the hybrid CNN-XGBoost model by efficiently navigating complex parameter spaces. Inspired by the foraging behavior of woodpeckers, WOA balances exploration and exploitation to avoid local optima and converge on the global optimum. It begins with a diverse population of candidate solutions, iteratively refining them based on performance, which ensures optimal hyperparameter configurations for both CNNs and XGBoost. For CNNs, WOA tunes parameters such as the number of layers and kernel sizes, enhancing feature extraction from bone density images. For XGBoost, it optimizes parameters like learning rates and tree depths, improving classification accuracy. Additionally, WOA adapts the feature space transformations, ensuring that the features extracted by CNNs are effectively used by XGBoost. By incorporating performance metrics such as recall, precision, and F1-score into the optimization process, WOA directly improves model performance on real-world tasks. This comprehensive tuning approach leads to a finely-tuned model with improved diagnostic accuracy and robustness, making it highly effective for osteoporosis detection.

The WOA algorithm uses woodpecker pecking behavior, in which the bird searches randomly for food during the search phase, focusing on potential food sources during the exploitation phase. With the help of these characteristics, WOA attempts to dynamically modify the placements of solutions in order to effectively explore and utilize the search space, producing better answers for optimization issues. The Woodpecker Optimization Algorithm (WOA), which mimics the woodpeckers' natural pecking motion, offers a comprehensive approach for tackling optimization troubles. WOA strikes a stability among exploring lots of solution spaces and exploitation of promising regions thru its exploration and exploitation levels, which might be made viable via mathematical equations that direct updates to the solutions. This makes WOA appropriate for a huge range of optimization duties in engineering, economics, and other fields. Fig. 2 represents Optimized CNN-XGBoost. Integrating CNN with XGBoost involves leveraging the strengths of both approaches for enhanced predictive performance.

WOA *Algorithm*

Initialize:
Generate initial population of solutions X within the feasible region
Evaluate objective function values for each solution in X

Repeat until termination criterion is met:
Update Positions:
For each solution X_i in population X:
Randomly select a solution X_{rand} from X
Update position for exploration:
$X_{new} = X_{rand} - A * D$
Where D is a vector representing distance between X_i and X_{rand}
Update position for exploitation:
X_{best} =find best solution(X)
C = random coefficient()
$X_{new} = X_{best} - A*abs(c*X_{best} - X_i)$
Apply boundary constraints to X_new to ensure feasibility
Evaluate fitness of X_new using objective function
Replace X_i with X_{new} if X_{new} is better (based on fitness)
Update algorithm parameters
Terminate when a stopping criterion is met

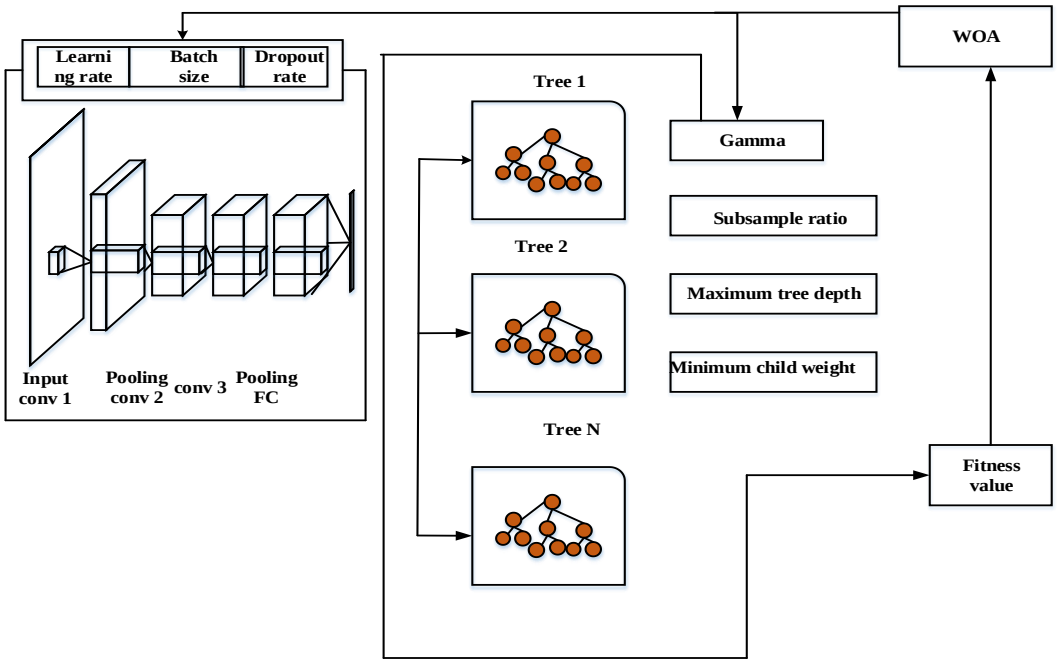


Fig. 2. Proposed figure.

V. Results and Discussion

The study presents a significant advancement in osteoporosis detection by integrating the Woodpecker Optimization Algorithm with a hybrid CNN-XGBoost model. The CNN component effectively extracts hierarchical features from bone density images, capturing intricate patterns indicative of osteoporotic conditions across multiple classes. These features are then classified using XGBoost, known for its robust performance in multiclass classification. Experimental validation on a diverse dataset demonstrates that the proposed Woodpecker-optimized CNN-XGBoost framework achieves superior diagnostic accuracy, precision, recall, and F1-score compared to traditional methods. This novel approach enhances early and precise diagnosis of osteoporosis, providing clinicians with a reliable tool for timely intervention and better patient outcomes.

A. Training and Testing Accuracy

Fig. 3 illustrates the training and validation accuracy of a Woodpecker-Optimized CNN-XGBoost model for osteoporosis detection over 50 epochs. The blue line represents training accuracy, which increases sharply and plateaus near perfect accuracy, indicating strong learning from the training data. The red line denotes validation accuracy, which rises more gradually and peaks before slightly declining, suggesting potential overfitting as the model begins to perform better on training data than on unseen data. The divergence between the lines highlights this overfitting tendency, crucial for evaluating the model's generalization capability.

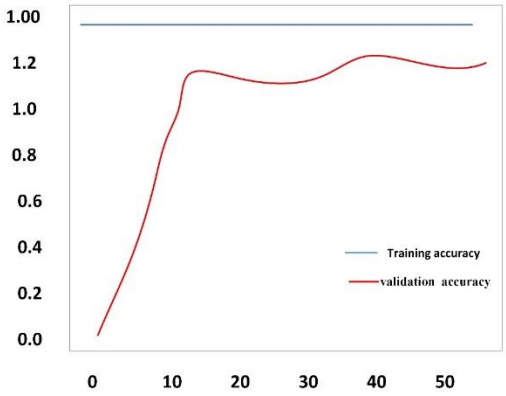


Fig. 3. Training and testing accuracy.

B. Training and Testing Loss

Fig. 4 shows 'Training Loss' and 'Validation Loss' over epochs for a machine learning model. “Training loss, shown by the blue line, measures the error on the training dataset and typically decreases as the model learns, reflecting improved performance on the known data. Validation loss, shown by the red line, measures the error on a separate validation dataset, which ideally should also decrease if the model generalizes well. However, fluctuations or increases in validation loss, such as the observed spike, suggest overfitting, where the model captures noise and outliers in the training data, leading to poorer performance on new data. Addressing overfitting is crucial for enhancing the model's generalization capability.

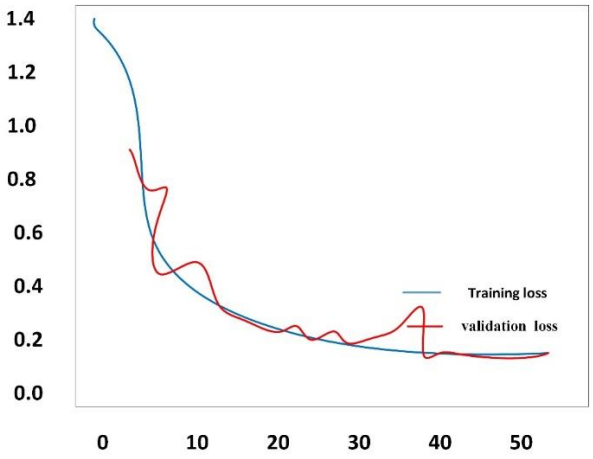


Fig. 4. Training and testing loss.

C. ROC Curve

Fig. 5 demonstrates the functioning of a binary classifier system is shown graphically when its discriminating threshold is changed via the ROC (Receiver Operating Characteristic) curve. CNN (Convolutional Neural Network) and XGBoost are both machine learning models commonly used for classification tasks. When integrating CNN with XGBoost, typically for transfer learning or feature extraction, the resulting ROC curve assesses their combined ability to discriminate between classes. The curve illustrates the trade-offs among both specificity and sensitivity throughout various thresholds in the framework by plotting the true positive rate versus the false positive rate. Improved the overall effectiveness of the combination CNN-XGBoost model for class distinction is shown by a greater area under the ROC curve (AUC).

Fig. 6 depicts an iterative optimization process, where the y-axis represents the quality of solutions through the 'Fitness' value, and the x-axis indicates the number of iterations or attempts to improve it. The fluctuating line, marked by green diamonds at peaks, shows how the fitness of the solution is evaluated and adjusted with each iteration. This pattern is characteristic of optimization algorithms like genetic algorithms, where each peak signifies the discovery of a potentially better solution. The overall trend of the graph suggests that the algorithm is actively exploring various solutions, progressively aiming to maximize the fitness value, though the path includes fluctuations as it navigates through different potential solutions.

D. Performance Assessment

Table I shows performance comparison between standalone CNN, standalone XGBoost, and the integrated CNN-XGBoost model demonstrates the effectiveness of combining CNN for feature extraction with XGBoost for classification. The standalone CNN achieved an accuracy of 88.5%, with a precision of 86.2%, recall of 85.9%, F1-score of 86.0%, and ROC-AUC of 89.7%. Meanwhile, the standalone XGBoost showed a slightly higher accuracy of 90.3%, but lower recall (82.5%) and F1-score (82.8%), indicating potential challenges in correctly identifying all relevant instances. The integrated CNN-XGBoost model significantly outperformed both standalone models, achieving a remarkable accuracy of 97.1%, with precision, recall, F1-score, and ROC-AUC values of 89.2%, 88.8%, 89.0%, and 92.3%, respectively. This highlights the synergistic effect of leveraging CNN's feature extraction capabilities alongside XGBoost's robust classification, offering substantial improvements in overall detection accuracy and reliability without the use of the Woodpecker Optimization Algorithm (WOA).

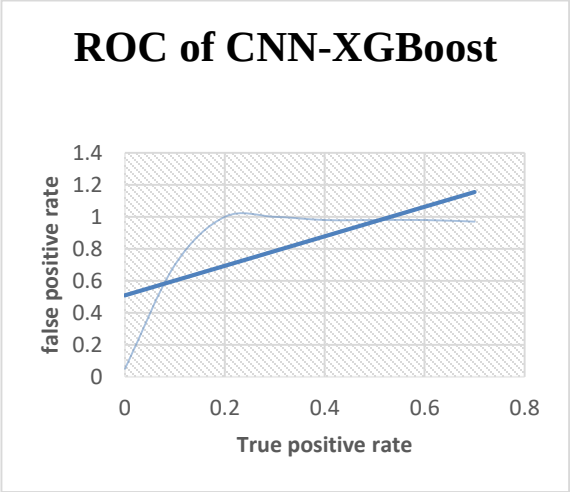


Fig. 5. ROC.

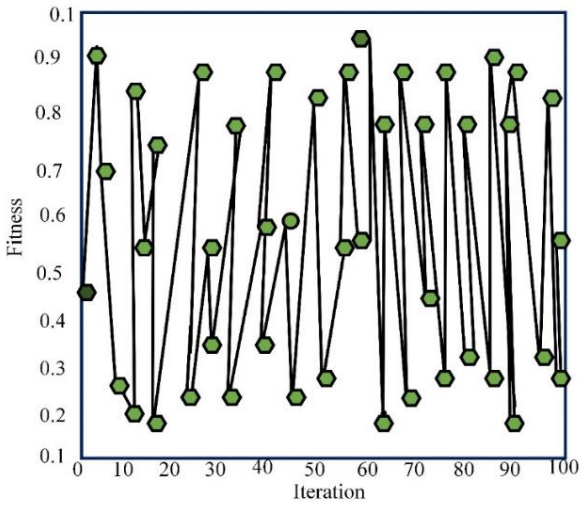


Fig. 6. Fitness improvement over iterations (WOA).

TABLE I. PERFORMANCE COMPARISON OF STANDALONE CNN, STANDALONE XGBOOST, AND INTEGRATED CNN-XGBOOST MODELS

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	ROC-AUC (%)
Standalone CNN	88.5	86.2	85.9	86.0	89.7
Standalone XGBoost	90.3	87.1	82.5	82.8	87.1
Integrated CNN-XGBoost (without WOA)	97.1	89.2	88.8	89.0	97

In comparing the performance of standalone CNN, standalone XGBoost, and the integrated CNN-XGBoost model, several key insights emerge is given in Table I and Fig. 7. Standalone XGBoost, while achieving the highest accuracy of 90.3%, has lower recall (82.5%) and F1-score (82.8%) compared to the integrated model. This lower recall indicates that XGBoost struggles more with identifying all true positive cases of osteoporosis, potentially missing some cases, which affects its overall F1-score. On the other hand, the standalone CNN shows slightly lower accuracy (88.5%) but higher recall (85.9%) and F1-score (86.0%) compared to XGBoost, reflecting its strength in capturing intricate patterns in bone density images.

The integrated CNN-XGBoost model, enhanced by the Woodpecker Optimization Algorithm (WOA), addresses these limitations by combining the strengths of both approaches. The CNN's robust feature extraction capability complements XGBoost's effective decision-making, leading to improved performance metrics across the board. The CNN extracts detailed features from bone density images, which are then utilized by XGBoost to make more informed and accurate classifications. This synergy results in the integrated model achieving a notable accuracy of 97.1%, with higher precision (89.2%), recall (88.8%), and F1-score (89.0%), and a ROC-AUC of 97%. Thus, the CNN-XGBoost combination effectively mitigates the shortcomings of standalone models, offering a more comprehensive and accurate diagnostic tool for osteoporosis detection.

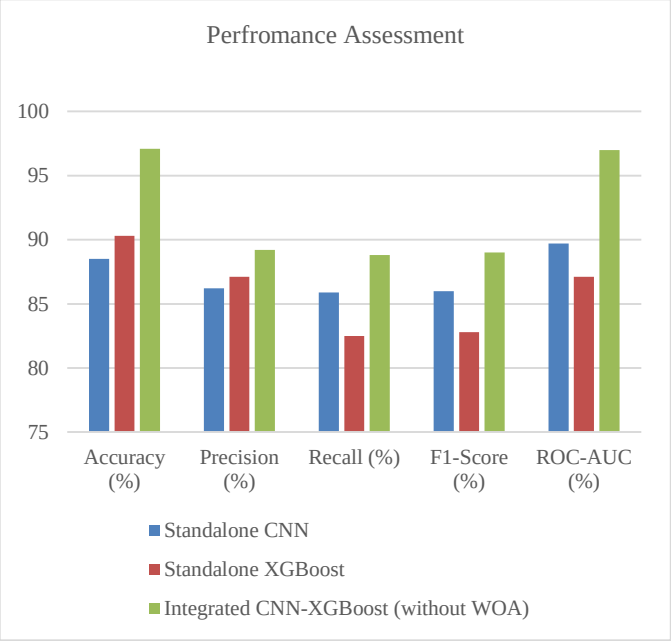


Fig. 7. Performance assessment of the suggested method.

E. Discussion

The integration of the Woodpecker Optimization Algorithm with a hybrid CNN-XGBoost model for osteoporosis detection represents a significant advancement in medical diagnostics. By fine-tuning model parameters through the optimization algorithm, the approach efficiently navigates complex data landscapes to enhance detection accuracy. The CNN component excels in extracting detailed, hierarchical features from bone density images, crucial for identifying subtle osteoporotic patterns across multiple classes [19]. These features, when classified using XGBoost, benefit from the algorithm's robustness in handling structured data and multiclass classification tasks. Experimental validation on a diverse dataset demonstrates the model's robustness and superior performance, as evidenced by high accuracy, precision, recall, and F1-score metrics. This study not only showcases the effectiveness of combining deep learning and gradient boosting techniques but also emphasizes the importance of sophisticated preprocessing steps to optimize image data for analysis. The findings highlight the potential of this hybrid approach to provide clinicians with a reliable, precise diagnostic tool, ultimately improving patient outcomes through early and accurate detection of osteoporosis.

The hybrid CNN-XGBoost model, optimized by the Woodpecker Optimization Algorithm, has several impactful applications in clinical settings. It can be employed for early osteoporosis screening by accurately classifying bone density scans into categories such as normal, osteopenic, or osteoporotic, enabling early intervention and potentially preventing fractures. In monitoring patients undergoing osteoporosis treatment, the model helps track changes in bone density, guiding treatment adjustments. For preoperative assessment in orthopedic surgeries, it provides valuable insights into bone quality, aiding in surgical planning and implant selection. Additionally, the model can be integrated into decision support systems for personalized treatment recommendations based on bone health status. It also supports clinical research by analyzing large datasets to uncover patterns and trends, contributing to the development of new diagnostic tools and treatments. Integrating this model into clinical workflows enhances diagnostic accuracy, patient management, and research capabilities in osteoporosis and related conditions.

This study has several limitations that should be acknowledged. Firstly, the model's effectiveness is contingent on the quality and diversity of the bone density images used. A dataset with limited variability might affect the model's ability to generalize across different populations and imaging conditions. Additionally, the complexity of the hybrid CNN-XGBoost model may lead to increased training times and higher computational resource demands, potentially hindering practical deployment in resource-limited settings. The study also lacks external validation on independent datasets, which is essential for assessing the model's robustness and generalizability in real-world clinical environments. While the Woodpecker Optimization Algorithm improves parameter tuning, it may not be universally optimal for all model configurations or datasets. Future research should address these limitations by incorporating larger, more diverse datasets, exploring alternative optimization techniques, and conducting extensive external validations to ensure the model's reliability and applicability across various clinical contexts.

VI. Conclusion and Future Work

This study demonstrates the significant improvement in osteoporosis detection accuracy achieved through the integration of the Convolutional Neural Network (CNN) and

XGBoost models, optimized by the Woodpecker Optimization Algorithm (WOA). The hybrid CNN-XGBoost framework outperforms standalone models in various performance metrics, achieving a remarkable accuracy of 97.1%, with enhanced precision, recall, F1-score, and ROC-AUC. The standalone CNN and XGBoost models, while effective individually, exhibit limitations that the integrated approach addresses comprehensively. Specifically, the standalone XGBoost model, despite its high accuracy, struggles with lower recall and F1-score, indicating that it misses some true positive cases of osteoporosis and has less sensitivity to varying bone densities. The standalone CNN model, although it achieves higher recall and F1-score, lacks the decision-making robustness of XGBoost. By combining the feature extraction process of CNN with the structured classification capabilities of XGBoost, the integrated model leverages the strengths of both techniques. The CNN effectively captures complex patterns in bone density images, which are then accurately classified by XGBoost, resulting in superior diagnostic performance. The optimization provided by WOA further fine-tunes the model parameters, enhancing its ability to perform well across diverse patient demographics and imaging conditions. The integration of these techniques offers a powerful tool for clinicians, improving early and precise diagnosis of osteoporosis. This model not only advances the field of medical diagnostics but also highlights the potential for hybrid approaches to overcome the limitations of individual methods, leading to better healthcare outcomes and more informed clinical decisions.

Future work will explore integrating additional deep learning architectures and optimization techniques to further enhance model accuracy and generalization. Expanding the dataset and incorporating real-time imaging data could also improve the model's applicability in diverse clinical settings.

REFERENCES

[1] J. Zhang et al., "Exploring deep learning radiomics for classifying osteoporotic vertebral fractures in X-ray images," Frontiers in Endocrinology, vol. 15, p. 1370838, 2024.

[2] R. Anantharaman, A. Bhandary, R. Nandakumar, R. R. Kumar, and P. Vajapeyam, "Utilizing Deep Learning to Opportunistically Screen for Osteoporosis from Dental Panoramic Radiographs," in 2022 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), IEEE, 2022, pp. 2969–2976.

[3] M. Alnaggar, M. Handosa, T. Medhat, and M. Z Rashad, "Thyroid disease multi-class classification based on optimized gradient boosting model," Egyptian Journal of Artificial Intelligence, vol. 2, no. 1, pp. 1–14, 2023.

[4] F. Xu et al., "Deep learning-based artificial intelligence model for classification of vertebral compression fractures: A multicenter diagnostic study," Frontiers in Endocrinology, vol. 14, p. 1025749, 2023.

[5] J. Oh, B. Kim, G. Oh, Y. Hwangbo, and J. C. Ye, "End-to-End Semi-Supervised Opportunistic Osteoporosis Screening Using Computed Tomography," Endocrinology and Metabolism (Seoul, Korea), 2024.

[6] H. Zhang et al., "Screening for osteoporosis based on IQon spectral CT virtual low monoenergetic images: Comparison with conventional 120 kVp images," Heliyon, vol. 9, no. 10, 2023.

[7] S. Ramkumar, M. R. Kumar, and G. Sasi, "Programmed for Automatic Bone Disorder Clustering Based on Cumulative Calcium Prediction for Feature Extraction.," Clinical Laboratory, vol. 68, no. 8, 2022.

[8] N. Dagan et al., "Automated opportunistic osteoporotic fracture risk assessment using computed tomography scans to aid in FRAX underutilization," Nature medicine, vol. 26, no. 1, pp. 77–82, 2020.

[9] E. R. Astuti, A. Z. Arifin, R. Indraswari, R. H. Putra, N. F. Ramadhani, and B. Pramatika, "Computer-aided system of the mandibular cortical bone porosity assessment on digital panoramic radiographs," European Journal of Dentistry, vol. 17, no. 02, pp. 464–471, 2023.

[10] C. Wang, Y. Liang, and G. Tan, "Periodic residual learning for crowd flow forecasting," in Proceedings of the 30th International Conference on Advances in Geographic Information Systems, Seattle Washington: ACM, Nov. 2022, pp. 1–10. doi: 10.1145/3557915.3560947.

[11] M. Liu et al., "Predicting fracture risk for elderly osteoporosis patients by hybrid machine learning model," Digital Health, vol. 10, p. 20552076241257456, 2024.

[12] C. Sethi, S. Singh, P. S. Chauhan, and B. Raj, "Bone Fracture Detection Using Machine Learning," in Distributed Intelligent Circuits and Systems, World Scientific, 2024, pp. 77–109.

[13] K. Zhang et al., "End to End Multitask Joint Learning Model for Osteoporosis Classification in CT Images," Computational Intelligence and Neuroscience, vol. 2023, pp. 1–18, Mar. 2023, doi: 10.1155/2023/3018320.

[14] D. H. Hwang, S. H. Bak, T.-J. Ha, Y. Kim, W. J. Kim, and H.-S. Choi, "Multi-View Computed Tomography Network for Osteoporosis Classification," IEEE Access, vol. 11, pp. 22297–22306, 2023, doi: 10.1109/ACCESS.2023.3252361.

[15] G. Cuaya-Simbro, A.-I. Perez-Sanpablo, E.-F. Morales, I. Quiñones Uriostegui, and L. Nuñez-Carrera, "Comparing machine learning methods to improve fall risk detection in elderly with osteoporosis from balance data," Journal of healthcare engineering, vol. 2021, no. 1, p. 8697805, 2021.

[16] R. Widyaningrum, E. I. Sela, R. Pulungan, and A. Septiarini, "Automatic segmentation of periapical radiograph using color histogram and machine learning for osteoporosis detection," International Journal of Dentistry, vol. 2023, no. 1, p. 6662911, 2023.

[17] K. Thawnashom, P. Pornsawad, and B. Makond, "Machine learning's performance in classifying postmenopausal osteoporosis Thai patients," Intelligence-Based Medicine, vol. 7, p. 100099, 2023.

[18] S. M. Ryu et al., "Diagnosis of osteoporotic vertebral compression fractures and fracture level detection using multitask learning with U-Net in lumbar spine lateral radiographs," Computational and Structural Biotechnology Journal, vol. 21, pp. 3452–3458, 2023.

[19] Z. Chen et al., "Osteoporosis diagnosis based on ultrasound radio frequency signal via multi-channel convolutional neural network," in 2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), IEEE, 2021, pp. 832–835.

[20] R. Fan, X. Li, S. Lee, T. Li, and H. L. Zhang, "Smart Image Enhancement Using CLAHE Based on an F-Shift Transformation during Decompression," Electronics, vol. 9, no. 9, p. 1374, Aug. 2020, doi: 10.3390/electronics9091374.

[21] M. Dhanushree, R. Priyadharsini, and T. Sree Sharmila, "Acoustic image denoising using various spatial filtering techniques," Int. j. inf. tecnol., vol. 11, no. 4, pp. 659–665, Dec. 2019, doi: 10.1007/s41870-018-0272-3.

[22] F. Yang et al., "Deep Learning for Smartphone-Based Malaria Parasite Detection in Thick Blood Smears," IEEE J. Biomed. Health Inform., vol. 24, no. 5, pp. 1427–1438, May 2020, doi: 10.1109/JBHI.2019.2939121.

[23] M. Zou, W.-G. Jiang, Q.-H. Qin, Y.-C. Liu, and M.-L. Li, "Optimized XGBoost model with small dataset for predicting relative density of Ti-6Al-4V parts manufactured by selective laser melting," Materials, vol. 15, no. 15, p. 5298, 2022.

Attention-Based Joint Learning for Intent Detection and Slot Filling Using Bidirectional Long Short-Term Memory and Convolutional Neural Networks

(AJLISBC)

Yusuf Idris Muhammad, Naomie Salim, Sharin Hazlin Huspi, Anazida Zainal
Faculty of Computing, Universiti Teknologi Malaysia, Skudai 81310, Malaysia

***Abstract*—Effective natural language understanding is crucial for dialogue systems, requiring precise intent detection and slot filling to facilitate interactions. Traditionally, these subtasks have been addressed separately, but their interconnection suggests that joint solutions yield better results. Recent neural network-based approaches have shown significant performance in joint intent detection and slot filling tasks. The two primary neural network structures used are recurrent neural networks (RNNs) and convolutional neural networks (CNNs). RNNs capture long-term dependencies and store previous information semantics in a fixed-size vector, but their ability to extract global semantics is limited. CNNs can capture n-gram features using convolutional filters, but their performance is constrained by filter width. To leverage the strengths and mitigate the weaknesses of both networks, this paper proposes an attention-based joint learning classification for intent detection and slot filling using BiLSTM and CNNs (AJLISBC). The BiLSTM encodes input sequences in both forward and backward directions, producing high-dimensional representations. It applies scalar and vectorial attention to obtain multichannel representations, with scalar attention calculating word-level importance and vectorial attention assessing feature-level importance. For classification, AJLISBC employs a CNN structure to capture word relations in the representations generated by the attention mechanism, effectively extracting n-gram features. Experimental results on the benchmark Airline Travel Information System (ATIS) dataset demonstrate that AJLISBC outperforms state-of-the-art methods.**

***Keywords*—Joint learning; intent detection; slot filling; multichannel**

I. Introduction

Owing to the integration of conversational agents into various applications, from virtual assistants to customer chatbots, the importance of accurately interpreting user inputs increases. In the field of Natural Language Understanding (NLU), two primary tasks are intent detection and slot filling [1, 2]. Intent detection is a classification problem involving the construction of features from a given utterance. These features are then subjected to a classification algorithm to predict the appropriate classes for utterances selected from the predefined classes [3].

Although intent detection is a classification problem, it differs from classical classification in that it addresses the spoken language. Therefore, engineered features must be oriented towards capturing the semantic meanings of these utterances [4]. This emphasis on semantics is crucial to understanding the underlying intent conveyed by the user. Recent approaches have expanded beyond the semantic content of individual words to internal aspects such as syntactic structures, word contextual relationships, and external information such as metadata [5].

Slot filling is a sequence-labeling problem that is used to identify the semantic constituents of a user's utterance and assign a semantic label to each word. The purpose of these labels is to describe the type of semantic information carried by the token, which can help identify the intent of the user [6, 7].

Traditionally, intent detection and slot filling tasks have been treated separately and assembled to form an entire system [8]. This type of methodology provides conceptual clarity, with each component independently addressing its specific challenges. However, there are some limitations in separating these models. It fails to leverage the interaction between the intent detection task and slot filling task, and this interaction plays a role in enhancing the overall system performance [9, 10]. Recent advances in Artificial Intelligence (AI), particularly deep learning, have opened the door to joint models. A joint model handles both intent detection and slot filling simultaneously by leveraging their interdependencies and shared representations to enhance overall performance and efficiency [11].

Encoder-decoder neural network architectures are generally used for the joint learning classification of intent detection and slot filling because of their powerful sequential processing capabilities. Early joint learning approaches were based on statistical models such as Support Vector Machines (SVMs) [12], maximum entropy models (MEM) [13], hidden Markov models (HMM) [14], and Conditional Random Fields (CRFs) [15], which require extensive feature engineering and struggle to capture the deep semantic nuances of language. The advent of deep learning has brought about a shift, enabling models to learn hierarchical representations from raw data. Convolutional Neural Networks (CNNs) [16], Recurrent Neural Networks (RNNs) [17], and transformer architecture [18] have been at the forefront of this revolution, offering powerful tools for sequence modeling.

A Recurrent Neural Network (RNN) is a widely used architecture for Natural Language Processing (NLP) tasks owing to its ability to maintain memory and capture

dependencies and patterns over time. This memory is implemented using recurrent connections within the network, allowing information to persist and update when new inputs are processed [19]. RNNs are particularly effective for intent detection and slot-filling tasks [20]. For example, in intent detection, understanding a user's intent requires considering the sequential nature of the dialogue. RNNs with recurrent connections can capture the context of previous words or phrases in a sentence, thereby enabling them to detect the intent of the user based on the entire input sequence. In slot-filling tasks, the presence and placement of entities within the input text are crucial for extracting slot values accurately [21]. RNNs can learn to recognize patterns in the sequential structure of sentences, allowing them to identify relevant slots based on their contextual relationships with other words or entities in a sentence [22]. Despite their capabilities, RNNs often face gradient vanishing or exploding issues. To address these challenges, Long Short-Term Memory (LSTM) networks and Gated Recurrent Units (GRUs) have been developed to improve memory handling. Another problem with RNNs is that long sentences tend to prioritize recent information over earlier information, which might be more significant.

To address the problem of input element selection, an attention mechanism was introduced to assign different weights to the output of the RNNs. The essence of this mechanism is to allow RNNs to combine outputs according to their assigned importance and retain variable-length memory. Attention mechanisms have proven to be effective in joint learning tasks, where different parts of the input may be relevant to intent detection and slot filling. By assigning different weights to different parts of the input, attention mechanisms help models prioritize the information that is crucial for each task. However, it has limitations in terms of capturing the order of the input sequence, which is crucial for NLP tasks. For instance, the sentences "I want flight from Baltimore to Dallas" and "I want flight from Dallas to Baltimore" will have identical weighted sums despite having opposite meanings.

Convolutional Neural Networks (CNNs) are another architecture for NLP tasks, known for their ability to learn spatial hierarchies and local correlations of features from input data using convolutional filters. For instance, 2-gram features can effectively be extracted from the given example such as "from Baltimore" and "from Dallas" likewise "to Dallas" and "to Baltimore" using CNNs. This type of representation provides better information than the sum of the RNN hidden states in the input sequence. Several studies have demonstrated the importance of CNNs for NLP tasks. In [23], it was demonstrated that a simple CNN consisting of a single convolutional layer applied to word vectors derived from an unsupervised neural language model achieved good performance in text classification. In addition, [24] illustrated that CNNs can be effectively utilized to extract morphological details such as word suffixes or prefixes and encode them into neural representations. However, CNNs are limited in preserving the sequential order [25].

Some researchers have developed hybrid frameworks that combine CNNs and RNNs to exploit their respective strengths. One such framework, the Recurrent Convolutional Neural Network (RCNN), captures contextual information using a recurrent convolutional structure [26]. Another framework, the Convolutional Recurrent Neural Network (CRNN), combines the benefits of both CNNs and RNNs to extract diverse linguistic features [27]. However, these hybrid models often fail to account for the varying semantic contributions of different words when treating all words with equal importance.

Motivated by the abovementioned issues, this study proposes a joint learning classification model that leverages the strengths of RNN and CNN architectures, enhanced with scalar and vectorial attention mechanisms. The major contributions of the proposed model are summarized as follows:

1) The model employs BiLSTM to encode the input sequence in both forward and backward directions, ensuring the retention of chronological features within sequences.

2) An attention mechanism is introduced to generate multiple channels, simulating the diversity of the input information. Scalar attention assesses word-level importance, whereas vectorial attention evaluates feature level significance. This representation allows the model to learn multiple representations of the semantics of an input sequence.

3) CNN is utilized to identify word relations using attention mechanisms rather than relying on weighted sum calculations. This approach enhances the ability of CNN to extract n-gram features.

4) A series of experiments conducted on the Airline Travel Information System (ATIS) dataset demonstrated that the proposed approach outperformed baseline methods.

The remainder of this paper is organized as follows: Section II reviews the related work, Section III outlines the methodology, Section IV describes the experimental setup, Section V discusses the experimental results, Section VI provides an in-depth analysis of the findings, and Section VII concludes the paper.

II. Related Work

Joint learning for intent detection and slot filling has evolved from classical models, such as triangular-chain CRFs [28] and Maximum Entropy Models (MEM) combined with CRFs [29], to capture the dependencies between the intent and slots of an utterance. However, they face scalability and manual feature engineering challenges [5].

Deep learning approaches have emerged as more scalable alternatives. Recently, attention-based joint learning for intent detection and slot filling has gained popularity owing to its ability to enhance task performance by improving feature extraction and the flow of information between these two interdependent tasks. By focusing on the most relevant parts of the input sequences, attention mechanisms allow models to capture fine-grained contextual relationships, making them highly effective for natural language understanding.

The research in [30] introduced an asynchronous joint extraction algorithm that combines a GRU network with a TextCNN-based feature representation layer. Their model incorporated a keyword attention mechanism to capture contextual semantics precisely, enhancing both intent detection and slot filling. Adding adversarial training further strengthens

the robustness of the model against adversarial attacks, thereby improving its reliability in real-world scenarios.

The study in [31] proposed a joint model leveraging graph neural networks (GNNs) fused with external knowledge and a graph attention mechanism. This model significantly enhances the semantic representation by facilitating the exchange of information between slots and intents, resulting in superior task performance. Similarly, [32] emphasized bidirectional information flow within GNN-based models, improving information exchange and interaction between intent recognition and slot labeling processes.

Using a different approach, [11] developed the JPIS model, which integrates user-specific profile information along with a slot-to-intent attention mechanism. This approach proved highly effective in scenarios where profile-based customization was required, substantially improving accuracy.

The study in [8] also explored the efficiency of attention mechanisms by developing a Fast Attention Network tailored to edge devices. This model balances accuracy with latency by utilizing a refined attention module to enhance semantic accuracy, while maintaining fast response times in real-time applications.

Although these models have demonstrated significant advancements in intent detection and slot filling, they highlight the need to balance model complexity with computational efficiency, especially for real-time applications. Enhanced scalar and vectorial attention mechanisms offer a potential solution by allowing the model to capture both the word- and feature-level importance in a more structured manner. Scalar attention assesses the significance of individual words, whereas vectorial attention evaluates feature-level relevance, enabling the model to generate multiple representations of input sequences that can be processed concurrently. This approach enhances the richness of the information captured, thereby improving the overall performance while maintaining the computational efficiency. Thus, the proposed enhanced scalar and vectorial attention mechanisms are justified by their ability to address these existing challenges while optimizing the interaction between intent detection and slot-filling tasks.

III. METHODOLOGY

The architecture of the proposed model is shown in Fig. 1. The proposed model comprises an input layer, a BiLSTM layer, a convolutional layer with subsequent max pooling, and two dense layers that implement softmax functions. These components jointly detect the intent of the user's input utterance and the associated slots by assigning them with multiclass labels (B, I, O), where "I," "O," and "B" signify Inside, Outside, and Beginning of slots, respectively. Details of the model are described in the following subsections.

A. Embedding Layer

First, the dialog must be transformed into a feature vector matrix to serve as the input layer of the model. In the proposed model, Google's word2vec [33] embedding technique is employed to translate each word feature into a word-embedding vector. As a result, dialog vectors are obtained as inputs $X = (x_1, x_2, \ldots, x_n)$.

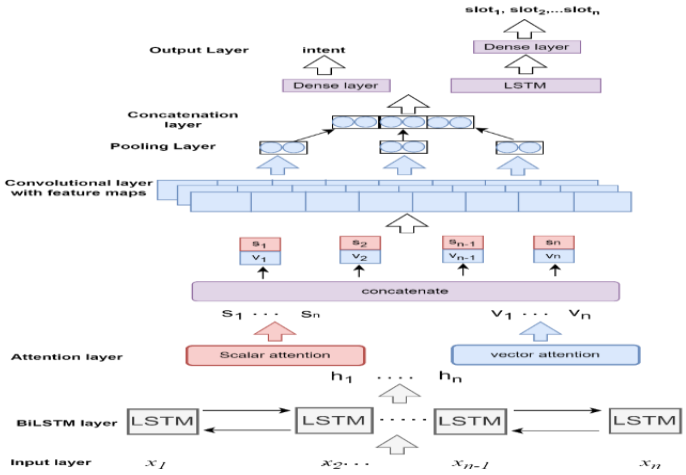


Fig. 1. The architecture of the proposed AJLISBC model.

B. Long Short-Term Memory Network

RNNs are widely used in NLP owing to their ability to handle sequential data and capture temporal dependencies. RNNs process a variable-length sequence at each time step t and updates its hidden state ht based on the current input x_t and previous hidden state h_{t-1}:

$$h_t = f(W[h_{t-1}, x_t] + b) \quad (1)$$

where, W is the weight matrix that combines the hidden states, and the current input vector and b is the bias vector.

However, basic RNNs have been avoided by researchers owing to issues such as the vanishing gradient problem. To address these problems, LSTM networks have been developed and have demonstrated good performance.

To convert a sentence consisting of n words, into a dense vector $x_{i,}$ an embedding matrix is used first. BiLSTM is then applied to generate word annotations by processing the sentence in both forward and backward directions. The forward LSTM process the sequence from x_i to x_n and produce $\overrightarrow{h_t}_i$,

whereas backward LSTM processes the sequence from x_n to x_i and produce $\overleftarrow{h_i}$

$$\overrightarrow{h_i} = LSTM\left(x_i, \overrightarrow{h_{i-1}}\right) \quad (2)$$

$$\overleftarrow{h_i} = LSTM\left(x_i, \overleftarrow{h_{i-1}}\right) \quad (3)$$

$$h_i = \overrightarrow{W_i}.\overrightarrow{h_i} + \overleftarrow{W_i}.\overleftarrow{h_i} + b \quad (4)$$

The hidden states from the forward $\overrightarrow{h_i}_i$ and backward $\overleftarrow{h_i}$ LSTMs are concatenated at each time step to provide a summary of the input sequence h_i.

C. Attention Mechanism

In NLP tasks such as intent detection and slot filling, not all words have the same significance in representing the input sequence. To address this, an attention mechanism was introduced to highlight the importance of each word by assigning greater weights to the crucial elements in the final output. However, the traditional attention mechanism struggles to preserve temporal order information. To resolve this, attention mechanisms are incorporated into the hidden states of BiLSTM, and these states are combined into a matrix that maintains the order information instead of relying on the weighted sum of vectors. In addition, by employing scalar and vectorial attention, multiple matrices were created, serving as multichannel inputs to the CNN.

1) Scalar attention mechanism: To determine the importance weights of all input sequences, scalar attention was employed. This attention is represented by a matrix M which captures the relationships between words in the sequence. The value in the ith row and the jth column of M indicates the level of association between the word in ith and jth column. In each channel L, a mask matrix V is applied, and the masked association matrix Mli, is calculated.

$$M_{l_{i,j}} = \tanh\left(\left[h_i, W_l.h_j\right] + b_l\right) \quad (5)$$

The i^{th} channel mask matrix $V_{l_{i,j}}$ obeys binomial distribution and is defined as:

$$V_{l_{i,j}} \sim B(1,p), i \in [1,n], j \in [1,n] \quad (6)$$

Given association matrix $M_{l_{i,j}}$ and mask matrix $V_{l_{i,j}}$, the channel is calculated as follows:

$$A_l = M_l \otimes V_l \quad (7)$$

$$s_{l_k} = \sum_x A_{l_{xk}} \quad (8)$$

$$p_k = \{^{-99999,}_{0,} \quad ^{if\ x_k\ is\ from\ pad}_{otherwise} \quad (9)$$

$$score_{l_k} = p_k + s_{l_k} \quad (10)$$

$$a_{l_k} = \frac{exp(score_k)}{\sum_i^n exp(score_i)} \quad (11)$$

$$c_{l_i} = a_{l_i}.h_i \quad (12)$$

where, $\otimes, \boldsymbol{c_{l_i}}, \boldsymbol{h_i}, \boldsymbol{l^{th}}, \boldsymbol{p_k}$ denotes element wise multiplication, updated hidden state, channel and padded mask respectively. To ensure that the padding symbol contains nearly zero attention $\boldsymbol{a_{l_k}}$, scalar attention $\boldsymbol{s_{l_k}}$ is subtracted from 99999 before applying the softmax function.

2) Vectorial attention mechanism: In NLP, words and sentences are transformed into *n*-dimensional vector to capture their meanings in a format suitable for computational models. Each dimension within this vector encodes a different aspect of a word or sentence's meaning, allowing for a rich and multifaceted representation of linguistic data. For example, consider the sentence "I want to book a flight from New York to Boston on July 20th". In intent detection task, the model can easily identify the intent as "book a flight" by focusing on dimensions related to booking and travel. In slot-filling tasks, the model can accurately fill slots using dimensions related to locations and dates.

A vectorial attention mechanism was proposed for the joint model based on the above assumptions.

$$score_{l_i} = W_a^T \sigma(W_b.h_i + b) \quad (13)$$

$$a_{vl_i} = \frac{exp\left(score_{l_i}\right)}{\Sigma_i\left(score_{l_i}\right)} \quad (14)$$

$$c_{vl_i} = a_{vl_i} \odot h_i \quad (15)$$

$$C_l = \left[c_{vl_1}, c_{vl_2}, c_{vl_3} \ldots., c_{vl_n}\right] \quad (16)$$

where W_a, W_b are weight matrices in the vectorial attention and b is the bias vector, $\odot$ is the element wise multiplication c_{l_i} denotes the output of h_i in l^{th} channel. Multichannel attention is generated by concatenating vector and scalar attention as follows:

$$c_{l_i} = a_{l_i}\left(a_{vl_i} \odot h_i\right) \quad (17)$$

Therefore, multichannel attention has the strengths of both vectorial attention and scalar attention.

D. Convolutional Neural Network Layer

To extract local features from the attention layer, convolution operations are employed on the combined attention layer. Typically, a zero-padding token is introduced before convolution to ensure uniform output sizes across different filters. Different filters and kernel sizes were applied to the multichannel attention c_{li}, to extract local features.

$$C = [c_1, c_2, \ldots c_n] \quad (18)$$

In the convolution operation, a filter $m \in \mathbb{R}^{l \times k}$ is applied to l consecutive words to generate a new feature. Here, $C \in \mathbb{R}^n$, where *k* and *n* are the embedded dimensions and input sequence length, respectively.

$$x_i = f(m.c_{i:i+l-1} + b) \quad (19)$$

where, $c_{i:i+l-1}$ is the concatenation of $c_i \ldots c_{i+l-1}$, f is a nonlinear activation function such as RELU, and $b \in \mathbb{R}$ is the bias term. After the filter *m* slides across, a feature map can be obtained as,

$$z = [z_1, z_2, \ldots, z_{n-l+1}] \quad (20)$$

Maxpooling is then applied to the feature map *z* to extract the most significant features for each filter *m*. To capture the different features of the input sequence, filters of various sizes are applied, resulting in a vector that is used at the output layer.

E. Output Layer

The proposed model consists of intent detection and slot filling outputs. The intent detection output is obtained using a fully connected layer with a softmax function to output the probability distribution over the intents. Therefore, the intent output vector is computed as follows:

$$y^i = Softmax(W.q + b) \quad (21)$$

For the slot-filling output, the feature vectors are passed to an LSTM decoder to capture the sequential nature of the slots and use the softmax function for the output.

$$d_i = LSTM(q_i, d_{i-1}) \quad (22)$$

$$y_i^s = softmax(W.d_i + b) \quad (23)$$

where, y^s is the slot label and W, b are the transformation matrix and bias vectors, respectively.

IV. Experimental Study

This section describes the datasets used in the experiments followed by a detailed experimental methodology to assess the effectiveness of the proposed approach. A comparative analysis of the baseline methods is presented. The performance of the model was evaluated using the widely adopted metrics of accuracy for the intent detection task and the F1-score for the slot-filling task.

A. Dataset

To validate the proposed model, experiments were conducted using the Airline Travel Information System (ATIS) dataset, which is one of the most widely recognized and historically significant datasets in Natural Language Understanding (NLU) research. The ATIS dataset has been a benchmark for Spoken Language Understanding (SLU) tasks for over three decades, making it an ideal choice for evaluating advancements in the field. The dataset focuses specifically on air travel-related queries and provides information on flights, fares, airlines, airports, cities, and ground services. It features 21 different intents and 128 slots, with a training set of 4478 samples, test set of 893 samples, and validation set of 500 samples [5].

The ATIS dataset presents several unique characteristics that support its use as a standard for model comparisons. One notable feature is the imbalanced distribution of intent types, with approximately 75% of intents belonging to a single class (atis_flight). This imbalance poses a challenge for intent detection models, making the dataset a rigorous test for the proposed approach. Moreover, the well-defined structure of the dataset allows for clear benchmarking and facilitates a direct comparison with existing models in the NLU domain. Its long-standing use in research ensures that the performance of the proposed model can be contextualized within the vast body of prior work, further validating its efficacy.

Table I gives an example of a semantic frame for an utterance from an ATIS dataset "I want fly from Baltimore to Dallas round trip." The slots adhere to the widely used IOB (in-out-begin) format for representing slot tags. This sentence pertains to airline travel with the intent of finding a flight. Notably, 'Baltimore' is tagged as departure city, 'Dallas' as arrival city and 'round trip' as round trip.

TABLE I. An Example of Frame

Entity	slots	Intent
I	O	atis_flight
want	O	
to	O	
fly	O	
from	O	
Baltimore	B-fromloc.city_name	
to	O	
Dallas	B-toloc.city_name	
round	B-round_trip	
trip	I-round_trip	

B. Experimental Settings

A grid search is employed to determine the optimal hyperparameters for the model. Specifically, three different filter sizes (2, 3, and 5) were tested and 128 feature maps were used. To prevent overfitting, a rate of 0.5 was applied to the feature maps. The shared encoder was configured with 200 hidden units and a rectified linear unit activation function was used. Additionally, a dropout rate of 0.5 was applied after the shared encoder, randomly dropping units to improve training was applied after the shared encoder. For intent detection classification and slot filling outputs, L2 regularization with a value of 0.001 was applied to the weights of the dense layers using a softmax activation function. The Adam optimizer and categorical cross-entropy loss functions were employed during training. Accuracy metrics were used to evaluate intent detection, and the F1-score was used for slot filling. A batch size of 32 was selected for the study. The input sequence was padded to a fixed length to fit the convolutional layer with a maximum length of 45 for the ATIS dataset. In the proposed model, the weights of the embedding layer were initialized with publicly available word2vec vectors, whereas words not included in the pretrained set were initialized with values from a uniform distribution to maintain consistent variance across all word vectors.

V. Experimental Results

The performance of the proposed model, AJLISBC-x, on the ATIS dataset is presented in Table I, where x represents the number of channels used during training. These channels enable the model to capture different representations of the input sequence, and the effectiveness of this multichannel representation is evident in the results.

As illustrated in Table II, all proposed AJLISBC models outperformed the baseline models. Specifically, AJLISBC-2, which utilizes two channels, demonstrated the highest accuracy

and F1-score. Among the configurations tested, AJLISBC-1, which operates with a single channel, showed inferior results compared with the multichannel models. This indicates that increasing the number of channels positively influences the model performance, although there may be diminishing returns as the number of channels increases beyond two.

TABLE II. COMPARISON OF AJLISBC WITH BASELINE RESULTS

Model	ATIS Dataset	
	Accuracy	F1-score
Bi-GRU + feature [34]	97.76	97.93
BiLSTM+Attention [35]	95.70	95.60
BC [36]	97.20	96.34
AJLISBC -1	97.89	98.32
AJLISBC -2	**98.19**	**98.61**
AJLISBC -3	97.99	98.38
AJLISBC -4	98.09	98.45
AJLISBC -5	97.89	98.56

In addition to channel variations, the effects of the attention mechanisms were evaluated. Table III presents the performance of AJLISBC-2 when using scalar attention, vectorial attention, or a combination of both. Scalar attention, which assigns importance weights to all elements of the input sequence, yields a slightly better accuracy than vectorial attention. However, both types of attention achieve the same F1-score, demonstrating that either mechanism is effective at improving the performance for this task. When scalar and vectorial attention are combined, the model achieves its highest F1-score and improved accuracy compared with either mechanism alone.

TABLE III. PROPOSED MODEL PERFORMANCE BASE ON ATTENTION MECHANISM

Model	Attention	ATIS Dataset	
		Accuracy	F1-score
AJLISBC -2	Scalar	97.09	98.42
AJLISBC -2	Vector	96.99	98.42
AJLISBC -2	Scalar + vector	97.89	98.56

VI. DISCUSSION

The results underscore the effectiveness of multichannel representation in enhancing the model performance. AJLISBC-2's superior performance compared to AJLISBC-1 suggests that using multiple channels helps the model capture diverse patterns within the input sequence. This is particularly relevant for complex tasks such as intent detection and slot filling, where different dimensions of the input can provide complementary information. The use of a single channel in AJLISBC-1 limits the ability of the model to process and leverage multiple facets of the input, leading to inferior results. Therefore, multichannel representation appears to be an effective strategy for improving the model performance.

However, it is worth noting that while increasing the number of channels generally improves the performance, the results show that the performance does not increase indefinitely. For example, AJLISBC-5 showed a slightly lower accuracy than AJLISBC-2, indicating that simply adding more channels may not necessarily result in better performance beyond a certain point. This may be due to the model encountering diminishing returns from the additional channels or because the increased complexity of the model requires more sophisticated optimization strategies. It is hypothesized that selecting the number of channels based on the number of informative words in a sentence can yield even better results, allowing the model to tailor the complexity of its representation to the specific needs of each input.

In examining attention mechanisms, scalar attention proves to be particularly effective for accuracy because of its ability to calculate the importance of all elements in the input sequence. This helps to identify the most relevant parts of the sequence for intent detection, which may explain its superior performance in this regard. Scalar attention is particularly useful in scenarios where the relationship between different elements in a sequence plays a crucial role, such as in slot-filling tasks. By contrast, vectorial attention selectively emphasizes features that are more relevant for specific tasks, thereby enhancing the robustness of the model. This mechanism introduces controlled perturbations in the hidden state, which allows the model to generalize more effectively to new inputs.

The combination of scalar and vectorial attention mechanisms leads to the best performance because it capitalizes on the strengths of both methods. Scalar attention helps to compute the overall importance of elements in the input, whereas vectorial attention fine-tunes the focus to specific dimensions of the input. This dual approach results in better performance in both intent detection and slot-filling tasks. The synergy between these two mechanisms also enables the model to indirectly assign varying learning rates to different dimensions of the hidden state, allowing more informative dimensions to be updated more rapidly than less informative ones do. This dynamic adjustment contributes to the observed performance improvement when both types of attention are used together.

VII. CONCLUSION

This paper presents an attention-based joint learning classification model for intent detection and slot-filling that combines BiLSTM and CNN (AJLISBC). The BiLSTM architecture captures contextual information, whereas scalar and vectorial attention mechanisms generate multichannel representations of the input sequence semantics. CNNs are applied to these multichannel representations to extract n-gram features and enhance performance in both intent detection and slot-filling tasks. Experimental results on the ATIS dataset show that the model outperforms baseline models, demonstrating the effectiveness of combining BiLSTM, CNN, and attention mechanisms for natural language understanding tasks. Despite the promising results, several limitations of this study should be acknowledged, as they may impact the validity, reliability, and generalizability of the findings. One limitation is the exclusive use of the ATIS dataset, which is relatively small, domain-specific, and focuses on flight-related queries.

This limited scope raises concerns about the generalizability of the findings to other domains or to larger, more diverse datasets. The model's performance might have been overestimated owing to the homogeneity of the dataset. Future work will involve testing the model on diverse datasets from various domains to better assess its generalizability and robustness across different natural language processing tasks. Another limitation is the manual selection of the number of channels used for multichannel representations. Although multichannel representations have shown effectiveness, the process of determining the optimal number of channels is empirical and not rigorously optimized. This could affect the reliability and consistency of the performance of the model across different datasets or tasks, as it may not generalize well to varying sentence lengths or input complexities. Future work will explore automated methods for determining the optimal number of channels, such as incorporating adaptive mechanisms based on input-data characteristics. This approach ensures that the model adapts more flexibly and consistently to diverse input scenarios. Future studies will address these limitations to further validate the effectiveness of the model and enhance its adaptability and applicability to a broader range of natural language understanding tasks.

ACKNOWLEDGMENT

This research was partly funded by the Ministry of Higher Education Malaysia under grant R.J130000.7851.5F568. The authors would also like to thank Universiti Teknologi Malaysia (UTM) for providing the resources used in this study.

REFERENCES

[1] P. Ni, Y. Li, G. Li, and V. Chang, "Natural language understanding approaches based on joint task of intent detection and slot filling for IoT voice interaction," Neural Computing and Applications, vol. 32, pp. 16149-16166, 2020.

[2] A. Algherairy and M. Ahmed, "A review of dialogue systems: current trends and future directions," Neural Computing and Applications, vol. 36, no. 12, pp. 6325-6351, 2024.

[3] S. Huang, P. Huang, Y. Xu, J. Liang, and J. Niu, "Exploring Label Hierarchy in Dialogue Intent Classification," in ICASSP 2024-2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2024: IEEE, pp. 11511-11515.

[4] T. Wu, M. Wang, Y. Xi and Z. Zhao, "Intent recognition model based on sequential information and sentence features," Neurocomputing, vol. 566, p. 127054, 2024.

[5] H. Weld, X. Huang, S. Long, J. Poon, and S. C. Han, "A survey of joint intent detection and slot filling models in natural language understanding," ACM Computing Surveys (CSUR), 2021, doi: https://doi.org/10.1145/3547138.

[6] M. Firdaus, A. Kumar, A. Ekbal, and P. Bhattacharyya, "A multi-task hierarchical approach for intent detection and slot filling," Knowledge-Based Systems, vol. 183, p. 104846, 2019, doi: https://doi.org/10.1016/j.knosys.2019.07.017.

[7] A. S. M. Zailan, N. H. I. Teo, N. A. S. Abdullah, and M. Joy, "State of the Art in Intent Detection and Slot Filling for Question Answering System: A Systematic Literature Review," International Journal of Advanced Computer Science & Applications, vol. 14, no. 11, 2023.

[8] L. Huang, S. Liang, F. Ye, and N. Gao, "A fast attention network for joint intent detection and slot filling on edge devices," IEEE Transactions on Artificial Intelligence, 2023.

[9] J. Wu, I. G. Harris, H. Zhao, and G. Ling, "A graph-to-sequence model for joint intent detection and slot filling," in 2023 IEEE 17th International Conference on Semantic Computing (ICSC), 2023: IEEE, pp. 131-138.

[10] M. Firdaus, H. Golchha, A. Ekbal, and P. Bhattacharyya, "A deep multi-task model for dialogue act classification, intent detection and slot filling," Cognitive Computation, vol. 13, no. 3, pp. 626-645, 2021, doi: https://doi.org/10.1007/s12559-020-09718-4.

[11] T. Pham and D. Q. Nguyen, "JPIS: A Joint Model for Profile-Based Intent Detection and Slot Filling with Slot-to-Intent Attention," in ICASSP 2024-2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2024: IEEE, pp. 10446-10450.

[12] F. Mairesse et al., "Spoken language understanding from unaligned data using discriminative classification models," in 2009 IEEE International Conference on Acoustics, Speech and Signal Processing, 2009: IEEE, pp. 4749-4752.

[13] Y.-Y. Wang, "Strategies for statistical spoken language understanding with small amount of data-an empirical study," in Eleventh Annual Conference of the International Speech Communication Association, 2010.

[14] A. Celikyilmaz and D. Hakkani-Tur, "A joint model for discovery of aspects in utterances," in Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), 2012, pp. 330-338.

[15] P. Xu and R. Sarikaya, "Convolutional neural network based triangular crf for joint intent detection and slot filling," in 2013 ieee workshop on automatic speech recognition and understanding, 2013: IEEE, pp. 78-83.

[16] M. Giménez, A. Fabregat-Hernández, R. Fabra-Boluda, J. Palanca, and V. Botti, "A detailed analysis of the interpretability of Convolutional Neural Networks for text classification," Logic Journal of the IGPL, p. jzae057, 2024, doi: https://doi.org/10.1093/jigpal/jzae057.

[17] A. Orvieto et al., "Resurrecting recurrent neural networks for long sequences," in International Conference on Machine Learning, 2023: PMLR, pp. 26670-26698.

[18] K. K. Jayanth, G. B. Mohan, R. P. Kumar, and M. Rithani, "Intent Recognition Leveraging XLM-RoBERTa for Effective NLU," in 2024 3rd International Conference on Applied Artificial Intelligence and Computing (ICAAIC), 2024: IEEE, pp. 877-882.

[19] U. Farooq, M. S. Mohd Rahim, and A. Abid, "A multi-stack RNN-based neural machine translation model for English to Pakistan sign language translation," Neural Computing and Applications, vol. 35, no. 18, pp. 13225-13238, 2023.

[20] W. A. Abro, G. Qi, Z. Ali, Y. Feng, and M. Aamir, "Multi-turn intent determination and slot filling with neural networks and regular expressions," Knowledge-Based Systems, vol. 208, p. 106428, 2020.

[21] M. Jbene, S. Tigani, R. Saadane, and A. Chehri, "A robust slot filling model based on lstm and crf for iot voice interaction," in 2022 IEEE Globecom Workshops (GC Wkshps), 2022: IEEE, pp. 922-926.

[22] S. Das, A. Tariq, T. Santos, S. S. Kantareddy, and I. Banerjee, "Recurrent neural networks (RNNs): architectures, training tricks, and introduction to influential research," Machine Learning for Brain Disorders, pp. 117-138, 2023.

[23] B. Kane, F. Rossi, O. Guinaudeau, V. Chiesa, I. Quénel, and S. Chau, "Joint Intent Detection and Slot Filling via CNN-LSTM-CRF," in 2020 6th IEEE Congress on Information Science and Technology (CiSt), 2021: IEEE, pp. 342-347.

[24] S. Cong and Y. Zhou, "A review of convolutional neural network architectures and their optimizations," Artificial Intelligence Review, vol. 56, no. 3, pp. 1905-1969, 2023.

[25] R. Patel and S. Patel, "Deep learning for natural language processing," in Information and Communication Technology for Competitive Strategies (ICTCS 2020) Intelligent Strategies for ICT, 2021: Springer, pp. 523-533.

[26] Y. Hui, J. Wang, N. Cheng, F. Yu, T. Wu, and J. Xiao, "Joint intent detection and slot filling based on continual learning model," in ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2021: IEEE, pp. 7643-7647.

[27] S. Yu, D. Liu, W. Zhu, Y. Zhang, and S. Zhao, "Attention-based LSTM, GRU and CNN for short text classification," Journal of Intelligent & Fuzzy Systems, vol. 39, no. 1, pp. 333-340, 2020.

[28] M. Jeong and G. G. Lee, "Triangular-chain conditional random fields," IEEE Transactions on Audio, Speech, and Language Processing, vol. 16, no. 7, pp. 1287-1302, 2008.

[29] D. Yu, S. Wang, and L. Deng, "Sequential labeling using deep-structured conditional random fields," IEEE Journal of Selected Topics in Signal Processing, vol. 4, no. 6, pp. 965-973, 2010.

[30] L. Zhang and M. Yang, "Asynchronous joint extraction algorithm based on intent-slot attention mechanism," in International Conference on Image, Signal Processing, and Pattern Recognition (ISPP 2024), 2024, vol. 13180: SPIE, pp. 824-831.

[31] H. Huang, X. Feng, and Z. Wan, "Joint Model of Intent Recognition and Slot Filling Based on Graph Neural Network fusion of external knowledge base," in 2024 36th Chinese Control and Decision Conference (CCDC), 2024: IEEE, pp. 323-329.

[32] J. Huang and H. Tang, "A Joint Model of Multiple Intent Recognition and Slot Filling Based on Graph Neural Network," 2024.

[33] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient estimation of word representations in vector space," arXiv preprint arXiv:1301.3781, 2013. [Online]. Available: https://arxiv.org/abs/1301.3781.

[34] M. Firdaus, S. Bhatnagar, A. Ekbal, and P. Bhattacharyya, "A deep learning based multi-task ensemble model for intent detection and slot filling in spoken language understanding," in Neural Information Processing: 25th International Conference, ICONIP 2018, Siem Reap, Cambodia, December 13-16, 2018, Proceedings, Part IV 25, 2018: Springer, pp. 647-658, doi: https://doi.org/10.1007/978-3-030-04212-7_57.

[35] W. Chao, Y. Ke, and W. Xiaofei, "POS Scaling Attention Model for Joint Slot Filling and Intent Classification," in 2020 IEEE 20th International Conference on Communication Technology (ICCT), 2020: IEEE, pp. 1483-1487, doi: 10.1109/ICCT50939.2020.9295901.

[36] C. Wang, Z. Huang, and M. Hu, "SASGBC: Improving sequence labeling performance for joint learning of slot filling and intent detection," in Proceedings of 2020 the 6th International Conference on Computing and Data Engineering, 2020, pp. 29-33, doi: https://doi.org/10.1109/CCDC62350.2024.10587455.

Attention-Based Deep Learning Approach for Pedestrian Detection in Self-Driving Cars

Wael Ahmad AlZoubi[1], Prof. Girish Bhagwant Desale[2], Dr. Sweety Bakyarani E[3],
Dr Uma Kumari C R[4], Dr. Divya Nimma[5], K Swetha[6], Dr B Kiran Bala[7]
Applied Science Department, Ajloun University College, Al-Balqa Applied University, Jordan[1]
HOD, Department of Computer Science & IT, JET'S Z. B. Patil College, Dhule. (M.S.), Jalgoan, India[2]
Department of Computer Science-Faculty of Science and Humanities,
SRM Institute of Science and Technology, Kattankulathur, India[3]
Associate Professor, Department of Electronics and Communication Engineering, KCG College of Technology, Chennai, India[4]
PhD in Computational Science, University of Southern Mississippi, Data Analyst in UMMC, USA[5]
Assistant Professor, Department of CSE, Koneru Lakshmaiah Education Foundation, Vaddeswaram, AP, India[6]
Head of the Department-Department of Artificial Intelligence and Data Science,
K. Ramakrishnan College of Engineering, Trichy, India[7]

Abstract*—Autonomous vehicle safety relies heavily on the ability to accurately detect pedestrians, as this capability is crucial for preventing accidents and saving lives. Pedestrian recognition is particularly challenging in the dynamic and complex environments of urban areas. Effective pedestrian detection is crucial for ensuring road safety in autonomous vehicles. Current pedestrian identification systems often fall short in capturing the nuances of pedestrian behavior and appearance, potentially leading to dangerous situations. These limitations are mainly due to difficulties in various conditions, such as low-light environments, occlusions, and intricate urban settings. This paper proposes a novel solution to these challenges by integrating an attention-based convolutional bi-GRU model with deep learning techniques for pedestrian recognition. This method leverages deep learning to provide a robust solution for pedestrian detection. Convolutional layers are utilized to extract spatial features, attention mechanisms highlight semantic details, and Bidirectional Gated Recurrent Units (Bi-GRU) capture the temporal context in the proposed model. The process begins with data collection to build a comprehensive pedestrian dataset, followed by preprocessing using min-max normalization. The key components of the model work together to enhance pedestrian detection, ensuring a more accurate and comprehensive understanding of dynamic pedestrian scenarios. The implementation of this unique approach was carried out using Python, employing libraries such as TensorFlow, Keras, and OpenCV. The proposed attention-based convolutional bi-GRU model outperforms previous models by an average of 17.1%, achieving an accuracy rate of 99.4%. The model significantly surpasses Random Forest, Faster R-CNN, and SVM in terms of pedestrian recognition accuracy, which is critical for autonomous vehicle safety.

Keywords*—Pedestrian recognition; autonomous vehicle safety; deep learning; attention mechanism; Bidirectional Gated Recurrent Units

I. INTRODUCTION

One of the most important aspects of road transport, and a key issue in the automotive industry, is vehicle safety. This includes a wide range of techniques and strategies aimed at reducing the likelihood of a collision and, if it occurs, the number of injuries to car occupants and other road users [1]. To help establish safety norms and standards, government agencies and independent agencies conduct extensive safety tests to assess how vehicles perform in collision situations Organizational development these tests include the Insurance Institute for Highway Safety (IIHS) and the National Highway Traffic Safety Administration (administered by the NHTSA) [2]. The use of advanced driver assistance systems (ADAS) in vehicles has become a necessity. With features such as parking assist, adaptive lighting, blind spot monitoring and collision avoidance technology, this system supports drivers and improves safety. While there are more hybrid electric vehicles their security concerns increase. It is important to guarantee that the high voltage electrical system is crashworthy, the battery is integrated and safe [3]. Car safety precautions include child safety. This includes instructing parents to check child rear seats before leaving the vehicle, using child safety seats, and integrated child safety doors Cybersecurity is a growing threat due to increased connectivity and carry reliance on software in vehicles therefore. Protecting vehicles from cyberattacks and ensuring the integrity of critical software systems is more important than ever [4].

Application of machine learning techniques in automotive safety brings a new era of aggressive crash avoidance and occupant protection Advanced Driver Assistance Systems (ADAS) are commonly used machine learning systems role results. These systems use algorithms to analyze sensor data from cameras, radar, lidar and ultrasonic sensors to detect potential hazards in crashes, turns and even drowsy driving [5]. This technology provides real-time warnings or interventions to help drivers avoid accidents. Machine learning is essential for autonomous vehicles to recognize and understand their surroundings, recognize traffic signs, people, and other moving objects and make decision calls to help them travel safely Predictive maintenance. which in turn uses machine learning to monitor the vehicles' own health and predict any problems before failure results [6]. This technology provides real-time warnings or interventions to help drivers avoid accidents.

Machine learning is essential for autonomous vehicles to recognize and understand their surroundings, recognize traffic signs, people, and other moving objects and make decision calls to help them travel safely Predictive maintenance. which in turn uses machine learning to monitor the health of the vehicles themselves [7]. By helping to identify and prevent any potential cyberattacks on the vehicle's systems, machine learning helps improve automotive cybersecurity. Vehicle software can be protected from unauthorized access and data breaches using machine learning algorithms that can detect anomalies and strange behavior [8]. Machine learning in automotive safety using real-time data analytics, predictive capabilities and continuous algorithmic improvements is an ongoing research area that can improve road traffic safety necessary to reduce traffic-related incidents , injuries and deaths to improve driving effectiveness and overall safety [9].

Deep learning has been shown to be a revolutionary force in automotive safety prediction, radically changing how we predict and avoid accidents on the road. With large amounts of data coming from many sources including as sensors, cameras and vehicle control systems, deep learning methods especially neural networks are used These methods can identify complex patterns, identify threats a it is possible and shows important information in real time. Deep learning algorithms play a key role with advanced driver assistance systems (ADAS) by testing the environment, detecting objects and predicting collision hazards. This enables the system to issue warnings in a timely manner or even preventive measures to prevent accidents. Deep learning in predictive maintenance is important because it can analyze sensor data to identify technical problems in the vehicle before they become dangerous.

Feature extraction is necessary to identify pedestrians from other features in and around an image or video frame Before manual feature engineering; However, with the development of deep learning, this approach is now widely incorporated into web design [10]. The conventional Histogram of Oriented Gradients (HOG) is one of the more popular extraction techniques. Unlike computer histograms that show gradient orientations in these cells by image segmentation to extract important edge information and shape information, local binary models (LBP) provide information about garment textures by encoding local texture patterns by comparing the pixel intensities with their neighbors is very useful in distinguishing between a pedestrians [10]. Also, a deep learning technique, convolutional neural networks (CNNs) are popular in feature extraction. These networks have greatly enhanced pedestrian recognition by learning and extracting features such as shapes, textures, and context [11].

Ensuring the safety of motorists and pedestrians sharing the road is of utmost importance. Accurate pedestrian detection and identification, enabled by state-of-the-art deep learning algorithms, is essential to the success of our efforts. By incorporating cognitive algorithms into a convolutional bi-directional gated recurrent unit (Bi-GRU) model for pedestrian recognition, this research aims to push the limits of autonomous vehicle safety [12]. This process improves model understanding in dynamic pedestrian scenarios using Bi-GRU to store time context in addition to the ability of convolutional neural networks (CNNs) to capture detailed scene details Feature those in ambiance, those in ambiance and crowds [13]. This ubiquitous approach is intended to greatly improve the accuracy and reliability of pedestrian detection methods, making autonomous vehicles safer and more responsive is more adept at dealing with real-world obstacles. In this work, we explore the key features and techniques we use in our model and create findings that reveal the potential to change the automotive safety feature of and about the whole new.

Academics and engineers have made great strides to improve autonomous vehicle safety as new techniques for deep learning model design and data preprocessing This review focuses on the use of Min-Max normalization, which provides the input data is better and increases the accuracy of the model in pedestrian detection Established it is optimized and standardized, resulting in more consistent and reliable results. In addition, the addition of the Bi-GRU (Bidirectional Gated Recurrent Unit) layer in the model framework introduces an additional period of pedestrian detection This layer provides the model with an important contextual understanding about dynamic pedestrian conditions because it can capture time dependence in input data Bidirectional information processing is possible, and it greatly improves the safety of the autonomous vehicle by enabling the image to carry protective information. In addition, the addition of focus improves the accuracy of the model in detecting pedestrians, especially in complex and multidimensional environments where pedestrian access or movement may be obstructed is unobservable, this dynamic distribution of weights across data points assures that the model focuses on the most relevant cases. Attention is a variable that enhances pedestrian visibility and contributes significantly to the overall safety of involved vehicles by reaching critical objects faster. If driven together, these developments make a significant contribution to the field and point to future advancements in autonomous driving safety and reliability.

Key contributions of this framework are,

- The proposed system increases the ability of the model to accurately detect pedestrians and enhance the input data quality by using Min-Max normalization. This pre-processing step provides reliable and consistent results.
- The two GRU layer uses the temporal relationships of input data to provide pedestrians with a description of the relevant context. The use of two-way information processing enhances the understanding of the dynamic pedestrian model and contributes to the safety of autonomous vehicles.
- The model focuses on the most pertinent information thanks to the attention layer, which dynamically distributes weights to different data regions. This is essential for precise pedestrian detection in complex scenarios. This technique enhances autonomous vehicle safety by more effectively acquiring and processing critical attributes.
- The convolutional layer improves the model's capacity to identify complex shapes, patterns, and textures in images, which is crucial for precise pedestrian identification in a variety of settings. By incorporating it, autonomous vehicle safety is further reinforced by

ensuring a more thorough grasp of dynamic pedestrian circumstances.

The research's remaining portions are listed below. Section II provides a review of the literature. The issue statement is covered in Section III. The suggested technique for detecting neurological diseases is then covered in Section IV. Section V discusses the results and discussions. Section VI discusses the conclusion.

II. Related Work

Chen et al. [14] proposes a thorough method for deep learning-based pedestrian recognition and tracking in challenging circumstances, with an emphasis on problems like small-target people and partially obstructed pedestrians in crowded areas. A notable improvement to the YOLO detection algorithm involves the addition of channel attention as well as spatial attention modules. The capacity of the model to represent crucial feature information is eventually improved by these modules, which aid in amplifying important feature data in both channels and spatial dimensions. These modules' incorporation into Darknet-53's backbone network exemplifies an organized method of feature extraction. It is a well-known choice to use the DeepSort tracking technique in conjunction using the Kalman filter algorithm to estimate the mobility status of pedestrians. The tracking procedure is further strengthened by the Mahalanobis distance and evident feature for similarity estimates and the Hungarian method for ideal target matching. This method strengthens the tracking system by integrating deep learning for identification with conventional tracking methods. It is necessary to do preliminary testing of the enhanced YOLO pedestrian detection system and DeepSort tracking technique in the same setting. The findings show a considerable decrease in missed detections and false positives, a more robust handling of tracking failures caused by occlusion, and an essential increase in detection precision for small-target pedestrians. These encouraging results point to the suggested approach's efficacy in enhancing the functionality of intelligent vehicle-pedestrian identification and tracking systems, particularly in challenging and congested environments.

Lu et al. [15] proposes an innovative method for predicting vehicle trajectory for connected and autonomous cars that operate in mixed traffic situations. For these vehicles to be safe and sustainable, accurate forecasting of trajectory is essential. This task is complicated by several variables, including individual vehicle motions, the state of the road, and interactions with other nearby cars. This problem is successfully addressed by the suggested method, Heterogeneous Context-Aware Graph Convolutional Networking with Encoder-Decoder architecture, which simultaneously extracts hidden contexts from each historical trajectory, the dynamic driving scene, and the interactions between vehicles. The design of the method, which uses 2-dimensional Convolutional Networks with temporal attention to represent the changing scene context from scene photos and Temporal Convolutional Networks to capture personal environments from previous vehicle trajectories, is praiseworthy. Spatial-temporal dynamic Graph Convolutional Networks, which incorporate both individual and scene settings as node representations, are a powerful option for modelling inter-vehicle interaction patterns. It is well thought out to combine these three circumstances and use them as input to the decoder to produce future trajectories. The credibility of the study is increased by the verification of the proposed framework on real-world datasets comprising various driving circumstances. The findings validate the assertion that the model beats state-of-the-art techniques in terms of prediction accuracy and stability regardless of vehicle conditions. The vehicle trajectory prediction field for connected and autonomous cars gains significantly from the work presented in this paper, which offers a thorough and practical solution to complicated real-world traffic situations [16].

Pustokhina et al. [17] addresses an important problem essential to improving the safety of vulnerable road users: the actual detection of anomalies in pedestrian areas. The expanded network of surveillance cameras and the amount of film recorded make manual verification and labeling of anomalies difficult and time-consuming As a result, surveillance systems automation, especially those based on deep learning theories, has become increasingly popular in computer vision. To overcome these challenges, our research focuses on the development of a deep learning anomaly detection technique (DLADT-PW) for pedestrian roads. The aim of the DLADT-PW method is to detect and classify different types of deformation in pedestrian traffic, such as the presence of cars, skateboarders, or jeeps Mask area convolutional neural network (Mask-RCNN) incorporating a dense connected mesh work in models after the pre-processing stage in order to reduce noise and detect. The image can be enhanced before phase. The study provides evidence of the effectiveness of the DLADT-PW method by running detailed simulations and analyzing the results from different angles. The findings confirm the excellent performance of the DLADT-PW model and demonstrate its potential for accurate identification. By developing a computerized anomaly detection algorithm that can quickly and segmentally detect anomalies in pedestrianized roadways, this research significantly enhances the field of computer vision, including pedestrian safety. In this context, deep learning models are used, especially masked RCNN and DenseNet, and emphasis is placed on implementing state-of-the-art techniques. The capabilities of the model are further confirmed through comprehensive analysis and empirical analysis, indicating the potential for useful application in real-world monitoring systems for providing vulnerable road users security has increased.

Islam et al. [18], it introduces a vision-based approach to customized safety messages (PSMs), which addresses an important part of passenger safety in Internet-connected vehicles in vehicular Internet-connected (V2P) communication is needed in situations out of this, but pedestrians face a major obstacle as it relates to the lack of low-level connectivity with nearby connected vehicles -which involve telephony or especially- long distance communication (DSRC) devices. The main contribution of the study is a vision-based system of real-time PSM using video feeds from roadside traffic cameras according to SAE J2945 According to the paper’s analysis, safety for pedestrians has been provided the planned netted vehicle is an example of the feasibility of this strategy. The presented results highlight the effectiveness of the vision-based

system in real-time collision warning to avoid future vehicular and pedestrian accidents, especially in time-to-collision (TTC) values in the emphasis. The study confirms that, in a connected vehicle context, our vision-based pedestrian safety notification system satisfies the latency criteria for the PSCW safety application. The important topic of pedestrian security within the environment of linked automobiles is addressed in this research, which is noteworthy. Its vision-based methodology provides a novel approach to the problem of pedestrian communication in the absence of specialized equipment. The creation of a useful safety application and the approach's proven effectiveness in real-time collision warning serves to further highlight its potential. Its legitimacy and application in the automobile sector are improved by adhering to SAE standards. To offer a more thorough understanding of the suggested approach, the study might utilize more information on the technical facets of the vision-based system, the specific data streams employed, and potential difficulties in real-world application.

The examined literature focuses on important advancements in deep learning and computer vision for problems relating to pedestrian safety and traffic safety. In the study, a better YOLO-based system for pedestrian tracking and recognition is introduced. By using attention modules to solve problems such as tiny targets and occluded pedestrians, the system improves detection accuracy and provides stable tracking in congested situations. In different research, a novel method for simulating complicated vehicle interactions and improving trajectory prediction accuracy is presented for predicting vehicle trajectories in mixed traffic settings using graph convolutional networks. A thorough analysis of current LSTM-based short-term traffic prediction algorithms is presented, together with a classification and assessment of their advantages and disadvantages, to help researchers choose the best method for certain traffic situations. Another study explores automated anomaly detection at pedestrian crossings and presents a Deep Learning anomaly detection technique that can identify a wide range of anomalies and improve pedestrian safety. Using roadside traffic cameras and adhering to industry standards, a vision-based system for generating Personal Safety Messages in connected vehicle contexts is discussed, enabling real-time collision warnings and ultimately enhancing pedestrian safety and communication in the absence of specialized equipment. Through creative approaches and useful solutions, these studies jointly increase traffic forecasts, intelligent transportation systems, and pedestrian safety.

III. Problem Statement

Pedestrian safety in the context of autonomous vehicles is still a major concern, according to the literature review, which draws from the discussion above. This is especially true in situations where there are small targets among the pedestrians, partial obstructions in crowded areas, and no personal communication devices. While recent research shows that deep learning, better detection algorithms, and vision-based techniques have potential, a holistic and cohesive solution is required to maximize pedestrian recognition and tracking for autonomous vehicle safety. Thus, Convolutional Bi-GRU, a concept-based model that combines traditional detection methods with deep learning for accurate and reliable pedestrian identification [19]. The main objective is to increase the detection accuracy of target pedestrians, and to reduce missed detections, false positives, and failures due to occlusion and tracking failures so especially the research objective of improving the safety of autonomous vehicles in harsh and crowded situations is maximized. The advantages of recurrent and convolutional neural networks are well combined in the Conv-Bi-GRU model. CNNs perform exceptionally well when extracting features from images, and this integration improves the model's ability to extract and analyze spatial and temporal complexity from video data, and provides improve our understanding of the temporal relationships observed in Bi-GRU walking trajectories.

IV. Proposed Framework of Attention Based Convolutional Bi-GRU Model

The approach adopted in this study is based on a structured design with several main components. In order to train and test the model, in the first phase of data collection, a large dataset of pedestrians should be obtained. The next step in processing the data is minimum-maximum normalization, which involves scaling the pixel values to the standard range. Concept-based convolutional two-GRU model, a deep learning system designed for pedestrian detection is the core of the methodology. This model has several layers, such as a conceptual approach for highlighting relevant semantic features in the data and bidirectional gated repetitive units (Bi-GRU) for capturing time descriptions. The data are extracted with convolutional layers of space objects to better understand pedestrians' events. A fully connected layer and SoftMax activation provide the final result, enabling the model to better classify pedestrians. Considering all things, this technology is designed to improve the safety of autonomous vehicles by preventing crashes and accidents, in addition to being exceptionally effective at detecting pedestrians and giving way it is possible to approach the front with an autonomous vehicle. Fig. 1 shows the block diagram of the proposed model.

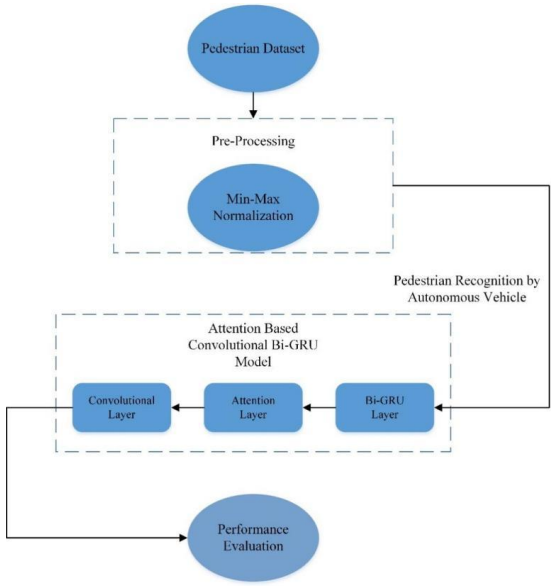


Fig. 1. Block diagram of proposed attention based convolutional Bi-GRU model.

A. Data Collection

The pedestrian dataset used in this study, sourced from Kaggle, consists of extensively documented videos capturing people using crosswalks in various contexts. The dataset is divided into three distinct segments, each providing valuable insights into pedestrian behavior. The first segment, titled "Crosswalk," is a 12-second clip showing individuals safely crossing the street using designated means, offering a clear understanding of compliance with pedestrian safety regulations. The second segment, a 25-second video titled "Night," highlights pedestrian behavior in low-light conditions, providing insights into how lighting influences crosswalk use. The dataset includes accurate bounding box annotations for pedestrians, meticulously encoded in a .csv file, making it highly suitable for analyzing and training machine learning models. For training and testing, the dataset was split into 70% for training and 30% for testing, ensuring a balanced evaluation of the model's performance [20].

A. Pre-Processing using Min-Max Normalization

Min-max scaling is a data processing technique, sometimes called normalizing, in which the pixel values of images in a data set are converted to a specific range, usually 0 to greater in computer vision, including safety-related applications for pedestrian detection and new autonomous vehicles. If you want to make sure that each pixel value falls within a defined, constrained range, min-max scaling comes in handy. This creates a model of deep learning and a smooth meeting [21]:

$$x_{sv} = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (1)$$

whereas, the scaled value is x_{sv}. The initial pixel value is represented by x. The smallest pixel value in the collection is denoted by x_{min} . The dataset's highest pixel value is represented by the value x_{max}.

This preprocessing step enhances model performance by reducing the impact of varying pixel intensities and ensuring that the model learns from data with consistent value ranges. In the context of pedestrian detection, this consistency improves the model's ability to accurately recognize and differentiate pedestrians from background elements, leading to better detection accuracy and robustness in safety-critical applications.

B. Pedestrian Recognition Using Attention-based Convolutional Bi-GRU Model to Improve Autonomous Vehicle Safety

Using the Attention-Based Convolutional Bi-GRU Model and deep learning for pedestrian recognition, this model optimizes autonomous vehicle safety by utilizing pre-trained feature vectors in the input layer to represent the important attributes related to pedestrian recognition. This study may effectively examine and comprehend the many attributes of pedestrians due to these vectors, which are pre-configured to encode pertinent information about them. Preprocessed data is fed into the input layer in the form of images. These images have usually been through operations like scaling, cropping, and potentially color normalization. The input layer is where the data enters the neural network; it doesn't do any computation.

$$X_i = (x_1, x_2, \dots \dots \dots x_{n-1}, x_n) \quad (2)$$

In Eq. (2), the feature vector of the i^{th} pedestrian data instance is represented by X_i, the attribute vector of the i^{th} characteristic inside the pedestrian data is represented by x_i, and the length or dimensionality of the input data is indicated by X. An image sequence's temporal dependencies are intended to be captured by the Bidirectional Gated Recurrent Unit (Bi-GRU) layer. It has the ability to analyze data both forward and backward, which allows it to gradually capture context. The temporal properties of the input data are represented as a series of feature vectors as this layer's output. As illustrated in Fig. 2, this model utilizes visual features that reflect pedestrian scenarios as input data. The Bi-GRU layer is split into two halves that each evaluate the image features in both forward and reverse directions at the same time. The GRU algorithm then processes these successive picture features and produces an output vector with specified dimensions. Four crucial computing components are involved in the GRU action. The reset gate is part of the initial element.

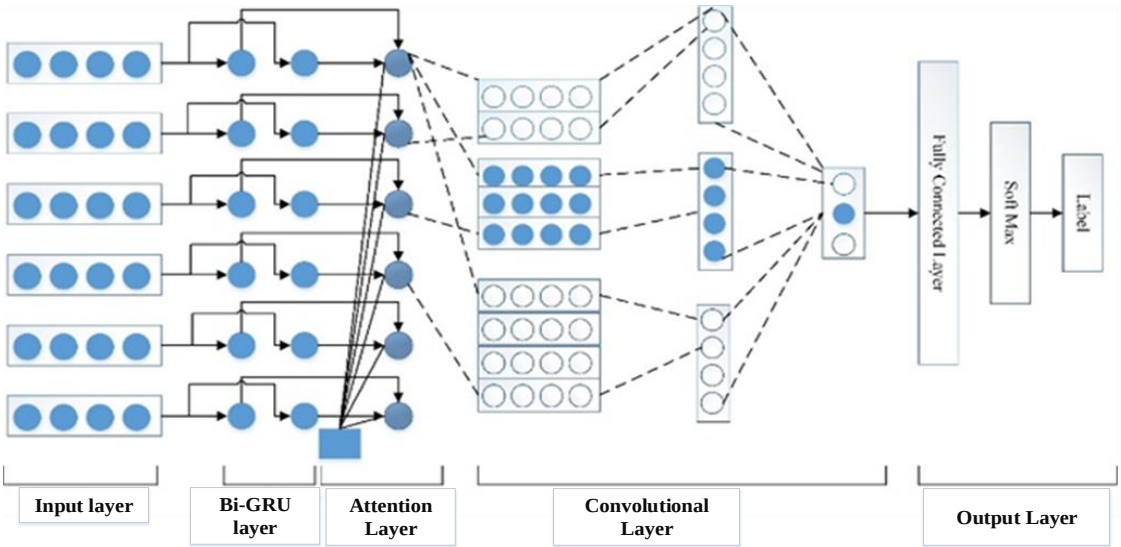


Fig. 2. Architecture of proposed attention-based convolutional Bi-GRU model.

$$Re_t = \tau(w_r x_t + u_r h_{t-1} + b_r) \quad (4)$$

$$\mathbb{Z}_t = \tau(w_z x_t + u_z h_{t-1} + b_z) \quad (5)$$

In order to determine the output of the current instant, GRU first computes the candidate memory content from Eq. (6), where w and u represent weight information and b represents bias.

$$h^*_t = \tanh(w x_t + u Re_t h_{t-1} + b) \quad (6)$$

The results from the preceding mentioned operations are the basis on which the GRU ultimately computes the output using Eq. (7). The Bi-GRU layer can record contextual information about every component in the input data for pedestrian identification.

$$h_t = (1 - \mathbb{Z}_t) h^*_t + \mathbb{Z}_t h_{t-1} \quad (7)$$

To concentrate on pertinent areas or frames within the input data, the attention mechanism is essential. In order to emphasize areas of interest, it dynamically assigns weights to various input sequence segments. This is especially crucial for spotting pedestrians in complex circumstances. The input is represented in an attention-weighted manner by this layer. Each Bi-GRU layer output is given weight by the attention mechanism, as shown in Fig. 2. The weight's magnitude indicates how important the semantic material is. To put it simply, locations with greater weights attract more "Attention," a sign that they have a significant influence on the final pedestrian recognition categorization result. For computing attention, the formula is denoted by Eq. (8) and Eq. (9). The attention weight a_{ik} and the related weight parameters (w_{a2} and w_{a1}) are represented in Eq. (10), together with the sequence data length t_h.

$$G_t = \sum_{t=1}^{t_h} a_{ik}\, h_t \quad (9)$$

$$a_{ik} = softmax(w_{a2} \tanh(w_{a1} h_t)) \quad (10)$$

Applying the Convolutional Layer to the spatial representation of the input data is common practice in Convolutional Neural Networks (CNNs). It is essential for recognizing pedestrians' visual traits, such as edges, forms, and textures, since it extracts pertinent aspects and patterns from the attention-weighted input. The computation of the convolution layer is carried out as follows:

The convolution layer receives its input from the attention layer's generated intermediate semantic information, where $G_{i:k}$ denotes the grouping of features pertaining to the i^{th} and k^{th} items in the context of pedestrian recognition.

$$G_{i:k} = G_i \oplus G_{i+1} \oplus \ldots \oplus G_k \quad (11)$$

Three different convolutional kernels are applied to obtain deep features. Throughout this procedure, 'f' stands for a hyperbolic tangent function, w for the weight parameters, ' m ' for the convolution kernel width, and ' b ' for the bias term.

$$t_i = f(w G_{i:k+m-1} + b) \quad (12)$$

After that, the model uses maximum pooling to merge the convolution results and identify important features. The final output of the entire convolution layer is then formed by combining the pooled results. 'n' stands for the number of convolution results, and 'j' for the number of convolution kernels. The computations are performed as follows.

$$t_k = [t_1, t_2, \ldots\ldots, t_{n-1}, t_n] \quad (13)$$

$$\widehat{t_k} = \max(t_k) \quad (14)$$

$$\hat{t} = [\widehat{t_1}, \widehat{t_2}, \widehat{t_3}] \quad (15)$$

In this study, the input of the fully connected layer is the output of the convolutional layer of the output layer. Then, use the SoftMax function to find as much as possible in each segment. The pedestrian detection part of the input has the highest potential. The calculations are written as the following formula, where WD represents the weighting parameters of the mesh and BD represents the bias term.

$$d_k = w_d \hat{t} + b_d \quad (16)$$

$$o_i = \frac{\exp(d_k)}{\sum_k \exp(d_k)} \quad (17)$$

To improve the safety of autonomous vehicles, a specific pedestrian detection method using a concept-based convolutional bi-GRU model approach is presented using deep learning methods and are mixed, such as convolutional layers, attention mechanisms, and bidirectional gated recurrent units (Bi- GRU). Important attributes are extracted by the convolutional layer using weights assigned by the model to key semantic descriptions by attention mechanism Pedestrians should be classified reliably by the model due to SoftMax activation and the fully connected layer it provides for the final output. This approach shows promise for increasing the safety of autonomous vehicles by helping to avoid collisions, and for better pedestrian detection.

V. Results and Discussion

Autonomous vehicle safety was experimentally optimized using a deep learning-based pedestrian recognition system, analyzed with a concept-based convolutional bi-GRU model. The approach in this study is organized with several key features. In the first phase of data collection, a large dataset of pedestrians was obtained. This dataset, used for benchmarking, includes extensively documented videos with accurate bounding box annotations, capturing pedestrian behavior in various scenarios such as crosswalks and low-light conditions. The next step in processing the data involved min-max normalization, which scales pixel values to a standard range. The key component of the method is the concept-based convolutional bi-GRU model, a deep learning system built for pedestrian recognition. This model consists of several layers, enabled by bidirectional gated recurrent units (Bi-GRU) to capture temporal patterns and important semantics in the data. Crucial output for the attention mechanism, which highlights essential sections and enables accurate pedestrian classification, is provided by a fully connected layer and SoftMax activation. The goal of this technology is to enhance the safety of autonomous vehicles by preventing collisions and accidents through highly effective pedestrian detection, offering a promising direction for the future of driverless transportation.

The Table I provides a comprehensive summary of a neural network model's architecture and parameters. It details each

layer's type, output shape, and the number of parameters involved. The model begins with a 1D convolutional layer (`Conv1D`) followed by a max pooling layer to down-sample the feature maps. It then uses a flattening layer to reshape the data for dense layers, which include two fully connected (`Dense`) layers with 128 units each and one final dense layer with a single output unit. The total number of parameters, including those in the convolutional and dense layers, is 131,905, which are all trainable, indicating the model's complexity and capacity for learning from data.

TABLE I. MODEL ARCHITECTURE AND PARAMETERS SUMMARY

Layer (type)	Output Shape	Param #
conv1d_1 (Conv1D)	(None, 28, 64)	448
max_pooling1d_1 (MaxPooling1D)	(None, 14, 64)	0
flatten_1 (Flatten)	(None, 896)	0
dense_2 (Dense)	(None, 128)	1,14,816
dense_3 (Dense)	(None, 128)	16,512
dense_4 (Dense)	(None, 1)	129
Total params	**1,31,905**	**(514 KB)**
Trainable params	**1,31,905**	**(514 KB)**
Non-trainable params	**0**	**(0.00 Byte)**

A. Performance Evaluation

For comparison the SVM, Random Forest Classifier and Faster R-CNN methods performance is compared with the proposed Attn-Based Convolutional Bi-GRU model. Precision, recall, F1-score, and accuracy were utilized as evaluation criteria for comparison. The model was evaluated using these parameters.

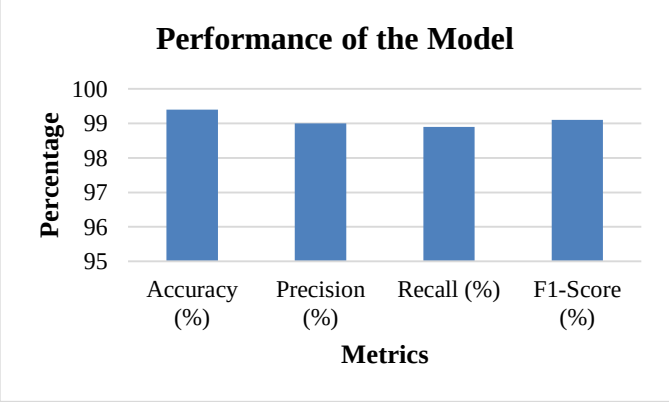


Fig. 3. Performance of the proposed attn-based convolutional Bi-GRU model.

A novel pedestrian recognition method is assessed using an Attn-Based Convolutional neural networks Bi-GRU model; the outcomes are displayed in Fig. 3 of the suggested model's performance metrics. The proposed model works remarkably well, demonstrating its high overall consistency in pedestrian recognition with an accuracy of 99.4%. It also shows exceptional accuracy at 99%, suggesting a minimal amount of false identifications, and remarkable recall at 98.9%, exhibiting its ability to reliably identify the majority of genuine pedestrians. The model's remarkable pedestrian recognition ability is validated by the F1-Score, which stands for an ideal balance between recall and precision and is notably high at 99.1% (see Table II).

TABLE II. PERFORMANCE METRICS OF ATTN.-BASED CONVOLUTIONAL BI-GRU MODEL WITH EXISTING METHODS

Method	Accuracy	Precision	Recall	F1-Score
SVM [22]	82.3(%)	88.16(%)	72.92(%)	79.82(%)
Faster R-CNN [23]	84(%)	96.6(%)	92.6(%)	94.6(%)
Random Forest [22]	90.2(%)	90(%)	80.8(%)	85.18(%)
Proposed Attn-Based Convolutional Bi-GRU Model	99.4(%)	99(%)	98.9(%)	99.1(%)

Accuracy, precision, recall, and F1-Score are only a few of the performance variables used in the Table I to compare different models from machine learning. Despite a modest recall of 72.92%, the Support Vector Machine, also known as the SVM, accurately classifies positive cases, achieving an accuracy of 82.3% and a precision of 88.16%. With an accuracy of 84%, a far better precision of 96.6%, as well as a much better recall of 92.6%, the faster R-CNN algorithm performs better than SVM, resulting in a strong F1-Score of 94.6%. With an F1-Score of 85.18%, the Random Forest algorithm performs exceptionally well, with accuracy of 90.2%, good precision of 90%, and recall of 80.8%. But the most effective model is known as the Proposed Attn-Based Convolutional Bi-GRU Model, which achieves an amazing 99.4% accuracy, 99% precision, 98.9% recall, and a remarkable 99.1% F1-Score, indicating that it can accurately classify either positive or negative instances. This shows that the suggested model performs better than any other approaches, which makes it a viable option for the specified classifying task.

The proposed attention-based convolutional bi-GRU model was compared with SVM, Faster R-CNN, and Random Forest due to their established use and performance in pedestrian detection tasks. SVM was chosen for its robustness in classification but is limited in capturing complex patterns and temporal dependencies. Faster R-CNN, a deep learning model known for high accuracy in object detection, served as a strong benchmark, though it struggles with temporal context. Random Forest, an ensemble learning method, was included as it performs well in various recognition tasks but may not handle the spatial and temporal complexities of urban environments effectively. These comparisons highlight the superior performance of the proposed model, particularly in accuracy, precision, recall, and F1-score, demonstrating its potential to significantly enhance pedestrian detection in autonomous vehicle systems.

The proposed Attn.-Based Convolutional Bi-GRU Model, Random Forest, Faster R-CNN, and Fig. 4 shows the Support Vector Machine (SVM) over all the remainder of the four machine learning models together with their accuracy, precision, recall, and overall F1-Score performance metrics. Notably, the SVM produces an F1-Score of 79.82% accompanied by a somewhat low recall of 72.92%, a strong

precision of 88.16%, and a reasonable accuracy of 82.3%. Faster R-CNN, on the other hand, obtains an incredible F1-Score of 94.6% with an increase in accuracy of 84%, a significantly higher precision of 96.6%, and an astounding recall of 92.6%. With an accurate rate of 90.2%, the approach known as Random Forest achieves a great F1-Score of 85.18% by balancing recall at 80.8% and precision at 90%. On the other hand, the Proposed Attn.-Based Convolutional Bi-GRU Model is clearly the strongest performance, with an astonishing F1-Score of 99.1%, near-perfect precision of 99%, recall of 98.9%, and astounding accuracy of 99.4%. This graph demonstrates how the proposed model outperforms the other methods across the board, suggesting that it has a lot of potential for solving the present categorization challenge.

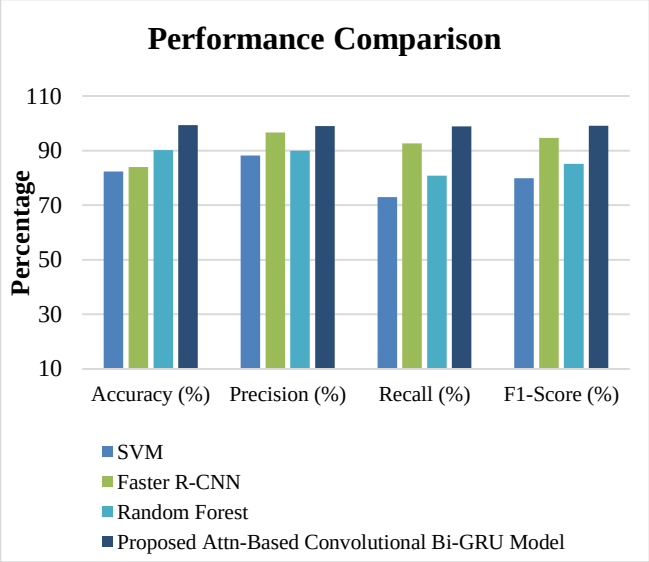


Fig. 4. Graphical depiction of the performance metrics of proposed Attn.-based convolutional Bi-GRU model with existing approaches.

Fig. 5 shows the training accuracy graph, within the framework of the proposed Attn-Based Convolutional Bi-GRU model, shows how well the model classifies pedestrians using the training dataset throughout its training iterations, or epochs. This graph shows how the model is learning and if it is becoming more accurate or if the training data may be overfitting. However, the effectiveness of the model on a separate, untested dataset that was not used for training is depicted on the testing accuracy graph. This graph sheds light on how well the model can use its knowledge of pedestrian detection to situations outside of the training set. A high testing accuracy shows the model's competence in consistently identifying pedestrians in a variety of real-world scenarios, and those is essential for improving protection for autonomous vehicle applications. A substantial training accuracy suggests that the framework successfully acquired from the training data.

Fig. 6 shows a graphical depiction of the training and testing loss for the proposed Attn-Based Convolutional Bi-GRU model, which illustrates how the model's loss function varies across the training and assessment stages.

Fig. 7 shows that the ROC curve depicts the performance of the model in binary classification tasks, especially for the proposed Attn-based Convolutional Bi-GRU model. When the threshold for classifying pedestrians is different, the sensitivity of the model changes, as detected by the ROC Curve. The model's ability to accurately identify pedestrians while avoiding false positives is reflected by a high number of true positives. In this case, the ROC curve generally exhibits a positive trend as the threshold increases from 0 to 0.7, indicating that the model is good at discriminating pedestrians from pedestrians, which is important for the safety of the application of the autonomous vehicle.

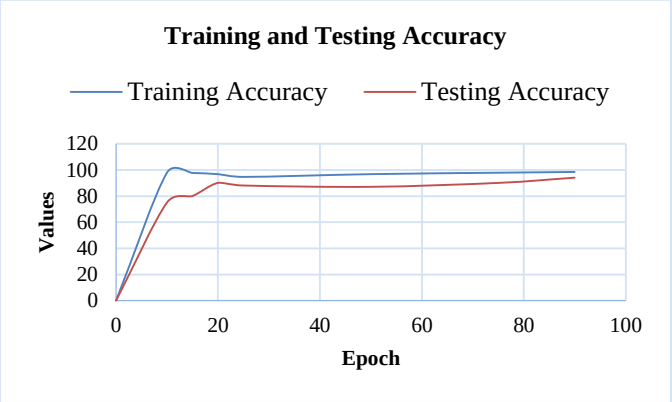


Fig. 5. Graphical depiction for training and testing accuracy of proposed Attn.-based convolutional Bi-GRU model.

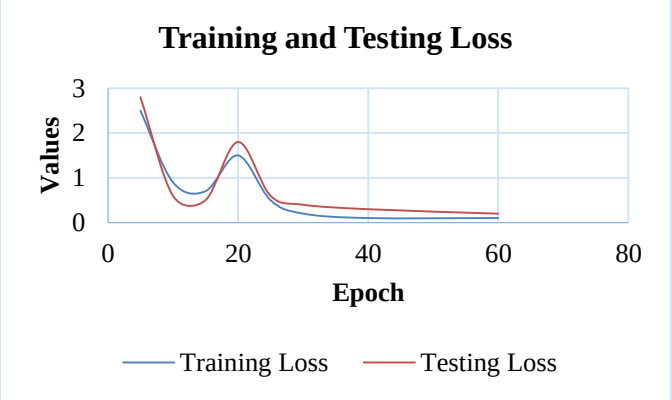


Fig. 6. Graphical depiction for training and testing loss of proposed Attn-based convolutional Bi-GRU model.

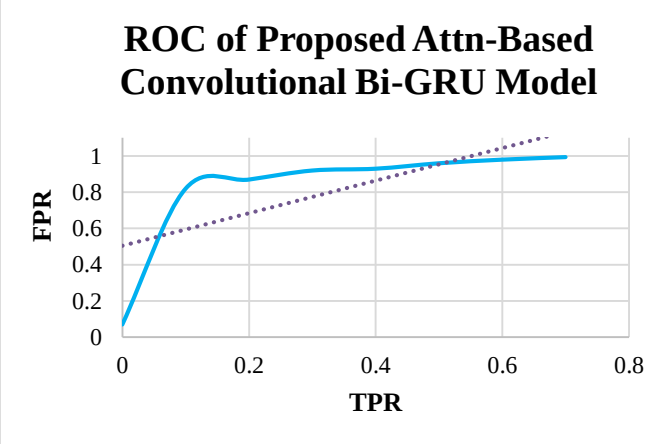


Fig. 7. ROC curve of proposed Attn-based convolutional Bi-GRU model.

B. Discussion

Improved road safety and autonomous vehicles were the two main outcomes of the study. When it comes to detecting pedestrians, the algorithm is exceptionally accurate at 99.4%, reducing the probability of accidents and increasing road safety Challenges in real-world urban environments, such as poorly lit areas, are prevented , and complex pedestrian behaviors, which are often too numerous for available algorithms to handle, and can pose risks to user safety [19]. Convolutional two-GRU architecture's attention methods allow the model to focus on relevant data, improving its performance in complex urban environments False alarm reduction with 99% accuracy ensures autonomous vehicles work more efficiently, safely and efficiently on public roads Advanced deep learning algorithms This project is a great illustration of how it can accelerate adoption and ultimately change the landscape of safe and efficient travel.

In pedestrian detection, attention mechanisms and Bidirectional Gated Recurrent Units (Bi-GRU) offer significant advantages by improving the model's ability to focus on relevant features and capture temporal dependencies. Attention mechanisms allow the model to emphasize critical regions of the input data, such as specific areas where pedestrians are likely to appear, enhancing the model's ability to differentiate between pedestrians and other objects or background noise. This selective focus leads to more accurate and reliable detection. Meanwhile, Bi-GRU units, which process data in both forward and backward directions, capture temporal patterns and contextual information from sequential data, such as video frames. This bidirectional approach helps the model understand the dynamic and evolving nature of pedestrian movements over time, improving its performance in complex and variable environments. Together, these techniques enhance the model's ability to accurately identify pedestrians and adapt to changing conditions, crucial for autonomous vehicle safety.

The proposed attention-based convolutional bi-GRU model outperforms existing models primarily due to its integrated approach, which leverages the strengths of both attention mechanisms and Bidirectional Gated Recurrent Units (Bi-GRU). The attention mechanism enhances the model's ability to focus on crucial parts of the input data, effectively distinguishing pedestrians from background elements and improving detection accuracy. Bi-GRU units, by capturing temporal dependencies from both directions, provide a comprehensive understanding of dynamic pedestrian movements, which is vital in real-world scenarios where pedestrians' positions and actions constantly evolve. However, potential weaknesses include the increased computational complexity and resource requirements associated with training and deploying such a sophisticated model. Additionally, while the model excels in controlled environments, its performance may degrade in highly unpredictable or extremely cluttered scenarios. Despite these limitations, the combined use of attention mechanisms and Bi-GRU units represents a significant advancement in pedestrian detection, offering a robust solution that enhances autonomous vehicle safety.

The research presented in this study, while demonstrating significant advancements in pedestrian detection accuracy, is subject to several limitations that warrant consideration. One primary limitation is the increased computational complexity associated with the attention-based convolutional bi-GRU model, which may restrict its real-time applicability in resource-constrained environments. Additionally, the model's robustness has primarily been tested in controlled scenarios, raising concerns about its performance in highly unpredictable or densely cluttered urban settings. Furthermore, the reliance on specific datasets for training and testing poses questions regarding the model's generalizability across different geographical regions and pedestrian behaviors. For future work, efforts should focus on optimizing the model's computational efficiency, potentially through the development of lightweight architectures or the use of hardware accelerators like GPUs or TPUs. Expanding the evaluation to include diverse environmental conditions and pedestrian behaviors will be crucial to ensure the model's robustness and applicability. Additionally, integrating the model with other sensor modalities, such as infrared or radar, and exploring multi-sensor fusion frameworks could further enhance its effectiveness and contribute to the overall safety of autonomous vehicles.

VI. Conclusion and Future Work

The use of Python as the implementation device is a crucial factor of the model's adaptability and applicability. This paper provides a comprehensive and beneficial technique to enhancing self-sustaining vehicle protection. The first and most important step in developing a strong pedestrian dataset is facts collecting. The subsequent level of pre-processing standardizes the information the use of Min-Max normalization to get it ready for the advanced model. The Attention-Based Convolutional Bi-GRU Model, a deep studying architecture with awesome pedestrian identity abilities, is the brains in the back of this tactic. The version consists of bidirectional gated recurrent units (Bi-GRU) for temporal context seize, attention mechanisms for emphasizing semantic records, and convolutional layers for extracting spatial traits. A 99.4% accuracy price is executed, that's an outstanding improvement over preceding models via a median of approximately 17.1%. The results are outstanding. The demanding situations of identifying pedestrians in elaborate and dynamic metropolitan contexts are greatly addressed by using this accomplishment, which will ultimately lead to safer self-sustaining vehicle operation. This method offers a promising first step in the direction of the introduction of safer and extra dependable self-reliant motors, which may lessen dangers and improve the security of pedestrians and other street customers as autonomous transportation structures maintain to broaden. To increase the Attention-Based Convolutional Bi-GRU Model's performance and adaptability to changing real-world settings, extra improvements and adjustments may be investigated in future study.

References

[1] J. Zhang et al., "An infrared pedestrian detection method based on segmentation and domain adaptation learning," Comput. Electr. Eng., vol. 99, p. 107781, 2022, doi: https://doi.org/10.1016/j.compeleceng.2022.107781.

[2] P. Jabłoński, J. Iwaniec, and W. Zabierowski, "Comparison of Pedestrian Detectors for LiDAR Sensor Trained on Custom Synthetic, Real and Mixed Datasets," Sensors, vol. 22, no. 18, Art. no. 18, Jan. 2022, doi: 10.3390/s22187014.

[3] S. Iftikhar, Z. Zhang, M. Asim, A. Muthanna, A. Koucheryavy, and A. A. Abd El-Latif, "Deep Learning-Based Pedestrian Detection in Autonomous Vehicles: Substantial Issues and Challenges," Electronics, vol. 11, no. 21, p. 3551, 2022.

[4] J. Kolluri and R. Das, "Intelligent multimodal pedestrian detection using hybrid metaheuristic optimization with deep learning model," Image Vis. Comput., vol. 131, p. 104628, 2023, doi: https://doi.org/10.1016/j.imavis.2023.104628.

[5] W. Wang, X. Li, X. Lyu, T. Zeng, J. Chen, and S. Chen, "Multi-Attribute NMS: An Enhanced Non-Maximum Suppression Algorithm for Pedestrian Detection in Crowded Scenes," Appl. Sci., vol. 13, no. 14, Art. no. 14, Jan. 2023, doi: 10.3390/app13148073.

[6] "NMS-Loss: Learning with Non-Maximum Suppression for Crowded Pedestrian Detection | Proceedings of the 2021 International Conference on Multimedia Retrieval," ACM Conferences. Accessed: Oct. 25, 2023. [Online]. Available: https://dl.acm.org/doi/10.1145/3460426.3463588

[7] H. Kulhandjian, J. Barron, M. Tamiyasu, M. Thompson, and M. Kulhandjian, "Pedestrian Detection and Avoidance at Night Using Multiple Sensors and Machine Learning," in 2023 International Conference on Computing, Networking and Communications (ICNC), 2023, pp. 165–169. doi: 10.1109/ICNC57223.2023.10074081.

[8] D. Yang, H. Zhang, E. Yurtsever, K. A. Redmill, and Ü. Özgüner, "Predicting Pedestrian Crossing Intention With Feature Fusion and Spatio-Temporal Attention," IEEE Trans. Intell. Veh., vol. 7, no. 2, pp. 221–230, 2022, doi: 10.1109/TIV.2022.3162719.

[9] U. Gawande, K. Hajari, and Y. Golhar, "Real-Time Deep Learning Approach for Pedestrian Detection and Suspicious Activity Recognition," Procedia Comput. Sci., vol. 218, pp. 2438–2447, 2023, doi: https://doi.org/10.1016/j.procs.2023.01.219.

[10] H. Xu, M. Guo, N. Nedjah, J. Zhang, and P. Li, "Vehicle and Pedestrian Detection Algorithm Based on Lightweight YOLOv3-Promote and Semi-Precision Acceleration," IEEE Trans. Intell. Transp. Syst., vol. 23, no. 10, pp. 19760–19771, 2022, doi: 10.1109/TITS.2021.3137253.

[11] A. M. Obeso, J. Benois-Pineau, M. S. G. Vázquez, and A. Á. R. Acosta, "Visual vs internal attention mechanisms in deep neural networks for image classification and object detection," Pattern Recognit., vol. 123, p. 108411, 2022, doi: https://doi.org/10.1016/j.patcog.2021.108411.

[12] H. Lv, H. Yan, K. Liu, Z. Zhou, and J. Jing, "YOLOv5-AC: Attention Mechanism-Based Lightweight YOLOv5 for Track Pedestrian Detection," Sensors, vol. 22, no. 15, Art. no. 15, Jan. 2022, doi: 10.3390/s22155903.

[13] X. Zhou and L. Zhang, "SA-FPN: An effective feature pyramid network for crowded human detection," Appl. Intell., vol. 52, no. 11, pp. 12556–12568, Sep. 2022, doi: 10.1007/s10489-021-03121-8.

[14] X. Chen, Y. Jia, X. Tong, and Z. Li, "Research on Pedestrian Detection and DeepSort Tracking in Front of Intelligent Vehicle Based on Deep Learning," Sustainability, vol. 14, no. 15, Art. no. 15, Jan. 2022, doi: 10.3390/su14159281.

[15] Y. Lu, W. Wang, X. Hu, P. Xu, S. Zhou, and M. Cai, "Vehicle trajectory prediction in connected environments via heterogeneous context-aware graph convolutional networks," IEEE Trans. Intell. Transp. Syst., vol. 24, pp. 8452–8464, Aug. 2023, doi: 10.1109/TITS.2022.3173944.

[16] A. Khan, M. M. Fouda, D.-T. Do, A. Almaleh, and A. U. Rahman, "Short-Term Traffic Prediction Using Deep Learning Long Short-Term Memory: Taxonomy, Applications, Challenges, and Future Trends," IEEE Access, vol. 11, pp. 94371–94391, 2023, doi: 10.1109/ACCESS.2023.3309601.

[17] I. V. Pustokhina, D. A. Pustokhin, T. Vaiyapuri, D. Gupta, S. Kumar, and K. Shankar, "An automated deep learning based anomaly detection in pedestrian walkways for vulnerable road users safety," Saf. Sci., vol. 142, p. 105356, Oct. 2021, doi: 10.1016/j.ssci.2021.105356.

[18] M. Islam, M. Rahman, M. Chowdhury, G. Comert, E. D. Sood, and A. Apon, "Vision-Based Personal Safety Messages (PSMs) Generation for Connected Vehicles," IEEE Trans. Veh. Technol., vol. 69, no. 9, pp. 9402–9416, 2020, doi: 10.1109/TVT.2020.2982189.

[19] T. Aminah, "FRAMEWORK FOR PEDESTRIAN WALKING BEHAVIOUR RECOGNITION TO MINIMIZE ROAD ACCIDENT".

[20] "Pedestrian Dataset." Accessed: Oct. 25, 2023. [Online]. Available: https://www.kaggle.com/datasets/smeschke/pedestrian-dataset

[21] J. Thaker, N. Jadav, S. Tanwar, P. Bhattacharya, and H. Shahinzadeh, Ensemble Learning-based Intrusion Detection System for Autonomous Vehicle. 2022. doi: 10.1109/SCIoT56583.2022.9953697.

[22] X. Peng and J. Shan, "Detection and Tracking of Pedestrians Using Doppler LiDAR," Remote Sens., vol. 13, no. 15, p. 2952, Jul. 2021, doi: 10.3390/rs13152952.

[23] C. Amisse, M. E. Jijón-Palma, and J. A. S. Centeno, "FINE-TUNING DEEP LEARNING MODELS FOR PEDESTRIAN DETECTION," Bol. Ciênc. Geodésicas, vol. 27, no. 2, p. e2021013, 2021, doi: 10.1590/s1982-21702021000200013.

Cryptographic Techniques in Digital Media Security: Current Practices and Future Directions

Gongling ZHANG
Luoyang Cultural Tourism Vocational College, Luoyang 471000, China

***Abstract*—Content privacy and unauthorized access to copyrighted digital media content are common in the dynamic, fast-paced digitalized media marketplace. Cryptographic methods are the foundation of modern digital media security, and they must ensure the security, integrity, and authenticity of digital media data. This article analyses cryptographic methods that are used to protect digital media content. The paper reviews the main cryptographic concepts, such as symmetric cryptography, asymmetric cryptography, hash functions, and digital signatures. The paper also discusses some popular approaches: encryption, Digital Rights Management (DRM), watermarking, and solutions based on blockchain. Finally, we highlight implementation challenges such as key management and scalability and identify emerging trends such as quantum-safe cryptography and privacy-preserving techniques. By presenting the current research results and discussing the directions for the future, the study aims to pave the way for secure, efficient, and robust cryptographic solutions for digital media protection, leading to sustainable development, innovation, and secure communication of digital content among users.**

***Keywords*—Digital media; cryptographic; content security; digital rights management; watermarking, blockchain**

I. Introduction

Digital content is now more dominant than tangible media in terms of dissemination and usage [1]. Inherent properties such as the ability to reproduce identical documents easily and disseminate them over the Internet through wired and wireless communication have relevant ramifications for intellectual property rights [2]. Thus, the sharing of the content generally takes place beyond the limits of copyright law [3]. Digital products have the potential to be copied, replicated, and distributed across the globe within minutes of issuance, making detection and enforcement very difficult [4]. The illegal redistribution and exploitation of intellectual information may result in substantial financial losses for content producers and owners, with industry estimates suggesting that these losses amount to billions of dollars each year [5].

Social media platforms facilitating global communication and information dissemination have become breeding grounds for image sharing [6]. With billions of images uploaded daily, concerns about illegal access, manipulation, and unauthorized distribution are growing. Applying digital image watermarks is emerging as a promising technique to address these security challenges [7]. However, effective watermarking depends on three crucial requirements: imperceptibility, robustness, and embedding capacity. The watermarked image should be visually indistinguishable from the original image so as not to impact the user experience [8]. The watermark should resist attacks such as compression, noise, or cropping to ensure its durability and accuracy and protect copyright. The watermark should embed sufficient information to reliably identify the owner or copyright holder. These requirements often have trade-offs. For example, spatial domain watermarking techniques provide high embedding capacity and imperceptibility but are not robust to manipulation. Conversely, spectral domain techniques achieve higher robustness but may result in visible artifacts that impact imperceptibility.

To achieve an optimal balance between these competing demands, researchers have explored hybrid approaches that leverage both spatial and spectral domains. Nature-inspired metaheuristic optimization algorithms were employed to optimize the embedding strength of the watermark, aiming to find a balance between imperceptibility, robustness, and embedding capacity [9, 10]. However, developing effective fitness functions that balance exploration (searching for novel solutions) and exploitation (refining promising solutions) remains an ongoing challenge.

Digital transformation, driven by the Internet of Things (IoT) and the increasing demand for secure and real-time communication, fundamentally changes how we interact with information and the world around us [11]. IoT envisions a future in which nearly every object is connected to the Internet and generates and transmits massive amounts of data, including personal, sensitive, and confidential information [12]. While these advances offer unprecedented opportunities, they pose significant security challenges. Smart cities and advances like cryptocurrencies hold enormous potential to reshape the coming decade, but the inherent vulnerabilities of these technologies raise concerns about privacy and communications security [13].

The purpose of this article is to discuss the principles of cryptographic technologies that can be used to prevent unauthorized access to digital media content and its piracy. The article begins with an introduction to the importance of digital media security and the function of cryptography. Basic cryptographic principles of symmetric and asymmetric encryption, as well as hash functions and digital signatures, are examined. Various security strategies are then discussed, including encryption, Digital Rights Management (DRM), watermarking, and blockchain security solutions. The discussion also addresses the difficulties and disadvantages of applying these approaches, new trends and promoting possible developments. Finally, the paper offers a brief conclusion with key findings and suggestions for where further research should be conducted.

II. BACKGROUNDS

Information security protects data from unauthorized access, theft, alteration, or destruction. This covers various information formats, including text, images, audio, and video [14]. Two main approaches are used to ensure information security: cryptography and steganography. Cryptography uses complex mathematical algorithms to encrypt or scramble data into an unreadable format (ciphertext). This makes the data unreadable even if unauthorized persons intercept it. However, encryption itself can sometimes signal valuable information. To counteract this limitation, steganography hides communication itself. Steganography embeds confidential messages in a seemingly innocuous cover object, such as an image or audio file. This allows data to be passed on unnoticed by eavesdroppers. Even if the cloak is intercepted, the information remains hidden, providing an additional layer of security. Fig. 1 illustrates the general process of digital media content encryption. A variety of multimedia encryption applications are also illustrated in Fig. 2.

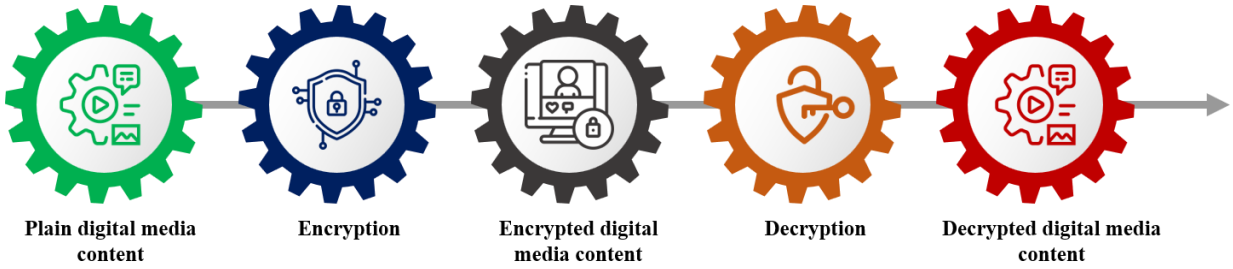


Fig. 1. General process of digital media content encryption.

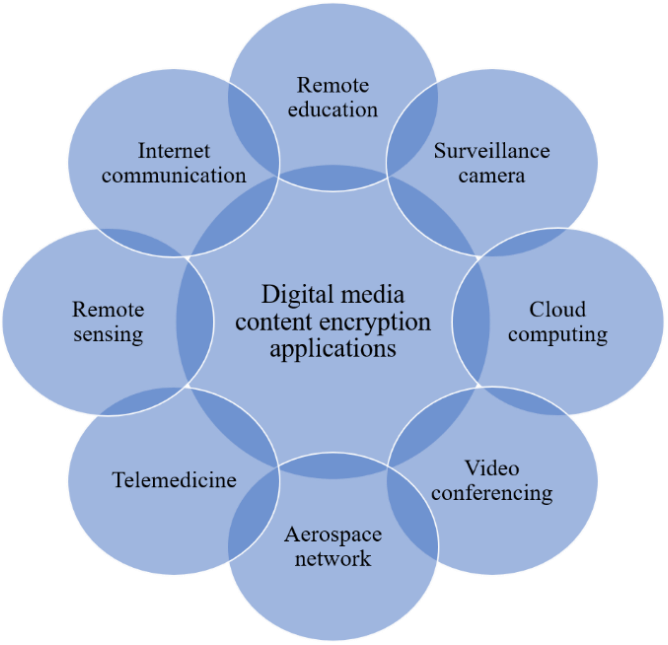


Fig. 2. Applications of multimedia encryption in various domains.

A. Symmetric Cryptography

This is also called secret-key cryptography, where the same secret key is used for encryption and decryption. This key should be kept secret by both the sender and the receiver to enhance the security of the passed information. The basis for the use of symmetric cryptography is the secrecy of the key and the choice of the algorithm used to encrypt the data. Symmetric encryption algorithms are generally classified into block ciphers and stream ciphers [15]. Block ciphers, like the Advanced Encryption Standard (AES), process data in fixed-size blocks during encryption. In contrast, stream ciphers, such as Rivest Cipher 4 (RC4), encrypt data one bit or byte at a time. The selection between block and stream ciphers depends on the application's requirements, such as prioritizing high-speed encryption or resistance to specific attack types. Symmetric cryptography underpins the security of various digital media content due to its efficiency and speed. A comparison of common symmetric encryption algorithms is presented in Table I.

- Content encryption: To prevent unauthorized access, digital media files (videos, audio, images) are often encrypted using symmetric algorithms. Content providers can ensure that only authorized users with the corresponding decryption key can access the media. For instance, streaming services like Netflix and Spotify might leverage the Advanced Encryption Standard (AES) to protect their media files during transmission and storage.

- DRM: Symmetric cryptography is critical in enforcing DRM systems' access controls and usage policies. The media content is encrypted with a symmetric key securely distributed only to authorized users. The DRM system manages key distribution and revocation, guaranteeing that content decryption and access are restricted to legitimate users.
- Secure streaming: Real-time encryption of media content before transmission is essential for live streaming scenarios. Due to their ability to encrypt data on the fly, stream ciphers are particularly well-suited for this purpose. Encrypted streams are transmitted over networks, ensuring intercepted data remains unintelligible without the decryption key.
- Storage encryption: Symmetric encryption safeguards digital media stored on servers or user devices. By encrypting media files at rest, service providers can prevent unauthorized access in the event of a data breach or physical theft of storage devices.

TABLE I. SYMMETRIC CRYPTOGRAPHY

Algorithm	Type	Key size (bits)	Block size (bits)	Example use cases
AES	Block cipher	128, 192, 256	128	Secure file transfer and encryption of sensitive data
DES	Block cipher	56	64	Legacy systems, previously used in banking transactions
RC4	Stream cipher	40-2048	Variable	Secure network protocols (e.g., WEP, SSL/TLS)
3DES	Block cipher	112, 168	64	Financial services and legacy systems requiring higher security

B. Asymmetric Cryptography

Asymmetric cryptography, also known as public-key cryptography, utilizes a mathematically elegant solution for secure communication by employing two distinct keys for encryption and decryption: a public key and a private key [16]. Unlike symmetric cryptography, the public key is openly distributed, allowing anyone to encrypt data intended for a specific recipient. The private key, conversely, is meticulously guarded by its owner and serves as the sole means to decrypt the data. This clear separation of keys addresses a fundamental challenge in symmetric cryptography: secure key distribution. Public-key cryptographic systems are built upon mathematical problems that are difficult to solve practically without possessing the corresponding private key. Well-established algorithms in this domain include RSA (Rivest-Shamir-Adleman), ECC (Elliptic Curve Cryptography), and DSA (Digital Signature Algorithm). These algorithms mathematically guarantee that data encrypted with a public key can only be decrypted by the holder of the corresponding private key, offering a secure method for exchanging information even over untrusted channels. Asymmetric cryptography is critical in securing digital media content, and enhancing security and user trust. It provides several key advantages over symmetric cryptography, particularly in areas where secure key distribution and verification of content authenticity are paramount. Table II presents a comparison of common asymmetric encryption algorithms.

TABLE II. ASYMMETRIC CRYPTOGRAPHY

Algorithm	Key type	Key size (bits)	Security level	Example use cases
RSA	Public/private	1024, 2048, 4096	High	Secure email, digital signatures, and SSL/TLS
ECC	Public/private	160-521	Very high	Mobile devices, digital signatures, and secure communications
DSA	Public/private	1024-3072	High	Digital signatures and secure document verification
ElGamal	Public/private	256-4096	High	Encrypted key exchange and digital signatures

- Secure key exchange: A primary application is a secure key exchange for symmetric encryption. Even if attackers intercept communication channels, the asymmetric system ensures the confidentiality of symmetric keys to encrypt media content. For instance, streaming services might leverage asymmetric cryptography to exchange encryption keys securely during the initial connection setup.
- Digital signatures and content authentication: Digital signatures, created using asymmetric cryptography, verify the authenticity and integrity of digital media. When content creators sign a media file with their private key, anyone can verify the signature using the corresponding public key. This ensures the content originates from the claimed source and has not been tampered with. This mechanism is widely used in software distribution and updates to guarantee the authenticity of digital media.
- Certificate-based authentication and secure communication: Digital certificates issued by trusted authorities rely on asymmetric cryptography to authenticate digital media distribution and consumption entities. These certificates ensure users connect to legitimate services and that media providers distribute content securely. For example, HTTPS protocols use certificates to secure communication between web browsers and streaming services.
- Enhanced DRM: Asymmetric cryptography strengthens DRM systems by facilitating secure key management and distribution. Public-key cryptography can encrypt content keys and distribute them only to authorized devices. These devices then use their private keys to decrypt the content keys and access the media. This restricts access to licensed users and devices, preventing unauthorized distribution and piracy.
- Secure content delivery with Content Distribution Networks (CDNs): CDNs leverage asymmetric cryptography to secure digital media distribution across

multiple servers. By encrypting media files with public keys and using private keys for decryption at endpoint servers, CDNs safeguard the integrity and confidentiality of media during transit, mitigating the risk of interception or tampering.

C. Hash Functions

Hash functions are cryptographic primitives that map arbitrary-length inputs to fixed-length outputs, typically represented as hexadecimal strings. These output values, or hash codes or values, possess the crucial uniqueness property. The corresponding hash value will be unique for any given input, and even minor alterations will result in a significantly dissimilar hash value. This phenomenon is known as the avalanche effect. Additionally, hash functions are deterministic, ensuring that identical inputs consistently produce the same hash value. Table III provides a comparison of common cryptographic hash functions.

TABLE III. HASH FUNCTIONS

Algorithm	Output size (bits)	Security level	Use cases
MD5	128	Insecure	Legacy systems and checksum verification
SHA-1	160	Insecure	Legacy systems and integrity checks
SHA-256	256	Secure	Digital signatures and integrity verification
SHA-512	512	Very secure	High-security applications and blockchain

A confluence of critical properties characterizes effective cryptographic hash functions. Determinism guarantees that identical inputs invariably produce the same hash value. Additionally, computation efficiency is paramount for real-world applications. Two fundamental properties ensure the robustness of hash functions against cryptographic attacks. Preimage resistance makes it computationally infeasible to retrieve the original input solely from the hash value. Collision resistance, on the other hand, safeguards against the possibility of finding two distinct inputs that generate the same hash output.

Furthermore, hash functions generate fixed-length outputs irrespective of the input size, enhancing their efficiency and facilitating comparisons. Common cryptographic hash functions include MD5 (though currently considered insecure) and SHA-1 (deprecated). The SHA-2 family, encompassing algorithms like SHA-256 and SHA-512, is now the recommended standard for secure hashing. Hash functions serve as a cornerstone for guaranteeing the integrity of digital media content. They provide a mechanism to verify that data remains unaltered or uncorrupted during transmission, storage, or manipulation.

- Data integrity verification: Hash functions safeguard digital media's integrity during transfers or storage across networks and diverse locations. By computing a hash value (unique digital fingerprint) for the original content, subsequent comparisons with the hash of the received or stored file can expose any modifications. This approach is essential for upholding trust in digital media distribution channels, cloud storage solutions, and content delivery networks.
- Digital signatures and certificates: Hash functions are instrumental in creating digital signatures, a cornerstone of digital media security. When a digital signature is generated, the hash of the content is encrypted with the sender's private key. This encrypted hash, often called a digital signature, is transmitted with the media. The recipient can decrypt the signed hash using the sender's public key and compare it to the hash value computed from the received content. If the hash values match, it confirms the content's integrity and authenticity, signifying that it remains unaltered since the signature was created.
- Content verification in blockchain: Blockchain technology, with its potential for secure digital media management, leverages hash functions to verify the integrity of content stored on the blockchain. Each block within a blockchain incorporates a hash value referencing the preceding block, effectively creating an immutable chain of records [17]. This immutability ensures that any attempt to tamper with a single piece of content would necessitate altering all subsequent blocks, rendering such efforts highly impractical.
- Deduplication and data management: Hash functions are valuable in identifying duplicate files within extensive digital media libraries. Systems can efficiently detect and manage duplicate content by generating and comparing hash values for various files. This optimization translates to improved storage utilization and streamlined retrieval processes.
- Integrity checks in DRM systems: DRM systems rely on hash functions to safeguard the integrity of protected media content. Verifying that the content has not been tampered with is crucial for maintaining the integrity of protected materials and ensuring compliance with licensing agreements.

D. Digital Signatures

Digital signatures are cryptographic mechanisms underpinning the verification of authenticity and integrity in digital messages and documents. Analogous in function to handwritten signatures, they offer a significantly enhanced level of security. The process of creating a digital signature leverages public-key cryptography. A unique signature is generated by encrypting a cryptographic hash of the message or document with the sender's private key. This signature can be validated by anyone possessing the corresponding public key, assuring that the document has not been tampered with and confirming the sender's identity. Common digital signature algorithms are listed in Table IV.

Digital signatures offer crucial properties: authentication, integrity, and non-repudiation. Authentication verifies the sender's identity, guaranteeing that the message or document originates from a legitimate source. Integrity, achieved through hash functions, ensures that the content has not been altered during transmission. Any modification to the content will result in a failed verification process due to a mismatch between the

calculated hash and the one embedded within the signature. Finally, non-repudiation prevents the sender from denying having sent a digitally signed document, thereby establishing legal validity and accountability. Digital signatures are a cornerstone for securing digital media content across various distribution and access channels. Their ability to verify authenticity and integrity fosters trust and combats potential security threats.

TABLE IV. DIGITAL SIGNATURES

Algorithm	Key type	Key size (bits)	Security level	Example use cases
RSA	Public/private	1024, 2048, 4096	High	Document signing and software distribution
ECDSA	Public/private	160-521	Very high	Secure transactions and blockchain
DSA	Public/private	1024-3072	High	Secure communications and legal document verification
EdDSA	Public/private	256-512	Very high	Secure messaging and financial transactions

- Content distribution verification: When digital media like music, videos, or e-books are distributed online, digital signatures safeguard their authenticity. Content creators or distributors can cryptographically sign their media files. Users can then verify these signatures using the publicly available content creator key. This process ensures the content hasn't been tampered with or corrupted during distribution, protecting users from unknowingly acquiring compromised media.
- Software and firmware updates: Digital signatures are critical in software and firmware updates for digital media devices like streaming boxes, smart TVs, and gaming consoles. Manufacturers sign update files with their private keys. Devices verify these signatures before installing the updates, guaranteeing that only legitimate and unaltered updates are applied. This mitigates the risk of installing malicious software disguised as updates.
- Blockchain and content authentication: With its potential for secure digital media management, Blockchain technology utilizes digital signatures for transaction authentication and ownership verification of digital assets. The content owner signs every transaction involving digital media content, and these signatures are recorded on the blockchain. This creates a transparent and tamper-proof record of ownership and distribution history.
- Secure online transactions: In e-commerce platforms selling digital media like music, videos, and software, digital signatures guarantee the authenticity of the purchased content. Customers can verify that the digital products they receive are genuine and unaltered. This enhances trust and transparency in the transaction process for consumers and vendors.
- Copyright protection and legal evidence: Digital signatures offer legal proof of ownership and authenticity, which is crucial for copyright protection and legal disputes. Content creators can leverage digital signatures to establish their rights over their creations. These signatures can then be presented as evidence in court cases related to copyright infringement.

III. REVIEW OF APPROACHES

The widespread adoption of image formats, especially JPEG, has opened avenues for embedding additional information within these files. While stenographic techniques utilizing the least significant bits have been dominant, this research proposes alternative methods. Harran, et al. [18] demonstrated the feasibility of incorporating a digital certificate alongside its corresponding metadata directly into an image file. This metadata references the issuing entity responsible for the certificate. Despite variations across devices, operating systems, and applications, JPEG files exhibit remarkable structural consistency. The proposed approach strategically inserts references to the issuing company within the file's metadata. This integration offers a distinct advantage: the digital certificate remains tethered to the file it applies to, ensuring it travels together throughout the file's lifecycle. The research ultimately establishes the potential of file metadata to house additional data that bolsters the integrity, authenticity, and provenance of the digital content embedded within the file. This approach paves the way for innovative methods to secure and verify digital content using existing file structures.

Table V summarizes the proposed approaches for embedding and securing digital media, highlighting key techniques, advantages, and use cases. Gurunathan and Rajagopalan [19] proposed a steganographic technique for embedding secret messages within a cover image using LSB substitution to evade detection by potential interceptors. This work builds upon the core concept of LSB substitution but introduces modifications to improve image quality and message capacity while maintaining security. Inspired by existing approaches that divide cover images into blocks, the proposed method partitions the cover image and the secret message into equal-sized blocks (typically 8x8 pixels). This strategy aims to achieve a balance between embedding capacity and image quality. A key innovation lies in utilizing the Cuckoo Search (CS) algorithm. Unlike prior methods that employ a single substitution matrix for the entire image, the proposed approach leverages CS to find an optimal substitution matrix for each block. This approach aims to achieve a more nuanced embedding process, optimizing message concealment within each block while minimizing visual artifacts in the resulting stego-image (the image containing the hidden message). The final stage involves evaluating the quality of the stego-image, the message capacity, and the security level of the proposed method. These metrics are then compared to existing techniques based on the Joint Photographic Experts Group (JPEG) standard and Joint Quantization Table Modification (JQTM). Experimental results, as reported by the authors, demonstrate that the proposed method surpasses both JPEG and JQTM-based methods in terms of image quality, security level, and message embedding capacity.

TABLE V. PROPOSED APPROACHES FOR EMBEDDING AND SECURING DIGITAL MEDIA

Authors	Approach	Key techniques	Advantages
Harran, et al. [18]	Embedding digital certificates in JPEG files	Metadata embedding	Ensures certificate travels with the file, enhances integrity and authenticity
Gurunathan and Rajagopalan [19]	Steganography using LSB substitution with Cuckoo Search algorithm	LSB substitution, Cuckoo Search	Improves image quality and message capacity, optimizes message concealment
Gafsi, et al. [20]	Hybrid image encryption using asymmetric and symmetric cryptography	RSA, AES-256, SHA-2	High security, robust against cryptanalysis attacks
Panchal, et al. [22]	Document security using fingerprint biometrics	Biometric feature extraction, convolution coding	High true positive rate, no need to store encryption keys
William, et al. [21]	Hybrid cryptographic solution merging AES, ECC, and SHA-256	AES, ECC, SHA-256	Enhanced efficiency for text encryption, secure data integrity
Yasser, et al. [23]	Multimedia encryption using chaotic dynamics and 2D alteration models	Chaotic maps, hybrid chaotification	Strong key sensitivity, high resistance to attacks
Sanivarapu, et al. [24]	Image watermarking scheme using cryptographic techniques and QR code scrambling	QR code, DWT, SVD, chaotic logistic map	Resilient to image processing attacks, maintains minimal visual distortion
Alarifi, et al. [25]	Hybrid cryptosystem for securing HEVC video streams	DNA sequences, Arnold chaotic map, Mandelbrot sets	Robust and enhanced security for HEVC video streaming

Gafsi, et al. [20] presented a novel image encryption system designed to achieve high security for digital images. Their approach leverages a combination of asymmetric and symmetric cryptography to provide robust protection. The asymmetric component utilizes the well-established RSA algorithm. RSA employs a public key for encryption and a private key for decryption, ensuring secure key distribution and management. However, image encryption relies on the AES-256 algorithm in Counter (CTR) mode. This combination offers a strong foundation for image data encryption. Furthermore, the system incorporates the SHA-2 hashing function. SHA-2 serves as a cryptographic hash function, generating a unique fingerprint of the original image data. This fingerprint can be used for integrity verification, ensuring the image has not been tampered with during encryption or decryption. The effectiveness of the proposed system was evaluated using various established tools and tests commonly employed within the image cryptography community. These tests utilized a diverse set of standard, non-compressed images. The experimental and analytical results indicate that the encryption scheme offers robustness and resistance against known cryptanalysis attacks. These positive results suggest the proposed method achieves high performance and efficiency, making it suitable for applications requiring strong image protection in various domains, such as military communication and securing sensitive data for personal privacy.

William, et al. [21] proposed a novel cryptographic solution that merges three distinct cryptographic primitives: a symmetric algorithm (AES), an asymmetric algorithm (ECC), and a hash function (SHA-256). SHA-256 is a cryptographic hash function that generates a unique message digest from the input data. This digest is a fingerprint to verify data integrity and expose potential tampering attempts. The proposed hybrid approach resembles existing techniques that leverage AES for encrypting textual and graphical data. However, the authors posit that their solution offers enhanced efficiency, particularly text encryption, compared to prior methods. While acknowledging the current limitations in image encryption speed, they suggest that future advancements could optimize the solution for improved image encryption performance.

Panchal, et al. [22] proposed a novel document security mechanism that leverages fingerprint biometrics. This system extracts unique features from a user's fingerprint captured by a biometric sensor. These features are then processed using convolution coding principles to generate a unique code. This unique code is the foundation for creating cryptographic keys for encrypting and decrypting user documents. A rigorous evaluation of the proposed approach, involving experimentation with various standard fingerprint images within a database, yielded impressive results. The system achieved a high true positive rate of 95%, indicating accurate identification of authorized users.

Furthermore, the system yielded a 0% false negative rate, ensuring no instances where authorized users were mistakenly denied access. This system offers several significant advantages. Firstly, it generates a unique key for each user, eliminating the need to store a central template of biometric data, which can be a security vulnerability. Secondly, the system avoids storing any encryption keys, further enhancing security. Finally, the reported efficiency suggests the system is suitable for real-world applications due to its speed and accuracy. These qualities make it a promising solution for developing robust data storage security systems.

Yasser, et al. [23] introduce novel multimedia encryption schemes that leverage chaotic dynamics and 2D alteration models to achieve high-security data transmission. Their approach revolves around a new perturbation-based data encryption method applicable to confusion and diffusion rounds. The core novelty lies in the hybrid chaotification structure, which combines multiple chaotic maps for enhanced media encryption. These blended maps generate control parameters for the permutation (shuffling) and diffusion (substitution) stages within the encryption process. The proposed schemes maintain the high encryption quality characteristic of chaotic systems and boast additional advantages. These include strong key sensitivity, resistance to unauthorized key derivation, and low residual clarity, minimizing the potential for intelligible information leakage from the encrypted media. Extensive security and differential analyses demonstrate the efficacy of the proposed schemes for securing multimedia transmissions. The encrypted media exhibits a high degree of resistance against various attacks. Additionally, statistical evaluations using established metrics for specific media types reveal that the schemes achieve low

residual intelligibility while maintaining statistically sound properties in the recovered data.

Sanivarapu, et al. [24] introduced a novel image watermarking scheme that utilizes cryptographic techniques to ensure copyright protection and content authentication. Their method centers around a watermark image containing a public-key/private-key pair generated through a cryptosystem. This watermark is then encoded into a quick response (QR) code. The QR code is scrambled using a chaotic logistic map to enhance security. The public and private keys serve a dual purpose: encrypting the data embedded within the watermark and facilitating its decryption during extraction. The scrambled QR watermark is then embedded into a color image using a single-level discrete wavelet transform (DWT) followed by singular value decomposition (SVD). The key plays a crucial role in this embedding process. Watermark extraction entails reversing the steps involved in embedding. The proposed method's effectiveness is evaluated through its resilience to various image-processing attacks commonly employed to remove watermarks. The authors compare their results with those achieved by state-of-the-art watermarking schemes, demonstrating that their method balances robustness (resistance to attacks) and imperceptibility (minimal visual distortion of the host image).

The burgeoning adoption of big data processing, cloud computing, and the IoT has fueled a surge in multimedia information consumption, particularly video. Within the Internet of Multimedia Things (IoMT), video is extensively streamed over communication networks, necessitating robust security measures. Unfortunately, existing methods for securing multimedia content transmission between cloud platforms and mobile devices often face limitations due to processing overhead, memory constraints, data size considerations, and battery power limitations on mobile devices. These limitations render such methods suboptimal for large multimedia files and unsuitable for the resource-restricted nature of mobile devices and cloud environments. High-Efficiency Video Coding (HEVC) is the latest video codec standard, enabling efficient storage and streaming of high-resolution videos while maintaining acceptable file sizes and superior quality. In this context, Alarifi, et al. [25] proposed a novel hybrid cryptosystem designed to safeguard the streaming of compressed HEVC video streams. This cryptosystem leverages a combination of Deoxyribonucleic Acid (DNA) sequences, the Arnold chaotic map, and Mandelbrot sets. The secure video transmission process commences with video encoding using the H.265/HEVC codec to achieve efficient compression. Subsequently, the proposed method employs the Arnold chaotic map for individual encryption of the three-color channels (Y, U, and V) within each compressed HEVC frame. Following this initial encryption step, DNA encoding sequences are established upon the resulting frames. Finally, a modified conditional shift process based on the Mandelbrot set is introduced to further obfuscate the encrypted data within the Y, U, and V channels. The authors conducted extensive simulations and security analyses to validate the proposed HEVC cryptosystem. The results demonstrate exceptional robustness and enhanced security compared to existing cryptosystems documented in the literature. This approach offers a promising solution for securing HEVC video streaming in resource-constrained environments.

IV. Results and Discussion

Because of its efficiency and speed, symmetric cryptography remains the cornerstone of digital media security. Symmetric algorithms such as AES and RC4 are widely used in content encryption, digital rights management, and secure streaming. Streaming services like Netflix use AES to protect media files during transmission and storage, ensuring that only authorized users can view the content. However, the biggest challenge is key management. The distribution and storage of secure keys represent a significant vulnerability, especially in large systems. Additionally, symmetric algorithms can be effective for real-time encryption, but their use of a single key raises security concerns if the key is compromised.

Asymmetric dual-key cryptography addresses some key management problems associated with symmetric cryptography. With public key cryptography, as demonstrated by RSA and ECC, encryption keys can be securely exchanged and content authenticity verified. Secure key exchange, digital signatures, and DRM enhancement rely heavily on asymmetric cryptography. Using asymmetric cryptography, digital certificates ensure users connect to legitimate services and media providers distribute content safely. Despite its advantages, asymmetric cryptography can be computationally intensive, which can be a problem for applications that require high-speed encryption.

Data integrity and authenticity are ensured in digital media security by hash functions. The SHA-256 feature is often used to create digital fingerprints for content, allowing unauthorized changes to be detected. Virtual signatures, blockchain content verification, and data deduplication in large media libraries are all based on hash functions. For example, blockchain technology secures digital media content by creating an immutable chain of records using hash functions. However, despite their effectiveness, hash functions have limitations. As computing power increases and new attack vectors emerge, collisions remain a problem, even with advanced algorithms like SHA-256.

The integrity and authenticity of digital content can be verified using digital signatures. In public key cryptography, digital signatures provide non-repudiation and ensure that the provenance and integrity of content can be independently verified. The review concluded that digital signatures make a significant contribution to verifying content distribution, software updates, and authentication of blockchain-based content. For example, a digital signature ensures that only legitimate software updates are distributed. This means there is no risk of devices becoming infected with malware. In resource-constrained environments, the implementation of digital signatures can be limited by the computational effort associated with signature generation and verification.

V. Challenges and Future Directions

Securing encryption keys remains one of the most significant challenges in cryptographic systems. In symmetric cryptography, securely distributing and storing keys can be problematic, especially as the number of users increases.

Although public keys can be distributed more freely in asymmetric systems, the private keys must be stored securely to prevent unauthorized access. The complexity of key management is further exacerbated in large-scale digital media systems where millions of users might be involved.

As digital media content and user bases grow, ensuring that cryptographic solutions scale effectively is crucial. High computational requirements for encryption and decryption can lead to performance bottlenecks, especially for real-time applications like live streaming. Finding a balance between robust security and system performance is essential for deploying cryptographic techniques in digital media systems.

Implementing strong cryptographic measures often introduces complexity for end-users. If accessing encrypted content or managing digital rights becomes too cumbersome, it can lead to poor user experience and lower adoption rates. Designing user-friendly cryptographic systems that provide robust security without compromising usability is a persistent challenge.

With many devices, platforms, and media formats, ensuring that cryptographic solutions are interoperable is challenging. Media content must be securely accessible across different devices and platforms without compromising security. Achieving interoperability while maintaining a high level of security requires standardization and widespread adoption of secure protocols.

Cryptographic techniques must evolve to stay ahead of emerging threats. As computational power increases, particularly with the advent of quantum computing, existing cryptographic algorithms may become vulnerable. Ensuring that cryptographic systems resist future threats is a significant challenge that requires ongoing research and adaptation.

Research into quantum-safe or post-quantum cryptography is crucial, given the potential threat of quantum computers rendering current cryptographic algorithms obsolete. Developing and standardizing cryptographic algorithms that can withstand quantum attacks will be essential for the long-term security of digital media content.

Using blockchain and other decentralized technologies can provide innovative solutions for digital media security. Blockchain can offer transparent and tamper-proof mechanisms for rights management, content distribution, and royalty payments. Smart contracts can automate and enforce access control and usage policies, reducing the reliance on centralized DRM systems.

As privacy concerns grow, incorporating privacy-preserving techniques such as homomorphic encryption, secure multi-party computation, and zero-knowledge proofs into digital media security solutions will become increasingly important. These techniques can ensure that user data is protected while still allowing necessary processing and verification.

Establishing and adopting interoperability standards for digital media cryptographic solutions can help address cross-platform compatibility challenges. Industry-wide collaboration is needed to develop and implement these standards to ensure seamless and secure access to digital media across different devices and platforms.

VI. Conclusion

In the ever-evolving digital media landscape, ensuring content security and integrity is paramount to prevent unauthorized access and piracy. This paper has provided a comprehensive overview of the cryptographic techniques for safeguarding digital media assets. The study delved into various methods to protect digital media, including encryption, DRM, watermarking, and blockchain-based solutions, starting with the fundamentals of symmetric and asymmetric cryptography, hash functions, and digital signatures. While these cryptographic techniques offer robust mechanisms to secure digital content, they also present several challenges, particularly in key management, scalability, usability, interoperability, and resistance to emerging threats. Addressing these challenges is crucial for the continued advancement and effectiveness of digital media security. Looking forward, the development of quantum-safe cryptography, enhanced key management systems, advanced DRM solutions, and privacy-preserving techniques will be essential. Integrating blockchain technology and establishing interoperability standards will further strengthen the security framework for digital media. Additionally, improving user education and simplifying the interface for secure content access will help bridge the gap between robust security measures and user convenience.

References

[1] L. Gastaldi, F. P. Appio, D. Trabucchi, T. Buganza, and M. Corso, "From mutualism to commensalism: Assessing the evolving relationship between complementors and digital platforms," Information Systems Journal, vol. 34, no. 4, pp. 1217-1263, 2024.

[2] S. Bonnet and F. Teuteberg, "Impact of blockchain and distributed ledger technology for the management, protection, enforcement and monetization of intellectual property: a systematic literature review," Information Systems and e-Business Management, vol. 21, no. 2, pp. 229-275, 2023.

[3] H. Song, N. Zhu, R. Xue, J. He, K. Zhang, and J. Wang, "Proof-of-Contribution consensus mechanism for blockchain and its application in intellectual property protection," Information processing & management, vol. 58, no. 3, p. 102507, 2021.

[4] K. Toshevska-Trpchevska, I. Kikerkova, E. M. Disoska, and L. Kocev, "The Importance of Intellectual Property Law in the Prevention of Selling Counterfeit Products Online," in Counterfeiting and Fraud in Supply Chains: Emerald Publishing Limited, 2022, pp. 147-169.

[5] S. Matted, G. Shankar, and B. B. Jain, "Enhanced image security using stenography and cryptography," in Computer Networks and Inventive Communication Technologies: Proceedings of Third ICCNCT 2020, 2021: Springer, pp. 1171-1182.

[6] I. Manor and E. Segev, "Social media mobility: Leveraging Twitter networks in online diplomacy," Global Policy, vol. 11, no. 2, pp. 233-244, 2020.

[7] A. K. Jain, S. R. Sahoo, and J. Kaubiyal, "Online social networks security and privacy: comprehensive review and analysis," Complex & Intelligent Systems, vol. 7, no. 5, pp. 2157-2177, 2021.

[8] F. Bertini, R. Sharma, and D. Montesi, "Are social networks watermarking us or are we (unawarely) watermarking ourself?," Journal of Imaging, vol. 8, no. 5, p. 132, 2022.

[9] E.-S. M. El-Kenawy et al., "Advanced dipper-throated meta-heuristic optimization algorithm for digital image watermarking," Applied Sciences, vol. 12, no. 20, p. 10642, 2022.

[10] P. Garg and R. Rama Kishore, "Comparative Analysis: Role of Meta-Heuristic Algorithms in Image Watermarking Optimization," in

Proceedings of Second Doctoral Symposium on Computational Intelligence: DoSCI 2021, 2022: Springer, pp. 315-327.

[11] B. Pourghebleh and V. Hayyolalam, "A comprehensive and systematic review of the load balancing mechanisms in the Internet of Things," Cluster Computing, pp. 1-21, 2019.

[12] B. Pourghebleh, N. Hekmati, Z. Davoudnia, and M. Sadeghi, "A roadmap towards energy - efficient data fusion methods in the Internet of Things," Concurrency and Computation: Practice and Experience, vol. 34, no. 15, p. e6959, 2022.

[13] V. Hayyolalam, B. Pourghebleh, and A. A. Pourhaji Kazem, "Trust management of services (TMoS): Investigating the current mechanisms," Transactions on Emerging Telecommunications Technologies, vol. 31, no. 10, p. e4063, 2020.

[14] F. Kamalov, B. Pourghebleh, M. Gheisari, Y. Liu, and S. Moussa, "Internet of medical things privacy and security: Challenges, solutions, and future trends from a new perspective," Sustainability, vol. 15, no. 4, p. 3317, 2023.

[15] M. A. Tofighi, B. Ousat, J. Zandi, E. Schafir, and A. Kharraz, "Constructs of Deceit: Exploring Nuances in Modern Social Engineering Attacks," in International Conference on Detection of Intrusions and Malware, and Vulnerability Assessment, 2024: Springer, pp. 107-127, doi: https://doi.org/10.1007/978-3-031-64171-8_6

[16] B. Pourghebleh, K. Wakil, and N. J. Navimipour, "A comprehensive study on the trust management techniques in the Internet of Things," IEEE Internet of Things Journal, vol. 6, no. 6, pp. 9326-9337, 2019.

[17] M. Anbari, H. Talebzadeh, M. Talebzadeh, A. Fattahiamin, M. Haghighatjoo, and A. M. Jafari, "Understanding the Drivers of Adoption for Blockchain-enabled Intelligent Transportation Systems," TEHNIČKI GLASNIK, vol. 18, no. 4, pp. 1-11, 2024.

[18] M. Harran, W. Farrelly, and K. Curran, "A method for verifying integrity & authenticating digital media," Applied computing and informatics, vol. 14, no. 2, pp. 145-158, 2018.

[19] K. Gurunathan and S. Rajagopalan, "A stegano-visual cryptography technique for multimedia security," Multimedia Tools and Applications, vol. 79, no. 5, pp. 3893-3911, 2020.

[20] M. Gafsi, S. Ajili, M. A. Hajjaji, J. Malek, and A. Mtibaa, "High securing cryptography system for digital image transmission," in Proceedings of the 8th International Conference on Sciences of Electronics, Technologies of Information and Telecommunications (SETIT'18), Vol. 1, 2020: Springer, pp. 311-322.

[21] P. William, A. Choubey, G. Chhabra, R. Bhattacharya, K. Vengatesan, and S. Choubey, "Assessment of hybrid cryptographic algorithm for secure sharing of textual and pictorial content," in 2022 International conference on electronics and renewable systems (ICEARS), 2022: IEEE, pp. 918-922.

[22] G. Panchal, D. Samanta, and S. Barman, "Biometric-based cryptography for digital content protection without any key storage," Multimedia Tools and Applications, vol. 78, pp. 26979-27000, 2019.

[23] I. Yasser, M. A. Mohamed, A. S. Samra, and F. Khalifa, "A chaotic-based encryption/decryption framework for secure multimedia communications," Entropy, vol. 22, no. 11, p. 1253, 2020.

[24] P. V. Sanivarapu, K. N. Rajesh, K. M. Hosny, and M. M. Fouda, "Digital watermarking system for copyright protection and authentication of images using cryptographic techniques," Applied Sciences, vol. 12, no. 17, p. 8724, 2022.

[25] A. Alarifi, S. Sankar, T. Altameem, K. Jithin, M. Amoon, and W. El-Shafai, "A novel hybrid cryptosystem for secure streaming of high efficiency H. 265 compressed videos in IoT multimedia applications," IEEE Access, vol. 8, pp. 128548-128573, 2020.

Detecting Online Gambling Promotions on Indonesian Twitter Using Text Mining Algorithm

Reza Bayu Perdana[1], Ardin[2], Indra Budi[3], Aris Budi Santoso[4], Amanah Ramadiah[5], Prabu Kresna Putra[6]
Faculty of Computer Science, University of Indonesia, Jakarta, Indonesia[1, 2, 3, 4, 5]
Research Center for Data and Information Science, National Research and Innovation Agency, Bandung, Indonesia[6]

***Abstract*—This study addresses the pressing challenge of detecting online gambling promotions on Indonesian Twitter using text mining algorithms for text classification and analytics. Amid limited research on this subject, especially in the Indonesian context, we aim to identify common textual features used in gambling promotions and determine the most effective classification models. By analyzing a dataset of 6038 tweets collected and using methods such as Random Forest, Logistic Regression, and Convolutional Neural Networks, complemented by a comparison analysis of text representation methods, we identified frequently occurring words such as 'link', 'situs', 'prediksi', 'jackpot', 'maxwin', and 'togel'. The results indicate that the combination of TF-IDF and Random Forest is the most effective method for detecting online gambling promotion content on Indonesian Twitter, achieving a recall value of 0.958 and a precision value of 0.966. These findings can contribute to cybersecurity and support law enforcement in mitigating the negative effects of such promotions, particularly on the Twitter platform in Indonesia.**

***Keywords*—Social media; analytics; online gambling; intention classification**

I. INTRODUCTION

In today's digital age, the landscape of social media platforms has undergone significant evolution, becoming an integral part of people's daily lives. Prominent social media accounts now have millions of followers [1], and they are utilized by various elements of society [2]. With billions of users worldwide, platforms such as Facebook, Twitter, Instagram serve as virtual hubs where individuals connect, share, and consume content. According to Digital 2024 Global Overview Report [3], as of January 2024, internet users in Indonesia spent an average of 3 hours 11 minutes a day accessing social media. This duration is above the global average of 2 hours 23 minutes, highlighting the strong attraction of social media for internet users in Indonesia.

Unfortunately, the strong appeal of social media in Indonesia has also led to a dramatic increase in exploitation of social media accounts for the promotion of online gambling in recent years, posing substantial challenges to regulatory authorities and law enforcement agencies. Online gambling promotion, particularly through social media platforms, has emerged as a crucial strategy for reaching potential players [4]. The widespread accessibility of online gambling sites, operating 24/7, is evident [5]. Furthermore, the prevalence of online gambling promotions on social media is notable, with influencers, social media personalities, and even celebrities actively engaging in such campaigns [6]. The ease of access and widespread reach of social media platforms have facilitated the dissemination of these illicit promotions, exacerbating concerns surrounding the detrimental effects of online gambling on vulnerable individuals, including minors and those with gambling addiction issues [7].

Over the past 30 years, the internet has sparked profound changes, presenting three primary challenges for gambling marketing and public policy: (1) significant transformations in the gambling industry's scale, scope, and nature; (2) a surge in gambling advertising on social media; and (3) inadequate methodologies for analyzing the extensive volume of online advertising data [8]. Detecting online gambling promotions on social media presents multifaceted challenges, ranging from identifying the numerous tactics employed by promoters to evade detection to pinpointing the specific accounts engaging in such activities. Moreover, discerning the nuanced interactions and strategies utilized by these accounts adds another layer of complexity to the detection process. Traditional gambling advertising techniques have historically been inspected through manual content analysis [9] [10] [11]. Despite offering detailed insights into specific content characteristics, this approach is laborious and resource-intensive, leading to significant constraints on sample sizes [12].

In the face of these challenges, there exists an opportunity to utilize cutting-edge technologies, especially artificial intelligence algorithms such as machine learning or deep learning with a text classification or text analytics approach, to effectively address the widespread dissemination of online gambling promotions [13]. By harnessing the power of machine learning, researchers can develop robust detection mechanisms capable of identifying and categorizing online gambling promotions with high accuracy. Therefore, the aim of this study is to answer these research questions (RQ):

- What textual features are commonly utilized in online gambling promotions on Indonesian Twitter?
- How do classification models effectively perform in detecting online gambling promotion content in Indonesian Twitter data?

This includes investigating the frequently used textual features in online gambling promotions on Indonesian Twitter and determining the effectiveness of various text mining classification methods for detecting such content. Through this study, we aim to provide significant insights into the common characteristics and patterns found in such promotional content,

as well as to identify the most effective classification algorithms for this context. We expect that the results will contribute to various aspects, including the development of prevention strategies and regulations, enhancing cybersecurity measures, and preventing the spread of illegal content on social media platforms. Furthermore, these findings are anticipated to aid law enforcement efforts and protect social media users from the negative impacts of online gambling promotion.

II. Related Works

Based on our literature review, discussions on the detection of online gambling promotion content in social media are scarcely found. Therefore, we have endeavored to compile knowledge related to the characteristics of similar content and techniques used in previous studies. The literature we gathered covers topics including online gambling, spam detection, as well as text classification and the detection of harmful content users.

In our exploration of online gambling, we have gained insights into textual features and common patterns found in gambling advertisements. Although studies [14] and [15] lack detailed evaluation reports on advertisement classification, study [14] reveals high-frequency words used in gambling ads, suggesting a trend of positive language that highlights benefits, bonuses, and special deals, as observed across studies [7], [16]–[18]. Complementing this, study [7] specifically sheds light on the characteristics of online gambling ads in Indonesia, pointing out features such as bonuses, luck, financial gains, and the ease of joining. This finding aligns with our earlier literature review, which identified a scarcity of discussions on the detection of online gambling content, especially in Indonesia, underscoring the need for our comprehensive approach to understanding these promotional strategies.

In our study, we have reviewed various approaches for text and spam classification. Predominant algorithms such as Random Forest, Support Vector Machine (SVM), and Naïve Bayes, particularly when integrated with N-Gram analysis [19][20], are favored for their reported high accuracy rates, often exceeding 95%. Furthermore, applications of LSTM and Word Sequence [21] are also considered effective alternatives, achieving an accuracy rate of 88%. Other studies [22]–[24] have explored RNN, manual analysis, and topic modeling, though their effectiveness evaluations are not explicitly reported.

Focusing on spam classification, especially content promoting gambling, many papers compare different algorithms, preprocessing stages, and their evaluations. Studies [25] and [20] advocate the use of the SVM algorithm for spam detection, with study [20] achieving an F-Score of 96% through bi-gram and TF-IDF preprocessing. Study [26] employs a Thai BERT derivative, reaching an F-Score of 0.8, while other research, like study [27] using neural networks and GloVe, and study [28] employing logistic regression and Word2Vec, report high F-Scores of 99.73% and 93%, respectively. An alternative approach in study [29] utilizes regex-based rules for spam detection, although it lacks detailed evaluation.

Most of the studies we analyzed use English text data. For text analysis in the Indonesian context, several studies have demonstrated the application of various classification models and feature extraction techniques on Indonesian-language datasets. One notable study [30] utilized a Convolutional Neural Network (CNN) alongside GloVe (Global Vectors) for intent classification within the ATIS (Airline Travel Information System) dataset, achieving a commendable accuracy of 95.84%. In another study [31], the implementation of Naive Bayes combined with TF-IDF was employed for the classification of terrorism-related content on Indonesian Twitter, attaining an F-measure of 77%. Setiawan's research [32] applied Logistic Regression with Word2Vec for feature extraction to detect spam posts on Indonesian Twitter, resulting in an accuracy of 93.67%. Additional research efforts in spam detection on Indonesian Twitter include [33], which used CountVectorizer and KNN to achieve an accuracy of 79%, and another research [34], which employed CountVectorizer and Random Forest, achieving an accuracy of 85.1%. These studies underscore the versatility and effectiveness of different text classification approaches within the Indonesian linguistic framework.

In advancing our research on the detection of online gambling promotion content in social media, we encounter several challenges that are crucial to address. These include the determination of effective keywords for data retrieval from social media queries, streamlining the labeling processes, and the selection of appropriate models and textual features. The need for precise keyword selection and labeling is particularly acute due to the limited data availability from Twitter following recent policy changes. To address this issue, we conducted descriptive analysis of frequently occurring words in online gambling promotions on Indonesian Twitter to provide insights into more effective search keyword identification. This analysis will precede the predictive analysis of training on classification models to detect such online gambling promotional content, aiming to develop an optimized approach for our text mining processes.

III. Methodology

In this research, we aim to conduct both descriptive and predictive analyses to gain a comprehensive understanding of online gambling promotion on Indonesian Twitter, thus enhancing detection knowledge and capabilities. Our research process, depicted in Fig. 1, initiates the identification of query keywords for Twitter data collection. Once keywords are selected, Twitter data retrieval is carried out. The accumulated data are labeled and then undergo pre-processing steps such as tokenization, normalization, noise removal, and stemming. Post-preprocessing, experiments are conducted as specified in Table II to identify the optimal model for detecting online gambling content on Indonesian Twitter. The experimental outcomes are further discussed to interpret the findings. Finally, conclusions and reports are drafted based on the discussions and analyses conducted, culminating in the research.

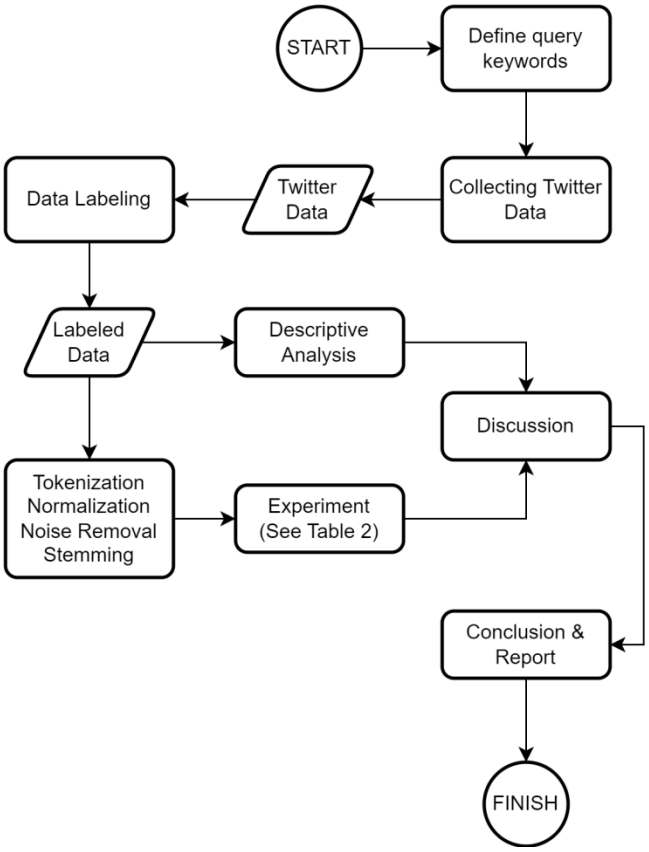


Fig. 1. Proposed workflow of this study.

Keyword query determination aims to ensure effective retrieval from Twitter searches. In identifying keywords, we refer to previous research on general online gambling promotion, characterized by a positive tone and suggestive of benefits and ease [7], [14]. Based on [7] and other studies published online [35], we use the query string *"(slot OR gacor OR untung OR cuan) lang:id"* to collect Indonesian Twitter samples containing online gambling content. We target a minimum of 5000 Indonesian language tweet samples for further analysis.

Upon sample data acquisition, data labeling ensues. We employ two annotators who have agreed upon the criteria for what constitutes online gambling promotion content, with examples outlined in Table I. The labeled data is then utilized for both descriptive and predictive analyses. In determining the label of a tweet, we provided clear labeling criteria to both annotators to minimize inconsistencies. A tweet was labeled as containing online gambling promotional content if it met one or more of the following criteria:

- Explicitly inviting individuals to participate in online gambling activities,
- Displaying links or names of online gambling platforms,
- Showing results from online gambling activities,
- Providing instructions on how to join online gambling activities.

TABLE I. EXAMPLE OF POSTS WITH ONLINE GAMBLING PROMOTION

No	Tweet
1	DIMANA LAGI CUMA MODAL KECIL DAPAT UNTUNG BESAR KALAU BUKAN DI SGA338, TUNGGU APA LAGI ? AYO ... BURUAN DAFTAR SEKARANG JUGA DAN DAPATKAN KEMENANGANNYA ! INFO LEBIH LANUT DAPAT HUBUNGI KONTAK KAMI NO HP WA : +6287815585xxx https://t.co/Y7OeIF7S35 https://t.co/qiJuNc88af
2	INFO SLOT GACOR HARI INI DAFTAR SLOT GACOR ➡□ https://t.co/hnMWbSLliD https://t.co/2bCt8SCFyB
3	388HERO - Slot Paling Gacor Di Indonesia !! ✅Bonus New Member 100% ✅Garansi Kekalahan 100% Daftar Slot Gacor Dan Claim Sekarang Juga □□□ https://t.co/jog4l9Jt0S https://t.co/RZTBrEyIYJ
4	Gotobet88 berbagi thr lewat maxwin yang di persembahkan spesial kepada member setia kami daftarkan diri anda segera di situs gacor gampang maxwin kami https://t.co/LgaCegaamG lewat @pinterest
5	Jadwal pertandingan CSGO. Ayo dukung jagoanmu dengan cara gabung di WINGSLOTS77, boskiuh. Raih kemenangan besar dengan cuan tiada batasnya. Link Alternatif : https://t.co/jdlSYjoZVS #wingslots77 #csgo #pgl #pglmajorcopenhagen https://t.co/YffbGd4Ebe

Furthermore, we calculated the inter-annotator agreement using Cohen's Kappa coefficient.

To address any disagreements between annotators, a reconciliation process was implemented. Disagreements were discussed in a joint session between the annotators, and if consensus could not be reached, a third expert annotator was brought in to make the final decision. This process ensured a thorough review and maintained consistency in labeling.

Descriptive analysis is conducted to depict the nature of online gambling content more comprehensively on Indonesian Twitter and aim to answer RQ 1. The steps for descriptive analysis derive from the study of online gambling content on Facebook [35], including temporal analysis, interaction count analysis, and examination of frequently occurring words besides search keywords, analyzed using a word cloud.

Predictive analysis aims to develop a classification model that detects online gambling promotional posts from the available dataset as the answer to RQ 2. After labeling, the data is tokenized, normalized, noise is removed, and stemming is conducted. The feature extraction and modeling phases, which determine the best algorithm for detecting online gambling promotion, are conducted through experiments listed in Table II. Algorithm selection for experimentation is based on previous research [30], [32], [34].

The selection of algorithms for this study was carefully considered to match the unique challenges presented by the task of detecting online gambling promotions on Twitter. Given the informal and varied nature of Twitter text, including the use of slang and abbreviations, Convolutional Neural Networks (CNNs) were chosen for their ability to capture complex patterns and contextual nuances within text. Random Forest was selected due to its effectiveness in handling high-dimensional data and its resistance to overfitting, which is essential when dealing with diverse textual features. Logistic Regression was included for its interpretability and as a

baseline model to compare performance, as it is widely used in text classification tasks. This combination of models was intended to leverage the strengths of each algorithm in addressing different facets of the data, ensuring a robust analysis of the factors contributing to the detection of online gambling promotions.

Evaluations are performed using a confusion matrix, focusing on recall and F-measure to minimize the potential oversight of online gambling content by the model. Due to the lack of reported training times in related studies, a direct comparison of computational efficiency between our model and those presented in [30], [32], and [34] cannot be conducted. Nevertheless, our results highlight the importance of including training time as a key metric for evaluating model performance, especially in contexts where computational resources are limited. We chose the F1 score and Recall as our primary evaluation metrics due to their relevance in contexts with imbalanced data. Recall is particularly important because it measures the model's ability to correctly identify all instances of the minority class, which is crucial in detecting online gambling promotions. The F1 score provides a balanced measure of both precision and recall, offering insight into the model's performance in minimizing false positives while still capturing true positives. These metrics were selected to ensure that the chosen model performs effectively even in the presence of class imbalance.

Once the optimal model is identified and descriptive analysis is complete, the report writing is finalized with discussions and conclusions.

TABLE II. EXPERIMENTAL LIST

Experiment	Text Representation	Algorithm
1	GloVe	CNN
2	GloVe	LR
3	GloVe	RF
4	Word2Vec	CNN
5	Word2Vec	LR
6	Word2Vec	RF
7	TF-IDF	CNN
8	TF-IDF	LR
9	TF-IDF	RF

These steps form the basis of our proposed research methodology, which aims to address the intricacies of detecting online gambling promotions in social media content effectively.

IV. RESULTS AND DISCUSSIONS

We successfully collected 6,038 Twitter data points based on the predetermined keywords. This result exceeds our target of 5,000 Twitter data points, allowing us to proceed with data annotation.

The data annotation was carried out by two annotators working independently. The annotation process involved assigning a label of "1" if the post contained online gambling promotion content, and a label of "0" if the post did not contain such content. The annotation results were then evaluated using Cohen's Kappa coefficient, with the matrix presented in Table III.

TABLE III. ANNOTATION EVALUATION MATRIX

Label	Label 0 (Annotator B)	Label 1 (Annotator B)
Label 0 (Annotator A)	4308	41
Label 1 (Annotator A)	42	1647

Based on Table III, the Cohen's Kappa coefficient from the evaluation is 0.9659, indicating an almost perfect agreement between the annotators [36]. This result suggests that the labeled data can be used for the next steps in the analysis, which include descriptive analysis and data preprocessing. The comparison of the label counts for each category is shown in Fig. 2 that shows the imbalanced dataset as labeling results.

Our approach reflects a realistic scenario where data imbalance is common, and practitioners may not always have the opportunity or resources to apply complex rebalancing techniques. By evaluating model performance on the raw, imbalanced data, we aimed to identify algorithms that are naturally resilient to such conditions. This decision was based on the need to assess the practical utility of the models in environments where preprocessing capabilities may be limited.

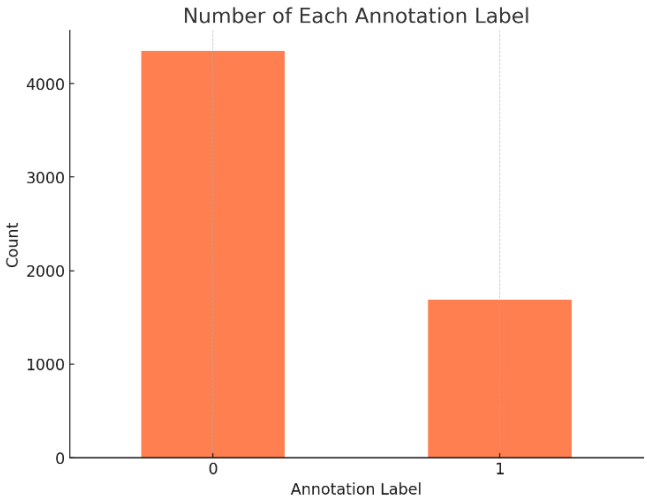


Fig. 2. Number of labels.

We then proceeded with temporal analysis by visualizing the number of tweets labeled "1" per hour. The visualization results, as shown in Fig. 3, indicate that online gambling promotion content is often posted during the early morning and late afternoon to evening hours.

Interactions with tweets containing online gambling promotion content are relatively high, averaging 1.68 interactions per tweet, including quotes, replies, retweets, or favorites. This is five times higher than the average interactions per tweet labeled "0," which is 0.38 interactions per tweet. Three accounts with usernames indicative of online gambling operators have average interactions of 36, 29, and 24 interactions per tweet. This indicates that accounts and posts

containing online gambling promotion content attract significant attention from Twitter users.

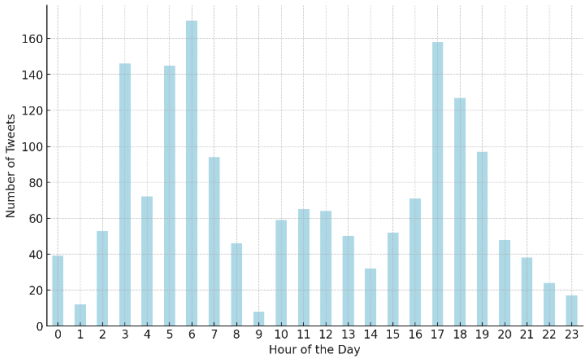


Fig. 3. Temporal distribution of labeled "1" data.

We then conducted a word cloud analysis to identify the most frequently occurring words in tweets containing online gambling promotion content. As shown in Fig. 4, in addition to the search keywords, words such as "link," "situs," "prediksi," "jackpot," "maxwin," and "togel" were identified as being affiliated with posts containing online gambling promotion content.



Fig. 4. Wordcloud of labeled "1" data.

We then proceeded with data preprocessing for modeling. We employed a back-translation method for data normalization [37]. In this process, we translated the documents into English and then retranslated them back into Indonesian using the Google Translate API. This approach ensures that words with the same meaning are standardized and helps avoid spelling errors in Indonesian. We present an example of the outcomes from the back-translation process in Table IV. This table includes an example of data labeled as '1' or 'online gambling promotion,' due to its explicit promotion of an online gambling site. Our back-translation process successfully reduced the noise present in the raw data.

TABLE IV. RESULT EXAMPLE OF BACK-TRANSLATION METHOD

Raw Data (Indonesian)	English Translation Data	Indonesian Back-Translation Data
Situs SABDA4D t3rbaik dan t3rperc4ya di 1ndon3sia m3ny3di4kan g4mes t3rlengkap d4n t3rpopuler s3rta RTP slot &amp pol4 g4cornya t3rupd4te s3tiap j4m ! bonu5 D3po 200k p4ka1 l1nk m4xwinny4 : {LINK} {LINK}	The best and most trusted SABDA4D site in Indonesia offers the most complete and popular games and RTP slots & pol4 g4cor every hour! D3po bonus 200k p4ka1 l1nk m4xwinny4: {LINK} {LINK}	Situs SABDA4D terbaik dan terpercaya di Indonesia menawarkan permainan dan RTP slot & pol4 g4cor terlengkap dan terpopuler setiap jamnya! Bonus D3po 200k p4ka1 l1nk m4xwinny4: {LINK} {LINK}

The next steps we undertook included removing stopwords, replacing slang words, removing extra characters in words, converting emojis to phrases, removing usernames, removing numbers, removing punctuation, and converting text to lowercase. During this data cleaning process, we encountered several challenges, including the use of slang, symbols, and local languages that were occasionally embedded in sentences written in Indonesian. As a result, we also conducted manual cleaning to ensure that the data included in the modeling process was accurate. Additionally, we applied stemming to revert words to their root forms. An example of the outcomes from these processes are presented in Table V.

TABLE V. RESULT EXAMPLE OF DATA PREPROCESSING

Initial Data	Cleaned Data
Situs SABDA4D terbaik dan terpercaya di Indonesia menawarkan permainan dan RTP slot & pol4 g4cor terlengkap dan terpopuler setiap jamnya! Bonus D3po 200k p4ka1 l1nk m4xwinny4: {LINK} {LINK}	situs sabda d baik percaya indonesia sedia game rtp slot pola gacor update per jam lengkap populer bonus

After completing these processes, we initiated the modeling process as outlined in Table II, starting with document vectorization.

The experimental results using nine scenarios, as shown in Table VI, indicate that the combination of GloVe and CNN demonstrates excellent performance across all evaluation metrics, albeit with a relatively high training time. The high recall and precision values indicate that this model is highly effective in detecting tweets that genuinely promote online gambling and minimizing false positives. In contrast, the combination of GloVe and Logistic Regression performs very poorly, especially in terms of recall and F1 score, indicating that this model is inadequate at detecting online gambling promotion tweets. Random Forest with GloVe shows good performance but is not as effective as CNN. This model has lower recall but is still reasonably good at detecting promotional tweets.

TABLE VI. RESULTS OF EXPERIMENT'S SCENARIO

Vectorizer	Algorithm	Accuracy	AUC	Recall	Precision	F1 Score	Kappa	MCC	Training Time
GloVe	CNN	0.97462	0.99517	0.94268	0.96493	0.95363	0.93612	0.93632	7.81112
GloVe	LR	0.72162	0.61804	0.08757	0.51136	0.14749	0.07424	0.11495	0.07246
GloVe	RF	0.93234	0.97619	0.80093	0.9492	0.86824	0.82316	0.82899	0.54267
TF-IDF	CNN	0.71799	0.74627	0.16888	0.48671	0.24517	0.12366	0.15008	8.037
TF-IDF	LR	0.97463	0.99434	0.94542	0.96326	0.95414	0.93655	0.93677	0.02478
TF-IDF	RF	0.97878	0.99419	0.95829	0.96572	0.96183	0.94711	0.94726	0.63892
Word2Vec	CNN	0.97082	0.9939	0.94344	0.95303	0.94759	0.92736	0.92798	4.12421
Word2Vec	LR	0.93732	0.98075	0.86085	0.911	0.88469	0.84168	0.8428	0.04422
Word2Vec	RF	0.96119	0.9902	0.92247	0.9376	0.92967	0.90286	0.9032	2.06679

The combination of TF-IDF and CNN shows poor performance, particularly in recall and F1 score, indicating that this model is less effective in detecting online gambling promotion tweets. The combination of TF-IDF and Logistic Regression demonstrates excellent performance, similar to GloVe-CNN, but with significantly lower training time, making it more efficient. Random Forest with TF-IDF shows the best performance among all combinations, with very high evaluation metrics and still relatively efficient training time.

The combination of Word2Vec and CNN demonstrates excellent performance, approaching the performance of GloVe-CNN but with lower training time. Logistic Regression with Word2Vec shows good performance, though not as strong as the other top combinations. It's very low training time makes this model highly efficient. Random Forest with Word2Vec shows excellent performance, slightly below the performance of TF-IDF with Random Forest. This model is still very effective in detecting promotional tweets, with lower training time compared to TF-IDF with CNN.

Overall, the combination of TF-IDF with the Random Forest algorithm provides the best performance in identifying online gambling promotion content. The Recall value of 95.8% indicates that the TF-IDF with RF combination can minimize false negatives. The high Precision value of 96.6% shows that this combination effectively avoids false positives. This is also supported by a high MCC value and moderate training time.

TF-IDF with Logistic Regression and Word2Vec with Logistic Regression show a good balance between high performance and very low training time. GloVe with Logistic Regression and TF-IDF with CNN demonstrate significantly lower performance and are not recommended for detecting online gambling promotion tweets.

We conducted hypothesis testing based on the results presented in Table IV to determine if there is a significant difference between the performance of the TF-IDF with Random Forest (RF) combination and the Word2Vec with Convolutional Neural Network (CNN) combination. This analysis was undertaken because both combinations demonstrated high Recall and F1 Score with acceptable training times, along with the difference in text vectorization's method. The null hypothesis (H0) was "There is no significant difference between the performance of TF-IDF with RF and Word2Vec with CNN," and the alternative hypothesis (H1) was "There is a significant difference between the performance of TF-IDF with RF and Word2Vec with CNN."

We performed hypothesis tests on two model evaluation metrics: Recall and F1 Score. The results of these hypothesis tests are presented in Table VII.

TABLE VII. HYPOTHESIS TEST RESULTS FOR RECALL AND F1 SCORE

	alpha	p-value	Results
Recall	0.05	0.130	Fail to reject H0: No significant difference
F1 Score	0.05	0.047	Reject H0: Significant difference

Table VII shows that the p-value for the recall metric is 0.130, which is greater than the significance level of 0.05. Therefore, we fail to reject the null hypothesis. This indicates that there is no statistically significant difference in recall between the TF-IDF + RF model and the Word2Vec + CNN model. In other words, both models perform similarly in terms of recall. The recall metric measures the ability of the model to correctly identify all relevant instances of online gambling promotion. The lack of a significant difference in recall suggests that both models are equally effective in identifying true positive instances of online gambling promotions on Indonesian Twitter. This could imply that both text representation methods (TF-IDF and Word2Vec) and classification algorithms (RF and CNN) are similarly proficient in capturing the relevant features needed for high recall in this context.

The p-value for the F1 score metric is 0.047, which is less than the significance level of 0.05. Therefore, we reject the null hypothesis. This indicates that there is a statistically significant difference in F1 score between the TF-IDF + RF model and the Word2Vec + CNN model. Specifically, the TF-IDF + RF model has a significantly better F1 score compared to the Word2Vec + CNN model. The F1 score is a harmonic means of precision and recall, providing a balance between the two. The significant difference in F1 score, favoring the TF-IDF + RF model, indicates that this model not only identifies true positives effectively (recall) but also minimizes false positives (precision). The higher F1 score suggests that TF-IDF + RF strikes a better balance between precision and recall compared

to Word2Vec + CNN, making it a more reliable model for this task.

Compared to the previous studies we use for reference, although the exact training times were not reported, the studies [30], [32], and [34] utilized models such as Random Forest and SVM, which are known for their differing computational complexities. Random Forest, for example, often requires more computational resources due to the ensemble nature of the algorithm, compared to a single-layer SVM model. By contrast, our study's use of a TF-IDF combined with Random Forest may offer a more balanced trade-off between accuracy and computational efficiency.

V. CONCLUSION

The conclusion of this study shows that, based on the analysis, textual features commonly utilized in online gambling promotions on Indonesian Twitter include words such as "link," "situs," "prediksi," "jackpot," "maxwin," and "togel." These words frequently appear in online gambling content on Twitter. This indicates that these words are effective textual features for detecting online gambling promotions on Indonesian Twitter, in addition to the initial textual features we used as keywords in our query string. Furthermore, the results of the classification model training indicate that the combination of TF-IDF and Random Forest is the most effective method for detecting online gambling promotion content on Indonesian Twitter. With a recall value of 95.8% and a precision value of 96.6%, this combination significantly minimizes false negative and positive detections. This research makes an important contribution to the development of strategies for preventing and regulating illegal content on social media and supports law enforcement efforts in addressing the negative impacts of online gambling promotion.

While our study focuses on detecting online gambling promotions on Indonesian Twitter, the proposed methodology could potentially be generalized to other languages and platforms. For instance, the tokenization and normalization processes would need adjustments to account for linguistic differences in languages like Japanese or Arabic. Furthermore, the high variability in content length and user behavior across platforms like Facebook and Instagram might require a reevaluation of the feature extraction techniques to maintain model accuracy.

For future research, it is recommended to apply and test this model on other social media platforms to broaden the generalization of the results, develop more efficient approaches for data annotation to enhance labeling accuracy and consistency, integrate sentiment analysis to gain a deeper understanding of the psychological impact of online gambling promotion content on social media users, and explore the use of other deep learning algorithms that may offer better performance with larger datasets.

With these suggestions, it is hoped that future research can be more comprehensive in addressing and preventing the spread of illegal content on social media.

ACKNOWLEDGMENT

This paper is the original work of the authors, Reza Bayu Perdana, Ardin, Aris Budi Santoso, Prabu Kresna Putra, Amanah Ramadiah, and Indra Budi, all of whom contributed significantly to the research, analysis, and writing of this study. We declare that no part of this work has been plagiarized, and it represents the authors' authentic efforts in addressing the challenge of detecting online gambling promotions on Indonesian Twitter.

The writing of this scientific paper was supported by the University of Indonesia and the Meteorology, Climatology, and Geophysics Agency (BMKG), with funding provided by BMKG as part of a scholarship to improve the quality of human resources managed by the BMKG Education and Training Center and also provided by Faculty of Computer Science University of Indonesia by Hibah Riset Internal Fakultas Ilmu Komputer UI TA 2024-2025 No. NKB-6/UN2.F2.D/NKP.05.00/2024.

REFERENCES

[1] S. Wu and D. Peng, “Pre-SMATS: A multi-task learning based prediction model for small multi-stage seasonal time series,” Expert Syst. Appl., vol. 201, 2022, doi: 10.1016/j.eswa.2022.117121.

[2] X. Y. Leung, B. Bai, and K. A. Stahura, “The marketing effectiveness of social media in the hotel industry: A comparison of Facebook and Twitter,” J. Hosp. Tour. Res., vol. 39, no. 2, pp. 147–169, 2015.

[3] Datareportal, “Digital 2024: Global Overview Report,” 2024. https://datareportal.com/reports/digital-2024-global-overview-report (accessed Jun. 01, 2024).

[4] A. Sudibyo, Media Massa Nasional Menghadapi Disrupsi Digital. Kepustakaan Populer Gramedia, 2023.

[5] I. Y. Nono, A. A. S. L. Dewi, and I. P. G. Seputra, “Penegakan Hukum Terhadap Selebgram yang Mempromosikan Situs Judi Online,” J. Analog. Huk., vol. 3, no. 2, pp. 235–239, 2021.

[6] S. Desriwaty, “Pertanggungjawaban Pidana terhadap Pelaku Promosi Judi Online yang dilakukan melalui Media Sosial ditinjau dari Perspektif Hukum Pidana.” Prodi Ilmu Hukum, 2023.

[7] M. S. Gunawan, N. Mujahidah, S. Sofyan, and M. A. M. A, “Pertanggungjawaban Platfom Media Sosial Terhadap Promosi Judi Online,” J. Plaza Huk. Indones., vol. 1, no. 19, pp. 1–15, 2023.

[8] O. Burkeman, “Forty years of the internet: How the world changed forever,” Guard., vol. 23, 2009.

[9] B. Abarbanel, S. M. Gainsbury, D. King, N. Hing, and P. H. Delfabbro, “Gambling games on social platforms: How do advertisements for social casino games target young adults?,” Policy & internet, vol. 9, no. 2, pp. 184–209, 2017.

[10] R. Cassidy and N. Ovenden, “Frequency, duration and medium of advertisements for gambling and other risky products in commercial and public service broadcasts of English Premier League football.,” 2017.

[11] S. M. Gainsbury, P. Delfabbro, D. L. King, and N. Hing, “An exploratory study of gambling operators’ use of social media and the latent messages conveyed,” J. Gambl. Stud., vol. 32, no. 1, pp. 125–141, 2016.

[12] C. Erlingsson and P. Brysiewicz, “A hands-on guide to doing content analysis,” African J. Emerg. Med., vol. 7, no. 3, pp. 93–99, 2017.

[13] R. J. E. James and A. Bradley, “The use of social media in research on gambling: A systematic review,” Curr. Addict. Reports, vol. 8, pp. 235–245, 2021.

[14] S. Choi, “Understanding Involuntary Illegal Online Gamblers in the U.S.: Framing in Misleading Information by Online Casino Reviews,” UNLV Gaming Res. Rev. J., vol. 27, no. 1, pp. 23–47, 2023, [Online]. Available: https://www.proquest.com/scholarly-journals/understanding-involuntary-illegal-online-gamblers/docview/2807105791/se-2?accountid=17242

[15] A. HernáNdez-Ruiz and Y. Gutiérrez, “Analysing the Twitter accounts of licensed Sports gambling operators in Spain: a space for responsible gambling?,” Commun. Soc., vol. 34, no. 4, pp. 65–79, 2021, doi: https://doi.org/10.15581/003.34.4.65-79.

[16] K. Kolandai-Matchett and M. Wenden Abbott, “Gaming-Gambling Convergence: Trends, Emerging Risks, and Legislative Responses,” Int. J. Ment. Health Addict., vol. 20, no. 4, pp. 2024–2056, Aug. 2022, doi: https://doi.org/10.1007/s11469-021-00498-y.

[17] A. Bradley and R. J. James, “How are major gambling brands using Twitter?,” Int. Gambl. Stud., vol. 19, no. 3, pp. 451–470, 2019.

[18] T. Teichert, A. Graf, T. B. Swanton, and S. M. Gainsbury, “The joint influence of regulatory and social cues on consumer choice of gambling websites: preliminary evidence from a discrete choice experiment,” Int. Gambl. Stud., vol. 21, no. 3, pp. 480–497, Sep. 2021, doi: 10.1080/14459795.2021.1921011.

[19] S. Aiyar and N. P. Shetty, “N-gram assisted youtube spam comment detection,” Procedia Comput. Sci., vol. 132, p. 1740182, 2018.

[20] Y. Zhang et al., “Lies in the Air: Characterizing Fake-base-station Spam Ecosystem in China,” in Proceedings of the 2020 ACM SIGSAC Conference on Computer and Communications Security, 2020, pp. 521–534. doi: 10.1145/3372297.3417257.

[21] E. Zhu, J. Wu, H. Liu, and K. Li, “A Sentiment Index of the Housing Market in China: Text Mining of Narratives on Social Media,” J. Real Estate Financ. Econ., vol. 66, no. 1, pp. 77–118, Jan. 2023, doi: https://doi.org/10.1007/s11146-022-09900-5.

[22] S. I. Alqahtani, W. M. S. Yafooz, A. Alsaeedi, L. Syed, and R. Alluhaibi, “Children’s Safety on YouTube: A Systematic Review,” Appl. Sci., vol. 13, no. 6, p. 4044, Mar. 2023, doi: 10.3390/app13064044.

[23] R. Rossi, A. Nairn, J. Smith, and C. Inskip, “Get a£ 10 Free Bet Every Week!”—gambling advertising on Twitter: volume, content, followers, engagement, and regulatory compliance,” J. public policy Mark., vol. 40, no. 4, pp. 487–504, 2021.

[24] J. Singer, V. Kufenko, A. Wöhr, M. Wuketich, and S. Otterbach, “How do Gambling Providers Use the Social Network Twitter in Germany? An Explorative Mixed-Methods Topic Modeling Approach,” J. Gambl. Stud., vol. 39, no. 3, pp. 1371–1398, 2023.

[25] S. Tang, X. Mi, Y. Li, X. Wang, and K. Chen, “Clues in Tweets: Twitter-Guided Discovery and Analysis of SMS Spam,” in Proceedings of the 2022 ACM SIGSAC Conference on Computer and Communications Security, 2022, pp. 2751–2764. doi: 10.1145/3548606.3559351.

[26] P. Chiawchansilp and P. Kantavat, “Spam Article Detection on Social Media Platform Using Deep Learning: Enhancing Content Integrity and User Experience,” 2023. doi: 10.1145/3628454.3628459.

[27] S. Kaddoura, S. A. Alex, M. Itani, S. Henno, A. AlNashash, and D. J. Hemanth, “Arabic spam tweets classification using deep learning,” Neural Comput. Appl., vol. 35, no. 23, pp. 17233–17246, Apr. 2023, doi: 10.1007/s00521-023-08614-w.

[28] M. Liu, Y. Zhang, B. Liu, Z. Li, H. Duan, and D. Sun, “Detecting and Characterizing SMS Spearphishing Attacks,” in Proceedings of the 37th Annual Computer Security Applications Conference, 2021, pp. 930–943. doi: 10.1145/3485832.3488012.

[29] N. Nasir, F. Iqbal, M. Zaheer, M. Shahjahan, and M. Javed, “Lures for Money: A First Look into YouTube Videos Promoting Money-Making Apps,” in Proceedings of the 2022 ACM on Asia Conference on Computer and Communications Security, 2022, pp. 1195–1206. doi: 10.1145/3488932.3517404.

[30] C. O. Bilah, T. B. Adji, and N. A. Setiawan, “Intent Detection on Indonesian Text Using Convolutional Neural Network,” 2022.

[31] M. Aryani, E. Miranda, Y. Fernando, and T. M. Kibtiah, “An Early Warning Detection System of Terrorism in Indonesia from Twitter Contents using Naïve Bayes Algorithm,” 2020.

[32] E. B. Setiawan, D. H. Widyantoro, and K. Surendro, “Detecting Indonesian Spammer on Twitter,” in 2018 6th International Conference on Information and Communication Technology (ICoICT), 2018, pp. 259–263. doi: 10.1109/ICoICT.2018.8528773.

[33] A. Pinandito, R. S. Perdana, M. C. Saputra, and H. M. Az-zahra, “Spam detection framework for Android Twitter application using Na\"{\i}ve Bayes and K-Nearest Neighbor classifiers,” in Proceedings of the 6th International Conference on Software and Computer Applications, 2017, pp. 77–82. doi: 10.1145/3056662.3056704.

[34] L. Alhaura and I. Budi, “Malicious Account Detection on Indonesian Twitter Account,” in 2020 3rd International Conference on Computer and Informatics Engineering (IC2IE), 2020, pp. 176–181. doi: 10.1109/IC2IE50715.2020.9274682.

[35] I. Fahmi, “Darurat Judi Online di Facebook Pages Indonesia (2) Periode Data: 1 Mei – 22 Agustus 2023,” 2023. https://threadreaderapp.com/thread/1704067338306543736.html

[36] D. Altman, “Inter-rater agreement,” Pract. Stat. Med. Res., pp. 403–409, 1991.

[37] R. I. Alhaqq, “Klasifikasi Ulasan Pengguna Aplikasi InfoBMKG di Google Play Store,” University of Indonesia, 2023.

Securing RPL Networks with Enhanced Routing Efficiency with Congestion Prediction and Load Balancing Strategy

Saumya Raj[1*], Rajesh R[2]
Research Scholar, Department of Computer Science, Bharathiar University, Coimbatore, India[1]
Associate Professor, CHRIST (Deemed to be university) Bangalore, Bharathiar University, Coimbatore, India[2]

***Abstract*—Low power and Lossy Networks (LLNs) are essential components of the Internet of Things (IoT) environment. In LNNs, the Routing Protocol for LLN (RPL)-based Internet Protocol Version 6 (IPv6) routing protocol is regarded as a standardized solution. However, the existing models did not account for the issues with congestion and security when modeling the RPL. Thus, to resolve these issues, this paper proposes a novel Exponential Poisson Distribution–Fuzzy (EPD-Fuzzy) model and Kullback Leibler Divergence-based Tunicate Swarm Algorithm (KLD-TSA) for developing a reliable RPL model. The hash codes are first generated for the registered nodes at the network end in order to achieve security; the hash codes are subsequently compared via requests with the immediate nodes. Each node sends a request to its neighbors using the hash value; if the hash value matches, a path is formed. The parent nodes are then chosen and ranked using the Pearson Correlation Coefficient-Spotted Hyena Optimization Algorithm (PCC-SHOA) technique to minimize latency. To avoid congestion, the EPD-Fuzzy is employed to predict congestion; then, a genitor node is introduced in the congested scenarios. The big data and videos are split, compressed, and sent via multiple paths to reduce the losses in the RPL. Moreover, to avoid network traffic, a novel KLD-TSA load balancing is introduced at the user end. The experiential outcomes exhibited the proposed technique's effectiveness regarding Packet delivery ratio (PDR).**

Keywords*—Low power and Lossy Network (LLN); Routing Protocol for LLN (RPL); load balancing; Internet of Things (IoT); Internet Protocol Version 6 (IPv6)

I. INTRODUCTION

A platform for extending the communication paradigm to novel along with varied levels is provided by the IoT for researchers. In the IoT, computing, as well as sensor devices, are related to the internet that provides services anywhere and anytime [1]. The devices in the IoT are connected over the internet via a gateway node. In various applications like smart homes, smart farming, smart healthcare, et cetera, the IoT is wielded [2]. A network layer in the IoT architecture that utilizes diverse standards and protocols is required by the devices in those applications. Such standards and protocols are Wireless Personal Area Network (WPAN), Internet Protocol Version 4 (IPv4), IPv6 over Low Power WPAN (6LoWPAN), IPv6 and Transmission Control Protocol (TCP) [3]. However, the devices utilized in IoT are deployed as LLNs. The LLN, which features restrictions on processing speed, power, and storage capacity make up an interconnected network of resource-constrained IoT devices [4].

Owing to the quality of radios and the minuscule size of LLN, the wireless links in LLN are lossy when analogized with other wireless networks; also, poor routing is provided by the weak routing protocols in LLN owing to the limitations like higher energy consumption as well as higher data loss within the network [5]. Thus, choosing the best routing protocol, which considers lower transmission range, lower power, along lower hardware capabilities, is significant in LLN [6]. Considering these issues, the IPv6 Routing Protocol for LLNs was standardized as a consequence of the working group efforts of the Internet Engineering Task Force (IETF), which identified RPL as the leading option to handle the routing requirements of a variety of LLN-centric applications [7].

A multi-hop routing tree rooted at a single LLN Border Router (LBR), also known as the sink node or gateway node, is constructed by the RPL, a distance-vector routing protocol, by creating Destination-Oriented Directed Acyclic Graphs (DODAGs) between nodes [8]. A sorted pair of nodes is chosen in the RPL to serve as a data packet source along with a target. Data packets are transmitted via intermediate nodes from one to another [9]; lastly, the data is passed to the internet via LBR. Although the RPL has the possible to enhance and prosper, it has limitations like load imbalance and disregard for stability [10]. Moreover, the network traffic is mounted by the load imbalance on the user side while accessing the data from the internet. Thus, to resolve these problems, a novel KL-TSA model is proposed for load balancing. Also, to enhance the RPL, a modified PCC-SHOA model is proposed for parent selection with an EPD-Fuzzy congestion prediction.

A. Problem Statement

Despite developing multiple measures for efficiently transferring data in the RPL, several problems are still unnoticed and need to be resolved. Some of such problems are,

- In prevailing works, energy efficiency is mostly concentrated on the RPL network and is not concentrated on node security and congestion.

*Corresponding Author.

- As reliability is affected by rate-limiting, optimum solutions are required by load balancing as well as congestion control.
- For example, collecting a substantial quantity of data leads to mounted traffic congestion in the network. The network's unpredictable and unreliable performance is yielded by the network traffic.

By considering these problems, the proposed technique aims to develop a reliable load-balancing model at the user end and develop a reliable congestion mitigation technique with optimal parent nodes at the network end. The major contributions of the work are as follows:

- The study introduces a genitor node-centric method with the PCC-SHOA-based parent node selection, offering a novel strategy for congestion control in the RPL.
- To identify parent node congestion during data transfer, a novel EPD-Fuzzy is used.
- Kullback-Leibler Divergence-Time Series Analysis (KLD-TSA) is a novel load-balancing model that is proposed to alleviate user congestion and enhance network performance.
- Effective network load balancing at the user results in a 4675 ms latency for 250 requests from users, demonstrating effective data handling and quick access time.

B. Motivation and Benefits of the Proposed Approach

The motivation of the proposed approach comes from the issues that currently exist in the RPL protocol: Energy efficiency, load balancing, congestion control, and reliability. Although various improvements have been made, most of the current solutions consider only one-by-one problems and ignore some very critical factors that bring performance degradation, latency, and instability, mainly in IoT environments with heavy traffic. This work gives a holistic solution to dynamic user-centric load balancing through the introduction of a KLD-TSA model, a method of selecting parent nodes using PCC-SHOA for the optimization of traffic distribution, and an EPD-Fuzzy congestion prediction mechanism in an effort to reduce energy waste and enhance stability. It reduces latency and improves access times, hence making the network more reliable with better energy efficiency to meet a more scalable, flexible, and sustainable IoT network that will help different types of applications, including mission-critical services like healthcare and smart city infrastructure.

The paper's formation is systematized as: Section II implies the recent related works of RPL for IoT. Section III states the proposed approaches. Section IV elaborates on the experimental outcomes. Section V ends the paper conclusion and a better suggestion for future enhancement.

II. RELATED WORKS

This section examines current research on load balancing, congestion, and the RPL network routing mechanism.

Safara *et al*., (2020) [11] established a priority-centric energy-efficient routing (PriNergy) technique for IoT systems. The RPL model developed its own routing protocol, which determined routing method through contents with an emphasis on energy consumption. The results showed that the PriNergy mechanism decreased the overhead on the use of energy. However, the energy consumption increased when the speed of nodes increased, this influenced the PriNergy model's performance.

Conti *et al.,* (2020) [12] presented a strong multicast communication protocol for LLNs. A lower-overhead cluster-centric multicast routing mechanism was wielded on the RPL protocol's top by the presented technique. The implementation outcomes proved the protocol's efficacy over conventional protocols regarding Packet Delivery Ratio (PDR) to 25%. But the model's overall energy consumption was more than the prevailing techniques.

Mutalemwa & Shin, (2020) [13] employed secure routing protocols for safeguarding source nodes in wireless networks with multiple hops of communication. Two phantom-centric source location privacy routing protocols were developed by the presented technique. The outcomes exhibited that the protocols had better performance features with controlled energy consumption as well as PDR. However, the model's complexity reduced data transmission reliability.

Hassan *et al.,* (2020) [14] introduced a Control layer-centered trust mechanism for supporting secure routing in RPL-grounded IoT applications. The technique was named CTrust-RPL, which assessed the nodes' trust grounded on the forwarding behaviors. The presented model's outcomes proved the superiority of the model with 35% more energy efficiency. Yet, CTrust-RPL could be confronted with energy preservation, scalability, along decentralization issues.

Preeth *et al.,* (2020) [15] deployed a proficient parent selection approach in the RPL by utilizing Ant Colony Optimization (ACO) along with coverage-centric dynamic trickle systems. For parent selection, an energy-efficient RPL protocol with ACO-grounded multi-factor optimization was generated by the study. The outcomes exposed that the E-RPL had 90% of PDR over 30 node topologies. Although it was a better model, the E-RPL could not achieve better routing overhead when the DODAG was increased.

Seyfollahi & Ghaffari, (2020) [16] explored a Lightweight Load balancing and Route Minimizing solution for RPL (L^2RMR). The L^2RMR scheme encompassed an Objective Function (OF) together with a routing metric grounded on the path route minimization. The outcomes exhibited that the developed model could enrich the energy consumption, End-to-End delay, along average Packet Loss Ratio (PLR). However, the L^2RMRscheme could not perform reliably during high traffic betwixt the nodes.

Manikannan & Nagarajan, (2020) [17] propounded a framework for the RPL/6LoWPAN-centric IoT network with the firefly approach. An RPL-based firefly optimization algorithm was developed to establish a stable and dependable protocol mobility management framework. The experiment proved that the mPRL-firefly optimizer enhanced the PDR by

an average of 2.31% more than the other prevailing algorithms. Nevertheless, the average power consumption in the developed system increased when contrasted with the conventional RPL model.

Chiti *et al.,* (2021) [18] implemented a green routing protocol with power transfer for IoT. An OF for RPL grounded on a composite metric, which considered the parent node's remaining power together with the child node, could handover to the parent node as per the Wireless Power Transfer (WPT) concept. The performance evaluation exhibited remarkable energy saving, which prolonged the network lifetime. Yet, for a long-range, the model could not perform routing efficiently.

Bidai, (2022) [19] enriched the RPL for supporting video traffic for Internet of Multimedia Things (IoMT) applications. A multi-Path version of RPL (MP-RPL), which leveraged the multi-parent feature provided by RPL for constructing various end-to-end paths of diverse qualities regarding radio link quality, was wielded by the enhanced model. The simulations exhibited that feasible and acceptable Quality of Service (QoS) was provided by the presented model when contrasted with the conventional single-path RPL. Since the conventional RPL performs better, the model is limited to the average end-to-end delay.

Karami and Derakhshanfard, (2020) [20] illustrated a Routing Protocol grounded on Remaining Time to encounter nodes with Destination nodes (RPRTD) utilizing an Artificial Neural Network (ANN). The routing was carried out by identifying the contact node with more effective conditions. The results showed that the RPRTD model efficiently and with higher accuracy anticipated the time needed for interacting nodes with the destination node while requiring less storage. However, with the ANN model, the RPRTD framework took more time for training in the LLNs.

Royaee *et al.,* (2021) [21] demonstrated a context-aware system for RPL load balancing of LLNs in the IoT. Therefore, load balancing and Automata-ant colony-centric Multiple Recursive RPL (AMRRPL) were developed to prevent congestion. The Cooja simulator experiments showed that the AMRRPL algorithm significantly improved with increased PDR and network lifetime. Nevertheless, the node ranking took more time to converge, which could cause a delay.

Sahraoui & Henni, (2021) [22] developed a Secure and Adaptive Multi-Path RPL (SAMP-RPL) for enriched security along with reliability in the heterogeneous IoT. For IPv6 RPL, the SAMP-RPL relied on three variants of adaptive together with safe multipath routing. The outcomes of the Cooja simulator exhibited the SAMP-RPL model's efficacy for enhanced dependability and security of communication at lower costs. The simulation on the Cooja platform triggered the inaccuracy issues.

Yassien *et al.,* (2021) [23] developed the RPL and Load Balancing Time-Based (LBTB) model to optimize the load balancing procedure with the capability for attaining superior network reliability as well as service time. The LBTB was employed with the modification of the trickle Timer algorithm. The outcomes displayed that higher performance was achieved regarding time-saving and power-saving. However, the imbalance among the nodes caused a congestion problem.

Musaddiq *et al.,* (2020) [24] employed an RPL for the heterogeneous traffic network. Here, various RPLs under heterogeneous traffic were evaluated; also, a protocol named Queue and Work-Load-based RPL (QWL-RPL) was introduced. The outcomes displayed that QWL-RPL could enhance the heterogeneous traffic network's performance concerning the amount of overhead, jitter, along average delay. However, for scheduling, the control messages as well as service discovery had issues associated with overhead and convergence time.

Hadaya & Alabady, (2021) [25] designed an enhanced RPL protocol for the IoT environment. An enhancement in the RPL OF is suggested by the presented work that considered 3 metrics, namely Expected Transmission count (ETX), residual energy, and load. The outcomes exposed that the RPL protocol was enhanced by the model concerning total power consumption, PLR, along PDR. However, the nodes' data security was not efficiently maintained since it was learned with only a limited number of Cognitive Packet Network's features.

A. Research Gap

Literature research gaps in the improvement of RPL protocols for IoT networks are based on some key challenges. While energy efficiency, scalability, and security have improved, there are related trade-offs that need to be addressed. On the other hand, energy efficiency improvements mostly come at a cost in terms of performance under different conditions, including high traffic or node mobility. On their part, scalability issues are manifested with increased network size, contributing to increased routing overhead. The reliability is poor, and security enhances the complexity under high traffic for the architecture. Current context-aware and adaptive protocols are overload with long convergence times, having computational overhead. Multimedia support suffers due to end-to-end delays. Decentralized approaches that extend the lifetime of the network remain underdeveloped. There is also a need for better support of multimedia services and QoS to reduce latency, particularly in real-time IoT applications. Other challenges include decentralize and efficiently manage over long distances. A critical research gap is thus the development of scalable, reliable, energy-efficient RPL-based solutions that ensure efficient handling of high traffic, enhanced security, support for real-time multimedia, and uniform performance across different IoT environments.

III. PROPOSED ROUTING APPROACHES IN THE RPL

The RPL concept was promoted by the rapid development of the IoT. However, in prevailing models, secure packet delivery and traffic congestion control were hardly performed; also, much importance was not given to the user-side traffic. Thus, to overcome these issues, a novel EPD-Fuzzy-centric congestion control in the RPL is proposed with the KL-TSA load balancing technique. Fig. 1 depicts the architecture of the proposed RPL.

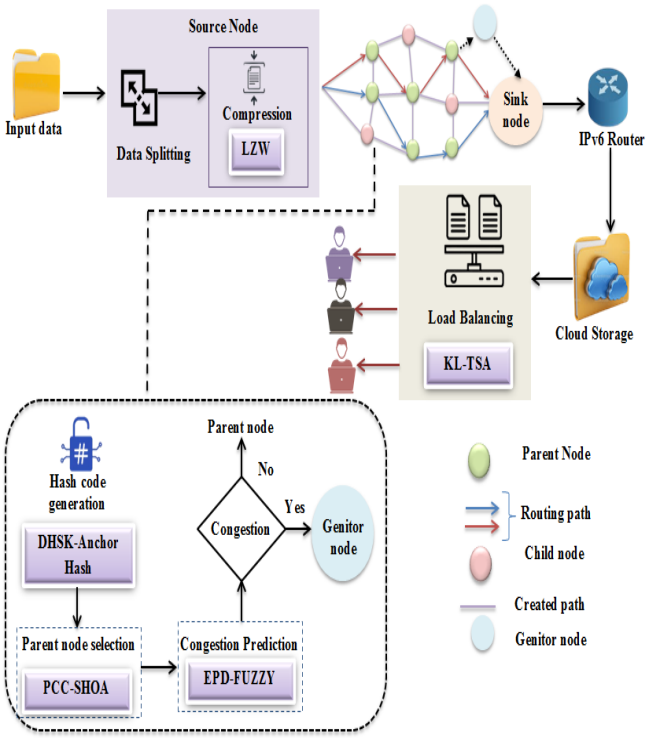


Fig. 1. Framework of the proposed RPL.

A. Node Registration

The proposed model begins with the registration of nodes participating in the network. All the nodes are registered in the network with their ID(ID), IP(IP), and MAC(M) Address. The registered details are mathematically represented in Eq. (1),

$$reg \leftarrow \langle ID, IP, M \rangle \tag{1}$$

Here, regspecifies node registration.

B. Hash Code Generation

During the registration, the hash code is generated for nodes utilizing the hybrid Diffie Hellman Secret Key-based Anchor Hashing (DHSK-Anchor Hash). In the DHSK-Anchor Hash, the keys are generated with the Diffie Hellman Algorithm (DHA), and then the generated keys are hashed with the Anchor hash. The DHSK-Anchor Hash procedure is explicated further.

1) Key generation: During the node registration, the keys are generated for every single node utilizing the DHA.

- Public and private key generation

The sender and the receiver side agree on a prime $(e)$and generator(r) in DHA. After that, the private keys$k$and $z$are chosen by the sender and the receiver side. With these values, the public keys generated at both sides are expressed in Eq. (2) and Eq. (3),

$$p = r^k \bmod e \tag{2}$$

$$a = y^z \bmod e \tag{3}$$

Where, p, aportray the public key generated at the sender side and receiver side, correspondingly. Afterward, $p, a$are shared betwixt the sender and receiver.

- Shared secret key calculation

After the public key is exchanged, the symmetric secret key(s) is generated at both sides, which is expressed in Eq. (4),

$$s = a^k \bmod e = p^z \bmod e \tag{4}$$

Here, $a^k \bmod e$ is assessed at the sender's side, $p^z \bmod e$and is assessed at the receiver's side.

2) Anchor hash: After the keys are generated, the hash value is computed utilizing the anchor hashing technique which $ID_{sender}, ID_{receiver}, p, aands$ are considered the key values(ϖ). Hence, the anchor hashing is given as follows,

a) Anchor representation: In the anchor hashing, a set of integer arrays(ϑ) is utilized for representing the anchors, which is mathematically represented as in Eq. (5),

$$\vartheta = [0,1,\ldots..,d] \tag{5}$$

where the size of the array is portrayed as$d$. After that, a bucket(B) of size $d-1$ is selected from the integer setϑ. Here, the bucket encloses$ID_{sender}, ID_{receiver}, p, a$and s. The current working buckets ϑ is symbolized as W, where $W \subseteq \vartheta$. Therefore, the bucket within the integer set is signified as$\vartheta[B]$, which is expressed as in Eq. (6),

$$\vartheta[B] = \begin{cases} 0 if B \in W \\ |W_B| if B \in \aleph \end{cases} \tag{6}$$

Where, $\aleph$represents the stack of the removed bucket and $W_B$indicates the size of the working set.

b) Hashing: In anchor hashing, a hashing function $H$is wielded to map the key values of the buckets, which is mathematically denoted in Eq. (7),

$$u_B(\varpi) \equiv \nabla(B, \varpi) \bmod \vartheta[B] \tag{7}$$

Here, $u_B(\varpi)$denotes the hashed output. During the path creation, each node sends a request to neighboring nodes with the $u_B(\varpi)$. A path will be created between such nodes if the neighboring nodes give the same hash value.

C. Optimal Parent Node Selection Using PCC-SHOA

During the path creation, the parent nodes are selected through which the data packets are forwarded. Here, utilizing the PCC-SHOA, the parent nodes get selected. In the conventional SHOA, position updation has more variation betwixt the prey and the hyena. Therefore, the Pearson Correlation Coefficient (PCC) technique is included in the SHOA model. Spotted hyena optimizer (SHO) is a recently created popular metaheuristic algorithm that draws its main inspiration from social ties between hyenas. The females in the family of spotted hyenas are the dominant ones. The spotted hyenas follow their prey using their inherent senses of sight, hearing, and scent. Spotted hyenas make a sound to interact with one another while searching for a new food source. They rely on a pack of about 100 hyenas who are their

closest companions for hunting. Thus, the working steps of PCC-SHOA are given further.

1) Initialization: The PCC-SHOA's input parameters are initialized in which the Spotted Hyena (SH) population(H) is the node involved in the path creation that can be mathematically formulated as in Eq. (8),

$$H = \{h_1, h_2, \ldots, h_l\} or h_x, x = 1,2, \ldots, l \quad (8)$$

Where, $h_l$specifies the position ofl^{th}SH, $l$signifies the population size. Also, the SH has four behaviors, namely Encircling, hunting, attacking, and searching for prey.

2) PCC-based encircling prey: In the PCC-SHOA algorithm, the best SH has obtained whose position is near the prey. The ability to locate their prey and encircle them is possessed by spotted hyenas. Since the search space is unknown in advance, the best contender at this time is assumed to be the spotted hyena that is closest to the target or prey. Once the optimal search solution has been determined, the locations of the other search agents are updated. By calculating the fitness function, the best position is attained. In the proposed model, the lower Residual energy, Transmission count, Distance, and bandwidth are considered as the fitness function. Afterward, during encircling, the distance between h_x and the prey position $(\alpha_{pos})$is calculated utilizing the PCC as in Eq. (9) and Eq. (10),

$$\lambda_{dist} = \left| \vec{C}. \frac{\Sigma(\vec{\alpha}'_{pos} - \alpha'_{pos})(\vec{h}'_x - h'_x)}{\sqrt{\Sigma(\vec{\alpha}'_{pos} - \alpha'_{pos})^2 \Sigma(\vec{h}'_x - h'_x)^2}} \right| \quad (9)$$

$$\vec{h}_x^{I+1} = \vec{\alpha}^I_{pos} - \vec{Q}.\lambda_{dist} \quad (10)$$

where, λ_{dist}specifies the distance betwixt SH and the prey, $\vec{C}, \vec{Q}$symbolizes the vector coefficients,$\vec{h}, h'$indicates the current and the mean position of SH,h_x^{I+1}implies the position of SH $x$in the iteration $I+1$, and iteration is signified as $I.\alpha, \alpha'$represent the current and the mean position of prey. The $\vec{C}, \vec{Q}$values are mathematically expressed as Eq. (11) and Eq. (12),

$$\vec{C} = 2 * \vec{R}_1 \quad (11)$$

$$\vec{Q} = 2 * \vec{\omega}.\vec{R}_2 - \vec{\omega} \quad (12)$$

Here, $\vec{R}_1, \vec{R}_2$symbolizes the random vectors and $\vec{\omega}$ portrays the reduction vector, which is computed as in Eq. (13),

$$\vec{\omega} = 5 - \left(I \times \frac{5}{I_{max}}()\right) \quad (13)$$

where the maximum iteration is notated as I_{max}. To ensure that exploration and exploitation are properly balanced, $\vec{\omega}$ falls linearly from 5 to 0 for the maximum iterations. With an increase in the number of iterations (MaxIteration), this method allows for further development. By modifying the values of $\vec{C}$and $\vec{Q}$, spotted hyenas can update their position in relation to the location of their prey.

3) Hunting prey: Spotted hyenas can detect prey, hunt in packs, and depend on a network of reliable companions. Assume that the prey is known to the best search agents, whichever is optimal, in order to define spotted hyena behaviour mathematically. Other search agents should update their location in accordance with the best solution and move in the direction of the best search agent. Here, the mathematical model is constructed by considering the best SH that knows the optimal position, whereas the other SHs update their corresponding position towards the best positions. This mathematical model is specified in Eq. (14),

$$\vec{\lambda}_{dist} = \left| \vec{C}.\vec{h}^*_x - \vec{h}_x \right| \quad (14)$$

$$\vec{h}_x = \vec{h}^*_x - \vec{Q}.\vec{\lambda}_{dist} \quad (15)$$

Where, $\vec{h}^*_x$specifies the first best spotted SH position, $\vec{h}_x$denotes the other SH positions near $\vec{h}^*_x$ which is defined in Eq. (15). Therefore, the cluster$(\vec{\Re})$with the number of the optimal solution is represented in Eq. (16),

$$\vec{\Re} = \vec{h}_x + \vec{h}_{x+1} + \ldots\ldots + \vec{h}_{x+\ell} \quad (16)$$

Here, ℓ indicates the number of SH in the best position and is defined in Eq. (17),

$$\ell = count_{nos}(\vec{h}^*_x, \vec{h}^*_{x+1}, \vec{h}^*_{x+2}, \ldots, (\vec{h}^*_x + \vec{M})) \quad (17)$$

where nos indicates the number of solutions and counts all candidate solutions after addition with $\vec{M}$, which are significantly close to the best optimal solution in the search space and $\vec{M}$ is a random vector with a value of [0.5, 1].

4) Attacking: For performing the attacking behavior, $\vec{\omega}$ is reduced. Moreover, the variation in the $\vec{Q}$is reduced to change the value of$\vec{\omega}$. The SH attacks the prey when $|Q| < 1$and the prey attacking is equated as in Eq. (18),

$$\vec{h}_x^{I+1} = \frac{\vec{\Re}}{\ell} \quad (18)$$

Update$\vec{h}_x^{I+1}$if the fitness of the current position $\left(\vec{h}_x^{I+1}\right)$ is better than the previous position, and by continuously updating $\vec{h}_x^{I+1}$, the optimal solution (parent nodes) is attained.

5) Prey search: The SHs search for their prey in the cluster vector$(\vec{\Re})$. Moreover, the SHs diverge from each other to attack and search the prey. The prey search is grounded on the changes in the $\vec{Q}$, which is utilized to randomly search the prey. If $(|Q| > 1)$, the SHs leave the prey and move to the next prey or else perform the attack on the selected prey. By this mechanism, global searches can be attained. Hence, the final parent nodes selected $\vec{h}_x^{I+1}$ $or n_\delta$ are signified as in Eq. (19),

$$P = \{n_1, n_2, \ldots, n_q\} or n_\delta \quad (19)$$

Where, P illustrates the parent node set and$n_q$represents the q^{th}selected parent node. The pseudocode of PCC-SHOA is given in Algorithm 1.

Algorithm 1: Pseudocode of PCC-SHOA

Input: Nodes$\{h_1, h_2, \ldots, h_l\} or h_x$,

Output: Selected parent node

Begin

Initialize *SH population,l,* $\vec{R}_1, \vec{R}_2$, *and maximum iteration* I_{max}

Set $\boldsymbol{I = 1}$

While $(\boldsymbol{I \leq I_{max}}())$***do***

Calculate *fitness*

Determine $\boldsymbol{\lambda_{dist}}$*using PCC*

Define $\vec{\Re} = \vec{h}_x + \vec{h}_{x+1} + \ldots\ldots + \vec{h}_{x+\ell}$

If *reducing factor* $(\vec{Q} > 1)_{\{}$

Update *position using* $\vec{h}_x^{I+1} = \vec{\alpha}_{pos}^{I} - \vec{Q}.\lambda_{dist}$

} ***Else*** *{*

Update *position using* $\vec{h}_x^{I+1} = \frac{\vec{\Re}}{\ell}$

}

End If

If the *fitness of* $\vec{h}_x^{I+1}$ *greater than* $\vec{h}_x^{I}$ ***Then***

Update $\vec{h}_x^{I+1}$

Else

$$I = I + 1$$

End If

End While

Return *optimal value*

End

After that, the rank is assigned to each selected parent node as per the fitness values of the parent nodes.

D. Data Splitting and Compression

After all the nodes are connected, the source node senses the data to be transferred to the destination node. If the sensed data size is huge, the files are split into small files, then compressed and sent to the destination node via multiple paths. This process is done to reduce the data loss in the LLN. The split parts are compressed with the Lempel–Ziv–Welch (LZW) lossless compression, then the data is transferred via nodes. The big file (D) is split into a small file as in Eq. (20),

$$D = \{v_1, v_2, \ldots, v_k\} or v_o \tag{20}$$

Hence, the k^{th}small file is illustrated asv_k. By utilizing the LZW algorithm, this small file $v_o$is compressed. The file$v_o$is compressed utilizing a table-centric lookup model in the LZW algorithm by performing encoding of the information in the file. The table formed is named dictionary or code table. The number of entries commonly accepted in the table is 4096; also, a single byte from the input file v_o is coded with the codes 0-255. While encoding is initiated, only the first 256 entries are present in the dictionary. The compression is attained by utilizing the 256 codes through 4095 entries for representing the sequence of bytes.

During compression, LZW identifies repeated sequences in v_o, then added to the dictionary. Suppose the string in the file $v_o$is represented as $babc$, which is compressed with LZW is given as in Eq. (21),

$$ababbabc = 12452 \tag{21}$$

This compressed value is sensed in the source node and transmitted to the server. Fig. 2 elucidates the flow diagram of the proposed system,

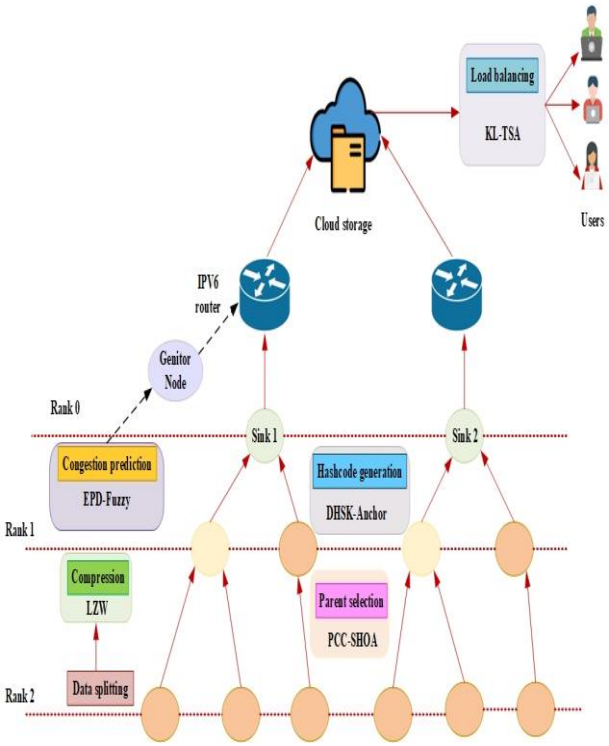


Fig. 2. Flow diagram of the proposed RPL.

E. Congestion Prediction Using Novel EPD-Fuzzy Model

During the data transfer, a novel EPD-Fuzzy detects the congestion among the parent nodes. Fuzzification, rule evaluation, and defuzzification are the three processes performed by the fuzzy algorithm. However, the fuzzy inference process has a lower level of rule generation processing than the prevailing Fuzzy algorithm. Thus, to resolve this issue, the Exponential Poisson Distribution technique is included in the prevailing Fuzzy algorithm. Hence, the congestion prediction with the EPD-Fuzzy is stated as follows,

1) Fuzzification: Primarily, the incoming number of packets (g), the number of outgoing packets (t) and the hop count (w)data $\{n_\delta\}$are given as the crisp set to the fuzzy control system, which gets mapped by a membership function for generating fuzzy sets. Hence, the membership function is represented in Eq. (22), Eq. (23) and Eq. (24),

$$m(g) = \frac{exp(-\varsigma\times\tau)(\varsigma\times\tau)^{g}}{g!} \quad (22)$$

$$m(w) = \frac{exp(-\varsigma\times\tau)(\varsigma\times\tau)^{w}}{w!} \quad (23)$$

$$m(t) = \frac{exp(-\varsigma\times\tau)(\varsigma\times\tau)^{t}}{t!} \quad (24)$$

Where, $m(\ )$signifies the EPD membership function andς, τelucidates the center and width of the fuzzy set.

2) Rule generation: After the membership function is defined, the $m(n_{\delta})$ is correlated to generating the fuzzy rules as in Eq. (25), (26) and (27),

$$\rho(g) = \begin{cases} exp\left(\frac{\psi[m(g)-m(g')]}{m(g^*)-m(g^*)}\right) if m(g) = \{g \in (g^*, \infty) \\ 1 else \end{cases} \quad (25)$$

$$\rho(w) = \begin{cases} exp\left(\frac{\psi[m(w)-m(w')]}{m(w')-m(w^*)}\right) if m(w) = \{w \in (w^*, \infty) \\ 1 else \end{cases} \quad (26)$$

$$\rho(t) = \begin{cases} exp\left(\frac{\psi[m(t)-m(t')]}{m(t')-m(t^*)}\right) if m(t) = \{t \in (t^*, \infty) \\ 1 else \end{cases} \quad (27)$$

where $\rho(\ )$implies the fuzzy rules generated in the inference, $*,'$ are the membership function's lower bound and upper bound. After that, the fuzzy rules are aggregated by utilizing IF-THEN statements. The aggregation method is given by max, which is also namedORoperator, which is expressed as in Eq. (28),

$$\Delta = max\big(\rho(g), \rho(w), \rho(t)\big) \quad (28)$$

Where, Δspecifies the aggregated outputs with the result of the implication technique.

3) Defuzzification: A process that converts fuzzy values to crisp values is named defuzzification. Hence, by computing the centroid technique, the crisp value is attained. Here, the center of the area of the fuzzy set is attained, which determines the crisp output (congestion rate)f.

F. Genitor Node

Here, a novel genitor node is included, which acts as the parent node for sending data if the $(f > Th)$is predicted by EPD-Fuzzy; where, Thindicates the threshold value. The sensed data is securely transferred to the cloud server via these processes.

G. KLD-TSA-based Novel Load Balancing

Conversely, users who want to access data from the cloud server give requests to access the data. But, multiple requests at the same time mount the network traffic. To avoid such congestion in the network, Load balancing is performed in the proposed model. Here, for load balancing, KLD-TSA is wielded. In the prevailing Tunicate Swarm Algorithm (TSA), the conflicts among the search agents are more, which affects the algorithm's performance. Hence, to avoid conflicts, Kullback Leibler Divergence (KLD) is introduced in the prevailing TSA. Tunicate is capable of locating food sources in the ocean. On the other hand, the food source in the specified search space is unknown. To locate the optimal food supply, tunicates use two different behaviors. Swarm intelligence and jet propulsion are these tendencies. Thus, the proposed load balancing is given further.

The users who request to access the resources from the server are considered as the initial population of tunicates and the position of the tunicate population is expressed as in Eq. (29),

$$J = \{j_1, j_2, \ldots, j_{\wp}\} or j_y \quad (29)$$

where the tunicate population is denoted as$J$, j_{ρ}represents the position of the tunicate$\wp$, and$\wp$indicates the population size. To attain the optimal solution, the tunicates perform jet-propulsion and swarm behavior. The mathematical model of jet propulsion satisfies three behaviors: Prevent conflicts, move toward the best search agent, and remain close to the best tunicate. Utilizing the fitness value, the best search agent is computed. Here, fitness is considered as less response and waiting time.

a) Prevent conflicts among agents: In the proposed KLD-TSA, the initialization of the new position of the search agent$(\vec{N})$ to avoid inter-agent conflict is given by utilizing the KLD in Eq. (30).

$$\vec{N} = \sum_{y=1}^{\wp} \Omega(\vec{V}_y)\, ln\, \frac{\Omega(\vec{V}_y)}{\Theta(\vec{S})} \quad (30)$$

Where,$\vec{V}_y$is the gravity force of tunicate y, and$\vec{S}$are the social forces betwixt tunicates. The gravity force is expressed in Eq. (31),

$$\vec{V} = r_2 + r_3 - \vec{G} \quad (31)$$

$$\vec{G} = 2.r_1 \quad (32)$$

Here,$r_1, r_2$and $r_3$epitomize the random values that lie in the range of 0 to 1. The water flow advection in the deep sea is symbolized by $\vec{G}$ and is defined in Eq. (32). $\vec{S}$ stands for the social dynamics among search agents. The vector $\vec{S}$ is computed as in Eq. (33),

$$\vec{S} = \left[A1min_{max_{min}}\right] \quad (33)$$

Here, the initial and subordinate speeds of social interaction are represented by A_{min} and A_{max}.

b) Move towards the best neighbor: After avoiding the conflicts betwixt the agents, the search agents move toward the direction of the best agent as in Eq. (34),

$$\vec{T} = \overrightarrow{\vec{F_l} - L(\vec{j_y^{It}})} \quad (34)$$

where$\vec{T}$indicates the distance between the tunicates and the food, $\vec{F_l}$symbolizes the food location, $L$is a random value between [0, 1], and $\vec{P_i}$signifies the tunicate positions.

c) Keeping close to the best agent: The search agent is able to stay in the direction of the optimal search agent (food source). Now, the tunicate move towards the prey is computed as in Eq. (35),

$$\vec{J}_y^{It} = \begin{cases} \left(\left|\vec{F}_l\right|^2 + |(N)(T)\sin\theta|^2\right)^{1/2} for L \geq 0.5 \\ \left(\left|\vec{F}_l\right|^2 - |(N)(T)\sin\theta|^2\right)^{1/2} for L < 0.5 \end{cases} \quad (35)$$

Where, θ signifies the angle between N and T. The updated position of tunicates in relation to the location of food sources is represented by $\vec{J}_y^{It}$.

d) Position update: The tunicate's swarm behavior is updated by updating the position of all search agents concerning the first two best search agents is revealed as follows in Eq. (36),

$$\left|\vec{J}_y^{It+1}\right| = \left(\frac{\left|\vec{J}_y^{It}\right|^2 + \left|\vec{J}_y^{It+1}\right|^2}{4+r^2}\right)^{1/2} \quad (36)$$

Where, $\left|\vec{J}_y^{It+1}\right|$ is the magnitude of the updated position of the tunicates. After that, the fitness $\left|\vec{J}_y^{It+1}\right|$ is evaluated. If the fitness $\left|\vec{J}_y^{It+1}\right|$ is greater than the $\left|\vec{J}_y^{It}\right|$, then the position is updated. Therefore, the optimal solution (i.e., optimal user) is obtained by updating the position. Thus, by selecting the optimal user, traffic is avoided. Hence, the network load is balanced. The pseudocode of the proposed KLD-TSA is given in Algorithm 2.

Algorithm 2: Pseudocode of KLD-TSA

Input: Users

Output: optimal user

Begin

Initialize *tunicate population, parameters* $\vec{N}, \vec{V}, \vec{S}$, *and maximum iterations* It_{max}

Calculate *fitness*

Set *Iteration* $It = 1$

While $(I_t \leq \omega)$***do***

Update *New position using KLD*

Move *toward the best search agent*

If $(L \geq 0.5)${

Update *tunicate position using* $\left(\left|\vec{F}_l\right|^2 + |(N)(T)\sin\theta|^2\right)^{1/2}$

} Else If $(L < 0.5)${

Update *tunicate position using* $\left(\left|\vec{F}_l\right|^2 - |(N)(T)\sin\theta|^2\right)^{1/2}$

}

End If

Update *the position of all tunicate*

End while

Set $I_t = I_t + 1$

Return $\left|\vec{J}_y^{It+1}\right|$

End

IV. RESULTS

Here, the proposed RPL methodologies' performance is experimentally evaluated with conventional techniques to demonstrate the reliability of the proposed protocol model. The performances are experimentally verified on the working platform of JAVA and the cloud sim simulation tool.

A. Performance Analysis

Here, the proposed protocol's performance is assessed in three phases, namely hash code generation, parent node selection, and load balancing. Here, regarding hash code generation time, the performance of the proposed hash code generation model DHSK-Anchor hash is comparatively analyzed with Anchor hash, SWIFFT, SHA512, and MD5 techniques.

The time taken to generate the hash value is named hash code generation time. The hash code generation time attained by the proposed protocol is 2163ms, which is 3051ms, 5178ms, and 5972ms lower than the prevailing SWIFFT, SHA-512, and MD5 techniques. This shows that the proposed RPL outperforms the conventional models. Fig. 3 elucidates the pictorial representation of hash code generation time.

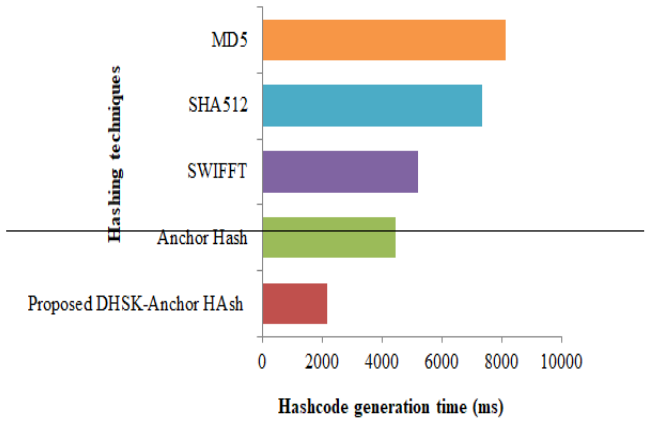


Fig. 3. Time analysis for hash code generation.

Fig. 4 depicts the graphical analysis of iteration vs. fitness for the proposed and existing algorithms.

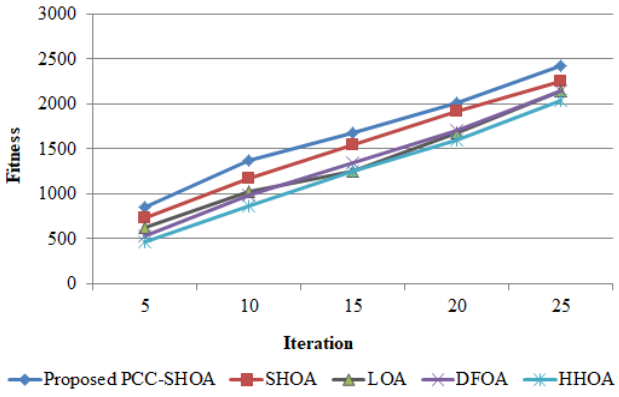


Fig. 4. Fitness vs. iteration analysis.

The Fitness value evaluation of the proposed PCC-SHOA approach and the conventional SHOA, LOA, DFOA, and HHOA selection algorithms is elucidated in Fig. 4. Here, at the 25th iteration, the proposed PCC-SHOA obtained an optimal parent node whose fitness is 2423, which is higher when contrasted with the fitness value achieved by the prevailing SHOA (2257), LOA (2148), DFOA (2144), and HHOA (2046). This proves that the proposed PCC-SHOA converged much faster than the prevailing techniques. Fig. 5 depicts the graphical analysis of throughput.

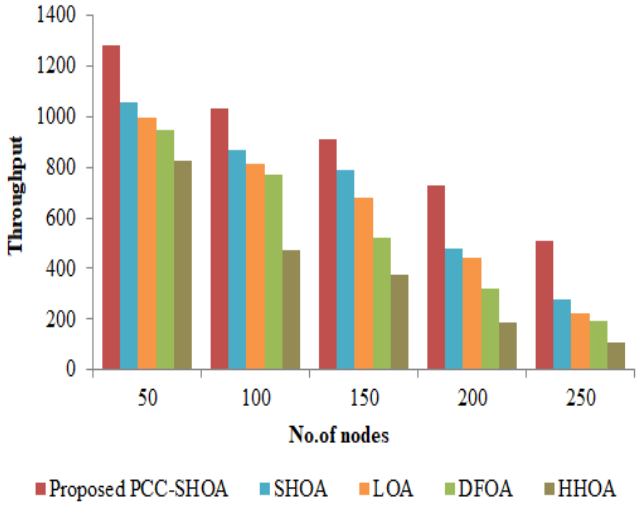
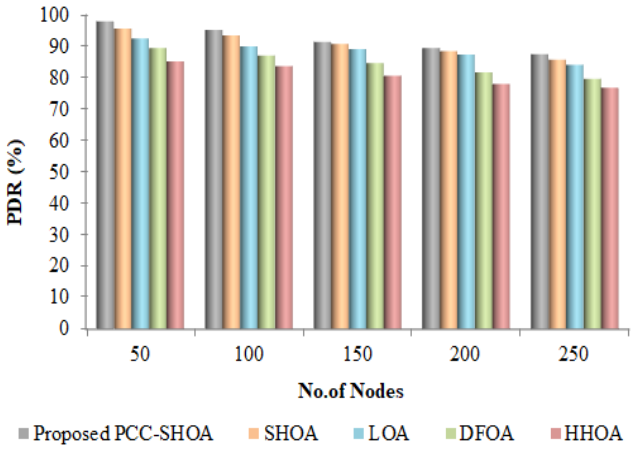
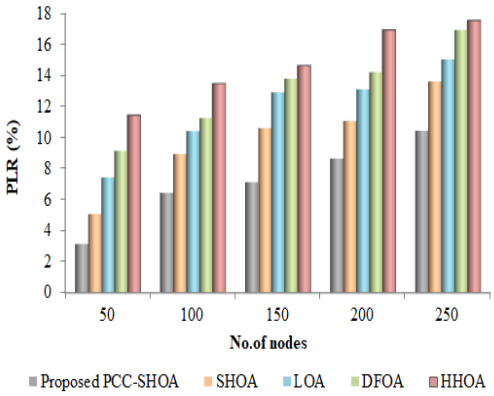


Fig. 5. Throughput analysis.

To determine how efficiently the algorithms achieved better data transmission with the selected parent nodes, the throughput is evaluated. Fig. 5 demonstrates that the throughput is analyzed for 50 to 100 nodes. Here, the proposed algorithm achieved the highest throughput of 1281 for 50 nodes, which is higher than the prevailing approaches that attained 1054 for SHOA, 993 for LOA, and 828 HHOA approaches. This concludes that with the proposed PCC-SHOA, the parent with lower Residual energy, Transmission count, Distance, and bandwidth is selected, which could enhance the proposed RPL. Fig. 6 (a) and (b) illustrate the analysis of PDR and PLR.



(a)



(b)

Fig. 6. (a) PDR and (b) PLR analysis.

The PDR and PLR are the metrics evaluated to determine the rate of packets delivered successfully and the rate of packets dropped during the data transmission. Fig. 6(a) displays that the rate of packets successfully delivered by the proposed PCC-SHOA for 100 nodes is 1.78%, 9.37%, and 13.70% higher than the prevailing SHOA, DFOA, and HHOA approaches. Then, from Fig. 6(b), it is revealed that the PLR of the proposed algorithm is 3.15%, 6.45%, 7.12%, and 10.45% for 50,100, 150, and 250 nodes, which are lower than the existing algorithms. This proves that with the use of PCC-SHOA-centric parent selection, more data is efficiently transferred with less loss, which is owing to the splitting and compression of large files.

Here, the latency, waiting time, and TAT performance of the proposed KLD-TSA approach are analyzed in comparison with the prevailing TSA, Cockroach Swarm Optimization Algorithm (CSOA), LOA, and DFOA approaches. Fig. 7 represents the latency attained by the proposed and the prevailing models.

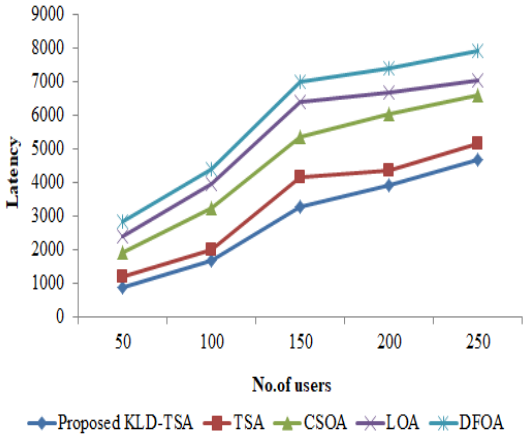


Fig. 7. Latency of the proposed framework.

Here, latency is the delay that occurs between when the user requests access and the response. The Fig. 7 displays that the Proposed KLD-TSA approach has lower latency than all other algorithms followed by TSA, CSOA, et cetera. However, the latency attained by the proposed KLD-TSA is 3289ms for 150 users, whereas 4176ms, 5347ms, and 6981ms

latency were attained by the existing TSA, CSOA, and DFOA schemes. Thus, the overall time efficiency of the proposed model is proved by this analysis.

B. Discussion

The performance of the proposed PCC-SHOA parent node selection algorithm and KLD-TSA load balancing algorithm is compared to that of current techniques in the discussion section.

1) Performance analysis of parent node selection: In this phase, the proposed PCC-SHOA algorithm's performance is comparatively analyzed with the prevailing SHOA, Lion Optimization Algorithm (LOA), Dragon Fly Optimization Algorithm (DFOA), and Harris Hawks Optimization Algorithm (HHOA) regarding the parent selection time, fitness value, throughput, response time, Turn-Around Time (TAT), PDR, and PLR. The time taken by various algorithms to select the parent node is illustrated in Table I.

TABLE I. TIME TAKEN TO SELECT PARENT NODES

Algorithms	Parent node selection time (ms)
Proposed PCC-SHOA	6034
SHOA	6513
LOA	8274
DFOA	9627
HHOA	10344

Among the prevailing algorithms, SHOA takes less time to choose the optimal parent node, which is 6513ms. Yet, with the implementation of the PCC technique in the SHOA, 479 ms lesser time is taken for choosing the optimal parent node, which displays the time effectiveness of the proposed PCC-SHOA approach.

The response and the turnaround time of the proposed and existing approaches for 50 to 250 nodes are illustrated in Table II.

TABLE II. RESPONSE TIME AND TAT

Metrics	Algorithms	Number of nodes				
		50	100	150	200	250
Response time (ms)	Proposed PCC-SHOA	3781	4796	5447	6145	6794
	SHOA	5142	6834	7402	9247	10375
	LOA	6753	7664	8314	9924	10852
	DFOA	7348	8016	9307	10267	11576
	HHOA	8457	9374	10493	11752	12055
TAT (ms)	Proposed PCC-SHOA	5423	6942	7581	9427	10524
	SHOA	7156	8123	9076	10072	11543
	LOA	8056	9365	10786	11898	12630
	DFOA	9546	10498	11966	12756	13277
	HHOA	10374	11863	13757	14624	15371

The time taken to send the data to the immediate node is named the response time, whereas the TAT is the time taken by the RPL to transmit data to the server. Here, the response time and TAT increase with the number of nodes. Here, for 250 nodes, the response time of the proposed PCC-SHOA is 6794ms, which is lower than the prevailing algorithms. Also, the best TAT is achieved by the proposed algorithm, which is 5423ms for 50 nodes.

2) Performance analysis of load balancing: The waiting and turnaround time determined for 250 users with the proposed KLD-TSA in comparison with the prevailing methodologies is illustrated in Table III.

TABLE III. WAITING TIME AND TAT OUTCOMES OF THE KLD-TSA APPROACH

Metrics	Algorithms	Number of users				
		50	100	150	200	250
waiting time (ms)	Proposed KLD-TSA	2781	4653	6447	7256	7649
	TSA	4133	6922	7464	8046	10953
	CSOA	5613	7914	8706	10264	11952
	LOA	7394	8672	9767	10527	11543
	DFOA	8857	9325	10335	11442	12594
TAT (ms)	Proposed PCC-SHOA	4423	6602	7921	9597	10554
	TSA	6969	8513	9276	10172	11643
	CSOA	7658	9265	10662	11658	12560
	LOA	9661	10598	11454	12656	13761
	DFOA	10456	11935	13746	14404	15879

The time taken by the users to access data after requesting is called waiting time, whereas TAT is the overall time taken to get data concerning the number of users. Here, the proposed model's waiting time for 50 users is 2781ms, which is lower than the prevailing CSOA (4133ms), LOA (7394ms), and DFOA (8857ms) approach. Moreover, the proposed model's TAT for 50 users is the least (4423ms). This exhibits that with the KLD-TSA data balancing, the data can be accessed from the server in the least time.

C. Comparative Analysis with the Related Works

Here, the PDR for 50 to 100 nodes is analyzed for the proposed routing protocol and the prevailing works of (Conti et al., 2020) [12], (Preeth et al., 2020) [15], and (Hadaya & Alabady, 2021) [25]. Table IV illustrates the comparative analysis of PDR with the proposed and existing algorithms.

TABLE IV. COMPARATIVE ANALYSIS WITH THE RELATED RESEARCH

Metric	Algorithms	Number of nodes		
		50	55	60
PDR (%)	Proposed EPD-Fuzzy	97.96	96.42	95.37
	(Hadaya & Alabady, 2021)	97.145	95.185	94.575
	(Preeth et al., 2020)	84.96	83.24	82.59
	(Conti et al., 2020)	82	81.16	79.60

The PDR metric is evaluated for determining the ratio of packets delivered successfully to the cloud server. The PDR is evaluated for 50, 55, and 60 nodes in Table IV, which displays that PDR is inversely proportional to the number of nodes that participated in the network. Here, the PDR is assessed for 55 nodes. With 55 nodes participating in the network, the PDR attained by the proposed routing protocol is 96.42%, which is 1.29%, 15.83%, and 18.80% higher than the prevailing works of [25], [12] and [15]. This proves that with the approaches introduced in the proposed routing protocol, more data packets are transmitted.

V. Conclusion

This paper proposes a genitor node-centric congestion control in the RPL with the PCC-SHOA-based parent node selection. The KLD-TSA-centric load-balancing model is proposed to avoid congestion among users. The proposed technique's experiments are performed on the Cloudsim simulator; also, the performance was assessed. The performance evaluation showed that the path betwixt the nodes is created in less time since the hash codes are generated in less time. Moreover, the optimal parent node is selected with the fitness of 2423 in lesser time; also, with the selected parent node, the PDR of the proposed model gets enhanced by 95.23% more than the prevailing algorithms. At the user end, the network load is balanced with a latency of 4675ms for 250 user requests. After that, the proposed RPL model's overall efficiency is proved by attaining higher PDR than the conventional systems. These outcomes proved that the proposed protocol was superior to other routing protocols. Several data are still lost even after utilizing the LZW compression in the proposed model. The research indicates that to further minimize data loss, future work may incorporate sophisticated deep-learning models for congestion prediction and make use of modified compression techniques.

Declarations

Conflict of interest: The authors declare that they have no conflict of interest.

Ethical approval: This article does not contain any studies with human participants or animals performed by any of the authors.

Consent of publication: Not applicable.

Availability of data and materials: Data sharing is not applicable to this article as no datasets were generated or analyzed during the current study.

Competing interests: The authors declare that they have no competing interests.

Funding: This work has no funding resource.

Author's contributions: All authors contributed to the study conception and design. Material preparation, data collection and analysis were performed by Saumya Raj, Dr. Rajesh R. The first draft of the manuscript was written by Saumya Raj and all authors commented on previous versions of the manuscript.

All authors read and approved the final manuscript.

Acknowledgment

We thank the anonymous referees for their useful suggestions.

References

[1] S. Sankar, S. Ramasubbareddy, A. K. Luhach, A. Nayyar, and B. Qureshi, "CT-RPL: Cluster tree based routing protocol to maximize the lifetime of internet of things," Sensors, vol. 20, no. 20, p. 5858, 2020.

[2] S. Sennan, R. Somula, A. K. Luhach, G. G. Deverajan, W. Alnumay, N. Jhanjhi, U. Ghosh, and P. Sharma, "Energy efficient optimal parent selection based routing protocol for internet of things using firefly optimization algorithm," Transactions on Emerging Telecommunica- tions Technologies, vol. 32, no. 8, p. e4171, 2021.

[3] Z. A. Almusaylim, A. Alhumam, and N. Jhanjhi, "Proposing a secure RPL based internet of things routing protocol: A review," Ad Hoc Networks, vol. 101, p. 102096, 2020.

[4] H. Farag and C˘. Stefanovic˘, "Congestion-aware routing in dynamic iot networks: A reinforcement learning ap- proach," in 2021 IEEE Global Communications Confer- ence (GLOBECOM). IEEE, 2021, pp. 1–6.

[5] H. Shreenidhi and N. S. Ramaiah, "Improving lifetime of IoT network by improvising routing protocol on low power and lossy network by using Contiki Cooja tool," in 2020 International Conference on Computational Intelli- gence for Smart Power System and Sustainable Energy (CISPSSE). IEEE, 2020, pp. 1–4.

[6] A. Touzene, A. Al Kalbani, K. Day, and N. Al Zidi, "Performance analysis of a new energy-aware RPL rout- ing objective function for internet of things," in 2020 International Conference on Electrical, Communication, and Computer Engineering (ICECCE). IEEE, 2020, pp. 1–6.

[7] M. Mahyoub, A. S. H. Mahmoud, M. Abu-Amara, and T. R. Sheltami, "An efficient RPL-based mechanism for node-to-node communications in IoT," IEEE internet of things journal, vol. 8, no. 9, pp. 7152–7169, 2020.

[8] Y. Kim and J. Paek, "NG-RPL for efficient P2P routing in low-power multihop wireless networks," IEEE Access, vol. 8, pp. 182 591–182 599, 2020.

[9] N. Azman, A. Syarif, J.-F. Dollinger, S. Ouchani, L. Idoumghar et al., "Performance analysis of RPL proto- cols in LLN network using Friedman's test," in 2020 7th International Conference on Internet of Things: Systems, Management and Security (IOTSMS). IEEE, 2020, pp. 1–6.

[10] *Pancaroglu and S. Sen, "Load balancing for RPL- based internet of things: A review," Ad Hoc Networks, vol. 116, p. 102491, 2021.*

[11] *Safara, A. Souri, T. Baker, I. Al Ridhawi, and M. Alo- qaily, "PriNergy: A priority-based energy-efficient rout- ing method for IoT systems," The Journal of Supercom- puting, vol. 76, no. 11, pp. 8609–8626, 2020.*

[12] M. Conti, P. Kaliyar, and C. Lal, "A robust multicast com- munication protocol for low power and lossy networks," Journal of Network and Computer Applications, vol. 164, p. 102675, 2020.

[13] L. C. Mutalemwa and S. Shin, "Secure routing protocols for source node privacy protection in multi-hop commu- nication wireless networks," Energies, vol. 13, no. 2, p. 292, 2020

[14] T. ul Hassan, M. Asim, T. Baker, J. Hassan, and N. Tariq, "CTrust-RPL: A control layer-based trust mechanism for supporting secure routing in routing protocol for low power and lossy networks-based internet of things appli- cations," Transactions on Emerging Telecommunications Technologies, vol. 32, no. 3, p. e4224, 2021.

[15] S. S. L. Preeth, R. Dhanalakshmi, R. Kumar, and S. Si, "Efficient parent selection for RPL using ACO and coverage based dynamic trickle techniques," Journal of Ambient Intelligence and Humanized Computing, vol. 11, pp. 4377–4391, 2020.

[16] A. Seyfollahi and A. Ghaffari, "A lightweight load bal- ancing and route minimizing solution for routing protocol for low-power and lossy networks," Computer networks, vol. 179, p. 107368, 2020.

[17] K. Manikannan and V. Nagarajan, "Optimized mobility management for RPL/6LoWPAN based IoT network architecture using the firefly

algorithm," Microprocessors and Microsystems, vol. 77, p. 103193, 2020.

[18] Chiti, R. Fantacci, and L. Pierucci, "A green rout- ing protocol with wireless power transfer for internet of things," Journal of Sensor and Actuator Networks, vol. 10, no. 1, p. 6, 2021.

[19] Z. Bidai, "RPL enhancement to support video traffic for IoMT applications," Wireless Personal Communications, vol. 122, no. 3, pp. 2367–2394, 2022.

[20] A. Karami and N. Derakhshanfard, "RPRTD: Routing protocol based on remaining time to encounter nodes with destination node in delay tolerant network using artificial neural network," Peer-to-Peer Networking and Applications, vol. 13, pp. 1406–1422, 2020.

[21] Z. Royaee, H. Mirvaziri, and A. Khatibi Bardsiri, "De- signing a context-aware model for rpl load balancing of low power and lossy networks in the internet of things," Journal of Ambient Intelligence and Humanized Computing, vol. 12, pp. 2449–2468, 2021.

[22] S. Sahraoui and N. Henni, "SAMP-RPL: secure and adaptive multipath rpl for enhanced security and reli- ability in heterogeneous iot-connected low power and lossy networks," Journal of Ambient Intelligence and Humanized Computing, vol. 14, no. 1, pp. 409–429, 2023.

[23] M. B. Yassien, S. A. Aljawarneh, M. Eyadat, and E. Eay- dat, "Routing protocol for low power and lossy network– load balancing time-based," International Journal of Ma- chine Learning and Cybernetics, vol. 12, no. 11, pp. 3101–3114, 2021.

[24] A. Musaddiq, Y. B. Zikria, Zulqarnain, and S. W. Kim, "Routing protocol for low-power and lossy networks for heterogeneous traffic network," EURASIP Journal on Wireless Communications and Networking, vol. 2020, pp. 1–23, 2020.

[25] N. N. Hadaya and S. A. Alabady, "Improved rpl protocol for low-power and lossy network for iot environment," SN Computer Science, vol. 2, no. 5, p. 341, 2021.

Optimizing Hyperparameters in Machine Learning Models for Accurate Fitness Activity Classification in School-Aged Children

Britsel Calluchi Arocutipa[1], Magaly Villegas Cahuana[2], Vanessa Huanca Hilachoque[3], Marco Cossio Bolaños[4]
Ingeniería de Sistemas, Universidad Nacional De San Agustín De Arequipa, Arequipa, Perú[1, 2, 3]
Universidad Católica Del Maule, Talca, Chile[4]

***Abstract*—Classification using machine learning algorithms in physical fitness tests carried out by students in educational centers can help prevent obesity and other related diseases. This research aims to evaluate physical fitness using percentiles of the tests and machine learning algorithms with hyperparameter optimization. The process followed was knowledge discovery in databases (KDD). Data were collected from 1525 students (784 women, 741 men) aged 6 to 17, selected non-probabilistically from five public schools. For the evaluation, anthropometric parameters such as age, weight, height, sitting height, abdominal circumference, relaxed arm circumference, oxygen saturation, resting heart rate, and maximum expiratory flow were considered. Physical Fitness tests included sitting flexibility, kangaroo horizontal jump, and 20-meter fly speed. Within the percentiles observed, we took three cut-off points as a basis for the present research: > P75 (above average), p25 to p75 (average), and < P25 (below average). The following machine learning algorithms were used for classification: Random Forest, Support Vector Machine, Decision tree, Logistic Regression, Naive Bayes, K-nearest neighbor, XGBboost, Neural network, Cat Boost, LGBM, and Gradient Boosting. The algorithms were hyperparameter optimized using GridSearchCV to find the best configurations. In conclusion, the importance of hyperparameter optimization in improving the accuracy of machine learning models is highlighted. Random Forest performs well in classifying the "High" and "Low" categories in most tests but struggles to correctly classify the "Normal" category for both male and female students.**

***Keywords*—Machine learning; classification; physical fitness; schoolchildren; hyperparameters**

I. Introduction

Machine learning (ML) is a subset of AI that involves building computer models capable of learning and making independent predictions or decisions based on the provided data [1]. In its operation, ML allows you to train a model to categorize data based on selected characteristics. It is classified into two broad categories: supervised and unsupervised. Unsupervised Machine Learning is used to conclude from data sets that contain input data without labeled responses. On the other hand, supervised machine learning attempts to discover the relationship between input attributes (independent variables) and a target attribute (dependent variable) [2]. This approach has a wide range of applications, including sectors like healthcare, education, and technological advancements. The applications are varied and can be integrated with the use of wearable technologies to track physical activity and monitor health conditions [3].

Physical activity is any body movement that results in an increase in energy expenditure above the resting level. Regular physical activity has been shown to help prevent and control non-communicable diseases, such as heart disease, stroke, diabetes, and several types of cancer. According to the WHO, more than 80% of adolescents worldwide have an insufficient level of physical activity, and it recommends that children between 5 and 17 years old dedicate at least an average of 60 minutes a day to moderate to intense physical activities, mainly aerobic, throughout the week [4].

Lack of physical activity, poor eating habits, and sedentary behaviors, such as excessive use of technology to watch television, play video games, or use cell phones, and even lack of sleep, have led to an increase in the prevalence of overweight in recent years [5].

According to a UNICEF report from 2023, in Latin America and the Caribbean, there are nearly 49 million children and adolescents between 5 and 19 years old who are overweight, which represents 30.6% of the population, above the global prevalence of 18.2 percent. South America has the highest number of people affected, with 30 million overweight, followed by Central America with 16 million and the Caribbean with three million. Argentina, Bahamas, Chile, and Mexico have the highest prevalence, with more than 35 percent. Peru also has a high prevalence, at 25 percent. Furthermore, the report shows differences by sex: the prevalence is 27 percent in men and 27.9 percent in women [6].

The analysis by Andermo et al. [7] underlines the effectiveness of school initiatives that promote physical activity among children and young people. According to these results, these actions reduce anxiety, strengthen resilience, improve well-being, and promote positive mental health.

The motivations of interest concern the Health of Students, the high prevalence of overweight, and the significant lack of physical activity among students. This public health problem requires innovative solutions that can be implemented on a large scale. The potential of machine learning with the application of specifically supervised ML algorithms can provide new insights and tools to classify and evaluate the physical fitness of schoolchildren. This will allow for a more targeted and personalized intervention. Innovation in Physical

Education: Integrating advanced technologies such as ML into physical fitness assessment can revolutionize how physical activity is understood and promoted in educational centers. Optimizing hyperparameters in ML models ensures that predictions and classifications are as accurate as possible, which is crucial for designing effective interventions.

Therefore, the paper aims to explore the level of physical fitness and the application of machine learning algorithms optimized by hyperparameters to optimally classify the physical fitness of schoolchildren from educational centers so that artificial intelligence techniques can contribute to health and academic contexts.

Creating a supervised machine learning model optimized for classifying the physical fitness level of schoolchildren is a significant contribution. This model can accurately evaluate different physical parameters and provide a detailed classification that facilitates personalized intervention. Implementing and analyzing hyperparameter optimization techniques will improve the accuracy and effectiveness of predictive models. This methodological approach can be applied in other areas of study that use machine learning, providing a framework to improve the quality of predictions. The research will provide a detailed analysis of how machine learning can evaluate and improve school initiatives that promote physical activity.

The article is organized as follows: Section I with the introduction, Section II has the literature review, Section III develops the methodology used, and Section IV presents the results obtained. Finally, the discussion and conclusion are given in Section V and Section VI respectively.

II. Literature Review

There is more than one way to measure physical activity levels (whether manually, with questionnaires, wearable technology, or smart devices), which will help classify them into levels. In this sense, we present the main works carried out with the topic under study.

Trejo et al. [8] indicate that obesity is a problem worldwide. Even more so, with the advancement of technology, many schoolchildren lead sedentary lifestyles due to being immersed in social networks and virtual video games. According to the results obtained, children with obesity spend an average of three hours watching television programs or playing video games.

According to the study by [9], Physical Activity, Diet Quality, and Physical Condition should be assessed early, considering it a physiological need to contribute to a healthy lifestyle and improve the child's future quality of life. The school framework is taken at an early average age of 8 - 12 since it is considered the ideal environment to promote good, healthy behaviors.

Zhou et al. [10] show that predicting adherence or commitment to physical activity is of utmost importance since it prevents a relapse in exercise, using automatic prediction messages with Logistic Regression and Support Vector Machine (SVM) models. This research conducted tests on sedentary people, testing their resistance to physical activity for a specific time. The Logistic Regression model demonstrated a slightly better performance than the Support Vector Machine (SVM). It should be noted that the precision in both models is high.

Alsareii et al. [11] mention physical activity is essential in controlling obesity and maintaining a healthy life. Tracking physical activities using state-of-the-art automatic techniques can promote healthy living and control obesity. This work introduces novel techniques to identify and record physical activities using machine learning techniques and wearable sensors.

Ahmadi, Pavey, and Trost [12] mention as the objective of the study the evaluation of the accuracy of Random Forest (RF) activity classification models for preschool children trained with data from free-living accelerometers in children in the range of four to nine years concluding that RF activity classification models trained with free-living accelerometer data provide accurate recognition of young children's movement behaviors under real-world conditions.

In study [13], students from public schools in Arequipa (Peru) were evaluated, and they performed the classification of motor competence based on the evaluation with the most popular machine learning algorithms optimized by their hyperparameters. The tests assessed using wearable technology were related to motor competence in schoolchildren aged 6 to 17. Different data was captured using a pedometer, accelerometer, and heart rate sensors. As a result, percentiles of schoolchildren were created. Regarding classification, the gradient boosting algorithm with its optimization of hyperparameters with the RandomizedSearchCV technique was the one that obtained the best precision of 0.95 and in the ROC-AUC curves with 0.98. Additionally, they developed software that implemented the model that was built.

Current research has identified several areas of intervention to optimize the classification of lack of physical activity in schoolchildren. Still, it has also highlighted significant gaps that must be addressed to improve the effectiveness of prevention and treatment strategies. Among them is the improvement of predictive models, which can be achieved by incorporating the configuration of hyperparameters. The comparison of tracking techniques, making an exhaustive comparison of different automatic fitness activity tracking techniques in terms of cost-effectiveness and accessibility, proposing practical solutions for implementation in school programs. The classification models can be integrated by incorporating Fitness Activity classification models into school programs and providing guidelines for their use by educators and parents. Algorithm optimization, performing a comparative analysis of the machine learning algorithms used to classify Fitness Activity, determining the most effective and accessible implementation in various contexts.

III. Methodology

It is necessary to perform data analysis as a fundamental process for classifying schoolchildren and obtaining valuable information on physical activity results. There are different methods and techniques to classify and perform data analysis. The KDD (Knowledge Discovery in Databases) process was

used for this research to extract knowledge from large volumes of data. It consists of a series of defined stages applied before using data mining techniques to search for hidden patterns in the data and analyze the patterns found. KDD uses a structured set of stages to address data mining projects, from understanding data to obtaining knowledge, as shown in Fig. 1; this represents how we use the methodology following the stages in the data process of physical activity, generating knowledge. The methodology used was KDD, which consists of six stages: Starting from Selection, followed by Preprocessing, Transformation, Mining, and Interpretation to reach the knowledge we want. Fig. 1 shows the KDD process.

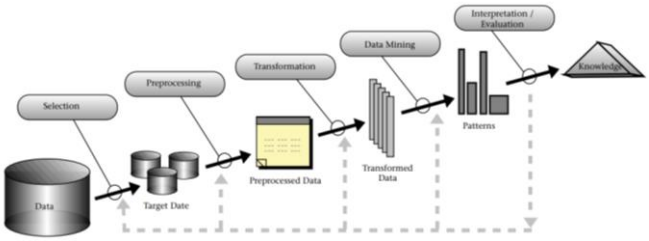


Fig. 1. KDD process [14].

A. Data Selection

The selection of data sources consisted of searching the data for appropriate input attributes of physical activity, obtaining 1525 students (784 women, 741 men) from 6 to 17 years old, including children and adolescents; the sample selection was non-probabilistic. This means knowing what you want to obtain and what data will facilitate this process to achieve the results.

The evaluations of the motor competence tests were carried out in public schools in the city of Arequipa (Peru). The tests were carried out during physical education sessions. The schoolchildren were previously informed about the evaluation to be carried out.

For this research, anthropometric measurements were carried out in the facilities of each school, working with children and adolescents with the authorization of their parents or guardians. Data collection and evaluation were carried out by two physical education teachers with experience in similar work. It was assessed according to the standardized protocol of Ross and Marfell-Jones to capture the standing height and weight measurements. Body weight (in kilograms) was measured using a BC-730 electronic scale, with a range of 0 to 150 kg and a precision of 100 grams. Standing height was measured according to the Frankfurt plane using a portable stadiometer with an accuracy of 0.1 mm. To divide abdominal fat (AF) into categories by sex and age, the suggestions described by Fernández in [15] were followed.

B. Pre-Processing

Data was collected during the different physical education sessions to pre-process this research. Once collected, the following steps were carried out, shown in Table I, necessary for use in the classification algorithms:

1) Data cleaning: In this step, the data was cleaned, including incomplete data (where there are missing attributes or attribute values), noise (incorrect or unexpected values), and inconsistent data (containing values and characteristics with different names). Conflicting data were eliminated because they would allow inadequate analysis and incorrect results.

The data cleaning tasks to be executed by the Jupyter panel were written in Python 3. The Pandas library, used for data manipulation and analysis, uses the Python Scikit-learn library [16], as shown in greater detail in Table II.

Fig. 2 shows the distribution of the students' classes: High, Normal, and Low physical activity for both sexes.

2) Data transformation: The data was normalized to be on the same scale since some machine learning algorithms are sensitive to scales. The physical fitness tests considered according to the specialists were:

a) Flexibility (cm): The dorsal-lumbar flexibility, sitting posture, and modified reach were measured.

b) Horizontal jump (cm): The horizontal jump was measured by the number of attempts in the "kangaroo" test.

c) Speed 20m (seconds): It was evaluated ten times in the 20-meter race and assessed in seconds with a stopwatch.

The percentile tables were used to consider the research [17]. These percentiles were divided into male and female ranges, and the parameters for each test taken for this investigation were shown.

TABLE I. DATA PRE-PROCESSING

Phases	Description
Data Cleaning	Errors in the data, such as empty spaces and punctuation marks not allowed, were detected and corrected.
Data integration	The data collected from the sessions was combined to provide a unified view by integrating them into a single format.
Data transformation	In this stage, categorical variables were normalized and standardized, and coding was carried out, transforming the data into a uniform format.
Data reduction	The data to be processed were identified for each record having ten input attributes.

TABLE II. APPLICATION OF TECHNOLOGIES

Technology	Description
Python 3.6	Python is the programming language we will use to analyze and process algorithms.
Colab	It is the environment that can run and program in Python. It provides a flexible environment for Python programming and other scientific computing tasks.
Scikit-learn	Python library includes various supervised and unsupervised machine learning algorithms, which are widely used for their ease of use, and a wide range of data analysis and modeling tools.

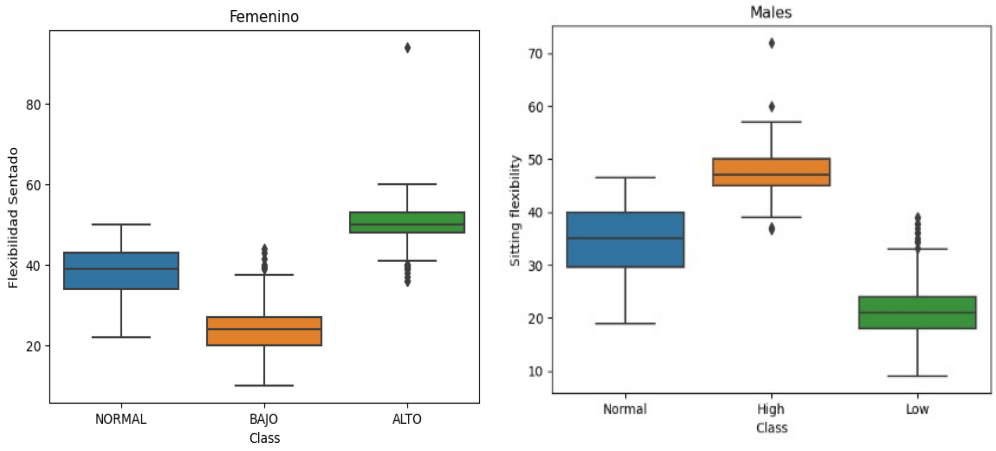


Fig. 2. Boxplot of the classes of female and male school children.

TABLE III. PERCENTILES OF THE FLEXIBILITY, HORIZONTAL JUMP, AND SPEED 20M TESTS FOR MEN AGES 6 TO 18

	Age												
Percentile/Test	6	7	8	9	10	11	12	13	14	15	16	17	18
Flexibility													
P25	39	18.5	24	21.8	24	26.5	42	35	24.8	25	26	28	29
P50	41.5	22	36	32.5	36	38	46	43.7	31	34	32	32	31
P75	43.1	36.1	44	43	43	44.5	47	47.6	42.5	45	43	38	36
Horizontal Jump													
P25	82.8	76.5	87	99.3	98	105	115	110	110	119	121	132	144
P50	91	93	110	110	115	120	120	123	130	134	143	150	155
P75	97	104	117	119	125	130	120	142	150	160	165	178	180
Speed 20m													
P25	5.02	4.32	4.22	4.15	4.2	4	3.32	4.08	3.6	3.4	3	2.98	2.9
P50	5.59	4.68	4.92	4.86	4.65	4.2	3.75	4.68	3.82	3.9	3.2	3.2	3
P75	6.14	5.1	5.38	5.43	4.89	4.59	3.8	4.97	4.28	4.31	3.5	3.6	3.2

TABLE IV. PERCENTILES OF THE FLEXIBILITY, HORIZONTAL JUMP, AND SPEED 20M TESTS FOR FEMALES AGES 6 TO 18

	Age												
Percentile/Test	6	7	8	9	10	11	12	13	14	15	16	17	18
Flexibility													
P25	36.5	21	22.3	25	29	27	44	40	27	33	29.8	27	30.5
P50	39.5	29.8	38	40	41	39.5	47	45	34.5	44	39.5	34	34
P75	43.8	39.4	46.8	47	47	46.5	53	50	48	51	49.3	39	36
Horizontal Jump													
P25	70	69.8	80	90	90	90	105	107	95.3	104	90	102	102
P50	81	82	95	102	105	105	111	113	110	112	109	117	110
P75	88	88.5	105	111	112	114	114	119	118	122	120	129	126
Speed 20m													
P25	5.4	4.58	4.64	4.3	4.44	3.9	4.1	4.52	3.8	4	3.78	3.6	3.65
P50	5.99	4.94	5.2	4.91	4.66	4.4	4.12	4.87	4.25	4.5	4.36	3.7	3.8
P75	6.2	5.4	5.67	5.54	5.29	4.94	4.63	5.35	4.81	5.19	5.12	4	4.05

Tables III and IV present the results of the evaluated tests and their corresponding percentiles: P25, P50, and P75. These percentiles were used as cut-off points in this research. Values close to the P75th percentile are considered excellent, while values at the P25th percentile are rated poor. The grade depends on the type of physical test and the objective. According to the studies, the following cut-off points were proposed for the diagnosis of physical fitness: 75 (High), 50 (Normal), and 25 (Low). The previously studied percentiles were fundamental for this research, providing a solid basis for establishing the cut-off points.

C. Transformation

For balance, the records were first sorted randomly, and then 80% were selected for training and 20% for testing. In summary, this process consisted of three phases: defining and determining the types of errors, finding and identifying instances containing errors, and correcting the discovered errors.

The registration of schoolchildren had the problem of imbalance in data sets, which is a significant problem in classification operations. When one class is significantly more frequent than others, machine learning algorithms can become biased toward the majority class, resulting in poor performance in predicting the minority class. Data balancing helps mitigate this problem; for this reason, data balancing was carried out to correctly predict the minority classes for a data set in the data analysis.

D. Classification of Data Mining

This stage consists of searching for patterns of interest that can be expressed as a model based on Machine learning algorithms applied to physical fitness tests in schoolchildren. Data analysis determined that the classification results from physical activity tests are labeled- Low, Normal, and High. Likewise, classification was the most appropriate type of prediction for this research. For the classification model, a comparison of supervised machine-learning techniques is made. The most used optimizers and algorithms, according to the literature, are:

TABLE V. DESCRIPTION OF CLASSIFICATION ALGORITHMS USED IN THE STUDY

Algorithm	Description	Advantages	Limitations
Decisión Tree	Separates the data of an ensemble into smaller subsets with an increase in the depth of the tree. The objective increases the prediction using decision nodes [18].	Visual interpretation of all possible results requires little data cleaning, is unaffected by different values, and uses numerical and categorical variables.	The calculation is complex, which implies more time is needed to train the model; a slight change in the data can cause a significant change in the tree's structure.
Random Forest	Models are made up of many decision trees; when training each tree, it learns from a random sample of the data points and a subset of features [19].	The final predictions of the Random Forest are made by averaging the predictions of each tree, reducing the problem of overfitting and variance.	Regression algorithms have a higher computational cost and longer training time than decision trees. They do not predict beyond the range in the training data.
Naive Bayes	Probabilistic learning algorithm that uses the rule of Bayes' theorem together with prior knowledge, whose characteristics depend on the independence provided by the class [20].	The advantage of Naive Bayes is when the cost of incorrectly classifying a result as positive is high, it is crucial to have high precision to minimize false results.	Naive Bayes is limited by underperforming when faced with splitting data sets with high dimensionality. As the number of features increases.
Support Vector Machine	SVMs find a line or hyperplane between different data and calculate a maximum margin that leads to a homogeneous division of all data points [21].	Efficient in memory usage when processing.	Choosing the correct kernel and parameters can be computationally expensive.
Logistic Regression	Used for binary classification problems, the basis of logistic regression is the logistic function (sigmoid) that takes any number with an accurate value and assigns it a value between 0 and 1 [22].	It is a simple but effective algorithm closely related to neural networks. The training time is less than that of other algorithms.	Predicting complex data is problematic because it has a linear decision surface; in high-dimensional data sets, this can generate overfitting.
Neural Network	It uses interconnected nodes with a layered structure that resembles the human brain. It creates an adaptive system that computers use to learn from their mistakes and continuously generate improvements [23].	Neural networks help computers make intelligent decisions with limited assistance. They can learn and model complex, non-linear input-output data relationships.	Its limitation is that it has a forward propagation network, which uses a feedback process to improve predictions over time.
k-Nearest Neighbors KN	It uses K-nearest neighbor to make classifications or predictions about the clustering of a data point. The main idea is that all data points are close to each other to belong to the same class [24].	Speeds training time by storing the training set and learning from it only when making predictions.	They are computationally expensive. They must be preprocessed and scaled, and the observations will be used only during prediction, so this step is costly.
XG Boost	It implements Gradient Boosting to maximize training speed and model performance [25].	Training optimizes memory resources and distributed computing, allowing for handling large data sets.	High flexibility produces many hyperparameters that strongly interact with the model's behavior.
Gradient Boos	With an ensemble algorithm with numerical optimization, the objective is to minimize the loss of the model by sequentially adding decision trees [26].	Good predictive accuracy, flexibility to adjust to different kinds of data, and predictions are made by a majority vote of weak learners.	Gradient boosting will continue to improve to minimize all errors that cause excessive overfitting.

Hyperparameters are defined as extra parameters or parameters that the learning algorithm does not memorize directly. Hyperparameters are external configuration variables that, when performing data analysis, help us manage the training of Machine Learning models. We also call model hyperparameters, which are manually configured before training a model. Optimization was used by configuring hyperparameters; among the main fields, the following were chosen:

Random Forest:

- n_estimators: Number of trees in the forest.
- max_depth: Maximum depth of each tree.
- min_samples_split: Minimum number of samples required to divide a node.
- min_samples_leaf: Minimum number of samples required in a leaf node.
- max_features: Number of features to consider for the best split.
- bootstrap: Whether to use bootstrap sampling when building trees.

Gradient Boosting:

- n_estimators: Number of impulse stages to perform.
- learning_rate: Learning rate.
- max_depth: Maximum depth of the trees.
- min_samples_split: Minimum number of samples required to divide a node.
- min_samples_leaf: Minimum number of samples required in a leaf node.
- max_features: Number of features to consider for the best split.
- subsample: The fraction of samples used to train each base tree.

IV. RESULTS

We worked with the Anaconda Navigator platform with Jupyter Notebook because it already had established, well-supported libraries. Table V shows the different machine-learning techniques applied in this research.

For the work of the data obtained already cleaned, it was used for shaping, 80% allocated for training, and 20% designated for testing, together with the Jupyter Notebook, a commonly used tool for machine learning, using the scikit-learning library.

The configuration used for the machine learning algorithms was made regarding the hyperparameters. The configuration achieved by the best results obtained in Random Forest and Gradient Boosting, respectively, is shown:

'n_estimators': 200,
'max_depth': 10,
'min_samples_split': 5,
'min_samples_leaf': 2,
'max_features': 'sqrt',
'bootstrap': False

'n_estimators': 100, 20,
'learning_rate': 0.1,
'max_depth': 3,
'min_samples_split': 5,
'min_samples_leaf': 1,
'max_features': 'sqrt',
'subsample': 1.0

Tables VI to VIII show the results obtained once processed. In male schoolchildren, they compare traditional and enhanced machine learning techniques with configured hyperparameters for flexibility, speed, and horizontal jump data. Accuracy metrics, such as F1 score, recall, and precision, are also shown for schoolchildren.

Tables IX to Table XI show the results of applying machine learning algorithms with hyperparameter optimization compared to traditional ones in female schoolchildren.

TABLE VI. COMPARISON OF RESULTS FOR THE MALE FLEXIBILITY TEST

Algorithm	Decision Tree	SVM	Random Forest	Naive Bayes	Logistic Regression	KNN	MLP	Gradient Boost	XGB	LGBM	CatBoost
Accuracy	0.92	0.79	0.94	0.85	0.86	0.73	0.81	0.96	0.96	0.98	0.95
Accuracy optimized Hyperparameter	0.95	0.89	0.97	0.85	0.91	0.82	0.83	0.95	0.89	0.98	0.95
F1-score	0.95	0.86	0.97	0.93	0.94	0.83	0.85	0.97	0.97	0.98	0.97
Recall	0.97	0.92	0.95	0.96	0.95	0.77	0.81	0.95	1.00	1.00	0.96
Precision	0.92	0.80	0.99	0.90	0.94	0.90	0.89	1.00	0.95	0.96	0.99

TABLE VII. COMPARISON OF RESULTS FOR THE MALE HORIZONTAL JUMP TEST

Algorithm	Decision Tree	SVM	Random Forest	Naive Bayes	Logistic Regression	KNN	MLP	Gradient Boost	XGB	LGBM	CatBoost
Accuracy	0.91	0.81	0.92	0.65	0.91	0.82	0.81	0.93	0.94	0.93	0.93
Accuracy optimized Hyperparameter	0.91	0.93	0.98	0.67	0.94	0.84	0.90	0.94	0.90	0.93	0.94
F1-score	0.94	0.82	0.94	0.71	0.93	0.85	0.83	0.96	0.96	0.96	0.95
Recall	0.94	0.78	0.93	0.70	0.96	0.83	0.80	0.97	0.95	0.95	0.98
Precision	0.94	0.87	0.95	0.73	0.91	0.88	0.86	0.95	0.98	0.98	0.93

TABLE VIII. COMPARISON OF RESULTS FOR THE MEN'S 20M SPEED TEST

Algorithm	Decision Tree	SVM	Random Forest	Naive Bayes	Logistic Regression	KNN	MLP	Gradient Boost	XGB	LGBM	CatBoost
Accuracy	0.85	0.54	0.81	0.61	0.77	0.48	0.45	0.91	0.90	0.93	0.90
Accuracy Optimized Hyperparameter	0.91	0.78	0.81	0.62	0.85	0.74	0.71	1.00	0.96	0.93	0.93
F1-score	0.88	0.62	0.81	0.70	0.85	0.60	0.61	0.90	0.88	0.88	0.91
Recall	0.92	0.56	0.76	0.64	0.88	0.56	0.68	0.88	0.93	0.93	0.88
Precision	0.83	0.70	0.87	0.76	0.83	0.64	0.56	0.93	0.85	0.85	0.95

TABLE IX. COMPARISON OF RESULTS FOR THE FEMALE FLEXIBILITY TEST

Algorithm	Decision Tree	SVM	Random Forest	Naive Bayes	Logistic Regression	KNN	MLP	Gradient Boost	XGB	LGBM	CatBoost
Accuracy	0.92	0.79	0.85	0.72	0.79	0.75	0.71	0.85	0.86	0.88	0.91
Accuracy Optimized Hyperparameter	0.91	0.88	0.91	0.75	0.85	0.78	0.72	0.82	0.85	0.89	0.85
F1-score	0.93	0.88	0.90	0.74	0.85	0.78	0.72	0.83	0.86	0.89	0.90
Recall	0.93	0.89	0.91	0.75	0.79	0.82	0.71	0.84	0.88	0.90	0.84
Precision	0.94	0.87	0.88	0.76	0.81	0.80	0.70	0.83	0.86	0.88	0.85

TABLE X. COMPARISON OF RESULTS FOR THE FEMALE HORIZONTAL JUMP TEST

Algorithm	Decision Tree	SVM	Random Forest	Naive Bayes	Logistic Regression	KNN	MLP	Gradient Boost	XGB	LGBM	CatBoost
Accuracy	0.90	0.83	0.87	0.67	0.81	0.74	0.71	0.84	0.87	0.88	0.90
Accuracy Optimized Hyperparameter	0.93	0.88	0.98	0.75	0.87	0.78	0.74	0.88	0.81	0.88	0.89
F1-score	0.94	0.89	0.91	0.75	0.89	0.78	0.82	0.87	0.92	0.90	0.93
Recall	0.93	0.90	0.94	0.85	0.82	0.73	0.86	0.86	0.89	0.91	0.94
Precision	0.94	0.88	0.88	0.67	0.70	0.85	0.78	0.89	0.94	0.89	0.92

TABLE XI. COMPARISON OF RESULTS FOR THE WOMEN'S 20M SPEED JUMP TEST

Algorithm	Decision Tree	SVM	Random Forest	Naive Bayes	Logistic Regression	KNN	MLP	Gradient Boost	XGB	LGBM	CatBoost
Accuracy	0.87	0.46	0.89	0.73	0.83	0.61	0.53	0.89	0.93	0.94	0.89
Accuracy Optimized Hyperparameter	0.89	0.82	0.97	0.85	0.84	0.77	0.79	0.90	0.89	0.94	0.89
F1-score	0.88	0.46	0.91	0.74	0.86	0.65	0.62	0.90	0.93	0.94	0.89
Recall	0.90	0.44	0.88	0.69	0.89	0.58	0.53	0.87	0.97	0.99	0.85
Precision	0.87	0.49	0.94	0.79	0.83	0.73	0.75	0.93	0.89	0.89	0.93

The results shown in the Tables show that, for the classification techniques, the Random Forest achieved the highest precision in Male Flexibility, whose value is 0.97, indicating a more significant adjustment of the estimated prediction. For the f1-score metrics, the result was 0.97, in the case of recall 0.95.

The choice of hyperparameters has a critical impact on the performance of machine learning models. In this case, the optimized parameters allowed Random Forest to achieve outstanding results in classifying flexibility data in male schoolchildren. The precision, F1-Score, and recall metrics indicate a well-fitted model with adequate generalization capacity.

The Random Forest algorithm obtained the best results for most of the tests. The following Fig. 3 to Fig. 5 show the ROC-AUC curves generated by the Random Forest algorithm for the Physical Fitness tests evaluated on male schoolchildren.

Fig. 3 shows the ROC-AUC curve for the flexibility test where the High class (AUC = 0.88): The model distinguishes between the "High" category and the other categories well. Low (AUC = 0.84), the model also distinguishes between the “Low” category and the different categories. Normal (AUC = 0.58), the model performs significantly lower in distinguishing between the “Normal” category and the others, indicating that the model is ineffective in this classification.

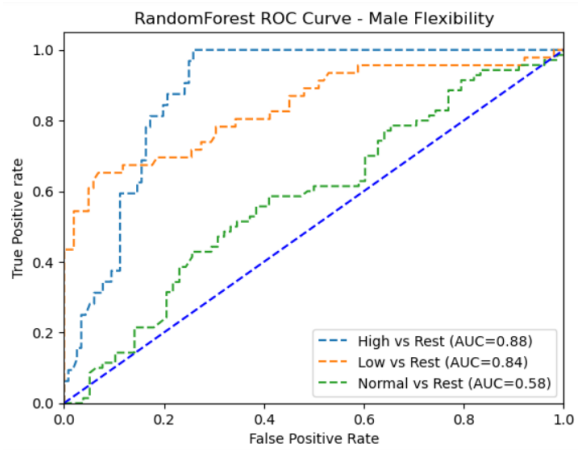


Fig. 3. Random forest ROC-AUC curve for the male flexibility test.

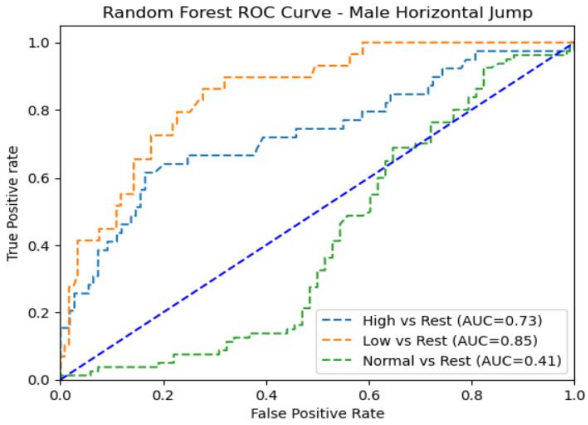


Fig. 4. Random forest ROC-AUC curve for the male horizontal jump test.

Fig. 4 shows the ROC-AUC curve for the horizontal jump test. The "High" class has an AUC of 0.73, indicating that the model performs moderately in distinguishing between this category and the others. The "Low" class has an AUC of 0.85, demonstrating the model's good performance in differentiating this category from the rest. In contrast, the "Normal" class has an AUC of 0.41, reflecting a relatively poor performance of the model in distinguishing between this category and the others.

Fig. 5 shows the ROC-AUC curve for the speed 20m test. For the High class (AUC = 0.83), the model performs well in distinguishing this category from the others. The model performs moderately for the Low class (AUC = 0.74). However, the model's performance is low for the Normal class (AUC = 0.53), indicating that it is ineffective for this classification.

In summary, from the tests of male schoolchildren, the model performs well for the "High" categories in all tests, with AUCs of 0.88, 0.73, and 0.83, respectively. The model has excellent performance for the "Low" category in flexibility and horizontal jump, with AUCs of 0.84 and 0.85, respectively, and moderate performance in speed with an AUC of 0.74. The model shows overall poor performance for the "Normal" category in all tests, with AUCs of 0.58, 0.41, and 0.53, respectively, indicating that the model's predictive ability is limited.

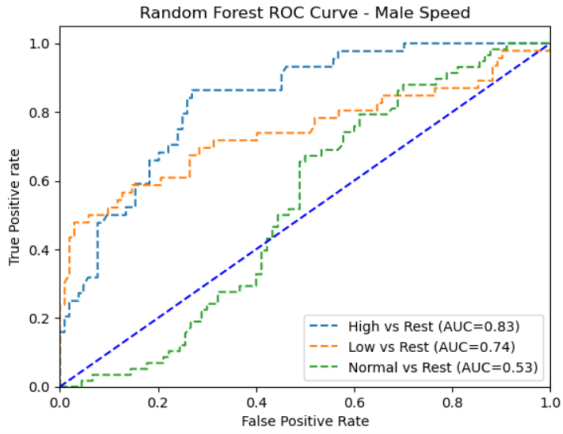


Fig. 5. Random forest ROC-AUC curve for the men's 20m speed test.

In conclusion, the Random Forest model performs well in classifying the “High” and “Low” categories in most tests but struggles to classify the “Normal” category correctly.

Fig. 6 to Fig. 8 show the ROC-AUC curves generated by the Random Forest algorithm for the Physical Fitness tests evaluated on female schoolchildren.

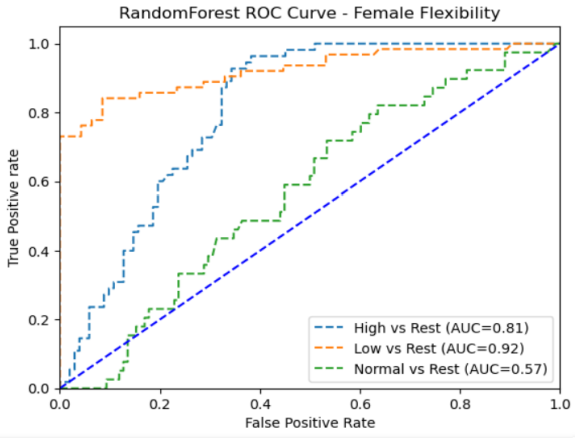


Fig. 6. Random forest ROC-AUC curve for the female flexibility test.

Fig. 6 shows the ROC-AUC curve for the flexibility test in female schoolchildren. The "High" class has an AUC of 0.81, indicating that the model performs well in distinguishing this category from the others. The "Low" class has an AUC of 0.92, demonstrating excellent model performance for this category. On the other hand, the "Normal" class has an AUC of 0.57, suggesting that the model's performance is moderately poor in distinguishing this category from the others.

The ROC-AUC curve for the "horizontal jump" test is shown in Fig. 7. The "High" class has an AUC of 0.71, indicating that the model has moderate performance in distinguishing between the "High" category and the others. The "Low" class has an AUC of 0.73 and moderately performs in distinguishing between the "Low" category and the others. However, for the "Normal" class, the AUC is 0.39, suggesting that the model performs poorly in distinguishing between the "Normal" category and the others, indicating that it is ineffective in this classification.

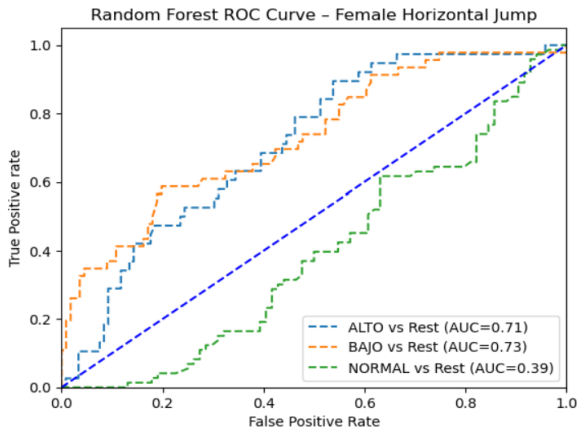


Fig. 7. Random forest ROC-AUC curve for the female horizontal jump test.

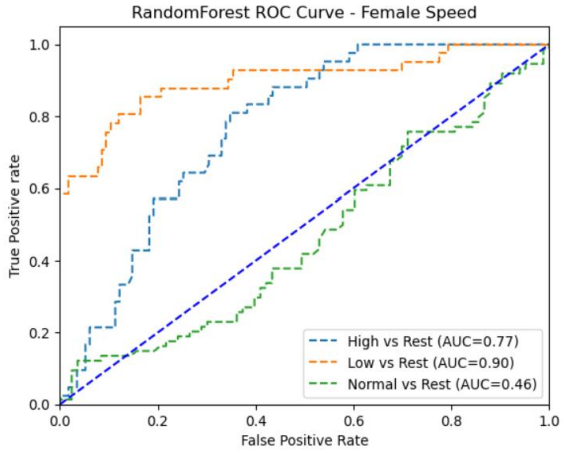


Fig. 8. Random forest ROC-AUC curve for the women's 20m speed test.

Fig. 8 shows the ROC-AUC curve for the Speed 20m High test (AUC = 0.77). The model distinguishes between the "High" category and the other categories. Low (AUC = 0.90): The model distinguishes between the “Low” category and the others. Normal (AUC = 0.46): The model performs poorly distinguishing between the “Normal” category and the others.

In summary, from the results of the tests of female schoolchildren, the model generally shows good performance for the "High" categories in all tests, with AUCs of 0.81, 0.71, and 0.77, respectively. The model has excellent performance for the "Low" category in flexibility and speed, with AUCs of 0.92 and 0.90, respectively, and moderate performance in horizontal jump with an AUC of 0.73. The model shows overall poor performance for the "Normal" category in all tests, with AUCs of 0.57, 0.39, and 0.46, respectively, indicating the model's limited predictive ability.

In conclusion, the Random Forest model performs well in classifying the “High” and “Low” categories in most tests but struggles to classify the “Normal” category correctly.

V. Discussion

This research used machine learning techniques to analyze the accuracy and effectiveness in classifying data related to male and female schoolchildren's flexibility, horizontal jump, and 20m speed. Algorithms evaluated included Decision Trees, Random Forest, Support Vector Machine (SVM), Naive Bayes, Logistic Regression, K-Nearest Neighbor (KNN), Multi-Layer Perceptron (MLP), Gradient Boosting, XGBoost, LightGBM and CatBoost [27]. The data was divided into 80% for training and 20% for testing. Hyperparameter optimization techniques were used to improve the accuracy of the models, as seen in Tables VI to XI. For example, the hyperparameters configured for Gradient Boost included criteria such as 'entropy,' 'max_depth,' and 'n_estimators,' which were manually tuned before model training.

The comparative results of the traditional and enhanced techniques with hyperparameters showed significant variations in accuracy, F1 score, recall, and precision for male and female schoolchildren. For the male group, the optimized Random Forest model showed outstanding accuracy in classifying flexibility, with an accuracy of 0.97, an F1-score of 0.97, and a recall of 0.95. This suggests the model can correctly predict physical activity classes with a low false positive rate. In contrast, the analysis of the female group revealed that the Random Forest model was also highly effective in classifying flexibility with an AUC of 0.92, indicating high sensitivity and specificity.

The ROC-AUC curves presented in Fig. 3 to Fig. 8 show the effectiveness of Random Forest models in classifying physical activities in schoolchildren, like studies of children aged 6 to 12 years [28]. Specifically, for male flexibility, the ROC-AUC curve showed a significant increase towards the upper left corner with an AUC of 0.88, confirming the high sensitivity of the model to detect the 'High' classification of physical activity. For female flexibility, the AUC was 0.92 for the 'Low' class, indicating high sensitivity and a low false positive rate.

Comparison tables showed that hyperparameter-optimized models significantly improved accuracy and other critical metrics compared to their non-optimized versions. For example, the SVM increased accuracy from 0.46 to 0.82 for the female group in the horizontal jump test. Similarly, the Gradient Boost model showed substantial improvements in accuracy and F1-score across multiple tests.

Comparing our results with those of other recent studies, we observed a congruence in the effectiveness of the Random Forest [29] and Gradient Boosting [30] algorithms. In our research and previous studies, these algorithms have repeatedly demonstrated their superiority in precision and recall in physical activity classification. This suggests that using advanced hyperparameter optimization techniques can significantly improve the accuracy of machine-learning models.

The consistency in results across multiple studies reinforces the validity of our findings and underscores the importance of selecting and optimizing appropriate algorithms for specific data classification tasks.

Limitations in the quality and quantity of data available for training machine learning models can significantly affect their performance. The challenge will be obtaining a large and representative data set, which can be difficult, especially in specific studies such as classifying physical activities in

schoolchildren. Another limitation is that incorrect selection of relevant features can decrease the effectiveness of the models. Determining which features are most informative requires deep domain knowledge and sometimes a thorough process of trial and error. Although hyperparameter optimization can significantly improve model performance, it is a resource-intensive process that requires time and computational power. The challenge is identifying the optimal combination of hyperparameters for each algorithm, which can be complicated and requires advanced search and cross-validation techniques.

VI. CONCLUSION

It is demonstrated that the application of machine learning algorithms, together with the optimization of hyperparameters, is an effective strategy for classifying students' physical condition in educational centers. Using the Knowledge Discovery in Databases (KDD) process and collecting anthropometric data and physical fitness tests could accurately assess the physical fitness of a representative sample of students.

The results indicated that the Random Forest and Gradient Boosting algorithms were particularly effective in classifying physical activities with high levels of precision, F1 score, recall, and specificity. These models' ability to differentiate between levels of physical fitness (below average, average, and above average) with a low false positive rate suggests their practical applicability in monitoring and evaluating students' physical health.

Future work could explore integrating these models into educational and health platforms and evaluating their long-term impact on student health. Other parameters and physical tests could also be considered for a more complete evaluation. The continued evolution of machine learning techniques and the availability of more granular data promise to further improve the accuracy and usefulness of these models.

ACKNOWLEDGMENT

To the 'Universidad Nacional de San Agustín de Arequipa', who has financed the project «Propuesta Normativa para valorar los niveles de actividad física de los escolares de la provincia de Arequipa», with contract number 15-2016-UNSA.

REFERENCES

[1] Kufel, J., Bargieł-Łączek, K., Kocot, S., Koźlik, M., Bartnikowska, W., Janik, M., ... & Gruszczyńska, K. (2023). What is machine learning, artificial neural networks and deep learning?—Examples of practical applications in medicine. Diagnostics, 13(15), 2582.

[2] Aized Amin Soofi and Arshad Awan, "Classification Techniques in Machine Learning: Applications and Issues," Journal of Basic & Applied Sciences, vol. 13, pp. 459–465, Jan. 2017, doi: 10.6000/1927-5129.2017.13.76.

[3] J. Sulla-Torres et al., "Quantification of the Number of Steps in a School Recess by Means of Smart Bands: Proposal of Referential Values for Children and Adolescents," Children, vol. 10, no. 6, p. 915, May 2023, doi: 10.3390/children10060915.

[4] WHO, "Physical activity," WHO. [Online]. Available: https://www.who.int/es/news-room/fact-sheets/detail/physical-activity

[5] H. N. C. Betancur, L. G. P. Canqui, Y. Y. R. Yapuchura, K. Pérez, S. Chura, and W. W. C. Castillo, "Obesidad infantil en estudiantes de educación primaria en Puno, Perú," Retos: nuevas tendencias en educación física, deporte y recreación, no. 54, pp. 466–477, 2024.

[6] "América Latina y el Caribe: Más de 4 millones de niños y niñas menores de 5 tienen sobrepeso." Accessed: Jun. 19, 2024. [Online]. Available: https://www.unicef.org/lac/comunicados-prensa/america-latina-caribe-mas-4-millones-ninos-ninas-menores-5-sobrepeso

[7] S. Andermo et al., "School-related physical activity interventions and mental health among children: a systematic review and meta-analysis," Sports Med Open, vol. 6, no. 1, p. 25, Dec. 2020, doi: 10.1186/s40798-020-00254-x.

[8] P. Trejo Ortiz, S. Jasso Chairez, F. Mollinedo Montaño, and L. Lugo Balderas, "Relación entre actividad física y obesidad en escolares," Revista Cubana de Medicina General Integral, vol. 28, no. 1, 2012.

[9] A. Rosa Guillamón, E. García-Cantó, P. L. Rodríguez García, J. J. Pérez Soto, M. L. Tárraga Marcos, and P. J. Tárraga López, "Physical activity, physical fitness and quality of diet in schoolchildren from 8 to 12 years," Nutr Hosp, vol. 34, no. 6, 2017.

[10] M. Zhou, Y. Fukuoka, K. Goldberg, E. Vittinghoff, and A. Aswani, "Applying machine learning to predict future adherence to physical activity programs," BMC Med Inform Decis Mak, vol. 19, no. 1, p. 169, Dec. 2019, doi: 10.1186/s12911-019-0890-0.

[11] S. A. Alsareii et al., "Physical Activity Monitoring and Classification Using Machine Learning Techniques," Life, vol. 12, no. 8, p. 1103, Jul. 2022, doi: 10.3390/life12081103.

[12] M. N. Ahmadi, T. G. Pavey, and S. G. Trost, "Machine Learning Models for Classifying Physical Activity in Free-Living Preschool Children," Sensors, vol. 20, no. 16, p. 4364, Aug. 2020, doi: 10.3390/s20164364.

[13] J. Sulla-Torres, A. Calla Gamboa, C. Avendaño Llanque, J. Angulo Osorio, and M. Zúñiga Carnero, "Classification of Motor Competence in Schoolchildren Using Wearable Technology and Machine Learning with Hyperparameter Optimization," Applied Sciences, vol. 14, no. 2, p. 707, Jan. 2024, doi: 10.3390/app14020707.

[14] U. Fayyad, G. Piatetsky-Shapiro, and P. Smyth, "From data mining to knowledge discovery in databases," AI Mag, vol. 17, no. 3, 1996.

[15] J. R. Fernández, D. T. Redden, A. Pietrobelli, and D. B. Allison, "Waist circumference percentiles in nationally representative samples of African-American, European-American, and Mexican-American children and adolescents," Journal of Pediatrics, vol. 145, no. 4, 2004, doi: 10.1016/j.jpeds.2004.06.044.

[16] F. Pedregosa et al., "Scikit-learn: Machine learning in Python," Journal of Machine Learning Research, vol. 12, 2011.

[17] J. Sulla-Torres, G. Luna-Luza, D. Ccama-Yana, J. Gallegos-Valdivia, and M. Cossio-Bolaños, "Neuro-fuzzy System with Particle Swarm Optimization for Classification of Physical Fitness in School Children," International Journal of Advanced Computer Science and Applications, vol. 11, no. 6, 2020, doi: 10.14569/IJACSA.2020.0110663.

[18] E. Engür and B. Soylu, "A linear multivariate decision tree with branch-and-bound components," Neurocomputing, vol. 576, 2024, doi: 10.1016/j.neucom.2024.127354.

[19] M. Schonlau and R. Y. Zou, "The random forest algorithm for statistical learning," Stata Journal, vol. 20, no. 1, 2020, doi: 10.1177/1536867X20909688.

[20] H. Kamel, D. Abdulah and J. M. Al-Tuwaijari, "Cancer Classification Using Gaussian Naive Bayes Algorithm," *2019 International Engineering Conference (IEC)*, Erbil, Iraq, 2019, pp. 165-170, doi: 10.1109/IEC47844.2019.8950650.

[21] JavaTpoint, "Support Vector Machine Algorithm," JavaTpoint, 2021.

[22] C. El Morr, M. Jammal, H. Ali-Hassan, and W. El-Hallak, "Logistic Regression," in International Series in Operations Research and Management Science, vol. 334, 2022. doi: 10.1007/978-3-031-16990-8_7.

[23] Y. chen Wu and J. wen Feng, "Development and Application of Artificial Neural Network," Wirel Pers Commun, vol. 102, no. 2, 2018, doi: 10.1007/s11277-017-5224-x.

[24] P. Cunningham and S. J. Delany, "K-Nearest Neighbour Classifiers-A Tutorial," ACM Computing Surveys, vol. 54, no. 6. 2021. doi: 10.1145/3459665.

[25] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2016. doi: 10.1145/2939672.2939785.

[26] A. Natekin and A. Knoll, "Gradient boosting machines, a tutorial," Front Neurorobot, vol. 7, no. DEC, 2013, doi: 10.3389/fnbot.2013.00021.

[27] C. Milanese, M. Sandri, V. Cavedon, and C. Zancanaro, "The role of age, sex, anthropometry, and body composition as determinants of physical fitness in nonobese children aged 6-12," PeerJ, vol. 2020, no. 3, 2020, doi: 10.7717/peerj.8657.

[28] S. R. Shakya, C. Zhang, and Z. Zhou, "Comparative study of machine learning and deep learning architecture for human activity recognition using accelerometer data," Int J Mach Learn Comput, vol. 8, no. 6, 2018, doi: 10.18178/ijmlc.2018.8.6.748.

[29] P. Probst, M. N. Wright, and A.-L. Boulesteix, 'Hyperparameters and tuning strategies for random forest', Wiley Interdisciplinary Reviews: data mining and knowledge discovery, vol. 9, no. 3, p. e1301, 2019.

[30] J. Guo et al., "An XGBoost-based physical fitness evaluation model using advanced feature selection and Bayesian hyper-parameter optimization for wearable running monitoring," Computer Networks, vol. 151, 2019, doi: 10.1016/j.comnet.2019.01.026.

Modeling Micro Traffic Flow Phenomena Based on Vehicle Types and Driver Characteristics Using Cellular Automata and Monte Carlo

Tri Harsono[1], Kohei Arai[2]
Dept. of Informatics and Computer Engineering, Politeknik Elektronika Negeri Surabaya (PENS), Surabaya, Indonesia[1]
Dept, of Information Science-Faculty of Science and Engineering, Saga University, Saga, Japan[2]

***Abstract*—The modeling of micro traffic flow on a highway has been extensively observed and studied in various aspects, such as driver characteristics in car-following and lane-changing behaviors. Regarding car-following and lane-changing, an interesting aspect is how to model the movement conditions of vehicles on a highway that exhibit unique characteristics regarding the speed of four-wheeled or more vehicles passing through it. This condition occurs on the Porong Highway in Sidoarjo, East Java, Indonesia. Based on these conditions, this study develops a microscopic traffic flow model incorporating driver characteristics categorized into three types: careful drivers, ordinary drivers, and skilled drivers, each with distinct vehicle speed traits. These driver characteristics are integrated into the Nagel-Schreckenberg Stochastic Traffic Cellular Automata (NaSch STCA) model, which we refer to as the Modified NaSch STCA. The Monte Carlo simulation is employed to generate events through random numbers for the Occupied Initial State, Slowdown Probability, and Probability of Lane Changing. These three components are integral parts of the Modified NaSch STCA model. Experiments (simulations) were conducted on the constructed vehicle movement model, and one of the outcomes is that the travel time obtained from the NaSch STCA model is significantly faster than that obtained from the Modified NaSch STCA model. This condition is attributed to the unique vehicle speed characteristics on the Porong Highway, where the average speed v_r = 38 km/h is relatively lower than the average speed typically observed on a highway.**

Keywords*—Micro traffic flow; driver characteristics; cellular automata; Monte Carlo

I. INTRODUCTION

Modeling and simulation of microscopic traffic flow have been extensively pursued by researchers, including the development of micro-traffic flow models based on intelligent transportation systems with wireless communication [1]; calibration of microscopic car-following (CF) models to accurately replicate and study traffic behavior and phenomena [2]; estimation and prediction of traffic states by integrating statistical data in both congested and uncongested scenarios [3]. Intelligent transportation systems offer an alternative to enhance traffic environments by integrating the Internet of Things and smart algorithms. These systems collect and process data from various sources to improve transportation efficiency. Research conducted by study [4] reviews the smart techniques employed for predicting traffic flow in urban areas. Additionally, it proposes a general taxonomy where the insights gained from traffic flow analysis merge with computational approaches. Microscopic urban traffic simulation using integrated modeling methods has been conducted, taking into account driver behavior characteristics related to car-following and lane changing. The results indicate that car-following behavior is more sensitive to variations in the status of adjacent vehicles and lane changes compared to lane-changing behavior during the lane-change process. This study also aids in analyzing travel characteristics and the impact mechanisms of vehicles in urban roads, serving as a guide for the development of sustainable transportation and autonomous vehicles in the future, and promoting efficient urban transportation system operations [5].

Microscopic traffic simulations are frequently employed to evaluate the effects of autonomous vehicles on safety and traffic flow. This study examines adaptive driver behavior by having drivers navigate the same route three times, each with a different level of autonomous vehicle penetration. The findings reveal that as the penetration level of autonomous vehicles increases, drivers adopt shorter waiting times, smaller following distances, and more consistent speeds closer to the maximum speed limit. These results indicate that driving behavior changes in response to variations in surrounding traffic composition [6].

Cellular automata have proven highly beneficial, not only in traffic flow simulation but also in diverse fields like pedestrian behavior for example: study a behavior-based cellular automata model that can represent heterogeneous crowd structures and explore the effects of different crowd compositions on pedestrian dynamics, particularly evacuation efficiency [7]. A multi-grid cellular automata model has been utilized to connect vehicle and pedestrian models. The enhanced Kerner-Klenov-Wolf (IKKW) model and a pedestrian movement model that incorporates Time to Collision (TTC) have been proposed. The application of these models to real-life scenarios has shown the impact of pedestrian intrusion behavior on traffic [8]. In the context of disaster mitigation, an expanded cellular automata model has been proposed for emergency evacuation dynamics involving pedestrians, utilizing parameters such as route change probability and group fields. Experiments were conducted to investigate the effects of this new extension, including the verification of related collective phenomena and the evaluation of safety performance metrics [9]. An LSTM-CA simulation for wildfire spread, combining Cellular Automata with Long

Short-Term Memory (LSTM), has been proposed. Real-world wildfire spread simulations have been conducted, and the accuracy of the wildfire spread predictions was verified using KAPPA coefficients, Hausdorff distances, and horizontal comparison experiments based on remote sensing imagery of wildfires [10].

Probabilistic Logistic Cellular Automata (LPCA) modeling has been carried out by integrating a basic logistic growth model with two-dimensional spatial dynamics to simulate the formation of regular patterns. The simulation results indicate that resource scarcity and environmental shape are the primary factors leading to the emergence of various regular patterns [11]. To prevent falling into local optima and to enhance convergence speed and global search potential, an advanced version of the Ant Colony Optimization (ACO) algorithm, known as the Cellular Automata-Based Enhanced Ant Colony Optimization Algorithm (CA-IACOA), has been explored. Simulation results suggest that this algorithm is effective in addressing DDoS attacks, as it achieves high-quality solutions for identifying optimal nodes and reliable routing paths [12]. The phenomena observed in modeling and simulating various aspects using cellular automata indicate its capability to provide solutions to both simple and complex problems.

In the use of cellular automata for a particular problem, the role of randomness and Monte Carlo simulation is crucial. Many researchers emphasize the incorporation of randomness into the rules of cellular automata, such as using randomness in the probability of lane switching [13], [14]. Monte Carlo simulation plays a significant role in modeling and simulating dynamic systems using cellular automata, including calculating simulation data for traffic queuing problems [15] and predicting real-time traffic flow based on normal distribution [16].

In micro traffic flow modeling, driver behavior is a critical parameter. One aspect involves defensive maneuvers by drivers to avoid obstacles and braking, such as encountering a pedestrian appearing suddenly on the road ahead of the vehicle [17]. The smart road stud (SRS) not only drastically alters microscopic driving characteristics but also significantly influences driver decision-making processes during overtaking maneuvers [18]. On the other hand, research on modeling driving behavior in developing countries is conducted using microsimulation approaches with multi-agents, deemed suitable for accurately replicating driving behaviors [19]. Adaptive driver behavior is observed through repeated driving of the same route three times with varying penetration rates of automated vehicles. It is demonstrated that driving behavior changes as the traffic composition around them changes [20]. It is noted that following behavior is more sensitive to variations in lateral vehicle movements and lane changes [21].

In the context of micro traffic flow modeling, many researchers utilize driver behavior as a key parameter. However, this driver behavior is rarely depicted based on the micro traffic flow phenomena observed on highways, especially the phenomenon of vehicle speeds passing through the roadway. Cellular automata, as a method of dynamic system specific to micro traffic flow modeling, employs rules that require processes of randomness, including determining initial density probabilities and lane-changing probabilities. This study employs Monte Carlo simulations for the randomness processes embedded within the cellular automata rules. A survey of vehicle speeds in micro traffic flow was conducted on the Porong Highway in Sidoarjo, East Java, Indonesia, focusing on four-wheeled vehicles: trucks/trailers, buses, public transportation, and private cars. The phenomenon of vehicle speeds passing through this highway served as the basis for categorizing driver characteristics. Based on normalized speed data from each vehicle type, drivers were categorized as follows: careful drivers for truck/trailer drivers, ordinary drivers for bus and public transportation drivers, and skilled drivers for private car drivers. A modified cellular automata model was developed based on these driver characteristic categories, and an analysis of traffic flow simulation results was conducted to assess the accuracy of the model. The results of this study are expected to benefit relevant stakeholders, such as government agencies involved in highway transportation. The findings, which include travel times based on vehicle speed characteristics, can provide valuable insights into the comfort and safety of driving on the Porong-Sidoarjo Highway in East Java, Indonesia.

This study continues with an explanation of the stochastic traffic cellular automata (STCA) model and Monte Carlo methods, which are discrete-time simulation models. It then addresses phenomena related to microscopic traffic flow characteristics observed on the Porong-Sidoarjo Highway in East Java, Indonesia, specifically focusing on driver and vehicle characteristics based on vehicle speed in Section II. The micro traffic flow phenomena is given in Section III. Proposed model is given in Section IV. This is followed by a section on testing the developed model and a discussion of the relevant results from these tests in Section V. The study concludes in Section VI with a summary of the research findings.

II. Discrete Time Simulation Model: Cellular Automata-Monte Carlo

A. Stochastic Model: Nagel–Schreckenberg STCA

In the realm of traffic simulation, modeling at the microscopic level has long been recognized as a intricate and time-intensive endeavor, requiring intricate models that portray the behaviors of individual vehicles. However, approximately ten years ago, a novel microscopic model emerged, drawing on the cellular automaton framework rooted in statistical physics. Its principal advantage lies in its efficient and swift performance during computer simulations, although it may exhibit slightly diminished accuracy at the microscopic level. These traffic models based on cellular automata (TCA) are dynamic systems characterized by discrete elements, where time progresses in distinct increments and space is represented in coarse units (for instance, roads divided into 7.5 meter per-cell, each either empty or occupied by a vehicle) [22].

In terms of randomness usage (probability), cellular automata models can be categorized into two types: deterministic models and stochastic models. One deterministic TCA model compares two acceleration models embedded within cellular automata, stating that the Acceleration Time Delay (ATD) model and Speed Adaptation (SA) model exhibit

spatiotemporal traffic congestion patterns consistent with empirical findings. In both models, the onset of congestion in free flow conditions on congested roads is associated with a first-order phase transition from free flow to synchronized flow; moving congestion spontaneously emerges only in synchronized flow [23].

Stochastic Traffic Cellular Automata (STCA) are widely used in modeling micro traffic flow. The basic model of STCA introduces a rule involving randomization, utilizing randomness within the TCA rules. This adaptation accounts for natural speed fluctuations caused by human behavior or various external conditions [24]. Incorporating this randomness into TCA transforms it into Stochastic Traffic Cellular Automata (STCA). Many researchers have developed STCA models from various aspects, focusing on updating acceleration in car-following or lane-changing behaviors.

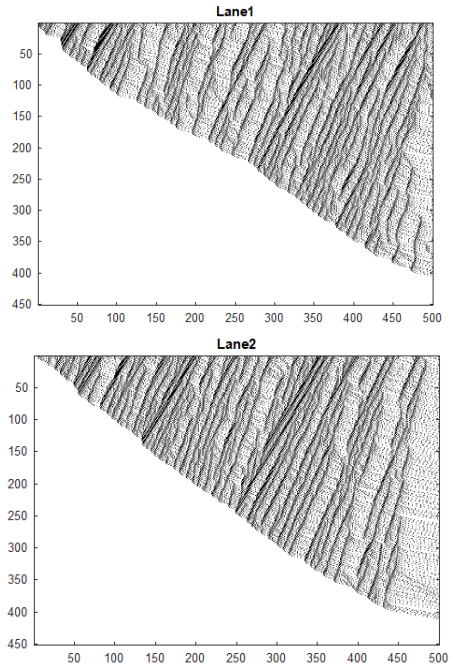


Fig. 1. Time-space diagram of NaSch STCA (two lanes), where the vertical axis represents time-steps and the horizontal axis represents space (cells), road length L = 500 cells, density k = 0.3, slowdown prob. P = 0.3, prob. of lane changing P_{lc} = 0. Top image for lane 1 and bottom image for lane 2.

The STCA developed by Nagel-Schreckenberg is defined on a one-dimensional array with a lattice length L and operates as an open-loop system. Each location (cell) can either be occupied by one vehicle or remain empty. The speed of each vehicle is represented by an integer between zero and $v_{\max}$. Updates in the NaSch STCA system consist of four sequential steps performed in parallel for all vehicles. The update rules in the NaSch STCA system (utilizing two lanes in this study) are as follows [22], [24]:

1) Acceleration: $v_{i,j}(t) \leftarrow v_{i,j}(t-1) + 1$

Iϕ $v_{i,j}(t-1) < v_{max}$ ανδ $gs_{i,j}(t-1) > v_{i,j}(t-1) + 1$

2) Braking: if $gs_{i,j}(t-1) \leq v_{i,j}(t-1)$,

$$v_{i,j}(t) \leftarrow gs_{i,j}(t-1) - 1$$

3) Randomization: with slowdown probability P and random number $\xi(t)$, if $\xi(t) < P \Longrightarrow v_{i,j}(t) \leftarrow v_{i,j}(t-1) - 1$.

4) Vehicle movement:

$$x_{i,j}(t) \leftarrow x_{i,j}(t-1) + v_{i,j}(t)$$

where, $v_{i,j}(t)$ is the speed of the vehicle in the i-th lane and j-th position, and $gs_{i,j}(t)$ is space gap, the distance between a vehicle and the vehicle in front of it. Fig. 1 shows one of the simulation results of NaSch STCA for two lanes with the specifications of a lattice length L = 500 cells, density k = 0.3, slowdown probability P = 0.3, probability of lane changing P_{lc} = 0. Vehicles move from left to right, and the system operates as an open-loop system.

B. Monte Carlo Simulation

The Monte Carlo simulation involved generating events through random numbers. This process comprised data collection, assigning random numbers, formulating models, and performing analysis. One reason for using a Monte Carlo simulation is that it typically applies to simulations utilizing stochastic methods to create new configurations of the system being studied [25]. On the other hand, the Monte Carlo Simulation procedure outlined by [15] and utilized in this research involves: (i) Step 1. Data collection, which uses a pseudo-random sequence; (ii) Step 2. Random-number assignment, where events are generated impartially by assigning random numbers in proportion to their probability of occurrence. The standard Monte Carlo method with pseudo-random sequences can achieve good convergence with N sample tests. In predicting traffic flow, the Monte Carlo Simulation serves as a mathematical tool to model risk or uncertainty in a system through the generation of random variables. The Monte Carlo Simulation model is designed to forecast traffic patterns. To create a new dataset with random probabilities, parameters such as the mean and standard deviation from the fitted normal distribution were utilized, as described by [16].

In other areas, the relationship between Cronbach's alpha and randomness was tested using Monte Carlo simulations, as opposed to issues with minimum sample width and bias. Simulation-generated artificial data were used to estimate the alpha coefficient for a K-item scale with a 5-point Likert-type format, answered randomly by 5K individuals, where K denotes the number of items. Each trial was conducted 5000 times, and one finding was that the probability of a Cronbach's alpha coefficient of 0.27 or higher for a K-item 5-point Likert scale, randomly answered by 5K people, is less than 5% [26].

The Monte Carlo simulation procedure, as generally outlined by [15], includes Step 1: Data collection and Step 2: Random-number assignment has been applied in this study. The Monte Carlo Simulation generates random numbers in three areas: the initial state of occupancy, randomization (slowdown probability), and lane-changing probability. These components are integrated into the micro traffic flow modeling addressed in this research. Below is an explanation of each section.

1) Occupied initial state: In this section, random numbers are generated using Monte Carlo, where the random value is generated to be less than a predetermined density percentage k of vehicles. If this condition is satisfied, the initial position of the vehicle is assigned to the corresponding cell, and its initial speed vij(t) is set according to a normal distribution. All vehicles will be moved in parallel per time-step after generating their initial positions and velocities. Based on the density percentage, random positions $x_{ij}(t)$ of vehicles are generated on a lattice length L as specified. The highway specifications used in this experiment replicate conditions on the Porong-Sidoarjo Highway in East Java, Indonesia. The traffic flow direction under study is from the cities of Sidoarjo/Surabaya towards Malang/Banyuwangi (one-way). This highway has two lanes (i = 1-2), and typical straight road length is 3750 meters, thus if 1 cell = 7.5 meters, the lattice length of the road is 500 cells (j = 1-500). In the Cellular Automata model, cells occupied by vehicles are identified with the number 1, while unoccupied cells are identified with 0. Here is the syntax for generating random numbers based on Monte Carlo:

```
for j=1:n
   for i=1:2
      r=rand;
      if(r<k)
         x(1,i,j)=1;
         v(1,i,j)=floor(vmax/2+0.69*randn);
      end
   end
end
```

The initial conditions of vehicle speeds $v_{ij}(t)$ are represented in matrix form with indices (1,i,j). Their values are computed using the floor function with the argument $\frac{v_{max}}{2} + 0{,}69\ x\ randn$, where randn is a function that generates random numbers from a standard normal distribution. Therefore, $v_{ij}(1,i,j)$ will contain the value from this expression after it has been rounded down to the nearest integer using the floor function.

2) Randomization (Slowdown probability): One of the rules in Nagel-Schreckenberg's STCA involves randomization, specifically reducing the speed by 1 cell per time-step for vehicles that satisfy the randomization condition stated in the rule. The value of the random number $\xi(t)$ is generated using Monte Carlo methods, where its magnitude is less than a predefined probability value p (referred to as the slowdown probability). Here is the syntax for generating random numbers based on Monte Carlo, applied from the initial simultaneous movement of vehicles until the end of the specified time period.

```
for j=1:n
   for i=1:2
      if(x(t,i,j)==1)
         r=rand;
         if(r<p)
            if(v(t,i,j)==5)
               ...
               ...
            end
         end
      end
   end
end
```

3) Probability of lane changing: In this study, we refer to the structure of the Porong Sidoarjo highway where each direction has two lanes. Lane changes occur simultaneously between lane 1 and lane 2 as vehicles move over time. Monte Carlo plays a role in generating random numbers to determine which vehicles will change lanes. The random number generated is smaller than a predetermined lane change probability p_{lc}. The syntax for lane change from lane 1 to lane 2 is described below. This condition is equivalent to lane change from lane 2 to lane 1.

```
for j=6:n
if(x(t,1,j)==1)
r=rand;
if (r<plc)
   if(v(t,1,j)==5)
      ...
   end
   ...
   if(v(t,1,j)==1)
      ...
   end
end
end
end
```

III. Micro Traffic Flow Phenomena, Vehicle Types, Driver Characteristics

Traffic flow survey has been conducted in the area of Porong Sidoarjo Highway, East Java, Indonesia, where the traffic direction is specifically from the cities of Sidoarjo/Surabaya towards Malang/Banyuwangi (one-way). This highway serves as a main road connecting major cities including Surabaya, Sidoarjo, Malang, Jember, and

Banyuwangi. The economic growth among these cities is linked to the existence of Porong Sidoarjo Highway. The presence of this highway is highly significant and gives rise to distinct micro-traffic flow phenomena. The survey was conducted from March 23rd to March 30th, 2010. Speed data of vehicles was collected hourly over 24 hours a day for eight days, totaling 190 data points. The surveyed vehicle types included four-wheeled or more vehicles such as trucks/trailers, buses, public transportation, and private cars. Vehicle speed is measured using a speed gun.

A. Micro Traffic Flow Phenomena

The survey of speeds from four types of vehicles conducted every hour yielded speed data, where each vehicle type has several speed data points per hour. The mean speed data for each vehicle type per hour was calculated, resulting in 190 mean speed data points per vehicle type. The phenomenon of these mean speeds shows variability in the data. It is desired that the speed data have a normal distribution so that descriptive statistics such as mean and standard deviation can accurately depict the patterns of mean speed and its variability. This includes the potential to determine driver characteristics based on mean speed data. The phenomenon of mean speed data from four types of vehicles over 190 consecutive hours is depicted in Fig. 2.

The results of the vehicle speed survey were analyzed to determine the form of the population distribution or its probability by testing the hypothesis that a specific distribution serves as the model for the speed population. This study conducted hypothesis testing on speed data to assess whether the speed data or population follows a normal distribution or not. The hypothesis test employed a formal goodness-of-fit test procedure based on the chi-square distribution.

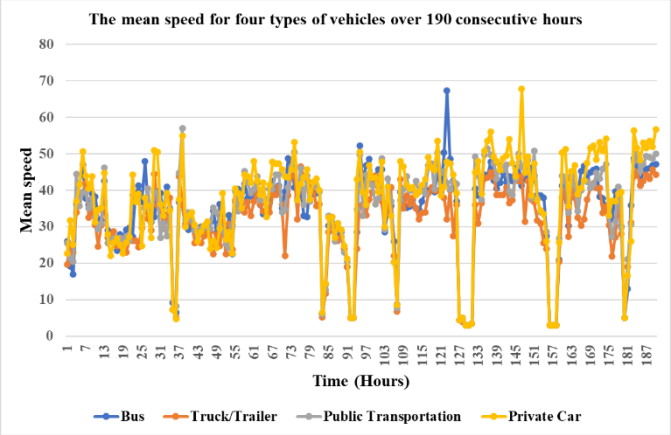


Fig. 2. The mean speed of four types of vehicles for 190 consecutive data points (190 hours).

The testing procedure requires several parameters: a random sample of size n; class interval c; observed frequency in class interval i, O_i; expected frequency in class interval i, E_i (given from the hypothesized probability distribution). Eq. (1) is used as the test statistic to analyze whether the speed data pattern conforms to a normal distribution or not.

$$X_0^2 = \sum_{i=1}^{c} \frac{(O_i - E_i)^2}{E_i} \quad (1)$$

A statistical test is conducted on the population to determine if it follows the hypothesized distribution. The test statistic, X_0^2 has approximately a chi-square distribution with $c - p - 1$ degrees of freedom, where p represents the number of parameters of the hypothesized distribution estimated by sample statistics. The hypothesis is rejected if the calculated value of the test statistic $X_0^2 > \chi^2_{\alpha,c-p-1}$.

In this hypothesis test, conducted on data concerning the mean speeds of all types of vehicles, the sample size n = 190 is utilized. With a significance level α = 0.05, the hypothesis test aims to determine whether traffic survey data (vehicle speeds) can be adequately modeled by a normal distribution. The test employs c = 8 class intervals, which, for a standard normal distribution, divide the distribution area into eight segments with equal probabilities: [0, 0.32), [0.32, 0.675), [0.675, 1.15), [1.15, ∞), and their mirrored intervals on the other side of zero.

Each interval has the probability p_i = 1/8 = 0.125, thus the expected cell frequencies $E_i = np_i$ = 190(0.125) = 23.75. Note that the parameter values specified are n = 190; $\alpha = 0.05$; c = 8 cells; p_i = 1/8 = 0.125; and $E_i = np_i$ = 23.75, which are consistent across all types of vehicles. Di sisi lain sebaran data kecepatan untuk semua jenis kendaraan memiliki mean μ = 35 dan standard deviation Std = 11.32.

TABLE I. TEST THE DISTRIBUTION OF SURVEY DATA (MEAN SPEED) FOR ALL TYPES OF VEHICLES (FOUR TYPES OF VEHICLES)

Class Interval	Observed Frequency o_i	Expected Frequency E_i	oi - Ei	(oi - Ei)²	(oi - Ei)²/Ei
x < 22	19	23,75	-4,75	22,56	0,95
22 ≤ x < 27	12	23,75	-11,75	138,06	5,81
27 ≤ x < 31	21	23,75	-2,75	7,56	0,32
31 ≤ x < 35	19	23,75	-4,75	22,56	0,95
35 ≤ x < 39	28	23,75	4,25	18,06	0,76
39 ≤ x < 43	41	23,75	17,25	297,56	12,53
43 ≤ x < 48	41	23,75	17,25	297,56	12,53
48 ≤ x	9	23,75	-14,75	217,56	9,16
Total	**190**	**190**			**43,01**

A hypothesis testing procedure is employed to determine if the sample data set of mean speeds follows a normal distribution.

H_0: The distribution takes on a normal form

H_1: The distribution does not adhere to a normal form

The test statistic is

$$X_0^2 = \sum_{i=1}^{c} \frac{(O_i - E_i)^2}{E_i}$$
$$= \frac{(19 - 23.75)^2}{23.75} + \frac{(12 - 23.75)^2}{23.75} + \cdots$$
$$+ \frac{(9 - 23.75)^2}{23.75} = 43.01$$

Table I shows observed and expected frequencies for each cell, as well as the results of the chi-square distribution calculation.

- The chi-square table with $\alpha = 0.05$ and degrees of freedom = $c - p - 1 = 8 - 2 - 1 = 5$ (p represents the number of parameters in the hypothesized distribution, which in this instance are the parameters of the normal distribution specifically, the mean μ and the variance σ^2. Thus, there are two parameters, so $p = 2$).

$$\chi^2_{\alpha,c-p-1} = \chi^2_{0.05,5} = 11.07$$

- Conclusion: since $X_0^2 = 43.01 > \chi^2_{0.05,5} = 11.07$, rejecting H_0 suggests that the speed data for all vehicles does not follow a normal distribution.

The results of the hypothesis test using a chi-square distribution indicate that the distribution is not normal, meaning that the mean speed data (survey results) from four types of vehicles do not exhibit a normal distribution. As mentioned above, transforming traffic flow survey data (vehicle speed data) into a normal distribution provides a strong basis for more in-depth statistical analysis, simplifies data interpretation, and enhances the validity of the analysis results to support decision-making.

Normalization of the mean speed data for four types of vehicles was conducted individually for each vehicle due to varying extreme data (outliers) they possess. Below is the normalization applied to the mean speed data of the truck/trailer.

Normalizing the speed of trucks/trailers.

Here is the normalization mechanism for the mean speed data of trucks/trailers.

- Transform the data using the natural logarithm function.
- Remove the outlier data. After removal, the dataset consists of 170 entries, down from the original 190.
- The mean is 3.53 and the standard deviation is 0.20.
- Appling the hypothesis-testing procedure:

  H_0: The distribution follows a normal form.

  H_1: The distribution does not follow a normal form

- The statistical test value is

$$X_0^2 = \sum_{i=1}^{c} \frac{(O_i - E_i)^2}{E_i}$$
$$= \frac{(27 - 21.25)^2}{21.25} + \frac{(17 - 21.25)^2}{21.25} + \cdots$$
$$+ \frac{(21 - 21.25)^2}{21.25} = 8.26$$

Each interval has a probability p_i = 1/8 = 0.125, thus the expected cell frequencies $E_i = np_i$ = 170(0.125) = 21.25 for each interval.

Table II shows observed and expected frequencies for each cell, as well as the results of the chi-square distribution calculation.

- The chi-square table with $\alpha = 0.05$ and degrees of freedom = $c - p - 1 = 8 - 2 - 1 = 5$ (p represents the number of parameters in the hypothesized distribution, which in this instance are the parameters of the normal distribution specifically, the mean μ and the variance σ^2. Thus, there are two parameters, so $p = 2$).

$$\chi^2_{\alpha,c-p-1} = \chi^2_{0.05,5} = 11.07$$

- Conclusion: Since $X_0^2 = 8.26 < \chi^2_{0.05,5} = 11.07$, accepting H_0 indicates that the speed data for the truck/trailer follows a normal distribution.

TABLE II. NORMALIZATION OF MEAN SPEED SURVEY DATA FOR TRUCK / TRAILER

Class Interval (Natural logarithmic numbers)	Observed Frequency o_i	Expected Frequency E_i	oi - Ei	(oi - Ei)^2	(oi - Ei)^2/Ei
x < 3,300	27	21,25	5,75	33,06	1,56
3,300 ≤ x < 3,395	17	21,25	-4,25	18,06	0,85
3,395 ≤ x < 3,466	13	21,25	-8,25	68,06	3,20
3,466 ≤ x < 3,530	18	21,25	-3,25	10,56	0,50
3,530 ≤ x < 3,594	22	21,25	0,75	0,56	0,03
3,594 ≤ x < 3,665	26	21,25	4,75	22,56	1,06
3,665 ≤ x < 3,760	26	21,25	4,75	22,56	1,06
3,760 ≤ x	21	21,25	-0,25	0,06	0,00
Total	**170**	**170**			**8,26**

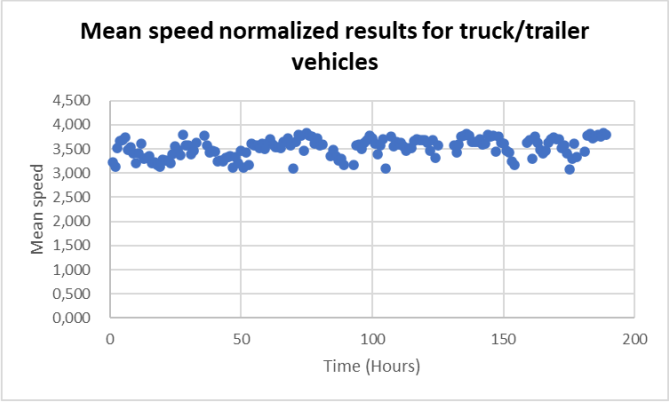


Fig. 3. Mean speed normalized results for truk/trailer vehicles.

Fig. 3 illustrates the normalized mean speed results for trucks/trailers, based on 170 data points (after removing outliers). The normalized mean speed data yielded a minimum of 3.086 and a maximum of 3.839, with respective actual mean speeds of 22 and 47.

Using the same method, normalization was also performed on the mean speed data for buses, public transportation, and private cars. Table III summarizes the normalization process for these three vehicle types. The normalized data counts for buses, public transportation, and private cars are 160, 170, and

160 respectively. The degrees of freedom used for all three types of vehicles are the same, which is 5. However, the significance levels (alpha) differ: 0.05 for buses, 0.005 for public transportation, and 0.010 for private cars. All chi-square distribution calculations for the mean speed data of these three vehicle types yielded values smaller than the chi-square values in the table. This condition indicates that the distribution of mean speed data for buses, public transportation, and private cars can be assumed to be normal.

TABLE III. SUMMARY OF MEAN SPEED SURVEY DATA NORMALIZATION FOR BUS, PUBLIC TRANSPORTATION, AND PRIVATE CAR VEHICLES

Vehicle	Steps	Data Specifications	Decision
Bus	▪Change the mean speed data to the function of natural logarithmic ▪Delete the extreme data (outlier's data) ▪Appling the hypothesis-testing procedure (H_o = normal distribution, H_1 = not normal distribution)	Normalized amount of data = 160; Degree of freedom = 5; alpha = 0.05; Mean = 3,66; Std = 0,14; Chi-square Dist. = 8,93; Chi-square Table = 11,07; Chi-square Dist. (=8,93) < Chi-square Table (=11,07)	H_o accepted; Normal Distribution
Public Transportation		Normalized amount of data = 170; Degree of freedom = 5; alpha = 0.005; Mean = 3,62; Std = 0,20; Chi-square Dist. = 15,98; Chi-square Table = 16,75; Chi-square Dist. (=15,98) < Chi-square Table (=16,75)	H_o accepted; Normal Distribution
Private Car		Normalized amount of data = 160; Degree of freedom = 5; alpha = 0.010; Mean = 3,70; Std = 0,21; Chi-square Dist. = 13,70; Chi-square Table = 15,09; Chi-square Dist. (=13,70) < Chi-square Table (=15,09)	H_o accepted; Normal Distribution

B. Driver Characteristics Based on Vehicle Speed Phenomena

The phenomenon of micro-traffic flow on Porong Sidoarjo Highway has been investigated. Based on speed surveys conducted on four types of vehicles (trucks/trailers, buses, public transportation, and private cars), mean speed data requiring normalization due to non-normal data distribution was obtained. The normalized mean speed data (in natural logarithm and real numbers) with min-max speed values are shown in Table IV. The vehicle speed phenomenon from the survey results and the normalization process of speed data are used as a basis to establish the characteristics of drivers passing through Porong Sidoarjo Highway. The normalized vehicle speed data obtained from the survey depict the characteristics of drivers crossing Porong Sidoarjo Highway more accurately. This condition considers how they regulate their speed according to their skill levels and safety preferences. Under these circumstances, driver characteristics are categorized into three groups: (i) Careful driver (for truck or trailer drivers); (ii) Ordinary driver (for bus and public transportation drivers); and (iii) Skilled driver (for private vehicle drivers).

TABLE IV. THE PHENOMENON OF VEHICLE SPEED TO DETERMINE DRIVER CHARACTERISTICS

Vehicle	Mean speed (natural logarithmic number)	Mean speed (real number)	(min - max) normalized speed	Driver characteristics
Truck/Trailer	3,53	34	22 - 47	Careful driver
Bus	3,66	39	27 - 52	Ordinary driver
Public Transportation	3,62	37	21 - 52	Ordinary driver
Private Car	3,7	41	26 - 57	Skilled driver

Characteristics of careful drivers regarding the speed phenomenon that occurs on Porong Sidoarjo Highway, East Java, Indonesia: (i) They are cautious, prioritizing safety over speed; (ii) They drive at a moderate speed, maintain distance from the vehicle ahead, and are ready to react quickly to changes in traffic; (iii) They drive at a speed lower than average to ensure safety and comfort. As for ordinary drivers, their characteristics are: (i) They typically follow basic traffic rules and drive at a comfortable speed that is appropriate for traffic conditions; (ii) They carefully follow the flow of traffic, adjusting their speed to the surrounding traffic conditions; (iii) They exhibit speeds close to the average. The characteristics of skilled drivers are: (i) They have a higher level of expertise in handling various driving situations; (ii) They can make quick decisions and maneuver vehicles effectively without compromising safety; (iii) They are capable of driving at speeds higher than average, yet within safe limits and well-controlled.

IV. THE PROPOSED MODEL OF MICRO TRAFFIC FLOW PHENOMENA

In the previous section, the characteristics of drivers were categorized based on the speed phenomena observed on Porong Highway in Sidoarjo, East Java, Indonesia. Using the results of a vehicle speed survey conducted on this highway, these characteristics were classified into three types: careful driver, ordinary driver, and skilled driver. This section involves mathematical modeling of the vehicle speeds for each driver type based on the phenomena observed on Porong Highway in Sidoarjo. Subsequently, these speed models are integrated into the NaSch STCA rules. Lane-changing behavior is also incorporated into the STCA model, considering two lanes as per the real conditions on Porong Highway in Sidoarjo.

A. Driver Characteristics Modeling

The three predefined driver characters exhibit fundamental differences in their speeds. Careful drivers maintain lower driving speeds below the average to prioritize safety and comfort. Ordinary drivers typically drive at speeds close to the average. Skilled drivers are capable of driving at speeds higher than the average, yet within safe limits and well-controlled. Here are the statements regarding the speed modeling for each driver character:

- careful driver: $1 \leq v_{cd} \leq \bar{v}_r$
- ordinary driver: $v_{od} = \bar{v}_r \pm 1$
- skilled driver: $\bar{v}_r \leq v_{sd} \leq v_{max}$

where, v_{cd} , v_{od} , and v_{sd} are the respective speeds of careful drivers, ordinary drivers, and skilled drivers in sequence, while $\bar{v}_r$ represents the average speed for all vehicle types passing through Porong Highway in Sidoarjo.

Based on the speed data from the survey conducted on Porong Highway in Sidoarjo over 190 consecutive hours (eight days), the average speed for all vehicle types (referring to Table IV) $\bar{v}_r$ is 38 km/h. Converting this to computational terms, where 1 cell equals 7.5 meters and 1 second corresponds to 1 time-step, the average vehicle speed passing through Porong Highway in Sidoarjo $\bar{v}_r$ becomes 1.4 cells/time-step. In computational calculations, this is rounded up to 2 cells/time-step.

B. Lane Changing

The Porong Sidoarjo Highway features two lanes in each direction. An illustration of two lanes in one direction is shown in Fig. 4. A vehicle is represented by a small box. If a vehicle intends to change lanes, it must satisfy the condition that on the new lane, there is a distance of *b* units between itself and the vehicle behind, and a distance of *a* units between itself and the vehicle in front. By meeting these conditions, collisions between vehicles can be avoided.

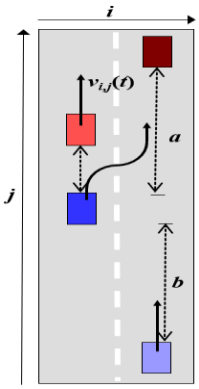


Fig. 4. Illustration of a two-lane highway with lane-changing dynamics.

Many studies have focused on lane-changing behavior in multi-lane and dual-lane scenarios. One such study examines lane-changing on a two-lane highway, as investigated by [27]. Based on the illustration in Fig. 4, the lane-changing used in this research must satisfy the following conditions:

$gs_{i=1,j}(t-1) < \min\{v_{i=1,j}(t-1), v_{max}\}$

$gs_{i=2,j-b}(t-1) > v_{max}$ and

$gs_{i=2,j+a}(t-1) > gs_{i=1,j}(t-1)$ with probability of lane changing P_{lc}.

C. Modified NaSch STCA

The vehicle movement model in the micro-traffic flow simulation conducted on Porong Sidoarjo Highway refers to the phenomenon of vehicle speeds observed on that highway. Based on the survey results of vehicle speed measurements, drivers' characteristics are categorized into three types: careful drivers, ordinary drivers, and skilled drivers. The fundamental difference among them lies in the typical driving speeds they maintain while traversing Porong Sidoarjo Highway.

In this study, the proposed vehicle movement is a modification of the NaSch STCA vehicle movement, where vehicle speeds are differentiated into three types according to predefined driver characteristics. The modified NaSch STCA follows these rules:

1) Acceleration: $v_{i,j}(t) \leftarrow v_{i,j}(t-1) + 1$

Iϕ $v_{i,j}(t-1) < v_{max}$ ανδ $gs_{i,j}(t-1) > v_{i,j}(t-1) + 1$

Where

- for careful drivers applies $1 \leq v_{i,j}(t-1) \leq \bar{v}_r$
- for ordinary drivers applies $\bar{v}_r - 1 \leq v_{i,j}(t-1) \leq \bar{v}_r + 1$
- for skilled drivers applies $\bar{v}_r \leq v_{i,j}(t-1) \leq v_{max}$

2) Braking: if $gs_{i,j}(t-1) \leq v_{i,j}(t-1)$,

$$v_{i,j}(t) \leftarrow gs_{i,j}(t-1) - 1$$

3) Randomization: with slowdown probability P and random number $\xi(t)$, if $\xi(t) < P \Longrightarrow v_{i,j}(t) \leftarrow v_{i,j}(t-1) - 1$.

4) Vehicle movement:

$$x_{i,j}(t) \leftarrow x_{i,j}(t-1) + v_{i,j}(t)$$

Together with the lane-changing rules stated in the previous session: $gs_{i=1,j}(t-1) < \min\{v_{i=1,j}(t-1), v_{max}\}$

$$gs_{i=2,j-b}(t-1) > v_{max} \text{ ανδ}$$

$$gs_{i=2,j+a}(t-1) > gs_{i=1,j}(t-1)$$

with probability of lane changing P_{lc}.

V. Experimental Results and Discussion

In this session, a trial was conducted on vehicle movement using a predetermined model known as the Modified NaSch STCA. As explained in the previous session, drivers are categorized into three types based on speed, which is characteristic of each driver (careful, ordinary, and skilled drivers). The determination of the number of each type of driver is based on probabilities (percentages) relative to the predetermined vehicle density. The probabilities for careful drivers, ordinary drivers, and skilled drivers are denoted as P_{cd}, P_{od}, and P_{sd} respectively.

A. Experimental Results

Fig. 5 illustrates a time-space diagram, one of the outcomes of the trial (simulation) of vehicle movement using the Modified NaSch STCA, replicating the phenomenon of vehicle movement on the Porong-Sidoarjo Highway. The specifications of this vehicle movement simulation are as follows: a two-lane highway with a lattice length of L = 500 cells, density k = 0.3, slowdown probability P = 0.1, probability of lane changing P_{lc} = 0.4, probability of careful drivers P_{cd} = 0.1, probability of ordinary drivers P_{od} = 0.3, and probability of skilled drivers P_{sd} = 0.6. The vehicles move in one direction, from left to right, replicating the movement from

the city of Sidoarjo / Surabaya towards Banyuwangi / Malang, and this movement system operates as an open-loop system.

Based on Fig. 5, it can be seen that with a vehicle density of k = 50% (0.5), where the probability of trucks/trailers (careful drivers) being present is 10% (0.1), the probability of buses and public transportation (ordinary drivers) is 30% (0.3), and the probability of private cars (skilled drivers) is 60% (0.6), all vehicles cover a distance of 500 cells = (500 x 7.5 meters) = 3750 meters. This results in a total travel time of t = 3161 time-steps ≈ 3161 seconds ≈ 52.68 minutes. This condition indicates that with the average speed per type of vehicle passing through the Porong Sidoarjo highway as shown in Table IV, where the minimum speed value of all types of vehicles = 22 km/h and the maximum = 57 km/h, the average speed for all types of vehicles = 38 km/h. This phenomenon highlights that with a vehicle density of k = 50% and an average speed $\bar{v}_r$ = 38 km/h, the travel time for all vehicles over a distance of 3750 meters is 3161 seconds (52.68 minutes).

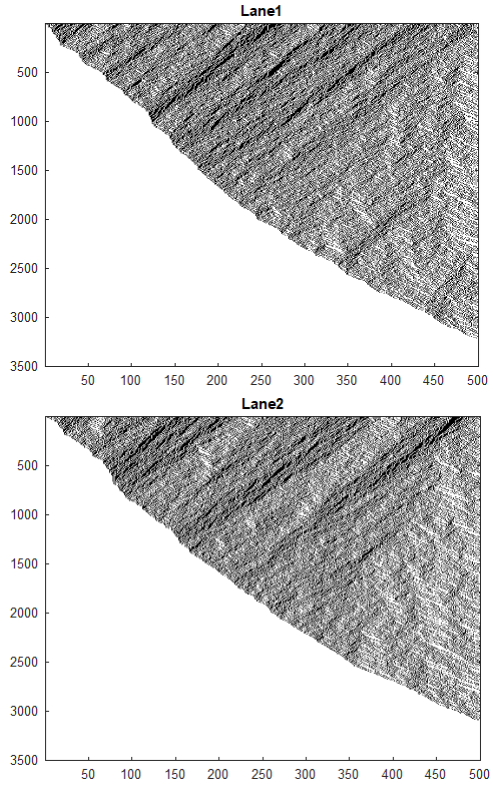


Fig. 5. Time-space diagram of Modified NaSch STCA (two lanes), where the vertical axis represents time-steps and the horizontal axis represents space (cells), road length L = 500 cells, density k = 0.5, slowdown prob. P = 0.3, prob. of lane changing P_{lc} = 0.4, prob. of careful drivers P_{cd} = 0.1, prob. of ordinary drivers P_{od} = 0.3, and prob. of skilled drivers P_{sd} = 0.6. Top image for lane 1 and bottom image for lane 2.

The comparison of the time-space diagram between NaSch STCA and Modified NaSch STCA is illustrated by one of the test results (simulations) as shown in Fig. 6. The test specifications employed are as follows: the road length 500 cells, with a vehicle density of k = 0.3, there is a 30% probability of slowdown (P = 0.3) and a 0% probability of lane changing (P_{lc} = 0.0). In the Modified NaSch STCA model, the probabilities P_{cd} = 0.1 for careful drivers, P_{od} = 0.3 for ordinary drivers, and P_{sd} = 0.6 for skilled drivers. It can be seen that with the same specifications, over a distance of 500 cells (3750 meters), the travel time for all vehicles in the NaSch STCA model is 409 time-steps (409 seconds), while the modified model takes 1934 time-steps (1934 seconds). This indicates that the travel time in the NaSch STCA model is significantly faster than in the modified model. This phenomenon occurs because the average speed of vehicles traveling on the Porong Highway tends to be lower than usual.

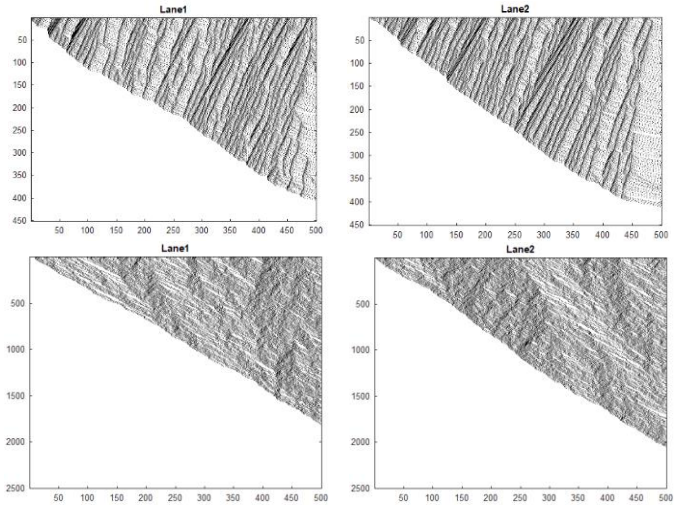


Fig. 6. A comparison of time-space diagrams for two lanes between the NaSch STCA model and the Modified NaSch STCA. In this visualization, the vertical axis corresponds to time-steps, while the horizontal axis represents space measured in cells. The road length is L = 500 cells, with a density of k = 0.3. The probability of slowdown is P = 0.3, and the probability of lane changing is P_{lc} = 0.0. For the Modified NaSch STCA model, the probabilities are as follows: P_{cd} = 0.1 for careful drivers, P_{od} = 0.3 for ordinary drivers, and P_{sd} = 0.6 for skilled drivers. The top image shows the time-space diagram for the NaSch STCA model, while the bottom image depicts the Modified NaSch STCA model.

A comparison of travel time versus vehicle density was conducted between the NaSch STCA model and the Modified NaSch STCA model. The simulation specifications include a road length of 500 cells, a slowdown probability P = 0.3, and a lane-changing probability P_{lc} = 0.3. For the Modified NaSch STCA model, the probabilities for careful drivers, ordinary drivers, and skilled drivers are P_{cd} = 0.1, P_{od} = 0.2, and P_{sd} = 0.7, respectively. Travel time for vehicles was calculated for each density value incrementing by 0.1, ranging from 0.1 to 0.9. Detailed travel time results for each density value k increasing by 0.1 are shown in Table V. It can be observed that, for both models, travel time increases with higher vehicle density. At the same density value, the NaSch STCA model exhibits significantly faster travel times compared to the Modified NaSch STCA model. This condition is consistent across all density values. For instance, in this simulation, the NaSch STCA model yields a travel time of 138 time-steps for a density of k = 0.1, increasing to 1196 time-steps for a density of k = 0.9. Meanwhile, for the Modified NaSch STCA model, the travel time is 1364 time-steps at a density of k = 0.1, increasing further with each increment in vehicle density, reaching 5641 time-steps at k = 0.9. It can be stated that the travel time for the NaSch STCA model is significantly faster than for the Modified NaSch STCA model. This condition is

attributed to the unique vehicle speed characteristics on the Porong Sidoarjo highway, where the average speed for all vehicle types $\bar{v}_r$ = 38 km/h is lower than the general average speed.

It is also noted the difference in travel time between the Modified NaSch STCA model and the NaSch STCA model. Table V shows that the difference for a density of k = 0.1 is 1226 time-steps. The travel time difference increases as vehicle density rises. For a density of k = 0.9, the travel time difference is 4445 time-steps. The simulation results are also generally illustrated in the line graph shown in Fig. 7.

TABLE V. VEHICLE TRAVEL TIME AND THE DIFFERENCES BETWEEN THE NSCH STCA MODEL AND THE MODIFIED NSCH STCA MODEL

Density k	The travel time (time-steps)		Difference (time-steps)
	NSch STCA	Modified NSch STCA	
0,1	138	1364	1226
0,2	252	1698	1446
0,3	394	1999	1605
0,4	509	2284	1775
0,5	672	3208	2536
0,6	802	3510	2708
0,7	939	4360	3421
0,8	1041	5094	4053
0,9	1196	5641	4445

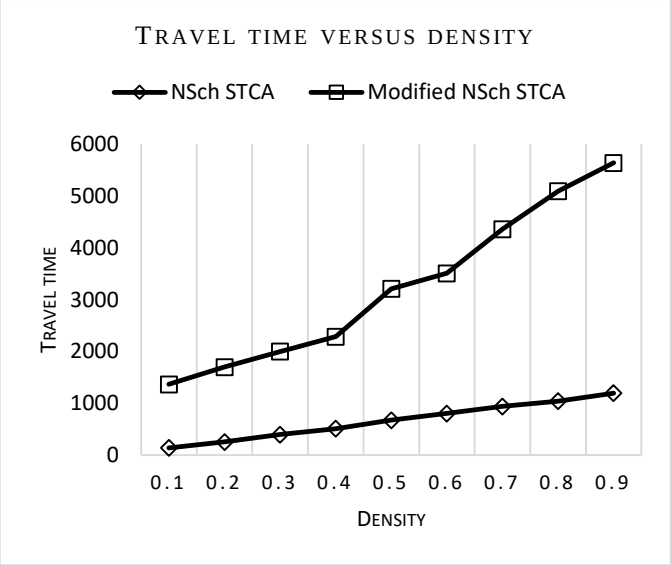


Fig. 7. Comparison of travel time with vehicle density between the NaSch STCA model and the Modified NaSch STCA model. The road length is L = 500 cells, the probability of slowdown is P = 0.3, and the probability of lane changing is P_{lc} = 0.3. Specifically for the Modified NaSch STCA model, the probabilities of the presence of careful drivers, ordinary drivers, and skilled drivers are P_{cd} = 0.1, P_{od} = 0.2, and P_{sd} = 0.7, respectively.

This study also examined vehicle travel time relative to the set travel distance, ranging from 100 cells to 500 cells. A comparison was made between the travel times of the NaSch STCA model and the Modified NaSch STCA model. Fig. 8 shows that the Modified NaSch STCA model exhibits significantly greater travel times for each specified distance compared to the NaSch STCA model.

The mean speed values produced by the modified NaSch STCA model were analyzed in relation to vehicle density. Simulation results with specifications of road length L = 500 cells; slowdown probabilities P = 0.1, 0.50, and 0.9 in sequence; and lane-changing probability P_{lc} = 0.3 are shown in Fig. 9. It is observed that higher vehicle densities lead to a decrease in mean speed. This condition aligns with the $(k, \bar{v}_r)$ diagram model described.

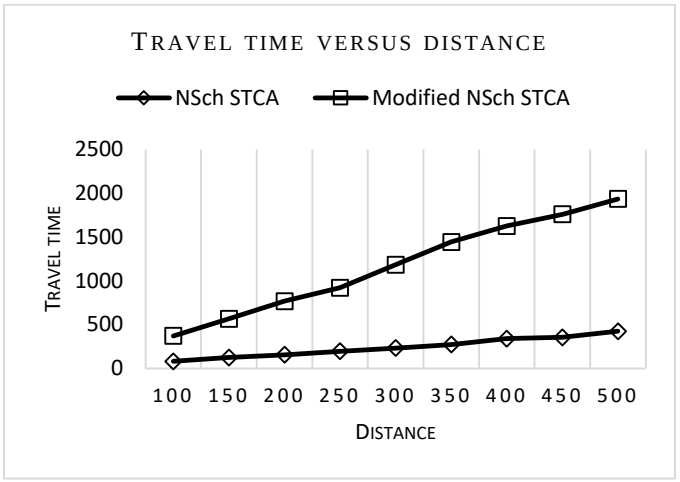


Fig. 8. Comparison of travel time with distance traveled between the NaSch STCA model and the Modified NaSch STCA model. The road length is L = 500 cells, density k = 0.3, the probability of slowdown is P = 0.3, and the probability of lane changing is P_{lc} = 0.3. Specifically for the Modified NaSch STCA model, the probabilities of the presence of careful drivers, ordinary drivers, and skilled drivers are P_{cd} = 0.1, P_{od} = 0.2, and P_{sd} = 0.7, respectively.

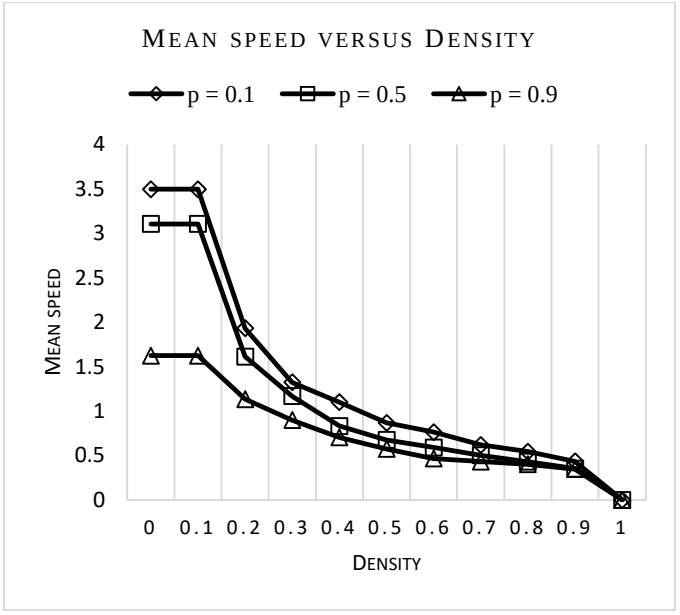


Fig. 9. Several $(k, \bar{v}_r)$ diagrams from the modified NaSch STCA model (mean speed versus density) with the specifications: road length L = 500 cells; slowdown probability P = 0.1, 0.5, and 0.9 in sequence; lane-changing probability P_{lc} = 0.3.

Delayed acceleration and slowdown probability have been incorporated into the modified NaSch STCA model. Regarding delayed acceleration, this study references, which is described by Takayasu–Takayasu TCA (T^2-*TCA*) in R2 as follows:

(R2) delayed acceleration:

$$v_i(t) = 0 \wedge g_{si}(t) \geq 2 \Longrightarrow v_i(t+1) \leftarrow 1$$

The presence of delayed acceleration and slowdown probability complements each other, as both impact the overall performance of the traffic system and contribute to increased travel time or congestion. When many vehicles experience delayed acceleration and there is also a high probability of slowdown, traffic can become more unstable and inefficient. This study demonstrates the effect of slowdown probability on travel time, with delayed acceleration incorporated into the Modified NaSch STCA model. It can be stated that the effect of slowdown probability on travel time is also influenced by the presence of delayed acceleration within the system. Simulation results for travel time in relation to slowdown probability are shown in Fig. 10. Observations were made for low, medium, and high densities, specifically k = 0.3, 0.5, and 0.9, with a lane-changing probability P_{lc} = 0.3.

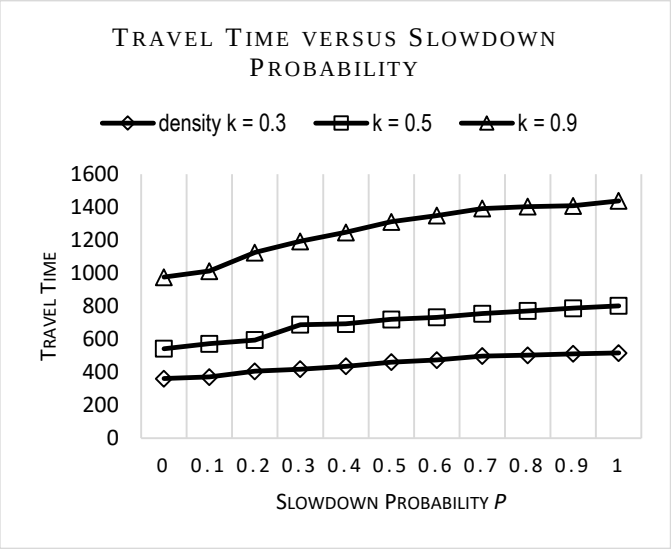


Fig. 10. Travel time observed as a function of slowdown probability P for densities k = 0.3, 0.5, and 0.9 sequentially; with a lane-changing probability P_{lc} = 0.3.

B. Discussion

The workflow of this study is as follows: (i) Conduct a survey of traffic phenomena on the Porong Highway in East Java, Indonesia. The survey data includes vehicle types, vehicle speeds, traffic density, and questionnaire responses about the conditions experienced by road users on Porong Highway. (ii) Process the survey data and analyze the observed phenomena. (iii) Examine the patterns emerging from the data analysis. (iv) Identify research topics that can be pursued based on the available data.

Based on the phenomena observed from the survey data, a unique finding is the variation in vehicle speeds, where different vehicle types exhibit different average speeds. Consequently, a microscopic traffic flow model was developed based on the speed characteristics of each vehicle type. Since vehicles are driven by individuals who influence their speed by accelerating or decelerating, the vehicle characteristics are inherently linked to the driver.

The discrete dynamic system model used is Cellular Automata, specifically the NaSch STCA model, which is a stochastic method utilizing random number generation. The random number generator employed in this model is the Monte Carlo method, integrated into the NaSch STCA with three components utilizing Monte Carlo simulation: the Occupied Initial State, Slowdown Probability, and Probability of Lane Changing. The innovation in this microscopic traffic flow modeling is the categorization of driver characteristics based on vehicle speeds observed on the Porong-Sidoarjo Highway, categorized into: careful driver, ordinary driver, and skilled driver, each with specific vehicle speed profiles.

The anticipated outcome is the alignment of the developed microscopic traffic flow model with the indications of vehicle travel times in relation to parameters such as density, travel distance, and slowdown probability. Traffic flow simulations on the Porong Highway have been conducted, producing vehicle travel times that replicate the actual traffic conditions on that road. The simulation specifications were adapted to match the conditions of the Porong Highway: the traffic flow under examination is from the Sidoarjo/Surabaya area towards Malang/Banyuwangi (one-way). This highway features two lanes (i = 1-2) and a typical straight road length of 3750 meters. Therefore, with each cell representing 7.5 meters, the lattice length of the road is 500 cells (j = 1-500).

For example, when using a slowdown probability P = 0.3 and a lane-changing probability P_{lc} = 0.3, the Modified NaSch STCA model assigns probabilities of P_{cd} = 0.1, P_{od} = 0.2, and P_{sd} = 0.7 for careful, ordinary, and skilled drivers, respectively. Vehicle travel times were computed for each density value incrementing by 0.1, from 0.1 to 0.9. Detailed results of travel times for each density increment are presented in Table V. For instance, the NaSch STCA model has a travel time of 138 time-steps at a density of k = 0.1, which increases to 1196 time-steps at k = 0.9. In contrast, the Modified NaSch STCA model shows a travel time of 1364 time-steps at k = 0.1, which grows to 5641 time-steps at k = 0.9. This indicates that the NaSch STCA model has significantly faster travel times compared to the Modified NaSch STCA model. This disparity is due to the unique vehicle speed characteristics on the Porong-Sidoarjo highway, where the average speed $\bar{v}_r$ =38 km/h is lower than the general average speed.

VI. CONCLUSION

The phenomenon of micro traffic flow on the Porong Highway in East Java, Indonesia, exhibits distinctive characteristics. A survey of vehicles with four or more wheels was conducted over eight days (190 hours) on this highway. The survey identified four types of vehicles: trucks/trailers, buses, public transportation, and private cars. The survey data revealed that these vehicles have average speeds of 34 km/h, 39 km/h, 37 km/h, and 41 km/h, respectively. Based on these speed characteristics, drivers of each vehicle type were categorized as careful drivers for trucks/trailers, ordinary

drivers for buses and public transportation, and skilled drivers for private cars.

Vehicle movement modeling was performed using the NaSch STCA (NaSch STCA) model, specifically the Modified NaSch STCA (Modified NaSch STCA), tailored to simulate conditions on the surveyed Porong Sidoarjo highway. The road length was 3750 meters (L = 500 cells), straight with two lanes in each direction. Simulation results with specific parameters (road length of 500 cells, density k = 0.3, slowdown probability P = 0.3, and probability of lane changing P_{lc} = 0.0) compared data between the NaSch STCA and Modified NaSch STCA models. The Modified NaSch STCA model included specific driver presence probabilities: 0.1 for careful drivers, 0.3 for ordinary drivers, and 0.6 for skilled drivers.

One of the comparisons between these models is their travel time, where the NaSch STCA model significantly outperformed the Modified NaSch STCA model. This outcome reflects the characteristic vehicle speeds on the Porong Sidoarjo highway, with an average speed for all vehicle types $\bar{v}_r$ = 38 km/h.

For future research, it is essential to develop micro traffic flow modeling based on driver characteristics on a particular roadway, in collaboration with local government authorities. A more effective approach would be to combine Cellular Automata with artificial intelligence, enhancing the modeling of dynamic micro traffic systems and driver characteristic classification.

ACKNOWLEDGMENT

Special thanks to the Knowledge Engineering research team at the Politeknik Elektronika Negeri Surabaya (PENS), East Java, Indonesia, for their support and motivation throughout the completion of this research. Particularly, heartfelt gratitude goes to Professor Kohei Arai from Saga University, Japan, for his comprehensive guidance and in-depth discussions on the cellular automata methods employed in the modeling and simulation aspects of this research.

REFERENCES

[1] Md. Anowar Hossain, Jun Tanimoto, A microscopic traffic flow model for sharing information from a vehicle to vehicle by considering system time delay effect, Physica A: Statistical Mechanics and its Applications, Vol 585, 2022, https://doi.org/10.1016/j.physa.2021. 126437.

[2] Zelin Wang, Zhiyuan Liu, Qixiu Cheng, Ziyuan Gu, Integrated self-consistent macro-micro traffic flow modeling and calibration framework based on trajectory data, Transportation Research Part C: Emerging Technologies, Volume 158, 2024, https://doi.org/10.1016/j.trc.2023.104439.

[3] Yuyan Annie Pan, Jifu Guo, Yanyan Chen, Qixiu Cheng, Wenhao Li, Yanyue Liu, A fundamental diagram based hybrid framework for traffic flow estimation and prediction by combining a Markovian model with deep learning, Expert Systems with Applications, Volume 238, Part E, 2024, https://doi.org/10.1016/j.eswa.2023.122219.

[4] Boris Medina-Salgado, Eddy Sánchez-DelaCruz, Pilar Pozos-Parra, Javier E. Sierra, Urban traffic flow prediction techniques: A review, Sustainable Computing: Informatics and Systems, Volume 35, 2022, https://doi.org/10.1016/j.suscom.2022.100739.

[5] Zhang L, Qu D, Zhang X, Dai S, Wang Q. Vehicle Driving Behavior Analysis and Unified Modeling in Urban Road Scenarios. Sustainability. 2024; 16(5):1956. https://doi.org/10.3390/su16051956.

[6] Rins de Zwart, Kas Kamphuis, Diane Cleij, Driver behavioural adaptations to simulated automated vehicles, potential implications for traffic microsimulation, Transportation Research Part F: Traffic Psychology and Behaviour, Volume 92, 2023, https://doi.org/10.1016/j.trf.2022.11.012.

[7] Yang Li, Maoyin Chen, Xiaoping Zheng, Zhan Dou, Yuan Cheng, Relationship between behavior aggressiveness and pedestrian dynamics using behavior-based cellular automata model, Applied Mathematics and Computation, Vol 371, 2020, https://doi.org/10.1016/j.amc.2019. 124941.

[8] Jinghui Wang, Wei Lv, Yajuan Jiang, Guangchen Huang, A cellular automata approach for modelling pedestrian-vehicle mixed traffic flow in urban city, Applied Mathematical Modelling, Volume 115, 2023, Pages 1-33, https://doi.org/10.1016/j.apm.2022.10.033.

[9] L.A. Pereira, D. Burgarelli, L.H. Duczmal, F.R.B. Cruz, Emergency evacuation models based on cellular automata with route changes and group fields, Physica A: Statistical Mechanics and its Applications, Volume 473, 2017, https://doi.org/10.1016/j.physa.2017.01.048.

[10] Li X, Zhang M, Zhang S, Liu J, Sun S, Hu T, Sun L. Simulating Forest Fire Spread with Cellular Automation Driven by a LSTM Based Speed Model. Fire. 2022; 5(1):13. https://doi.org/10.3390/fire5010013

[11] Jingyao Sun, Xinrong Li, Ning Chen, Yanli Wang, Guang Song, Regular pattern formation regulates population dynamics: Logistic growth in cellular automata, Ecological Modelling, Volume 418, 2020, https://doi.org/10.1016/j.ecolmodel.2019.108878.

[12] K. Deepa Thilak, A. Amuthan, Cellular Automata-based Improved Ant Colony-based Optimization Algorithm for mitigating DDoS attacks in VANETs, Future Generation Computer Systems, Volume 82, 2018, Pages 304-314, https://doi.org/10.1016/j.future.2017.11.043.

[13] Han, Z.; Xie, G.; Zhou, Y.; Zhuo, Y.;Wang, Y.; Shen, L. "Dynamic Response Analysis of Long-Span Bridges under Random Traffic Flow Based on Sieving Method". Buildings, Vol. 13, No. 2389, 2023.

[14] Lu, H.; Sun, D.; Hao, J. "Random Traffic Flow Simulation of Heavy Vehicles Based on R-Vine Copula Model and Improved Latin Hypercube Sampling Method". Sensors, Vol. 23, No. 2795, pp. 1-13, 2023.

[15] Huang, W.-T.; Dang, J.-F. "The Dynamic Adjusting Model of Traffic Queuing Time—A Monte Carlo Simulation Study", Applied Sciences, 10, 6364, 2020.

[16] Niraj Kumar Shah, Prena Chaudhary, Gopi Chandra Kaphle, "Real-Time Traffic Prediction Using Monte Carlo Simulation: A Case Study of Kantipath Road, Kathmandu, Nepal", BIBECHANA Vol. 19, No. (1-2), pp. 83-89, September 2022.

[17] Jurecki, R.S.; Stańczyk, T.L. "Modelling Driver's Behaviour While Avoiding Obstacles". Appl. Sci., 13, 616, 2023.

[18] Li, M.; Luo, Q.; Fan, J.; Ning, Q. "Impact Analysis of Smart Road Stud on Driving Behavior and Traffic Flow in Two-Lane Two-Way Highway". Sustainability, 15, 11559, 2023.

[19] Gracian, V.A., Galland, S., Lombard, A., "Behavioral models of drivers in developing countries with an agent-based perspective: a literature review". Autonomous Intelligent Systems, Vol. 4, No.5, April 2024.

[20] Rins de Zwart, Kas Kamphuis, Diane Cleij, "Driver behavioural adaptations to simulated automated vehicles, potential implications for traffic microsimulation", Transportation Research Part F: Traffic Psychology and Behaviour, Vol. 92, pp. 255-265, 2023.

[21] Zhang L, Qu D, Zhang X, Dai S, Wang Q. "Vehicle Driving Behavior Analysis and Unified Modeling in Urban Road Scenarios". Sustainability, 16(5):1956, 2024.

[22] Sven Maerivoet, Bart De Moor. "Cellular automata models of road traffic", Physics Reports 419: 1 – 64, 2005.

[23] Boris S. Kerner and Sergey L. Klenov, "Deterministic microscopic three-phase traffic flow models", Journal of Physics A: Mathematical and General, Vol. 39, pp. 1775-1809, 2005.

[24] Kai Nagel, Michael Schreckenberg. "A cellular automaton model for freeway traffic". Journal de Physique I, Vol. 2 (12), pp.2221-2229, 1992.

[25] Wen-Tso Huang, Ping-Shun Chen, John J. Liu, Yi-Ru Chen, Yen-Hsin Chen, Dynamic configuration scheduling problem for stochastic medical

resources, Journal of Biomedical Informatics, Vol. 80, pp. 96-105, 2018. https://doi.org/10.1016/j.jbi.2018.03.005.

[26] Recep Bindak, Relationship between Randomness and Coefficient Alpha: A Monte Carlo Simulation Study, Journal of Data Analysis and Information Processing, Vol.1 No.2, pp. 13-17, 2013.

[27] S. Rajeswaran, S. Rajasekaran, "Analyzing of Two-Lane Traffic Flow Simulation Model using Cellular Automata", International Journal of Computational Science and Mathematics, ISSN 0974-3189, Vol. 4, No. 2, pp. 77-90, 2012.

AUTHORS' PROFILE

Tri Harsono, He earned his Bachelor's degree in Mathematics from Sepuluh Nopember Institute of Technology, Surabaya in 1993, followed by a Master's degree in Information Technology from the same institution in 2005. He pursued his doctoral studies at Saga University, Japan, in the Department of Information Science, Faculty of Science and Engineering, from 2008 to 2011, and completed his PhD in September 2011. Since October 1993, he has been a lecturer in the Computer Engineering program at the Electronics Engineering Polytechnic Institute of Surabaya (EEPIS) or Politeknik Elektronika Negeri Surabaya (PENS). His research primarily focuses on modeling and simulation, particularly in micro traffic flow, evacuation systems, and prediction systems, with a specific interest in the behavior of objects within these systems. He is also focused on data analysis, particularly in recognizing patterns and classifying data using cellular automata techniques and learning algorithms. His most recent research involves classifying medical imaging data with cellular learning automata.

Kohei Arai, He received BS, MS and PhD degrees in 1972, 1974 and 1982, respectively. He was with The Institute for Industrial Science and Technology of the University of Tokyo from April 1974 to December 1978 also was with National Space Development Agency of Japan (Current JAXA) from January, 1979 to March, 1990. During from 1985 to 1987, he was with Canada Centre for Remote Sensing as a Post Doctoral Fellow of National Science and Engineering Research Council of Canada. He moved to Saga University as a Professor in Department of Information Science on April 1990. He was a councilor for the Aeronautics and Space related to the Technology Committee of the Ministry of Science and Technology during from 1998 to 2000. He was a councilor of Saga University for 2002 and 2003. He also was an executive councilor for the Remote Sensing Society of Japan for 2003 to 2005. He is a Science Council of Japan Special Member since 2012. He is an Adjunct Professor of Brawijaya University, Kurume Insitute of Technology and Nishi-Kyushu University. He also was Vice Chairman of the Science Commission "A" of ICSU/COSPAR for 2008-2016 then he is now award committee member of ICSU/COSPAR. He wrote 87 books and published 710 journal papers as well as 650 conference papers. He received 66 of awards including ICSU/COSPAR Vikram Sarabhai Medal in 2016, and Science award of Ministry of Mister of Education of Japan in 2015. He is now Editor-in-Chief of IJACSA and IJISA. http://teagis.ip.is.saga-u.ac.jp/index.html

Convolutional Neural Network Model for Cacao Phytophthora Palmivora Disease Recognition

Jude B. Rola[1], Jomari Joseph A. Barrera[2], Maricel V. Calhoun[3], Jonah Flor Oraño – Maaghop[4], Magdalene C. Unajan[5], Joshua Mhel Boncalon[6], Elizabeth T. Sebios[7], Joy S. Espinosa[8]
Department of Computer Science and Technology, Visayas State University, Baybay City, Philippines[1, 2, 3, 4, 5, 6, 8]
Department of Information Technology Education, Abuyog Community College, Philippines[7]

***Abstract*—Cacao, scientifically known as Theobroma cacao, is a highly nutritious food and is extensively utilized in multiple sectors, including agriculture and health. Nevertheless, the agricultural sector encounters notable obstacles as a result of Cacao disease such as pod rot, predominantly attributed to the Phytophthora genus. The objective of this work is to conduct a comparative analysis to determine the most effective machine-learning technique for the detection of P. palmivora infection in Cacao pods. Few studies have delved into this topic previously, but this study focuses in utilizing a little larger dataset, achieving better model, and attaining higher accuracy. A total of 2000 images of cacao pods, both healthy and disease-infected were collected. Subsequently, the images were subjected to manual classification by a domain expert based on the discernible presence or absence of the disease. The study examined six machine learning algorithms, specifically Naïve Bayes, Random Forest, Hoeffding Tree, Multilayer Neural Network, and Convolutional Neural Network (CNN). The CNN model had 99% level of accuracy, the highest among the five machine learning algorithms in the testing phase. The methodology has the potential to significantly advance sustainable agricultural practices and disease management. To enhance the model's recognition capabilities, additional datasets encompassing a broader range of Cacao varieties is necessary.**

Keywords*—Machine-learning; Convolutional Neural Network; detection of P. palmivora

I. Introduction

Cacao (Theobroma cacao), also known as “superfood” or “the food of the gods”, has gained full attention by farmers, researchers, and even health enthusiasts [1]. Consumption of this product worldwide is estimated to be over 4.5 million tons annually and is still growing. The innovation of new products that requires cacao as an ingredient has increased the demand for such. Examples of these are not limited to cosmetics, pharmaceuticals, food and beverages, and health food supplement. The Philippines' geolocation and the humid temperature make it conducive for cacao production to flourish and possibly attenuate poverty; thus, local farmers' interest scaled up, and exporters push for a more productive cacao industry that can participate in the worldwide supply competition [2].

However, the yield of cacao production is affected by several diseases, with pod rots as the most common adverse conditions of the tree. Phytophthora, a genus of straminipilous organisms (formerly classified as oomycetes), is responsible for the devastating black pod rot, the most widespread disease affecting cacao crops. Depending on environmental conditions, this disease can lead to annual losses of up to 90% in pod production. The entire plant can become infected, causing severe damage [3].

Traditionally, diagnosing infections relies on human experts. However, aside from the high cost and time-consuming nature, this approach can be impractical due to the scarcity of experts and potential geographic barriers, especially if the farm is remote. In such cases, the disease might affect a large number of pods. To address these challenges, researchers are turning to machine learning techniques for plant disease recognition as a viable solution.

Although few studies have previously explored this topic, this research aims to use a larger dataset to develop a better model and achieve higher accuracy.

Moreover, this study aimed to apply different machine learning techniques to the dataset for comparative analysis; and evaluate the accuracy of the models in recognizing the incidence of P. palmivora disease on Cacao pods. The dataset covers the cultivars Criollo, Forastero, and Trinitario; and the common variety of Cacao found at the Molave hill of the Visayas State University, Leyte, Philippines.

This research could play a crucial role in the early recognition of P. palmivora disease, which would allow for the prompt application of treatments. Consequently, this would lead to an improvement in Cacao production. The method proposed in this research represents a potentially significant advancement in eco-friendly and sustainable farming techniques, as well as in disease control within the Cacao industry. By promoting such practices, this research supports the United Nations' Sustainable Development Goals related to agriculture and innovation. Adopting these advanced methods would not only enhance crop yields but also foster more sustainable and environmentally responsible agricultural practices.

II. Review of Literature

A number of studies have already been conducted to recognize pests and diseases in plants and fruit trees using computer vision for the past years. These endeavors have greatly helped the agricultural sector in achieving better harvest to feed the world.

A. Non-Cacao Plants / Trees

The study by [4] proposed a system for identifying leaf diseases using Complete Local Binary Pattern and K-means

Visayas State University, Philippines - https://www.vsu.edu.ph (*sponsor*)

clustering methods. The system's true positive and false positive rates were also measured. The authors suggest that their system will enable farmers to detect significant diseases and pest infestations in crops, thereby allowing them to take necessary preventive measures. On the other hand, [5] employed image processing and fuzzy logic classifier to detect prevalence of P. palmivora disease on jackfruit. The model effectively recognized and classified the infection with an accuracy rate of 90%. Also, [6] enhanced the application of convolutional neural network on detection of tomato fruit common physiological diseases by data augmentation technique, by adding grayscale processor and foreground extractor components, and by utilizing k-means clustering algorithm. The mean Average Precision of the model was 97.24%. Similarly, [7] created a strawberry grading system based on 3 attributes namely: color, size, and shape. The developed technology used K-means clustering method, multi-attribute Decision Making Theory, and a single-chip-microcomputer (SCM). The size processing error is below 6%, its color marking precision is 88.8%, the shape categorizing accuracy is 90%, and the strawberry fruit grading takes 3s. Relatively, [8] implemented a hybrid system using Artificial Neural Network (ANN), Fourier descriptors (FD) and spatial domain analysis (SDA) for identifying fruits and sorting. The 3 different angles of camera inclination were used in the testing and evaluation. The experimental results showed a 99.10% accuracy rate. Also, [9] developed a system to detect and classify Pomegranate fruit diseases employing k-means clustering segmentation, GLCM method, and Artificial Neural Network. Based on the evaluation, the system yielded an accuracy rate of 90%. The author in [10] implemented a system that identifies the deformity in orange fruits utilizing multi-class SVM with K-means clustering for disease classification and Fuzzy logic to calculate the level of infection severity. The system assessment outcome revealed 90% accuracy. Another study by [11] integrated advanced defect segmentation techniques and texture, combined color, and shape-based features with the Histogram of Oriented Gradients (HOG) feature descriptor and a Bagged Decision Trees classifier. This approach successfully distinguished between healthy and defective apples, achieving an accuracy rate of 96%. Also, [12] employed an enhanced fuzzy C-means (FCM) algorithm and the marked-watershed algorithm to improve the extraction of cucumber leaf spot disease from images, even under complex backgrounds. The study evaluated 129 cucumber disease images from a vegetable disease database, revealing an average segmentation error of just 0.12%. This method offers a reliable and robust segmentation approach for the classification and grading of cucumber diseases in agriculture, with potential applicability to other imaging-based agricultural assessments.

B. Cacao Tree

The author in [13] utilized the K-means algorithm in conjunction with a Support Vector Machine (SVM), leveraging color information in the L*a*b* color space as features to recognize and segment the affected areas. The study outcome revealed an 89.2% accuracy rate. The method can be employed in enhancing the performance of the SVM classifier particularly in differentiating clearly the healthy reddish color of the cacao pod from a disease of the same color. On the following year, Tan et al. developed a mobile app called Automated Tool for Disease Detection and Assessment for Cacao (AuToDiDAC) that automates the detection, separation, and examination of the level of Black Pod Rot (BPR) infection in the fruit. K-means clustering algorithm and Support Vector Machine (SVM) were still utilized. Fifty (50) pod images were used in training the system while 10 were used in testing. The app showed only an accuracy mean of 85%. The author in [14] implemented another mobile app to identify cocoa diseases on cocoa plant utilizing digital image processing methods. The canker infection on cocoa pods was emphasized as a sample disease. Ten (10) images were utilized in training while 20 for testing the automated identifier. The experiment revealed a 100% accuracy in identifying the existence of canker disease on the pods. In the testing phase, the user had to feed infected pods only, though the article did not mention the type of infection. Another study detecting the condition of cacao pods by [15] got a 94% accuracy. In relation, [16] built a MobileNet model to identify pod diseases which yielded an 86.04% accuracy. In 2024, [17] developed a deep learning-based computational model to identify cocoa pod diseases with an accuracy of 34%.

This study aimed to achieve a higher accuracy rate in identifying the P. palmivora infection occurrence on cacao pod and would explore the following machine learning techniques: Naive Bayes Classification, Decision Stump, Random Forest, Hoeffding Tree, Multilayer Neural Network, and Convolutional Neural Network.

III. METHODOLOGY

The research framework adapted is shown in Fig. 1. Each phase of the framework is an integral part in the success of this endeavor:

1) Image acquisition: The dataset was captured using Vivo Y91 with 13 MP rear camera with an image resolution of 3120 x 4160 pixels. Collection of sample images was done between 10:00 in the morning until 3:00 in the afternoon at the cacao farm of the Visayas State University, Philippines. The camera was positioned 9 cm away from the cacao pod for both infected and not infected pods. The presence of P. palmivora on the Cacao pods was the focus of this study. These images were then manually classified by the domain expert, Dr. Arsenio D. Ramos of the VSU Horticulture Department.

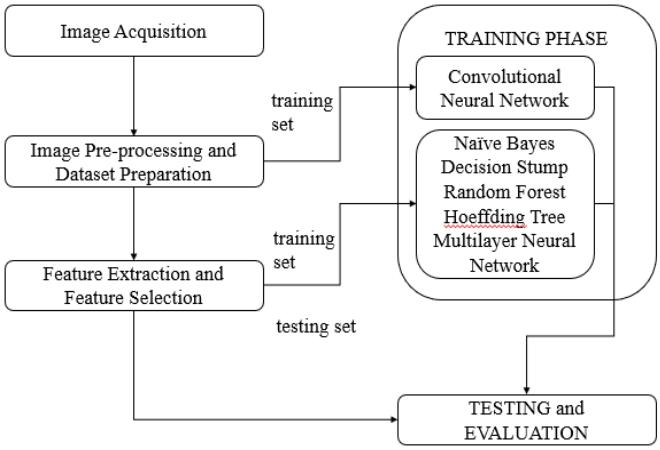


Fig. 1. Conceptual framework of the study.

The classified images were grouped and further divided into three subsets: 80.00% for training and 20.00% testing. The distribution of images among the classes in each subset is shown in Table I.

TABLE I. DATASET FOR THE STUDY

Class	Training	Testing	Total
Healthy	800	200	1000
Unhealthy	800	200	1000
Total (Dataset)	1600	400	2000

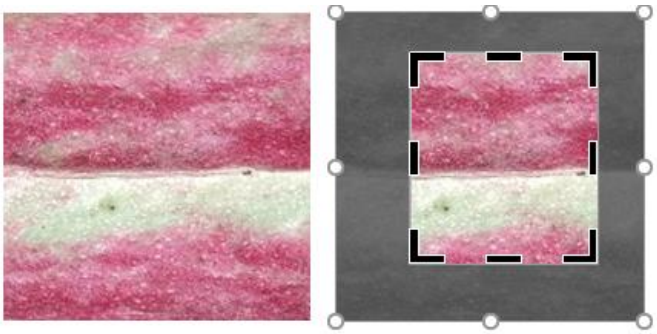


Fig. 2. Cropped image.

2) Image Pre-processing: The sample images for the training and testing phases were cropped to 150 x 150 pixels dimensions emphasizing the cacao texture. This applies to both healthy and infected datasets. Fig. 2 shows the cropping of an image of an infected pod.

These datasets were intended for use in all six (6) machine learning techniques.

3) Feature extraction and feature selection: Features such as color, shape and texture were extracted to characterize images using Haralick algorithm. Feature selection was carried out to determine the most suitable features for training the model and for attaining the best classification.

4) Model training: Training procedures were performed using different supervised machine learning algorithms that recognizes patterns and relationships from labeled training dataset. The different algorithms are the following: Naive Bayes Classification, Decision Stump, Random Forest, Hoeffding Tree, Multilayer Neural Network, and Convolutional Neural Network (CNN).

5) Model testing and evaluation: The generated models were verified to predict the untrained dataset. The performance of the models was evaluated and compared using the basic metric, the accuracy.

IV. RESULTS AND DISCUSSION

The Convolutional Neural Network (CNN) model exhibited outstanding performance by obtaining a 99% accuracy rate in detecting and classifying data throughout the testing phase, as evidenced in Table II, which presents a comparative comparison of six distinct machine learning algorithms. The CNN model demonstrated superior performance compared to the other tested techniques, indicating its effectiveness and reliability for the given task.

A. Convolutional Neural Network (CNN)Architecture

Fig. 3 illustrates the overarching architecture of the convolutional neural network employed in this study. The network consists of a sequence of interconnected layers that have been tailored to obtain features from the input image, decrease dimensionality, and to finally categorize the input data. More precisely, the architecture employs a sequential model structure, consisting of three convolutional layers. The initial layer is assigned a total of 32 filters, the succeeding layer has 64 filters, and the third layer has 128 filters.

Following each convolutional operation, a Rectified Linear Unit (ReLU) activation function was applied. Subsequently, max pooling layers with a 2x2 filter size and a stride of 2 are employed to down sample the feature maps, effectively halving their dimensions. The resulting pooled feature maps were then flattened into a singular vector, facilitating their transition into the fully connected (dense) layers for further processing.

The first dense layer is comprised of 128 nodes, with each node being triggered by a Rectified Linear Unit (ReLU) function. In addition, a batch normalization layer is included to normalize inputs for the following layers, while dropout regularization was used to reduce overestimation by inhibiting complex co-adaptations during training.

Conversely, the output layer employs a softmax activation function with 2 units, allowing the model to predict the likelihood of each input belonging to one of the two distinct classes based on the highest probability assignment.

Table III illustrates the model's synthesis, showing the resulting output form and the parameters that were learned automatically throughout the training process. The calculation of parameters inside each layer follows the equation [18]:

Number of parameters = weights + biases

Where: weights = input maps × (filter size) × output maps

The initial convolutional layer, with a 3x3 filter, applied to an input image size of 150x150x3 (150 pixels wide, 150 pixels high, and 3 color channels), results in an output shape of 148. In this layer, 3 feature maps are used as input, while 32 feature maps are produced as output. This requires the use of 32 separate filters, each with dimensions of 3x3x3. By including a bias term for every attribute map, the total number of parameters adds up to 896.

TABLE II. MACHINE LEARNING TECHNIQUES COMPARATIVE ANALYSIS

Machine Learning Technique	Training Accuracy	Testing Accuracy
Naïve Bayes	75.31%	73.00%
Decision Stump	74.80%	68.70%
Random Forest	99.00%	89.00%
Hoeffding Tree	78.00%	77.25%
Multilayer Neural Network	86.25%	86.50%
Convolutional Neural Network	98.06%	99.00%

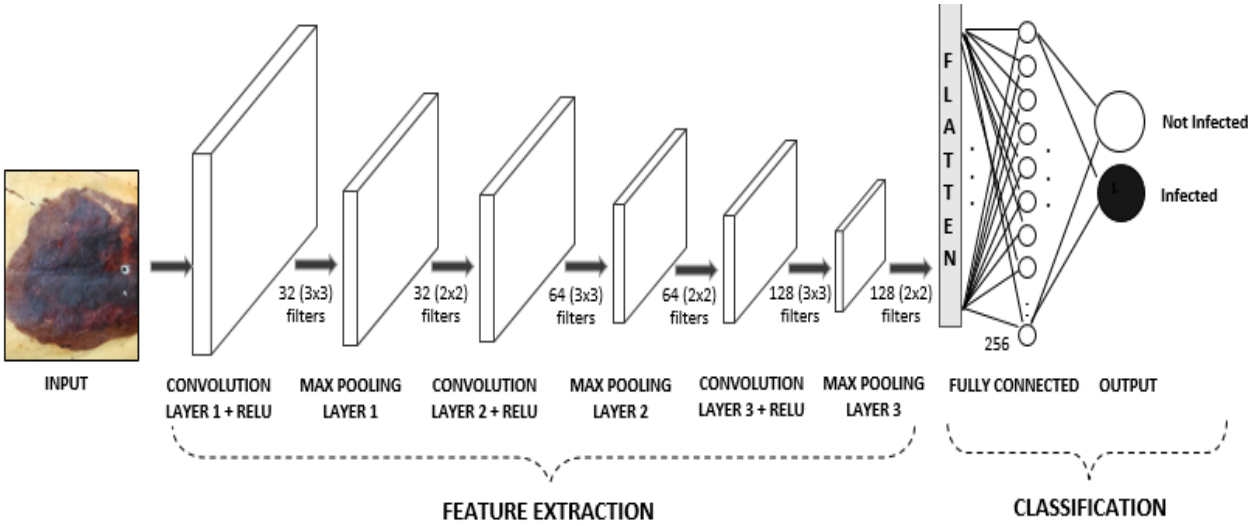


Fig. 3. CNN Architecture.

Consequently, the pooling layer conducts a straightforward substitution of a 2x2 neighborhood with its maximum value, thereby evading the inclusion of learnable parameters within this stratum.

When moving to the fully linked layer, the number of parameters is calculated by multiplying the number of input and output maps, and then adding an extra bias for each output. As a result, the two dense layers contain 73856 and 9470208 parameters, respectively.

Fig. 4 and Fig. 5 depict the visualization of model accuracy and loss per epoch, respectively. The line graphs indicate that the model was effectively learned, with consistent performance observed across both the training and validation datasets.

TABLE III. SUMMARY OF THE CNN MODEL

Layer (type)	Output shape	Param
conv2d_1 (Conv2D)	(None, 148, 148, 32)	896
activation_1 (Activation)	(None, 148, 148, 32)	0
max_pooling2d_1(Maxpooling2)	(None, 74, 74, 32)	0
conv2d_2 (Conv2D)	(None, 72, 72, 64)	18496
activation_2 (Activation)	(None, 72, 72, 64)	0
max_pooling2d_2(Maxpooling2)	(None, 36, 36, 64)	0
conv2d_3 (Conv2D)	(None, 34, 34, 128)	73856
activation_3 (Activation)	(None, 34, 34, 128)	0
max_pooling2d_3(Maxpooling2)	(None, 17, 17, 128)	0
flatten_1 (flatten)	(None, 36992)	0
dense_1 (Dense)	(None, 256)	9470208
activation_4 (Activation)	(None, 256)	0
dense_2 (Dense)	(None, 2)	514
activation_5 (Activation)	(None, 2)	0
Total params: 9,563,970 Trainable params: 9,563,970 Non-Trainable params: 0		

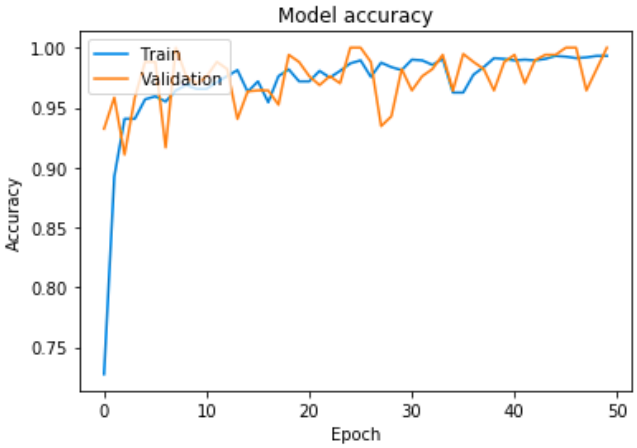


Fig. 4. The plot illustrating the training accuracy of the CNN model.

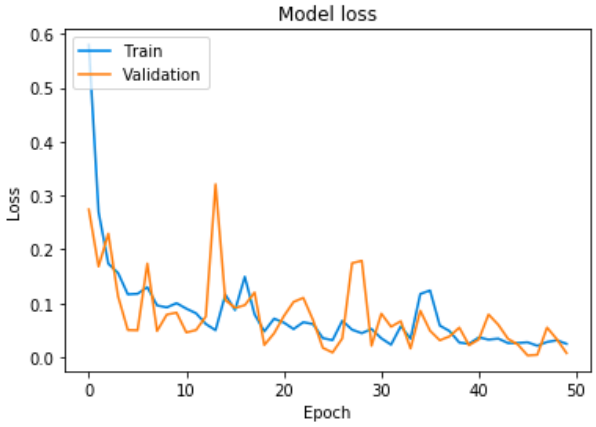


Fig. 5. The plot depicting the training loss of the CNN model.

V. CONCLUSION

This study successfully extracted color, shape, and texture features from digital images of cacao pods using the Haralick algorithm. These features were then partitioned and utilized to build and test classification models using various machine

learning techniques, including Naive Bayes Classification, Decision Stump, Random Forest, Hoeffding Tree, and Multilayer Neural Network. Additionally, the same dataset was employed to develop a Convolutional Neural Network (CNN) model with different feature extraction method. Among the six models developed, the CNN model achieved the highest accuracy at 99%, outperforming the other five machine learning algorithms during testing. This methodology holds significant potential for advancing sustainable agricultural practices and disease management.

VI. Future Work

Cacao cultivation encompasses a wide range of varieties, each with its unique characteristics and susceptibilities to diseases and pests. Expanding the dataset to encompass more cacao varieties ensures its applicability across more diverse agricultural contexts, catering to the specific needs and challenges faced by farmers cultivating different varieties. Also, training a dataset with artificial lights at any time of the day or in the absence of natural light may be done. The CNN model may be integrated in mobile app so that intended users can try the system while inputs from them may be solicited to improve the modeling.

Acknowledgment

The researchers wish to express their sincere appreciation to several key contributors, whose support was critical to the success of this project. They are especially grateful to Dr. Arsenio D. Ramos and Prof. Nestor I. Gaurana for their invaluable guidance and expertise. They also extend their thanks to the Visayas State University (VSU) for providing essential resources and institutional support. We recognize the Baybay City Agriculture Office for their assistance in local agricultural matters, and the Department of Computer Science and Technology at VSU for its guidance. Each of these individuals and organizations played a vital role in achieving the project's goals.

References

[1] S. D. Coe and M. D. Coe, The true history of chocolates, New York: Thames and Hudson, 2019.

[2] Department of Agriculture (DA) and Department of Trade and Industry (DTI), "2021-2025 Philippine cacao industry roadmap," 2022. [Online]. Available: https://www.da.gov.ph/wp-content/uploads/2023/05/Philippine-Cacao-Industry-Roadmap.pdf. [Accessed 1 October 2023].

[3] R. E. Hanada, A. W. Pomella, W. Soberanis, L. L. Loguercio and J. Pereira, "Biocontrol potential of Trichoderma martiale against the black-pod disease (Phytophthora palmivora) of cacao," Biological Control, vol. 50, pp. 143-149, 2009.

[4] P. N. Wankhade and G. G. Chiddarwar, "An overview of different mechanisms to detect plant leaf disease infected area," International Journal of Innovative Research in Computer and Communication Engineering, vol. 5, no. 5, 2017.

[5] J. V. Oraño, E. A. Maravillas, C. G. Aliac and J. V. Oraño, "Jackfruit Phytophthora palmivora (Butler) disease recognizer using Mamdani fuzzy logic," in 6th ICPEP National Conference 2018, Baguio, 2018.

[6] J. Zhao and J. Qu, "A detection method for tomato fruit common physiological diseases based on regression model," in 10th International Conference on Information Technology in Medicine and Education (ITME), Qingdao, 2019.

[7] X. Liming and Z. Yanchao, "Automated strawberry grading system based on image processing," Computers and Electronics in Agriculture, Vols. 71, Supplement 1, p. S32–S39, 2010.

[8] A. M. Aibinu, M. E. Salami, A. A. Shafie, N. Hazali and N. Termidzi, "Automatic fruits identification system using hybrid technique," in 2011 Sixth IEEE International Symposium on Electronic Design, Test and Application, Queenstown, 2011.

[9] M. Dhakate and A. B. Ingole, "Diagnosis of pomegranate plant diseases using neural network," in Fifth National Conference on Computer Vision, Pattern Recognition, Image Processing and Graphics (NCVPRIPG), Patna, 2015.

[10] S. Behera, L. Jena, A. Rath and P. Sethy, "Disease classification and grading of orange," in International Conference on Communication and Signal Processing, Patna, 2018.

[11] P. Sujatha, J. Sandhya, J. Chaitanya and R. Subashini.

[12] X. Bai, X. Li, Z. Fu, X. Lv and L. Zhang, "A fuzzy clustering segmentation method based on neighborhood," Computers and Electronics in Agriculture, vol. 136, p. 157–165, 2017.

[13] D. Tan, R. Leong, A. Laguna, C. Ngo, A. Lao, A. Amalin and D. Alvindia, "A method for detecting and segmenting," in Proceedings of the DLSU Research Congress, Vol 4, Manila, 2016.

[14] N. Harivinod, P. Pooja, H. K. Nithesh, B. S. Ashritha and G. G. Hegde, "Cocoa care - an android application for cocoa disease identification," International Journal on Recent and Innovation Trends in Computing and Communication, vol. 5, no. 6, pp. 440-445, 2017.

[15] R. Godmalin, C. Aliac and L. Feliscuzo, "Classification of cacao pod if healthy or attack by pest or black pod disease using deep learning algorithm," in 2022 IEEE International Conference on Artificial Intelligence in Engineering and Technology (IICAIET), Malaysia, 2022.

[16] D. Mamadou, K. Ayikpa, A. Ballo and B. Kouassi, "Cocoa pods diseases detection by MobileNet Confluence and classification," International Journal of Advanced Computer Science and Applications, vol. 14, no. 9, 2023.

[17] D. Vera, B. Oviedo, W. Casanova and C. Zambrano-Vega, "Deep learning-based computational model for disease identification in cocoa pods (Theobroma cacao L.)," Cornell University, Los Rios, 2024.

[18] J. Brownlee, "Machine learning mastery," 2019. [Online]. Available: https://machinelearningmastery.com/. [Accessed 2022].

Leveraging Mechanomyography Signal for Quantitative Muscle Spasticity Assessment of Upper Limb in Neurological Disorders Using Machine Learning

Muhamad Aliff Imran Daud[1], Asmarani Ahmad Puzi[2]*, Shahrul Na'im Sidek[3],
Ahmad Anwar Zainuddin[4], Ismail Mohd Khairuddin[5], Mohd Azri Abdul Mutalib[6]
Dept. of Computer Science, International Islamic University Malaysia, Kuala Lumpur, Malaysia[1, 2, 4]
Dept. of Mechatronics Engineering, International Islamic University Malaysia, Kuala Lumpur, Malaysia[3]
Faculty of Manufacturing and Mechatronic Engineering, Universiti Malaysia Pahang Al-Sultan Abdullah, Pekan, Malaysia[5]
Dept. of Machine Design, SIRIM Berhad, Hulu Selangor, Malaysia[6]

Abstract*—Upper motor neuron syndrome is characterised by spasticity, which represents a neurological disability that can be found in several disorders such as cerebral palsy, amyotrophic lateral sclerosis, stroke, brain injury, and spinal cord injury. Muscle spasticity is always assessed by therapists using conventional methods involving passive movement and assigning spasticity grades to the relevant joints based on the degree of muscle resistance which leads to inconsistency in assessment and could affect the efficiency of the rehabilitation process. To address this problem, the study proposed to develop a muscle spasticity model using Mechanomyography (MMG) signals from the forearm muscles. The muscle spasticity model leveraged based on the Modified Ashworth Scale and focus on flexion and extension movements of the forearm. Thirty subjects who satisfied the requirements and provided consent were recruited to participate in the data collection. The data underwent a pre-processing stage and was subsequently analysed prior to the extraction of features. The dataset consists of forty-eight extracted features from the three-direction x, y, z axes (for both biceps and triceps muscle), corresponding to the longitudinal, lateral, and transverse orientations relative to the muscle fibers. Significant features selection was conducted to analyse if overall significant difference showed in the combined set of these features across the different spasticity levels. The test results determined the selection of twenty-five features from a total of forty-eight which be used to train an optimum classifier algorithm for the purpose of quantifying the level of muscle spasticity. Linear Discriminant Analysis (LDA), Decision Trees (DTs), Support Vector Machine (SVM), and K-Nearest Neighbour (KNN) algorithms have been employed to achieve better accuracy in quantifying the muscle spasticity level. The KNN-based classifier achieved the highest performance, with an accuracy of 91.29% with k=15, surpassing the accuracy of other classifiers. This leads to consistency in spasticity evaluation, hence offering optimum rehabilitation strategies.

Keywords*—Spasticity; mechanomyography; Modified Ashworth Scale; machine learning

I. Introduction

A stroke is a sudden and chronic loss of neurological function caused by infarction or haemorrhage in the brain, spinal cord, or retina, resulting in impaired motor function and significant restrictions in performing everyday tasks and overall well-being [1]. Most stroke patients experience movement difficulties, with approximately 30% of those affected experiencing spasticity [2]. Moreover, stroke remains the second most prevalent cause of fatality and the primary cause of impairment on a global scale, making it the third leading cause of death and disability worldwide [3], [4], [5]. On a global scale the prevalence of stroke has grown in correlation with the process of modemisation, modifications in lifestyle, and a growing population of older individuals [6].

Upper motor neuron syndrome has been defined by spasticity, a neurological impairment that occurs in a variety of conditions, including cerebral palsy, amyotrophic lateral sclerosis, stroke, brain injury, and spinal cord injury [7]. Lance introduced the term "spasticity" in 1980 to define the upper motor neuron syndrome, a motor disorder marked by increased muscle tone and exaggerated tendon jerks that rely on movement velocity and result from the hyperexcitability of the stretch reaction [8], [9]. This description exclusively emphasises the impact of spasticity on involuntary movements, disregarding its effect on deliberate behaviours. The Modified Ashworth Scale (MAS) and the Australian Spasticity Assessment Scale (ASAS) are widely recognised as the most reliable methods for evaluating spasticity in clinical settings, with the MAS being a frequently employed tool in stroke rehabilitation [10], [11].

In addition, there are several other clinical tools available for assessing spasticity, such as Spinal Cord Assessment Tool for Spastic Reflexes (SCATS), Fugl-Meyer Assessment (FMA), Penn Spasm Frequency Scale (PSFS), and Modified Tardieu Scale (MTS) [12], [13]. However, these tools are less accurate compared to the MAS and ASAS. Furthermore, the conventional method that has been used to assess spasticity nowadays involves subjective measurement by the therapists [14]. Although, the therapists already been trained well in assessing the spasticity using MAS tool measurement, there might be a possibility of difference in identify the spasticity level. The variability can disturb the effectiveness of the rehabilitation process for the neurological disorder patients.

During the procedural application of the MAS, the therapist executes passive movement and assigns spasticity grades to the relevant joints based on the degree of muscle resistance experienced during passive stretching [15], [16].

The main objective of this study is to validate Mechanomyography (MMG) as a reliable signal by comparing the accuracy of various machine learning algorithms and demonstrate its clinical applicability in objective measurement. This study highlights the effectiveness of combining mechanomyography with machine learning as a superior approach for evaluating muscular spasticity in patients with upper limb neurological disorders.

The main contributions of this study include the introduction of MMG as a new and unbiased instrument for evaluating muscular spasticity, which has the potential to enhance current subjective approaches. The research also evaluates various machine learning algorithms to determine the best models for analysing MMG data, thereby improving the accuracy and consistency of spasticity assessments. Additionally, the study illustrates the practical applicability of MMG in clinical settings, highlighting its potential to standardize evaluations, optimize rehabilitation strategies, and ultimately improve patient outcomes. These contributions have substantial significance for the field of neurorehabilitation as it establishes an accurate and unbiased technique for evaluating spasticity, which can lead to enhanced diagnostic precision, individualized treatment strategies, and potentially improved long-term results for patients with upper limb neurological diseases.

The structure of the article is as follow: Section II describe on characteristics of electromyography and mechanomyography on clinical evaluation. Section III presents a comprehensive summary of the research carried out by researchers in the topic throughout the years. Section IV provide detailed explanations on the selection of subjects, the experimental setup, and the pre-processing and analysis of the data. Section V provides an explanation of the machine learning algorithms. Section VI provided and deliberated upon the experimental findings and the subsequent section explain the conclusion of the study findings.

II. Electromyography and Mechanomyography

The utilization of electromyography (EMG) in routine therapeutic procedures represents a contemporary and pioneering approach to neurorehabilitation for individuals recovering from a stroke [17]. EMG has been used to record electrical muscle activity for quite some time, though it's currently limited to therapeutic purposes [18]. Additionally, EMG can be highly susceptible to interference from noise and variations in resistance, rendering it unreliable in diverse settings or during prolonged data collection, such as when an individual starts sweating [19], [20]. At the same time, the use of EMG sensors necessitates time-consuming skin preparation, including disinfection and abrasive paste application, along with electrode placement on multiple leg muscles which requires an expert environment for accurate sensor positioning and signal interpretation [21]. Mechanomyography (MMG) serves as an alternative or mechanical counterpart to EMG by quantifying muscle vibrations, or mechanical activity generated by active muscle, using sensors such as microphones or accelerometers [22], [23]. The invention of piezoelectric, microphone, and accelerometers demonstrated the adequate detection of mechanical signals from the surface of skeletal muscles at low frequencies, known as MMG signals that tend to be contaminated by electrical noise [24], [25]. MMG provides a method that enables the detection and measurement of vibrations resulting from muscle contractions and stretching [26]. These vibrations propagate through the tissue and can be detected on the surface of the skin.

The characteristics of a reliable MMG transducer typically include high sensitivity within the muscle vibrational frequency range of 2 Hz to 100 Hz, low sensitivity to random noise, ease and standardization of sensor attachment, biocompatibility, suitability for clinical environments, and cost-effectiveness compared to other clinical assessment techniques [27]. Compared to EMG, MMG has not yet gained widespread acceptance, particularly in clinical settings. Despite not yet achieving widespread acceptance, particularly in clinical settings, MMG holds significant potential for various applications. These include controlling prosthetic devices, recognizing gestures in human-machine interfaces (HMIs), and studying the underlying physiological mechanisms of the neuromuscular system in scientific research [28]. Additionally, MMG offers greater convenience than EMG as it being highly responsive to skin conditions and reliable performance in dynamic settings, reducing the necessity for frequent cleaning, drying, and optimal skin condition maintenance throughout usage [29], [30]. MMG responses can be utilized in various medical contexts, including the clinical evaluation of neuromuscular tissue, biofeedback rehabilitation, and neural/myoelectric prosthetic control.

III. Related Works on Machine Learning

Machine learning has become a popular data analytics technology that uses statistical methods to analyze observed data and make predictions or classify new data [31]. Multiple studies have investigated the effectiveness of machine learning in improving the provision of rehabilitation services, showcasing its capacity to enhance patient outcomes and improve clinical procedures.

For instance, Puzi et al. [32] developed An Automatic Muscle Spasticity Assessment System (AMSAS) to evaluate the muscle spasticity, specifically emphasizing the utilization of machine learning methods. The torque and angle signals generated by the arm muscles were examined to classify levels of spasticity according to the Modified Ashworth Scale (MAS). Twenty-five patients with varied degrees of spasticity were analysed, and seven features were retrieved. A Linear Support Vector Machine (SVM) classifier with four specified characteristics got the maximum accuracy of 84% when classifying spasticity levels. The variables of Three-Way Decision (TWD) including the first and second halves of the region, catch position, and post-catch stiffness, were found to have a significant association with MAS levels.

In another study, Puzi et al. [33] presented a new classifier that uses clinical data from the affected upper limb to accurately measure levels of muscle spasticity. The study proposes a methodical quantification strategy that utilizes the Modified Ashworth Scale (MAS) in conjunction with a one-way ANOVA test to assess the extent to which these features accurately

predicted test scores. subsequently, four important features were determined as the most significant for creating efficient classification models and were employed in the training process. The study showcased that the Support Vector Machine (SVM) classifier surpassed the Adaptive Neuro-Fuzzy Inference System (ANFIS) classifier, with an accuracy rate of 88.0%.

Additionally, Liu et al. [34] investigated on muscle spasticity, specifically targeting the wrist flexor and extensor muscles. The methods employed in this study utilise MMG signals to identify periods of muscular activity in real-time gesture recognition, which is essential for diagnosing muscle spasticity. Additionally, it has been utilised to offer significant observations on muscle exhaustion and torque, proving their potential in evaluating muscle spasticity. The study involved assessing eight distinct and atypical gestures, which included clapping, flicking the index finger, snapping the finger, flipping a coin, shooting, extending the wrist, bending the wrist, and creating a fist. The K-nearest neighbours (KNN) algorithm, with a value of K set to 7, achieved the maximum classification accuracy of 94.56% for the eight gestures.

Furthermore, Kim et al. [31] conducted a study aimed at assessing elbow spasticity by the application of machine learning techniques. This was achieved by employing sophisticated machine learning algorithms to meticulously analyse acceleration and rotation characteristics derived from the injured elbow's side. The acceleration and rotation properties of the elbows of affected patients have been examined to determine the degree of spastic movement, similar to the way the modified Ashworth scale (MAS) score was used. Achieving an accuracy of up to 95.4%, a random forest (RF) algorithm was used to classify spasticity. The learning problem was classified as supervised since the signals correlated with MAS scores, as evaluated by therapists. Additional features were extracted and incorporated into the existing feature set, resulting in enhanced classification performance.

These studies collectively illustrate the potential of machine learning in the precise assessment and classification of muscle spasticity, offering significant advancements in the field of rehabilitation.

IV. Methodology

A. Subjects

30 post-stroke subjects with upper limb spasticity participated in the study. This study has obtained approval by the Research Ethics Committee of the International Islamic University Malaysia (IIUM) under the identification number IREC 2023-025. Specifically, the subjects were diagnosed with spasticity in upper limbs (UL), with an age range of 18 to 80 years recruited from Sultan Ahmad Shah Medical Centre (SASMEC) and National Stroke Association of Malaysia (NASAM). The informed consent has been provided by the subjects prior to participation, and the study adhered to strict data protection procedures, ensuring that all subject information was managed in accordance with applicable data privacy regulations. The subjects recruited for this research were chosen from MAS levels 0, 1, 1+, 2, and 3. The MAS level 4 was omitted due to the absence of any noticeable bending and straightening movements during the evaluation. The pilot examination was conducted by experienced therapists to evaluate the subjects movement capability and identify potential issues that can be addressed for the upcoming data collection. The demographic characteristics of the research subjects were detailed in Table I. MAS scores ranging from 0 to 3 were determined for the participants' affected muscles. Five groups were formed from the volunteers: MAS-0 (N=5), MAS-1 (N=16), MAS-1+ (N=3), MAS-2 (N=4) and MAS-3(N=2).

TABLE I. Demographic Data of the Patients (Divided into Four Groups)

Mas Level	Numbers (N)	Genders (M/F)	Affected Hand (Left/Right)	Age (Year)
0	6	3/3	2/4	44.3 ± 18.4
1	15	11/5	7/8	62.7 ± 10.1
1+	3	3/0	2/1	56.0 ± 7.5
2	4	3/1	1/3	50.0 ± 9.5
3	2	2/0	2/0	38.7 ± 19.1

B. QSAT Platform

A new platform called as Quantitative Spasticity Assessment Technology (QSAT) has been developed based on the Mechanomyography (MMG) technique to overcome the inconsistency measurement of spasticity as depicted in Fig. 1. The platform incorporates two primary sensors: an accelerometer Mechanomyography (ACC-MMG), which measures muscle vibrations in the biceps and triceps, and a potentiometer, which assesses the angular position of the upper limb during flexion and extension movements. Through the measurement of patients' biological signals, the extracted features have been examined for their correlation with MAS using machine learning. The utilization of platform measurements mapped to MAS levels enhances the evaluation of spasticity and streamlines the clinical workflows of therapists.

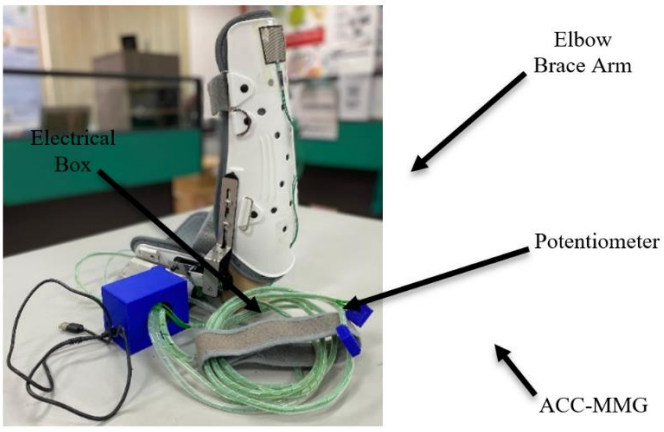
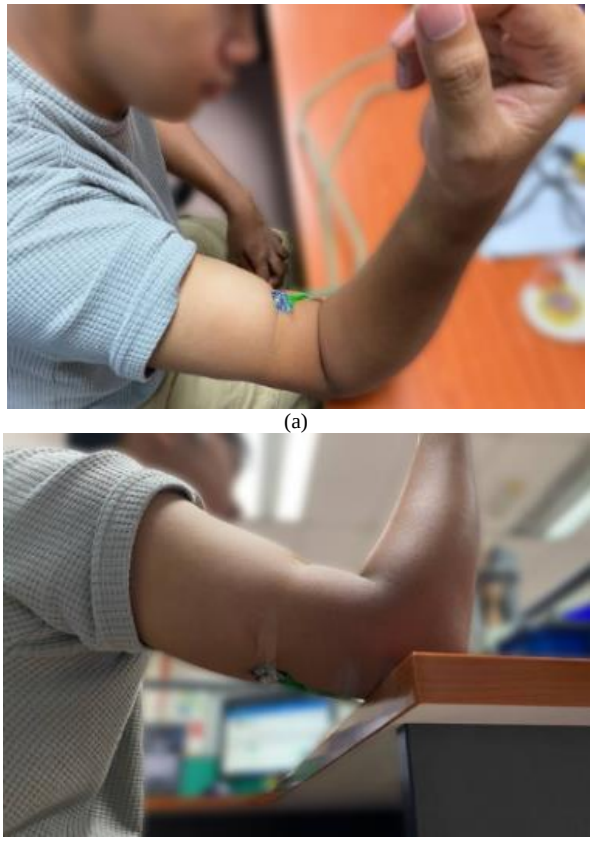


Fig. 1. QSAT System with labels.

C. Data Acquisition and Experimental Setup

In this study, a commercial biological signal acquisition system (Raspberry Pi Pico) was used to record ACC-MMG signals (ADXL345, Digital Devices, full-scale range = ± 2 g to ± 16 g; typical frequency responses = 0.1 to 3200 Hz; sensitivity = 3.9 mg/LSB; size = 3 mm x 5 mm x 1 mm)) and potentiometer

data, both sampled at a rate of 166.7 Hz. Each muscle group, including the biceps and triceps, was equipped with a tri-axial ACC-MMG accelerometer and a potentiometer, which were integrated within an elbow brace arm. This configuration enabled single-channel potentiometer recording alongside simultaneous two-channel ACC-MMG recording. The ACC-MMG signal, captured three-dimensionally by the accelerometers, included three distinct sub-signals corresponding to the x, y, and z axes. Consequently, one channel was designated for potentiometer data, while two channels were dedicated to ACC-MMG signal acquisition, with all data recorded concurrently. The ACC-MMG signals along the muscle axes were captured using three distinct tri-axial accelerometers. These accelerometers were oriented along the x, y, and z axes, corresponding to the longitudinal, lateral, and transverse orientations relative to the muscle fibers, respectively, as illustrated in Fig. 2.



(a)

(b)

Fig. 2. ACC-MMG Sensor placement: (a) On the Biceps and (b) On the triceps.

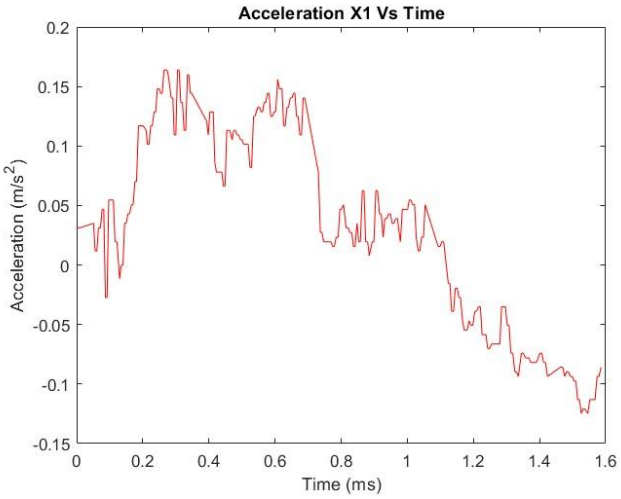
The experimental protocol began with each subject directed to lie down in the supine position with their arm positioned alongside their body. This evaluation was carried out using the Modified Ashworth Scale (MAS) as the clinical tool for assessment. During the implementation of the MAS, the therapist performs passive movements and assigns spasticity grades to the corresponding joints depending on the level of muscular resistance observed during passive stretching. After the session, the ACC-MMG signal of biceps and triceps was recorded. The sensors were affixed to the skin in a secure manner through the utilization of double-sided tape. "Sensor 1" was placed on the biceps muscle's belly, while "Sensor 2" was positioned on the triceps muscle's belly. The sensor's x-axis was aligned with the direction of muscle fiber contraction while z-axis was touched directly to the skin surface. At the same time, the potentiometer with elbow brace arm support was attached to the elbow joint. The QSAT experiment began by assessing the subject's arm through the placement of one therapist's hand beneath the lower arm in proximity to the wrist, while the other hand gave stability to the upper arm near the shoulder, as illustrated in Fig. 3. The subject's arm underwent three repetitions of a movement, transitioning from full extension (0°) to full flexion (135°) for a duration of two seconds each time. All results have been recorded and organized in an Excel datasheet.



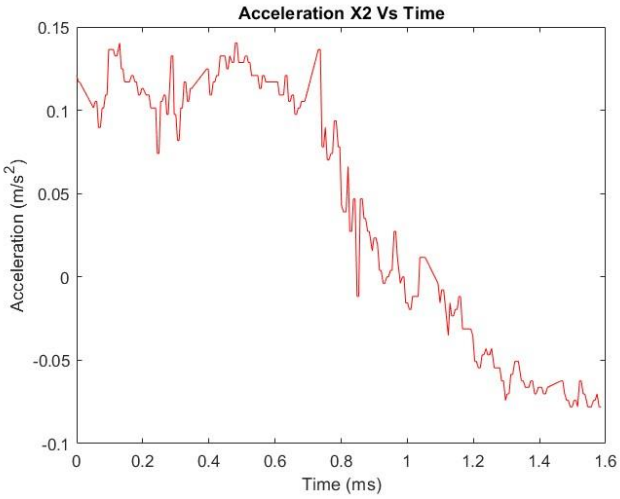
Fig. 3. Setup of QSAT Platform measures for upper limb.

D. Data Analysis and Feature Extraction

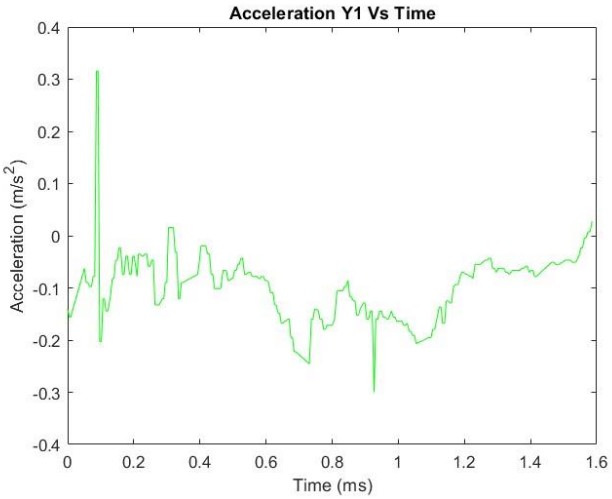
Signal preprocessing was conducted using MATLAB R2023a software (MathWorks Inc.). The ACC-MMG and potentiometer data collected throughout the experiments underwent initial preprocessing to ensure precision and reliability. Based on the analysis of the raw data, the signals exhibited minimal noise interference, indicating that filtering was unnecessary. The continuous data was then divided into epochs corresponding to each movement cycle, ranging from full extension (0°) to complete flexion (135°) of the elbow. The potentiometer data provided distinct markers indicating the beginning and end of each movement cycle. Fig. 4 shows a graph that presents the time-series plots of the ACC-MMG signals along the x_1, y_1, z_1 axes (for the biceps) and the x_2, y_2, z_2 axes (for the triceps), together with the potentiometer readings. The resulting graph visually represents the patterns of muscle fibres vibration along all three spatial directions and joint motions during flexion. The ACC-MMG signals had distinct amplitude and frequency characteristics for each axis, aligning with the longitudinal, lateral, and transverse orientations of muscle fibres. Significant differences in muscle activity patterns can be observed when comparing the ACC-MMG signals of the biceps and triceps while the potentiometer readings, indicating joint angles, align with the ACC-MMG signals, validating the observed movement cycles.


Acceleration X1 Vs Time
Acceleration (m/s^2)
Time (ms)

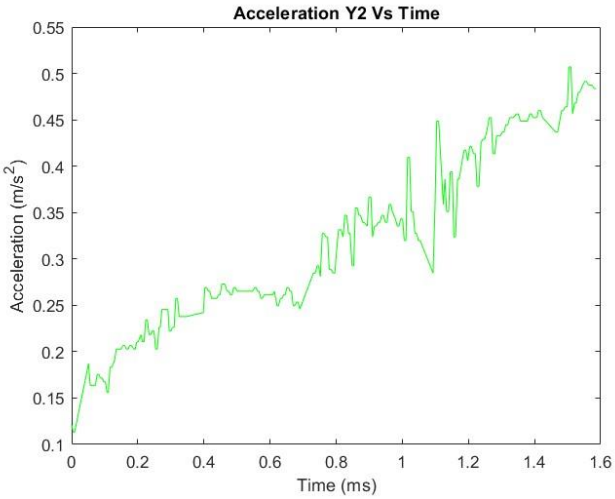
(a)


Acceleration X2 Vs Time
Acceleration (m/s^2)
Time (ms)

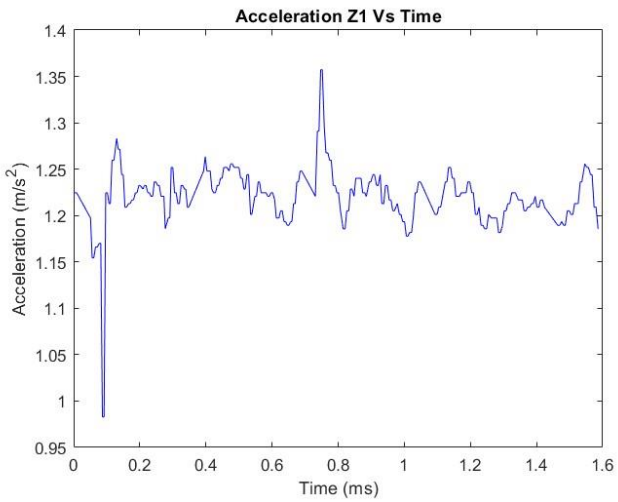
(d)


Acceleration Y1 Vs Time
Acceleration (m/s^2)
Time (ms)

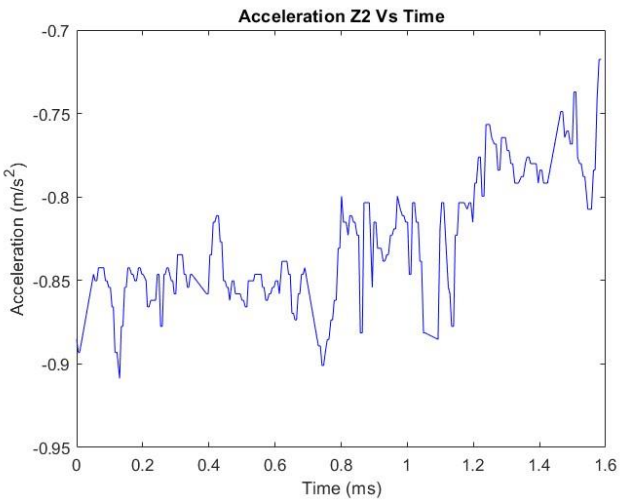
(b)


Acceleration Y2 Vs Time
Acceleration (m/s^2)
Time (ms)

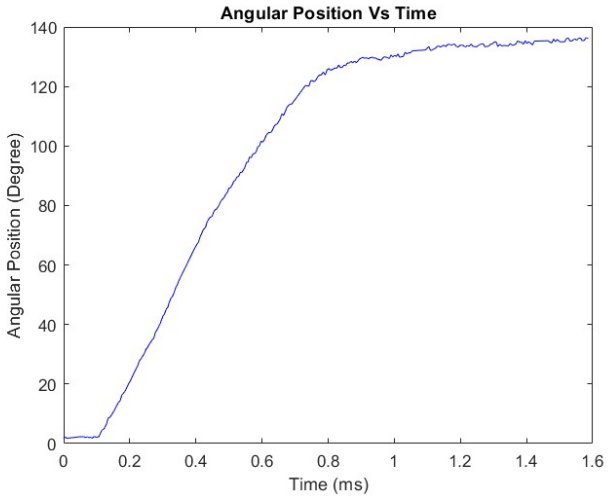
(e)


Acceleration Z1 Vs Time
Acceleration (m/s^2)
Time (ms)

(c)


Acceleration Z2 Vs Time
Acceleration (m/s^2)
Time (ms)

(f)



(g)

Fig. 4. Muscle vibrations of the biceps (a, b, c) and triceps (d, e, f), and the angular position (g) during flexion for patient 30.

A specialised algorithm was developed to extract significant features from the ACC-MMG signals. The features selected for analysis included Root Mean Square (RMS), Peak to Peak Amplitude (PTP), Max, Min, Mean Average Value (MAV), Standard Deviation (SD), Skewness (S), and Kurtosis (K). During the feature extraction stage, time-domain features were extracted for the x_1, y_1, z_1 axes (for the biceps) and the x_2, y_2, z_2 axes (for the triceps), corresponding to the longitudinal, lateral, and transverse orientations relative to the muscle fibers. The obtained features were tabulated into a dataset. The equation that determines each extracted feature are as follows:

$$RMS = \sqrt{\frac{1}{n}\sum_i x_i^2} \quad (1)$$

$$PTP = Max - Min \quad (2)$$

$$Max = maximum\ of\ the\ terms \quad (3)$$

$$Min = minimum\ of\ the\ terms \quad (4)$$

$$MAV = \frac{sum\ of\ the\ terms}{number\ of\ terms} \quad (5)$$

$$SD = \frac{\sum(x_i - MAV)^2}{n} \quad (6)$$

$$S = \frac{\sum_i^n (x_i - MAV)^3}{(n-1)\ SD^3} \quad (7)$$

$$K = \frac{n\ (n+1)}{(n-1)(n-2)(n-3)} \times \sum\left(\frac{x_i - MAV}{SD}\right)^4 - \frac{3(n-1)^2}{(n-2)(n-3)} \quad (8)$$

The RMS is a metric that quantifies the amplitude of a signal, serving as an indicator of the intensity of muscular contractions or tension of the upper limbs [2], [35]. MAV of MMG signals serves as an indicator of the muscular strength and endurance of the specific muscle in concern [36]. It symbolizes the power, and the energy generated by the muscle. SD were computed for the ACC-MMG signals in the x, y, and z axes, providing insights into the average muscle activity and its variability. Complementing these metrics, the PTP value measures the extent of a signal by determining the disparity between its highest positive peak and its lowest negative peak of vibration amplitude during a specific timeframe, thereby representing the magnitude spectrum of muscular oscillations or motions within the ACC-MMG signals. Additionally, skewness and kurtosis were calculated to offer a deeper understanding of the properties of muscle signal distributions, thus enhancing the comprehension of muscle activity and its variability.

For this study, a total of 90 datasets were collected from 30 subjects in order to train the muscular spasticity classifier. The one-way MANOVA test was utilised with SPSS 27.0.1 (IBM Inc.) to minimise dependent and redundant features through significant feature analysis. The statistical test selected was a one-way MANOVA due to the presence of multiple continuous dependent variables in independent groups [37]. The features were testing using the technique to examine the significant difference in mean value between the groups. The results from the one-way MANOVA test, including significant values and corresponding p-values, are presented in Table II. A rejection threshold was set at $p < 0.05$ to identify significant differences in the dependent variables. The null hypothesis for ANOVA posited that no difference existed in mean values among the groups. As the p-values of the features listed in Table II were less than 0.05, the null hypothesis was effectively rejected. Consequently, these twenty-five optimal features were selected to train the classifiers for classifying the level of muscle spasticity.

TABLE II. SIGNIFICANT VALUE OF FEATURES

Features	P Values
$MAVx_1$	0.000
$MAVy_1$	0.000
$MAVz_1$	0.000
SDz_1	0.001
$PTPy_1$	0.013
$Maxx_1$	0.000
$Maxy_1$	0.003
$Maxz_1$	0.000
$Minx_1$	0.000
$Miny_1$	0.000
$Minz_1$	0.000
Sx_1	0.001
Ky_1	0.020
$RMSx_1$	0.004
$RMSy_1$	0.000
$RMSz_1$	0.000
$MAVy_2$	0.000
$MAVz_2$	0.006
SDz_2	0.042
$Maxy_2$	0.000
$Maxz_2$	0.011
$Miny_2$	0.000
$Minz_2$	0.003
$RMSy_2$	0.000
$RMSz_2$	0.002

E. Machine Learning Algorithm

Machine-learning classifiers were utilised to automatically infer a prediction function from labelled data derived from inertial signals obtained during passive stretching. The therapist provided MAS ratings to annotate these inertial signals, thus structuring the task as a supervised learning problem. Most offline approaches rely on supervised Machine Learning (ML) models for activity recognition, such as Support Vector Machine (SVM), Decision Trees (DTs), K-Nearest Neighbors (KNN) and Linear Discriminant Analysis (LDA) [31], [38]. Supervised machine-learning algorithms have the benefit over unsupervised methods of being able to assign appropriate labels to training data based on preset classes, thereby avoiding the need to create "artificial" groups [39].

SVM known as a collection of supervised learning techniques employed for the purposes of classification and regression [40], [41]. For a classification problem, SVM seeks to identify the separating hyperplanes that maximize the margin between sets of data points in an n-dimensional space, where each data point belongs to one of the available classes. This will guarantee a strong ability to make accurate predictions in various situations, assuming that the target function remains stable between the training and testing data. SVM are primarily used when the data cannot be separated by a straight line in their current domain [42]. SVM applies a transformation to the input data points, mapping them to a feature space where they can be separated by a linear boundary. Essentially, it separates the classes by incorporating support vectors to optimize the separation between samples belonging to distinct classes. Therefore, it is also known as large-margin categorization.

DTs is a structured representation of a decision-making process used to determine the class of a given instance [43]. Every node in the tree represents either a class label or a particular test that divides the instance space according to the potential results of that test. Every subset of partitions corresponds to a subproblem of classification, which is then resolved by a subtree. The terminal nodes of the decision tree include the class labels. To categorize an instance, one must follow a path from the starting point of the tree to one of its end nodes, taking into account the results of the tests at each step of the process.

KNN classifier is a method used to categorise unlabelled data by assigning them to the class of the most comparable labelled samples. Observational characteristics are gathered for both the training and test datasets [44]. The intuition behind Nearest Neighbor Classification is straightforward. It often proves beneficial to consider multiple neighbors, leading to the more commonly utilized K-Nearest Neighbor (KNN) Classification, where the class of an instance is determined based on the k nearest neighbors [45]. Besides that, LDA is also highly popular technique used to extract distinctive features for the purpose of pattern classification [46]. Linear Discriminant Analysis (LDA) leverages label information to acquire a discriminant projection that effectively increases the separation between different classes and decreases the distance within each class, hence enhancing the accuracy of classification. Several extensions of LDA have been established to improve performance and efficiency. The traditional Linear Discriminant Analysis (LDA) model typically assigns a Gaussian density to each class, assuming that all classes have an identical covariance matrix [47]. LDA is closely associated with ANOVA (analysis of variance) and regression analysis, since each attempt to represent a dependent variable as a linear combination of other traits or data.

Two dataset has been prepared and structured for muscle spasticity model development. The first data utilized all available features, while the second dataset incorporated only the significant features identified through a one-way MANOVA test. Each dataset datasets were divided into training and testing sets with ratios of 90/10, 80/20, and 70/30. In the 90/10 split, 90% of the data was used for training the model while the remaining 10% was reserved for testing. Similar procedures were followed for the 80/20 and 70/30 splits. These different partitions were used to evaluate the models' robustness and generalization capabilities. The optimal algorithm underwent k-fold cross-validation, where it was trained and tested with k values of 5, 10 and 15 to evaluate the stability of the models across different partitioning schemes. The performance of the machine learning algorithms was assessed using confusion matrices, accuracy, and training length. The percentage of correctly predicted samples to the total number of samples represents the definition of accuracy. A True Positive (TP) outcome occurs when the model accurately predicts the positive class while a True Negative (TN) refers to an outcome where the model accurately predicts the negative class. likewise, a False Positive (FP) occurs when the model wrongly predicts the positive class, whereas a False Negative (FN) occurs when the model incorrectly predicts the negative class. Equation 9 can be used to compute the accuracy by considering the values of TP (true positive), TN (true negative), FN (false negative), and FP (false positive) [48].

$$Accuracy = \frac{T_p + T_N}{T_p + T_N + FP + FN} \quad (9)$$

A confusion matrix comprises a square matrix displaying the general classification model performance. The rows of the confusion matrix show actual instances of class labels, whereas the columns show instances of predicted class labels. For each trial, the diagonal components of this matrix will indicate how many times the predicted label matches the actual label. For assessing how effectively the model classified data, the confusion matrix acts as a useful indicator.

V. Results and Discussion

The accuracy performance of the algorithms influenced by different training and testing splits which presented in Table III. It also compares the accuracy based on features selected for training and testing, yielding varying results. Dataset with all features and significant features for the 90/10 split showed highest accuracy compared to other data ratios. However, most researchers recommend using a 70/30 split for smaller datasets [49]. Dataset with all features shown that KNN algorithm achieved an accuracy of 83.95%, outperforms the others algorithm. Notably, the KNN algorithm demonstrated an even higher accuracy of 90.12% when using significant features, indicating that the use of significant features enhances the model's performance.

TABLE III. PERCENTAGE OF DATA SET AND ACCURACY WITH ALL FEATURES AND SIGNIFICANT FEATURES

Algorithm	Training and Testing Split Percentage of Accuracy					
	All Features			*Significant Features*		
	90-10	80-20	70-30	90-10	80-20	70-30
DT	64.20	65.28	55.56	65.43	63.89	65.08
LDA	69.14	59.72	46.03	76.54	70.83	66.67
SVM	65.43	69.44	60.32	72.84	68.06	69.84
KNN	83.95	80.56	66.67	90.12	86.11	84.13

The confusion matrix findings shows that the accuracy of KNN algorithm for all features and significant features which presented in Fig. 5 and Fig. 6. The True Positive Rate (TPR) and the False Negative Rate (FNR) are shown in a confusion matrix in KNN algorithms. The rows and columns of the matrix represent the predicted and actual classes for the MAS levels 0, 1, 1.5 (1+), 2, and 3, respectively. Comparing the accuracy of the KNN classifier using all features and significant features reveals important insights. KNN algorithm with all features achieves an overall accuracy of 83.95%, demonstrating superior performance, particularly in correctly identifying the extreme classes (0 and 3) and maintaining high true positive rates across most classes. However, when using the dataset with significant features, the KNN algorithm's overall accuracy increases to 90.12%. This improvement is evident in its enhanced ability to correctly identify not only the extreme classes (0 and 3) but also intermediate classes 1. It also showed increase in true positive rates across classes 1.5 compared to KNN algorithm using all features. This comparison highlights the effectiveness of using significant features in improving the classification accuracy of the KNN model. Futhermore, comparing with both dataset, significant features proving to be most optimum accuracy.

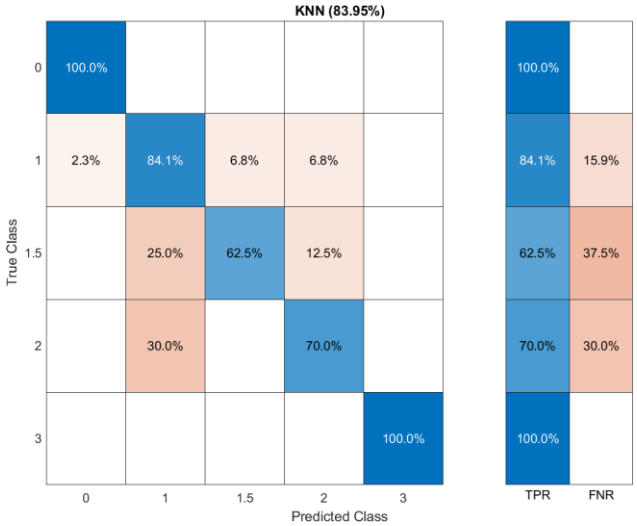


Fig. 5. Confusion matrix for KNN using all features.

Table IV presents the training durations for several machine learning algorithms, illustrating that the use of significant features consistently reduces training time compared to using all features. Specifically, the training time for the Decision Trees (DTs) algorithm decreased from 4.27 seconds to 3.24 seconds. The Linear Discriminant Analysis (LDA) algorithm's training time was reduced from 1.13 seconds to 0.91 seconds. The Support Vector Machine (SVM) algorithm showed a reduction in training time from 4.35 seconds to 3.75 seconds. Similarly, the K-Nearest Neighbors (KNN) algorithm experienced a decrease in training time from 3.03 seconds to 2.46 seconds. Among the algorithms evaluated, the Decision Trees (DTs) algorithm exhibited the most significant reduction in training time based on the percentage difference with 24.12% due to algorithm structured. The notable decrease in training time emphasises the efficiency improvements obtained by prioritising the most pertinent features, thereby illustrating the advantages of feature selection in machine learning models. Moreover, utilising crucial features simplifies the training process by decreasing the complexity and size of the dataset, enabling machine learning algorithms to function more effectively and attain quicker convergence [47].

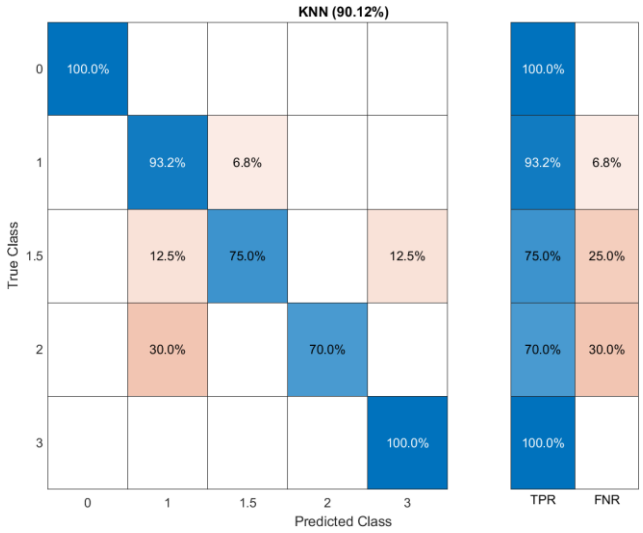


Fig. 6. Confusion matrix for KNN using significant features.

TABLE IV. TRAINING TIME OF ALL ALGORITHM WITH ALL FEATURES AND SIGNIFICANT FEATURES

Table Head	Training Time (seconds)		Percentage Difference (%)
	All Features	*Significant Features*	
DT	4.27	3.24	24.12
LDA	1.13	0.91	19.47
SVM	4.35	3.75	13.79
KNN	3.03	2.46	18.81

Machine learning models trained on datasets with significant features identified through a one-way MANOVA test consistently showed higher accuracy across all data split tests compared to those using all available features. This underscores the impact of feature selection on model performance, demonstrating that models trained on statistically significant features often outperform those using all features. This highlights the importance of feature selection in enhancing model accuracy and efficiency. Furthermore, the efficacy and efficiency of a machine learning solution are contingent upon the inherent qualities and attributes of the data, as well as the proficiency of the learning algorithms [47]. Among the algorithms, KNN exhibited the highest accuracy across all

datasets. The KNN classifier proves to be the optimal method for classifying biomechanical parameter features, particularly in scenarios with limited datasets and low dimensionality [50], [51].

A. Comparative Performance Analysis of Classifier Algorithms

The KNN algorithm using significant features was evaluated and compared in a 90/10 split using k-fold cross-validation with k values of 5, 10, and 15. This methodology enabled a thorough evaluation of the performance of KNN algorithm by dividing the dataset into k subsets, or folds, in a systematic manner. The algorithm underwent k-fold cross-validation, where it was trained and tested k times. In each iteration, a different fold was used as the validation set, while the remaining k-1 folds were utilised for training. Varying the value of k allowed for an examination of the models' stability and robustness across different partitioning schemes, providing a thorough evaluation of their predictive capabilities and overall performance in diverse scenarios. Table V illustrates the comprehensive comparison of accuracy in classifying various levels of spasticity, based on the output of MAS levels. The accuracy of KNN algorithm showed a decreased when the number of folds increased from 5 to 10. However, KNN algorithm demonstrates optimum accuracy with k= 15 at 91.29%.

Based on the result, there is no direct correlation between adjusting the value of K in k-fold cross-validation and the accuracy of machine learning algorithms [52]. Hence, while choosing the value of k, it is important to exercise caution as a lower k value entails decreased computing cost, reduced variance, but increased bias. Conversely, a larger value of k is more computationally demanding but exhibits greater variability and reduced bias. Therefore, the value of k must be chosen such that the size of each validation set is sufficient to ensure a reliable assessment of the model's performance. In conclusion, the KNN algorithm with significant features demonstrated superior performance in objectively evaluating the level of muscle spasticity.

TABLE V. PERFORMANCE OF KNN ALGORITHM WITH DIFFERENT VALUE OF K-FOLDS

***k*-folds**	**Percentage of Accuracy**
5	90.12
10	88.89
15	91.29

B. Clinical Implementation and Integration

A systematic approach would be beneficial for the successful incorporation of mechanomyography (MMG) technology into current spasticity management treatments. MMG evaluations can serve as an addition to older methods like the Modified Ashworth Scale (MAS) and the Australian Spasticity Assessment Scale (ASAS). MMG can enhance the reliability and consistency of spasticity level evaluations by offering unbiased data that can validate and improve the subjective assessments currently employed.

As trust in the technology increases, MMG might be progressively integrated as a principal evaluation tool. This would require the development of standardised protocols that integrate MMG measurements into clinical decision-making processes. For instance, MMG data can be utilised to modify treatment strategies, track the development of spasticity over a period, and assess the efficacy of therapies. Integrating MMG with current electronic health record (EHR) systems could enhance efficiency by enabling doctors to conveniently access and analyse MMG data in conjunction with other patient information.

Comprehensive training for therapists is crucial for ensuring the effective utilisation of MMG technology in clinical contexts. This training should include both the technical aspects of utilising MMG devices, and the analysis of data produced by the machine learning models. It is important for therapists to receive training to comprehend the importance of MMG signals, specifically the time and frequency domain characteristics considered essential for precise assessment of spasticity.

Furthermore, therapists must acquire knowledge of the machine learning algorithms employed to analyse MMG data. This entails comprehending the mechanisms by which these models generate predictions, interpreting the significance of the primary output metrics, and incorporating these insights into clinical practice. Hands-on training, supported by user-friendly software interfaces, will further enhance therapists' proficiency in utilizing MMG technology effectively.

C. Limitation of Study

While this study demonstrates the potential of mechanomyography (MMG) in assessing muscle spasticity, several limitations should be considered. Initially, while MMG is proficient in assessing muscle vibrations within the frequency range of 2 to 100 Hz, enhancing the sampling rate could enhance the precision of the data. Increasing the sampling rates can catch finer details of the muscle signals, perhaps improving the accuracy of the assessments and offering a more thorough comprehension of muscle spasticity.

Moreover, the study primarily utilises time domain variables for analysis. Although these features provide information, it may not comprehensively capture all the complexities that comprise MMG signals. In contrast, frequency domain features may identify additional patterns and behaviours that are not evident in the temporal domain. By including frequency domain analysis, a more comprehensive and precise depiction of muscle vibrations can be achieved. This has the potential to enhance the performance of models and enable more dependable evaluations of spasticity.

D. Future Research

A precise assessment of muscle spasticity is essential for the effective treatment and control of neurological diseases in patients. Although mechanomyography (MMG) has potential as a technique for assessing spasticity, its present uses mostly rely on extracting time-domain characteristics from MMG data. To completely maximise the potential of MMG and enhance its practical application in clinical settings, future research should prioritise several crucial areas of development.

Exploring advanced optimisation approaches and ensemble methods can greatly improve the accuracy and reliability of predictive models used in spasticity assessment. Techniques

such as hyperparameter tuning, ensemble learning, and deep learning approaches could offer significant improvements in interpreting MMG data and assessing spasticity levels. Besides that, the exploration of ensemble methods offers another promising avenue for improving model performance. By combining the predictions of multiple algorithms, ensemble techniques such as bagging, boosting, or stacking could reduce variance, mitigate overfitting, and increase the overall predictive power of the models. These methods could enhance the model's ability to generalize across different patient populations and clinical settings, thereby improving the robustness of spasticity predictions.

Moreover, the integration of advanced deep learning techniques, such as convolutional neural networks (CNNs) or recurrent neural networks (RNNs), could enhance the automated extraction of detailed and significant characteristics from the MMG signals. These models have demonstrated considerable efficacy in various biomedical signal processing contexts, as they are adept at identifying intricate patterns and relationships within the data that might be overlooked by traditional feature extraction methods. While this study primarily utilizes time domain features extracted from MMG signals, there remains significant potential to enhance model performance through more sophisticated feature engineering approaches. For instance, increasing the sampling rate of MMG data could capture finer details of the signal, thereby enabling the application of frequency domain extraction methods. Such techniques would provide a more nuanced analysis of the MMG signals, potentially uncovering features that are not detectable in the time domain alone.

By integrating these advanced techniques, the precision and robustness of spasticity assessments could be substantially improved, leading to more accurate and reliable predictions. Future research should investigate these avenues to further enhance the clinical utility of MMG in the objective assessment of muscle spasticity.

VI. Conclusion

Essentially, the purpose of this study was to address the problem of subjective and inconsistent evaluation of muscle spasticity in patients with neurological diseases. The objective was to validate MMG as a reliable signal by comparing the accuracy of various machine learning algorithms and demonstrate its clinal applicability in objective measurement. The study demonstrated the efficacy of employing different machine learning algorithms, such as Decision Trees (DTs), Support Vector Machines (SVM), and K-Nearest Neighbours (KNN), for accurately predicting degrees of spasticity. The KNN algorithm, using both all features and significant features, achieved optimal accuracy in the 90/10 split. Specifically, KNN with significant features demonstrated the highest accuracy at 91.29% with k=15, outperforming the use of all features, highlighting its effectiveness in categorizing biomechanical parameters. This technological development has the potential to greatly improve rehabilitation processes by offering more accurate and unbiased evaluations of spasticity. Moreover, it has the potential to decrease related expenses and time, ultimately resulting in an enhancement in the standard of treatment for impacted patients.

Acknowledgment

The research was conducted in the Biomechatronics Research Laboratory at the International Islamic University Malaysia. The author wishes to gratefully acknowledge the Ministry of Education (MOE) through Fundamental Research Grant Scheme (FRGS/1/2022/TK07/UIAM/02/6).

References

[1] J. C. Chacon-Barba, J. A. Moral-Munoz, A. De Miguel-Rubio, and D. Lucena-Anton, “Effects of Resistance Training on Spasticity in People with Stroke: A Systematic Review,” *Brain Sciences*, vol. 14, no. 1. Multidisciplinary Digital Publishing Institute (MDPI), Jan. 01, 2024. doi: 10.3390/brainsci14010057.

[2] H. Wang *et al.*, “Assessment of elbow spasticity with surface electromyography and mechanomyography based on support vector machine,” *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, no. Table 1, pp. 3860–3863, 2017, doi: 10.1109/EMBC.2017.8037699.

[3] M. Starosta, K. Marek, J. Redlicka, and E. Miller, “Extracorporeal Shockwave Treatment as Additional Therapy in Patients with Post-Stroke Spasticity of Upper Limb—A Narrative Review,” *Journal of Clinical Medicine*, vol. 13, no. 7. Multidisciplinary Digital Publishing Institute (MDPI), Apr. 01, 2024. doi: 10.3390/jcm13072017.

[4] J. H. Kim *et al.*, “Prospects of therapeutic target and directions for ischemic stroke,” *Pharmaceuticals*, vol. 14, no. 4, Apr. 2021, doi: 10.3390/ph14040321.

[5] V. L. Feigin *et al.*, “World Stroke Organization (WSO): Global Stroke Fact Sheet 2022,” *International Journal of Stroke*, vol. 17, no. 1. SAGE Publications Inc., pp. 18–29, Jan. 01, 2022. doi: 10.1177/17474930211065917.

[6] A. Popa-Wagner *et al.*, “Dietary habits, lifestyle factors and neurodegenerative diseases,” *Neural Regeneration Research*, vol. 15, no. 3. Wolters Kluwer Medknow Publications, pp. 394–400, Mar. 01, 2020. doi: 10.4103/1673-5374.266045.

[7] E. Krueger, E. Mendonça Scheeren, G. Nogueira-Neto, V. Lúcia da Silveira Nantes Button, and P. Nohama, *A New Approach to Assess the Spasticity in Hamstrings Muscles Using Mechanomyography Antagonist Muscular Group*. 2012. doi: 10.0/Linux-x86_64.

[8] J. W. Lance, “The control of muscle tone, reflexes, and movernent: Robert Wartenbeg lecture,” *Neurology*, vol. 30, no. 12, pp. 1303–1313, 1980, doi: 10.1212/wnl.30.12.1303.

[9] T. A. Whitten, A. Loyola Sanchez, B. Gyawali, E. D. E. Papathanassoglou, J. A. Bakal, and J. A. Krysa, “Predicting inpatient rehabilitation length of stay for adults with traumatic spinal cord injury,” *Journal of Spinal Cord Medicine*, 2024, doi: 10.1080/10790268.2024.2325165.

[10] C. Wang *et al.*, “Quantitative Elbow Spasticity Measurement Based on Muscle Activation Estimation Using Maximal Voluntary Contraction,” *IEEE Trans Instrum Meas*, vol. 71, 2022, doi: 10.1109/TIM.2022.3173273.

[11] S. Yu, Y. Chen, Q. Cai, K. Ma, H. Zheng, and L. Xie, “A Novel Quantitative Spasticity Evaluation Method Based on Surface Electromyogram Signals and Adaptive Neuro Fuzzy Inference System,” *Front Neurosci*, vol. 14, May 2020, doi: 10.3389/fnins.2020.00462.

[12] Z. J. Billington, A. M. Henke, and D. R. Gater, “Spasticity Management after Spinal Cord Injury: The Here and Now,” *J Pers Med*, vol. 12, no. 5, May 2022, doi: 10.3390/jpm12050808.

[13] A. Ahmad Puzi, S. N. Sidek, H. Mat Rosly, N. Daud, and H. Md Yusof, “Modified Ashworth Scale (MAS) Model based on Clinical Data Measurement towards Quantitative Evaluation of Upper Limb Spasticity,” in *IOP Conference Series: Materials Science and Engineering*, Institute of Physics Publishing, Nov. 2017. doi: 10.1088/1757-899X/260/1/012024.

[14] M. S. Erden, W. McColl, D. Abassebay, and S. Haldane, “Hand Exoskeleton to Assess Hand Spasticity,” in *Proceedings of the IEEE RAS and EMBS International Conference on Biomedical Robotics and Biomechatronics*, IEEE Computer Society, Nov. 2020, pp. 1004–1009. doi: 10.1109/BioRob49111.2020.9224329.

[15] E. Santos, E. Krueger, G. N. Nogueira-Neto, and P. Nohama, "Comparison of Modified Ashworth Scale with Systems and Techniques for Quantitative Assessment of Spasticity- Literature Review," *J Neurol Disord Stroke*, vol. 5, no. 2, pp. 1–9, 2017, [Online]. Available: https://pdfs.semanticscholar.org/cb25/a36e71801913a17d513865f6b18e0e6ade6e.pdf

[16] K. Fujimura *et al.*, "Requirements for Eliciting a Spastic Response With Passive Joint Movements and the Influence of Velocity on Response Patterns: An Experimental Study of Velocity-Response Relationships in Mild Spasticity With Repeated-Measures Analysis," *Front Neurol*, vol. 13, Mar. 2022, doi: 10.3389/fneur.2022.854125.

[17] P. Lewandowska-Sroka *et al.*, "The influence of emg-triggered robotic movement on walking, muscle force and spasticity after an ischemic stroke," *Medicina (Lithuania)*, vol. 57, no. 3, pp. 1–11, Mar. 2021, doi: 10.3390/medicina57030227.

[18] M. Aliff *et al.*, "MEKATRONIKA JOURNAL OF MECHATRONICS AND INTELLIGENT MANUFACTURING Mechanomyography in Assessing Muscle Spasticity: A Systematic Literature Review," vol. 6, pp. 92–103, 2023, doi: 10.15282/mekatronikajintellmanufmechatron.v6i1.10204.

[19] M. Correa, M. Projetti, I. A. Siegler, and N. Vignais, "Mechanomyographic Analysis for Muscle Activity Assessment during a Load-Lifting Task," *Sensors*, vol. 23, no. 18, Sep. 2023, doi: 10.3390/s23187969.

[20] S. W. Jun, S. J. Yong, M. Jo, Y. H. Kim, and S. H. Kim, "Brief report: Preliminary study on evaluation of spasticity in patients with brain lesions using mechanomyography," *Clinical Biomechanics*, vol. 54, pp. 16–21, May 2018, doi: 10.1016/j.clinbiomech.2018.02.020.

[21] E. L. Spieker *et al.*, "Targeting Transcutaneous Spinal Cord Stimulation Using a Supervised Machine Learning Approach Based on Mechanomyography," *Sensors*, vol. 24, no. 2, Jan. 2024, doi: 10.3390/s24020634.

[22] C. Meagher *et al.*, "New advances in mechanomyography sensor technology and signal processing: Validity and intrarater reliability of recordings from muscle," *J Rehabil Assist Technol Eng*, vol. 7, p. 205566832091611, Jan. 2020, doi: 10.1177/2055668320916116.

[23] D. Esposito *et al.*, "A piezoresistive sensor to measure muscle contraction and mechanomyography," *Sensors (Switzerland)*, vol. 18, no. 8, pp. 1–12, 2018, doi: 10.3390/s18082553.

[24] R. Uwamahoro, K. Sundaraj, and I. D. Subramaniam, "Assessment of muscle activity using electrical stimulation and mechanomyography: a systematic review," *BioMedical Engineering Online*, vol. 20, no. 1. BioMed Central Ltd, Dec. 01, 2021. doi: 10.1186/s12938-020-00840-w.

[25] T. Hazem, H. Soubra, and H. Othman, "MMG Signal Analysis for Muscle Performance Assessment," in *Procedia Computer Science*, Elsevier B.V., 2023, pp. 1412–1419. doi: 10.1016/j.procs.2023.01.430.

[26] E. L. Santos, M. C. Santos, E. Krueger, G. N. Nogueira-Neto, and P. Nohama, "Mechanomyography signals in spastic muscle and the correlation with the modified Ashworth scale," *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, vol. 2016-Octob, no. Ll, pp. 3789–3792, 2016, doi: 10.1109/EMBC.2016.7591553.

[27] M. O. Ibitoye, N. A. Hamzaid, J. M. Zuniga, N. Hasnan, and A. K. A. Wahab, "Mechanomyographic parameter extraction methods: An appraisal for clinical applications," *Sensors (Switzerland)*, vol. 14, no. 12, pp. 22940–22970, Dec. 2014, doi: 10.3390/s141222940.

[28] M. Szumilas, M. Władziński, and K. Wildner, "A coupled piezoelectric sensor for mmg-based human-machine interfaces," *Sensors*, vol. 21, no. 24, Dec. 2021, doi: 10.3390/s21248380.

[29] C. S. M. Castillo, S. Wilson, R. Vaidyanathan, and S. F. Atashzar, "Wearable MMG-Plus-One Armband: Evaluation of Normal Force on Mechanomyography (MMG) to Enhance Human-Machine Interfacing," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 29, pp. 196–205, 2021, doi: 10.1109/TNSRE.2020.3043368.

[30] M. A. I. Daud *et al.*, "Recent Studies of Human Limbs Rehabilitation Using Mechanomyography Signal: A Survey," in *Lecture Notes in Networks and Systems*, Springer Science and Business Media Deutschland GmbH, 2024, pp. 263–273. doi: 10.1007/978-981-99-8819-8_21.

[31] J. Y. Kim, G. Park, S. A. Lee, and Y. Nam, "Analysis of machine learning-based assessment for elbow spasticity using inertial sensors," *Sensors (Switzerland)*, vol. 20, no. 6, pp. 1–15, 2020, doi: 10.3390/s20061622.

[32] A. A. Puzi, S. N. Sidek, H. M. Yusof, and I. Khairuddin, "Objective analysis of muscle spasticity level in rehabilitation assessment," *International Journal of Integrated Engineering*, vol. 11, no. 3, pp. 223–231, 2019, doi: 10.30880/ijie.2019.11.03.023.

[33] A. A. Puzi, S. N. Sidek, I. M. Khairuddin, and H. M. Yusof, "Objective assessment for classification of muscle spasticity level," *ACM International Conference Proceeding Series*, pp. 4–9, 2020, doi: 10.1145/3440084.3441181.

[34] M. K. Liu, Y. T. Lin, Z. W. Qiu, C. K. Kuo, and C. K. Wu, "Hand Gesture Recognition by a MMG-Based Wearable Device," *IEEE Sens J*, vol. 20, no. 24, pp. 14703–14712, Dec. 2020, doi: 10.1109/JSEN.2020.3011825.

[35] T. Xie *et al.*, "Increased Muscle Activity Accompanying With Decreased Complexity as Spasticity Appears: High-Density EMG-Based Case Studies on Stroke Patients," *Front Bioeng Biotechnol*, vol. 8, Nov. 2020, doi: 10.3389/fbioe.2020.589321.

[36] M. International Functional Electrical Stimulation Society. Annual Conference (19th : 2014 : Kuala Lumpur and Institute of Electrical and Electronics Engineers, *2014 IEEE 19th International Functional Electrical Stimulation Society Annual Conference (IFESS) : conference proceedings : 17th-19th September 2014, Impiana Hotel KLCC, Kuala Lumpur, Malaysia.*

[37] B. B. Etana, B. Malengier, J. Krishnamoorthy, and L. Van Langenhove, "Integrating Wearable Textiles Sensors and IoT for Continuous sEMG Monitoring," *Sensors*, vol. 24, no. 6, Mar. 2024, doi: 10.3390/s24061834.

[38] L. M. Martins, N. F. Ribeiro, F. Soares, and C. P. Santos, "Inertial Data-Based AI Approaches for ADL and Fall Recognition," *Sensors*, vol. 22, no. 11, Jun. 2022, doi: 10.3390/s22114028.

[39] Y. Zhang and Y. Ma, "Application of supervised machine learning algorithms in the classification of sagittal gait patterns of cerebral palsy children with spastic diplegia," *Comput Biol Med*, vol. 106, pp. 33–39, Mar. 2019, doi: 10.1016/j.compbiomed.2019.01.009.

[40] V. N. Vapnik, "An Overview of Statistical Learning Theory," 1999.

[41] M. Castelli, L. Vanneschi, and Á. R. Largo, "Supervised learning: Classification," in *Encyclopedia of Bioinformatics and Computational Biology: ABC of Bioinformatics*, vol. 1–3, Elsevier, 2018, pp. 342–349. doi: 10.1016/B978-0-12-809633-8.20332-4.

[42] C. Mokri, M. Bamdad, and V. Abolghasemi, "Muscle force estimation from lower limb EMG signals using novel optimised machine learning techniques," *Med Biol Eng Comput*, vol. 60, no. 3, pp. 683–699, Mar. 2022, doi: 10.1007/s11517-021-02466-z.

[43] D. S. Stokic, M. Bohanec, M. M. Priebe, and A. M. Sherwood, "Relating clinical and neurophysiological assessment of spasticity by machine learning," 1998.

[44] Z. Zhang, "Introduction to machine learning: K-nearest neighbors," *Ann Transl Med*, vol. 4, no. 11, Jun. 2016, doi: 10.21037/atm.2016.03.37.

[45] P. Cunningham and S. J. Delany, "K-Nearest Neighbour Classifiers-A Tutorial," *ACM Computing Surveys*, vol. 54, no. 6. Association for Computing Machinery, Jul. 01, 2021. doi: 10.1145/3459665.

[46] J. Wen *et al.*, "Robust Sparse Linear Discriminant Analysis," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 29, no. 2, pp. 390–403, Feb. 2019, doi: 10.1109/TCSVT.2018.2799214.

[47] I. H. Sarker, "Machine Learning: Algorithms, Real-World Applications and Research Directions," *SN Computer Science*, vol. 2, no. 3. Springer, May 01, 2021. doi: 10.1007/s42979-021-00592-x.

[48] A. Sarkar, S. K. S. Hossain, and R. Sarkar, "Human activity recognition from sensor data using spatial attention-aided CNN with genetic algorithm," *Neural Comput Appl*, vol. 35, no. 7, pp. 5165–5191, Mar. 2023, doi: 10.1007/s00521-022-07911-0.

[49] I. O. Muraina, "IDEAL DATASET SPLITTING RATIOS IN MACHINE LEARNING ALGORITHMS: GENERAL CONCERNS FOR DATA SCIENTISTS AND DATA ANALYSTS." [Online]. Available: https://www.researchgate.net/publication/358284895

[50] N. Seth, D. Johnson, G. W. Taylor, O. B. Allen, and H. A. Abdullah, "Robotic pilot study for analysing spasticity: Clinical data versus healthy

controls," *J Neuroeng Rehabil*, vol. 12, no. 1, Dec. 2015, doi: 10.1186/s12984-015-0103-8.

[51] S. Sharma and V. Sharma, "Performance of Various Machine Learning Classifiers on Small Datasets with Varying Dimensionalities: A Study," *Circulation in Computer Science*, vol. 1, no. 1, pp. 30–35, Jul. 2016, doi: 10.22632/ccs-2016-251-23.

[52] I. K. Nti, O. Nyarko-Boateng, and J. Aning, "Performance of Machine Learning Algorithms with Different K Values in K-fold CrossValidation," *International Journal of Information Technology and Computer Science*, vol. 13, no. 6, pp. 61–71, Dec. 2021, doi: 10.5815/ijitcs.2021.06.05.

Interactive Color Design Based on AR Virtual Implantation Technology Between Users and Artificial Intelligence

Jun Ma[1], Ying Chen[2]*
School of Visual Communication Design, Luxun Academy of Fine Arts, Dalian City, 116650, China[1]
School of Media and Animation, Luxun Academy of Fine Arts, Dalian, 116650, China[2]

***Abstract*—To achieve user interactive color design, this study takes furniture color interactive design as an example, introduces artificial intelligence and augmented reality virtual implantation technology, allowing users to design furniture colors and styles according to their own ideas. By improving cyclic consistency to generate adversarial networks, furniture image style transfer is carried out, and indoor feature point classification and virtual model registration are carried out through density based spatial clustering and other methods to design an unlabeled augmented reality furniture system. The results showed that compared to other methods, the improved cyclic consistency generation adversarial network had a higher structural similarity value. In the zebra to horse image, the structural similarity value was 0.987, which was 0.018 higher than the algorithm before improvement. The registration effect of density-based spatial clustering algorithm was good, with a shorter time consumption in different scenarios, and a maximum time consumption of 0.308 seconds in occluded composite scenes. The performance of the drawing component is good, with each process of tracking threads taking less than 20ms. The research method not only satisfies users in designing furniture colors and styles, but also enhances their experience.**

Keywords*—Artificial intelligence; AR virtual implantation technology; color; style transfer; furniture

I. Introduction

The rapid development of information technology has advanced the field of human-computer interaction (HCI) research and enhanced the user experience. However, in the home furnishing industry, there are still shortcomings in the development of human-machine interaction. Users have not been able to achieve interactive design for furniture colors and styles according to their own needs, resulting in furniture colors and other aspects that do not match the home environment [1-3]. Augmented reality (AR) virtual implantation technology (AR-VIT) can solve this problem, providing users with a better interactive experience and significantly improving their sense of participation [4-6]. The emergence of artificial intelligence (AI) technology has driven the development of image style transfer (IST), enabling personalized design to be realized. When designing indoor furniture, choosing this technology is beneficial for real-time migration and adjustment of furniture style, allowing users to adjust colors, textures, and other conditions according to their preferences to meet their needs. In this regard, in furniture color design, to achieve user interaction, this study introduces AR-VIT and uses Cycle-Consistent generative adversarial networks (CycleGAN) to transfer home style to meet the needs of user interactive color design and improve user experience. The study is divided into five sections. Section II is a literature review that introduces the research status of domestic and foreign scholars on furniture design, CycleGAN algorithm, and AR-VIT. Section III involves IST, virtual model registration, fusion of 3D reconstruction algorithms, and virtual real fusion of furniture models. Section IV conducts result analysis to study the image clustering effect and the application of AR-VIT. Section V summarizes the research methods, pointing out shortcomings and future research directions. The contributions of the research are as follows: (1) the study introduced artificial intelligence and augmented reality virtual implantation technology, enabling users to interactively design furniture colors and styles based on personalized needs. (2) By optimizing loop consistency, the improved network proposed in this study exhibits higher structural similarity values in furniture image style transfer. (3) We have researched, designed, and implemented a label free augmented reality furniture system that can integrate user designed furniture styles into the actual environment in real-time, enhancing the user experience. The abbreviation table used in the study is shown in Table I.

TABLE I. Related Time Consumption

Full name	Abbreviation
Human-computer interaction	HCI
Augmented reality	AR
Augmented reality virtual implantation technology	AR-VIT
Artificial intelligence	AI
Image style transfer	IST
Cycle-Consistent generative adversarial networks	CycleGAN
Features from accelerated segment test	FAST
Random sample consensus	RANSAC
Parallel tracking and mapping	PTAM
Bundle adjustment	BA
Sum of squared differences	SSD
Multi-modal unsupervised image to image translation	MUNIT
Peak signal-to-noise ratio	PSNR
Structural similarity	SSIM
Inception score	IS
Normalized mutual information	NMI

II. Related Work

In the field of HCI, color interactive design is a part of it. However, in the field of furniture design, the user experience in furniture color and other aspects still needs to be improved. Ge S et al. conducted research on digital design methods for furniture cultural tourism exhibition platforms to improve their design level, based on multimedia networks. They collected and summarized data on Jinzuo furniture from different platforms, and divided furniture genres. After analysis, it was found that design methods can promote the improvement of platform design level [7]. Jiang L et al. studied adolescents and children to understand whether they were influenced by their own color preferences when choosing furniture. They analyzed the color preferences of different functional furniture at different ages. Tests have found that their color preferences can affect the selection of different types of furniture [8]. In the process of designing living room furniture, Nasir EB conducted a specific analysis of the application of design thinking in the design of living room furniture for students. The results showed that under the influence of design thinking, it was beneficial to understand the vague needs of users and had a positive promoting effect on design practice [9]. In the process of studying library furniture, Parvez M S et al. faced the problem of matching its size with student body measurements, conducted experimental settings, collected relevant data on students and home design, and compared them. The mismatch between these two types of data was more pronounced, which would be detrimental to the growth and development of students [10]. Lee I J et al. introduced AR technology and conducted comparative experiments to promote the understanding of 3D space for novice carpenters. With the help of AR, novice carpenters could better understand 3D space and have a higher level of mastery of complex mortise and tenon structures [11].

Liu X et al. designed and optimized the CycleGAN algorithm for unsupervised training in image dehazing, based on the relevant generator. Comparative analysis showed that this method had a better visual effect on image processing [12]. Chen et al. introduced the CycleGAN algorithm in spatio-temporal image fusion to address the issue of insufficient spatial information in images, and based on this, proposed a corresponding fusion framework. This model simulated the generation and processing of images, and input them into the constructed framework. After verification, the application effect of the proposed method was good [13]. In the process of studying Earth observation applications, Soto P J et al., faced with insufficient training data, chose the CycleGAN algorithm and improved the adaptability of the research object through nonlinear mapping functions. The proposed method could effectively solve the domain transfer problem in remote sensing applications [14]. Angrini L M and others faced the problem of poor mathematical thinking among students and conducted experiments on flat shape design using AR technology based on Unity 3D software. It randomly selected students for testing and interviews, and conducted result analysis. With the assistance of AR, students' mathematical and computational abilities had been significantly improved [15]. Ahmad H et al. focused on consumers and selected respondents to analyze their potential travel destinations during the pandemic, collecting their views on the impact of AR on tourism behavior. Statistical data showed that AR technology had a certain degree of impact on the travel intentions of consumers [16].

In summary, in the design of furniture styles and other aspects, most of them tend to focus on the specific design situation of furniture, with less emphasis on combining AI and AR-VIT, and there has been no research on user interaction color design. The CycleGAN algorithm has shown good performance in IST, which helps users design furniture styles. In addition, AR-VIT can overlay virtual objects into the actual environment, which is beneficial for furniture display. Therefore, to achieve interactive design of furniture styles such as color and texture for users, this study cites the CycleGAN algorithm and AR-VIT for furniture interactive design. Compared to previous research, this study has developed an unlabeled augmented reality system, overcoming the limitations of using artificial markers and applying it to the furniture field. The advantages and disadvantages analysis of different methods are shown in Table II.

TABLE II. Analysis of Advantages and Disadvantages of Different Methods

Author	Technique/Method	Advantages	Disadvantages
Ge et al.	Digital Design	Enhances platform design	Limited application in interactive design
Jiang et al.	Color Preference Study	Identifies impact of color preferences	Limited generalizability
Nasir EB	Design Thinking Application	Positively promotes design practice	Lacks technical application details
Parvez et al.	Furniture Size Matching	Highlights mismatch issues beneficial for student development	Restricted to specific settings
Lee et al.	AR Technology Application	Improves mastery of complex structures	Limited to novice carpenters
Liu et al.	CycleGAN Optimization	Better visual effects in image processing	Application scenarios not detailed
Chen et al.	CycleGAN in Image Fusion	Proposes a fusion framework with good application effects	Performance in practical scenarios not specified
Soto et al.	CycleGAN in Remote Sensing	Solves domain transfer issues effectively	Limited to remote sensing domain
Angrini et al.	AR for Computational Thinking	Significantly enhances students' abilities	Limited scope and generalizability
Ahmad et al.	Impact of AR on Tourism Behavior	Shows impact of AR on travel intentions	Limited to pandemic context

III. Interactive Design Based on AI and AR-VIT

To achieve interactive design of furniture styles such as color and texture, this study improves the CycleGAN algorithm to assist users in personalized design of furniture styles. Using AR-VIT, indoor feature point classification and virtual model registration are carried out through density-based spatial clustering of applications with noise DBSCAN and other methods, and an unlabeled augmented reality furniture system design is carried out.

A. IST based on Improved CycleGAN Algorithm

The development of the real estate market has led to the rapid prosperity of the furniture industry, but users still choose furniture using traditional methods. When users are not satisfied with the color, texture, etc. of furniture, it often means that the product transaction has failed, or when the purchased furniture arrives at home, there is a situation where it does not meet the expected effect. Therefore, to achieve interactive design of furniture color among users, this study conducts furniture IST based on user needs to change furniture color. The CycleGAN algorithm is introduced in this study, and its generator structure is Fig. 1.

In Fig. 1, for the CycleGAN algorithm generator, its transformation network contains six residual blocks, and its encoding network has three convolutional layers. Due to the incomplete style transfer and other issues in generating furniture images using this algorithm, this study improves the algorithm by optimizing its generator, discriminator, and loss function. The main architecture of the improved algorithm is the U-Net network, and nine residual blocks are added to the bottom of the network to promote sufficient transformation of image style. In improving the CycleGAN algorithm, its discriminative network has two discriminators. Different discriminators discriminate from different receptive fields to achieve higher network discrimination performance. In the improvement of the loss function, the L1 norm of the identity loss function is added, and the Wasserstein distance with gradient penalty term is introduced to rewrite the original adversarial loss function and improve the performance of the algorithm. The overall architecture of the improved CycleGAN algorithm is Fig. 2.

In Fig. 2, X and Y represent two different domains, with X being the content domain and Y being the style domain. Sample x in X and sample y in Y. x obtains sample $\hat{y}$ through the generator. D_Y is a discriminative network for Y, used to distinguish between y and $\hat{y}$ and determine the probability value of the true sample. F represents the generator from Y to X. $\hat{y}$ generates sample $\hat{x}$ through F. D_x represents the discriminative network of X, and x and $\hat{x}$ can be distinguished through D_x. To maintain the relative balance between the generator and the discriminative network, it is necessary to alternate training between the two, with the images being trained from X to Y and then to X, thus forming a loop. Specifically, to design two generators to improve the CycleGAN algorithm, with the same structure. By adopting this symmetrical structure, feature maps that are completely aligned with the content domain and style domain can be obtained. Overall, the structure of the generator is Fig. 3.

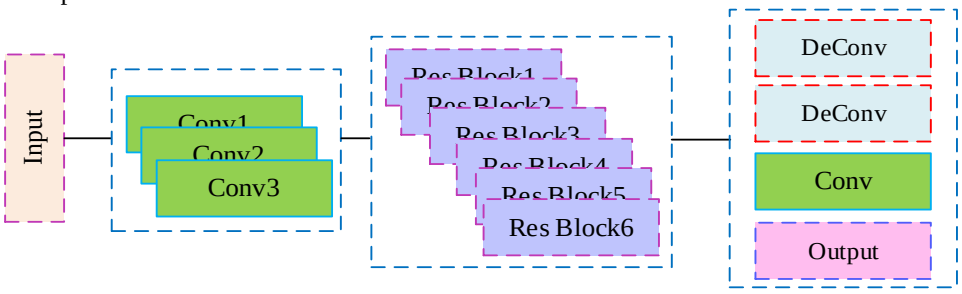


Fig. 1. Related structures.

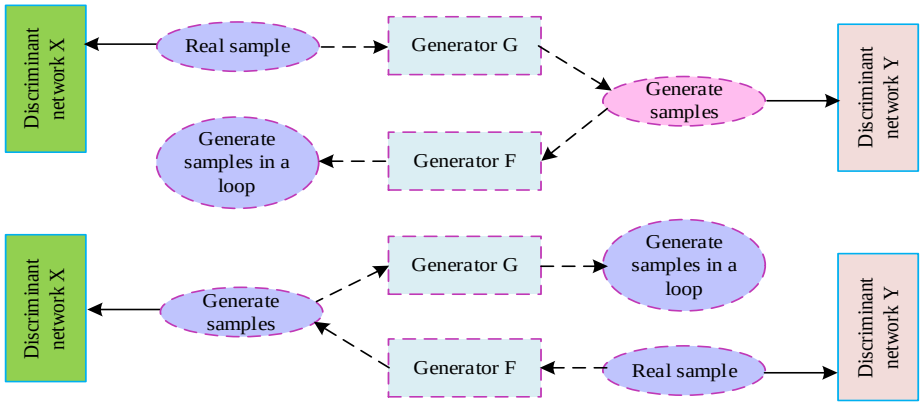


Fig. 2. The overall architecture of the algorithm.

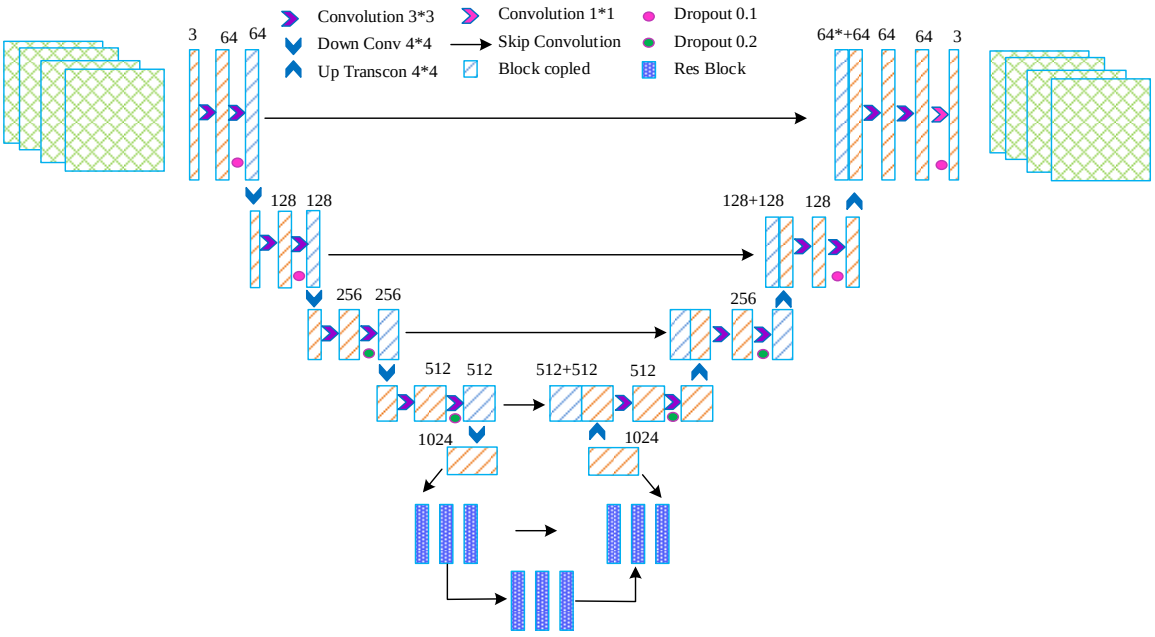


Fig. 3. Encoding network and decoding network.

In Fig. 3, the encoding network and decoding network represent the contraction path and expansion path of the U-Net network, respectively. In an encoding network, the size of the input image is 3 * 256 * 256, and the output is a feature map with a size of 1024 * 30 * 30. The sliding step size of the convolutional layer is set to 1, its filling amplitude is 1, and its convolution kernel is 3 * 3. Under the influence of this convolutional layer, the input image is processed into a feature map of 64 * 256 * 256. Due to the fact that selecting either the maximum pooling layer or the average pooling layer during down-sampling can result in significant loss of feature information. In the face of this problem, this study sets the convolution kernel to a size of 4 * 4, with a sliding step size of 2 and a filling amplitude of 1, respectively. In this case, it can be ensured that the number of output channels remains unchanged and the feature map size is halved. To encode and process images according to the two designed convolution kernels. The convolutional kernel size of the last down-sampling convolutional layer is 5 * 5, with a sliding step size and filling amplitude of 1 to obtain the final feature map. The image serves as the input of the transformation network, and the final output feature map size remains unchanged. Among them, in the residual block, a mirror is filled with a filling amplitude of 1, and the edge information of the feature map is saved as much as possible, making the feature map size 1024 * 32 * 32. Under the influence of a convolutional layer containing 3 * 3 convolution kernels, the size of the feature map changes to 1024 * 30 * 30. Repeating the image filling and convolutional layer processing, with the same parameter settings, to make the feature map size 1024 * 30 * 30 again. Using ReLu activation function and instance normalization layer to input the obtained feature map into the decoding network. Its output is the image after style transfer. During this period, feature maps and two different sizes of convolutional layers are concatenated using skip connections, ultimately enabling image reconstruction. The sizes of these two types of convolutional layers are 5 * 5 and 3 * 3, respectively.

Design of discriminator: It involves adding discriminator D_1 to the original discriminator PatchGAN D_2. The two discriminators have the same structure, but there is a difference in their input feature sizes. The former is 3 * 256 * 256, while the latter is 3 * 128 * 128. Calculating the mean probability value of two discriminators, and the result obtained is the final output probability value. To design the loss function, and the loss function L_{GAN} from X to Y is Eq. (1).

$$L_{GAN}(G,D_Y,X,Y)=E_{x\sim P_G}[D_Y(x)]-E_{x\sim P_{data}}[D_Y(y)]+\lambda \underset{\hat{x}\sim P_{\hat{x}}}{E}[(\|\nabla_{\hat{x}}D_Y(\hat{x})\|_2-1)^2] \quad (1)$$

In Eq. (1), the distribution of X and Y is set to $x\sim P_{data}$ and $y\sim P_{data}$, respectively. P represents the smallest rectangular plane of the feature points, which belongs to the initial plane. After passing G, the potential distribution $y\sim P_G$ of X 's image can be obtained. $\underset{\hat{x}-P_{\hat{x}}}{E}$ represents the gradient penalty term. Constraining the gradient around 1 to satisfy the Lipschitz condition for the discriminator. The penalty coefficient is set to λ, and the corresponding distribution $\hat{x}-P_{data}$ can be obtained through $x\sim P_{data}$ and $y\sim P_{data}$.

$$L_{GAN}(G,D_X,Y,X)=E_{y\sim P_G}[D_X(y)]-E_{y\sim P_{data}}[D_Y(y)]+\lambda \underset{\hat{y}\sim P_{\hat{y}}}{E}[(\|\nabla_{\hat{y}}D_X(\hat{y})\|_2-1)^2] \quad (2)$$

Eq. (2) is the adversarial loss function from Y to X.

$$L_{identity}(G,F)=E_{y\sim P_{data}}[\|G(y)-y\|_1]+E_{x\sim P_{data}}[\|F(x)-x\|_1] \quad (3)$$

In Eq. (3), $L_{identity}(G,F)$ represents the identity loss function.

$$L(G,F_x,D_X,D_Y)=L_{GAN}(G,D_Y,X,Y)+L_{GAN}(F,D_X,Y,X)+\lambda_i L_{identity}+\lambda_c L_{cyc}(G,F) \quad (4)$$

In Eq. (4), $L(G,F_x,D_x,D_Y)$ is the total loss function. $L_{cyc}(G,F)$ represents the cyclic consistency loss function. λ_i and λ_c are hyper-parameters. Adjusting the proportion of the loss function to the total loss function through these parameters.

B. Furniture System Based on DBSCAN Algorithm and AR-VIT

After completing the furniture style design, AR-VIT is introduced to conduct research on furniture stacking in actual environments. The first step is to register the virtual model, as shown in Fig. 4.

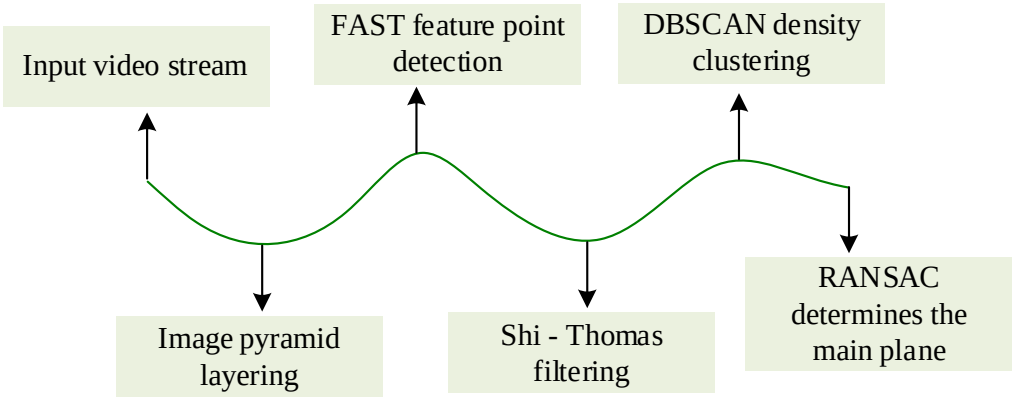


Fig. 4. Registration process.

In Fig. 4, the camera captures the video stream and performs pyramid layering on it. Different image layers correspond to different scales. The Features from accelerated segment test (FAST) algorithm has been selected to detect layered image features. Definition of corner: If the difference between a pixel and a large number of neighboring pixels exceeds a predetermined threshold, there is a certain probability that the pixel belongs to the corner. The relevant formula is Eq. (5).

$$N = \sum_{x' \forall (circle(p))} \left| I(x') - I(p) \right| > \varepsilon_d \tag{5}$$

In Eq. (5), x' represents the pixels on the circumference, and $I(x')$ represents the grayscale value of x'. The center point is p and the radius is r. The threshold for grayscale difference is set to ε_d, the adjustable parameter for the algorithm is set to N, and N is set to 10. When there is more than one feature point in the region, to delete the feature points with smaller response values. Calculating the feature point score S: In the feature point set V, assuming the existence of point c, its score is S_c. At point c, there is a neighborhood with a size of $s*s$. There is an arbitrary feature point l in this region, and if S_l exceeds S_c, point l is treated as a local maximum. Iterating other feature points in V in this way. The selection of the area with the densest indoor feature points is carried out by constructing a virtual object display main plane in that area, and then classifying feature points using the DBSCAN algorithm. To achieve higher operational efficiency, the size of the clustering dataset is reduced by setting its feature point data to 500. Datasets with a scale less than 500 are directly clustered, otherwise the score of feature points is calculated using the Shi-Thomas algorithm. According to the order of scores from high to low, the top 500 points are selected for clustering. During the operation of the DBSCAN algorithm, all sample points are scanned first. If the number of points within the threshold Eps radius range of a certain sample point exceeds the threshold $MinPts$ of the number of samples in the region centered on that point with a radius of Eps, then that point is selected into the core point list. Summarizing the points with the highest density to obtain the corresponding temporary cluster. Merging all temporary clusters to determine the final cluster. In the dataset, filtering out error points and treating them as noise points to avoid their impact on clustering results. After the feature point clustering is completed, the main plane of the virtual object is established for display.

$$q = \begin{cases} reserve, if\ q \in C_{\max} \\ delete, otherwise \end{cases} \tag{6}$$

In Eq. (6), q represents the feature points, and the cluster with the highest number of feature points after clustering is set to $C_{\max}$. At the location of $C_{\max}$, to establish the main plane using random sample consensus (RANSAC), which is a rectangle. Choosing the parallel tracking and mapping (PTAM) algorithm to ensure real-time tracking of virtual objects under changing indoor environments. Before this, camera calibration is performed. This study selects a pinhole camera imaging model and calibrates the camera using the Zhang calibration method to obtain the internal and external parameters of the camera. The calibration template is in black and white chessboard format. During the calibration process, 4 frames of images need to be collected from different angles. The main module of PTAM consists of two parts: drawing thread and tracking thread. The drawing thread process is Fig. 5.

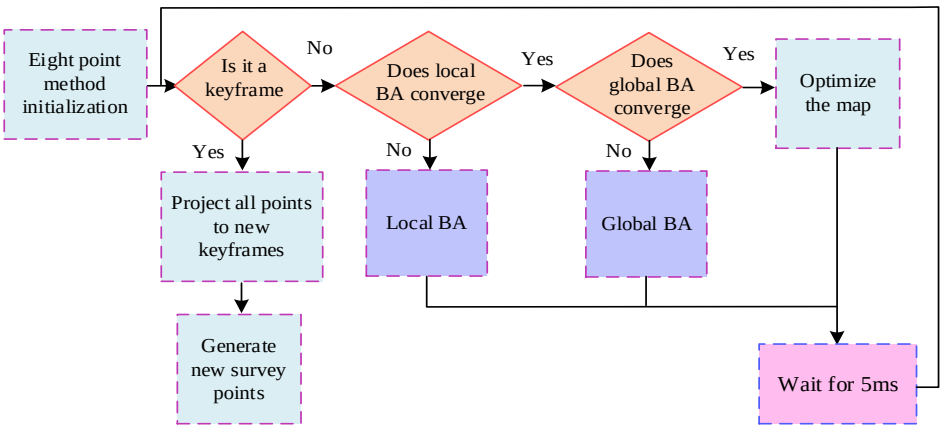


Fig. 5. Drawing thread process.

In Fig. 5, the first step is to manually set two keyframes and initialize them. When the camera deviates from the initial pose, the camera determines the current keyframe. It needs to have good tracking performance, contain a large number of feature points, and have sufficient duration. The interval between the image frame and the previous keyframe should be greater than 20 frames. In addition, the current frame should have sufficient distance in spatial position. When a keyframe is determined, the map information is updated based on the world points in that frame, combined with the original world points. Conversely, the map information is optimized using the bundle adjustment (BA) method. In local BA optimization, only the newly added last 5 keyframes and their world points are considered. The constraint is that other keyframes that can observe the world point are not optimized. In global BA optimization, all keyframes and their map points are optimized. During idle time, the drawing thread optimizes the map through old keyframes. To process outliers, treating them as new world points when they can be observed and converge, and adding them to the map. Through this approach, it is possible to prevent the loss of virtual furniture tracking and improve system accuracy. The process of tracking threads is Fig. 6.

In Fig. 6, the input image is layered into four pyramid layers, and the feature points of each layer are extracted using the FAST algorithm. The DBSCAN algorithm is used to cluster the feature points. After initializing the map, based on the initial two keyframes determined, feature points are selected through sum of squared differences (SSD) block matching (two frame images). Calculating the homography matrix between two frames, decomposing it into a rotation translation matrix, and treating it as the initial pose.

$$SSD(u,v) = Sum\left\{\left[left(u,v) - right(u,v)\right]^2\right\} \tag{7}$$

In Eq. (7), u and v represent the u-axis and v-axis of the pixel coordinate system, respectively. These two coordinate axes are parallel to the x"-axis and y"-axis of the image's physical coordinate system, respectively. The smallest difference in this equation is the best matching block. The normalized coordinates of the feature points in the initial two frames are set to $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$, and the formulas involved in these two coordinates are shown in Eq. (8).

$$s_1 x^*_1 = s_2 R x^*_2 + t^* \tag{8}$$

In Eq. (8), s_1 and s_2 represent the parameters of x^*_1 and x^*_2, respectively. R and t^* are the rotation translation matrix.

$$s_1 x^{*\wedge}_1 x^*_1 = s_2 x^{*\wedge}_1 R x^*_2 + x^{*\wedge}_1 t^* \tag{9}$$

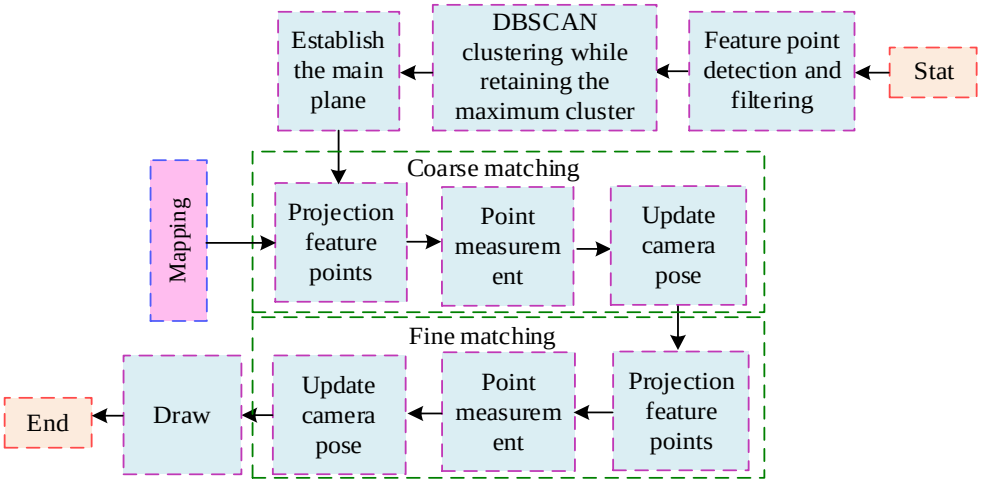


Fig. 6. Track the process of threads.

On the left side of Eq. (9), the equal sign is 0, and s_2 can be directly obtained. Based on this, s_1 can be calculated. Due to the influence of noise, R and t^* may not accurately make the left side of the equation equal to 0, so optimization is carried out through BA. The minimum objective function for local BA optimization is Eq. (10).

$$\left\{\left\{\mu_{x'} \in X'\right\}\left\{p'_z \in Z\right\}\right\} = \arg\min_{\{\{\mu\}\{p\}\}} \sum_{i \in X' \cup Y'} \sum_{i \in Z \cup S_i} Obj(i, j) \quad (10)$$

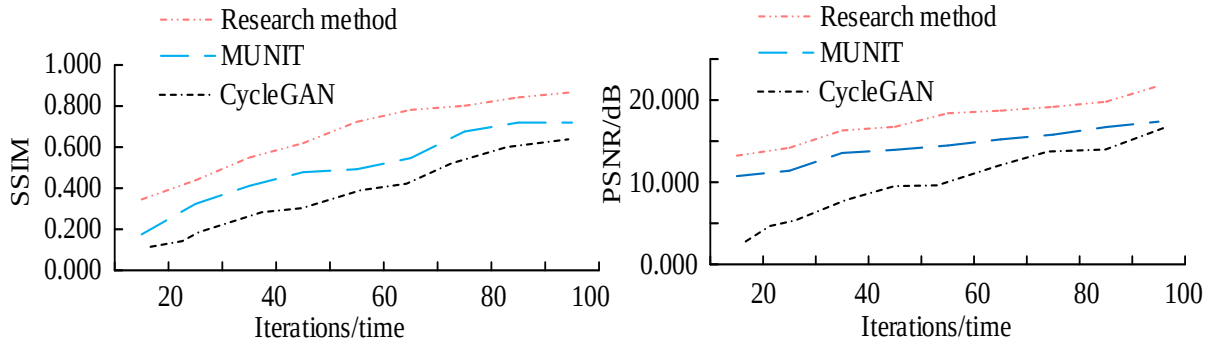
In Eq. (10), set X' consists of the current frame and its closest 4 other keyframes. In these keyframes, the observable world points are set as set Z. p'_z belongs to the world point in Z. The set of keyframes with observable world points in Z is represented as Y'. i and j are the serial number. $\mu_{x'}$ represents the element in X'. Local BA only optimizes elements in Z and X'. Calculating the initial pose of the camera. After obtaining the coordinates of the world point through triangulation, establishing the display main plane. 30-60 points in the top layer of the image pyramid are selected for coarse matching, and the world points based on the camera's pose are mapped. It is projected into the current frame to calculate the error with the current feature point. Based on the obtained results, an error optimization function is established and the camera pose is calculated. Based on the updated camera pose, to re select image points with a quantity of around 1000. Performing pose detection: The process is the same as coarse matching. Optimizing the pose by combining all points. The PTAM component completes the 3D registration of virtual furniture models to the real environment's 3D coordinate system and real-time tracking through the above process. Virtual real fusion uses user input to paste stylized images onto the surface of the constructed virtual furniture 3D model, and obtains the correct occlusion relationship. This can minimize the number of model vertices and patches, and reduce scene drawing calls, thereby achieving system performance optimization.

IV. Application Analysis of Furniture Systems Based on IST and AR-VIT

To analyze the effectiveness of IST and improve the performance of the CycleGAN algorithm, this study compared multiple algorithms with CycleGAN. By analyzing the performance of DBSCAN and PTAM algorithms, the application of the system could be evaluated.

A. Analysis of IST Results

To verify the performance of the improved CycleGAN algorithm, the Ubuntu operating system was selected with a learning rate of 0.0002. The experiment used the Horse2Zebra public dataset, $\lambda_c = 10$, $\lambda = 10$, and $\lambda_i = 1$. The comparative algorithms were the CycleGAN algorithm and the multi-modal unsupervised image to image translation (MUNIT) algorithm. The evaluation indicators were peak signal-to-noise ratio (PSNR), structural similarity (SSIM), and inception score (IS). The style transfer effect of the algorithm under different iterations analyzed is Fig. 7.

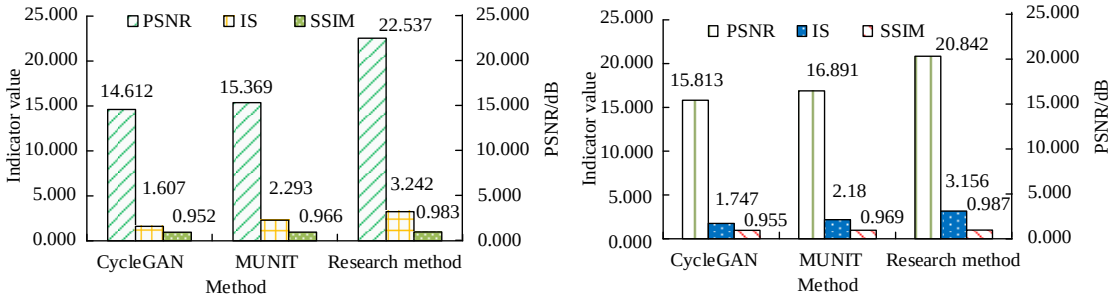


(a) SSIM values under different iteration numbers (b) PSNR values under different iteration numbers

Fig. 7. Style transfer effect.

In Fig. 7 (a), as the number of iterations increased, the SSIM value of the iterative feature effect map also increased, and the quality of the iterative map obtained by the research method was better. When the number of iterations was 35, the SSIM value of the research method was 0.547, which was 0.135 higher than the MUNIT algorithm, while the SSIM value of the CycleGAN algorithm was the smallest. When the iteration was 75, the SSIM values of the research method and MUNIT algorithm were 0.805 and 0.679, respectively. In Fig. 7 (b), the line where the research method was located was above the line where other algorithms were located. When the iteration was 45, the PSNR value of the research method was 16.201dB, which was 6.204dB higher than the CycleGAN algorithm. The results of analyzing the conversion effect between horse and zebra images are shown in Fig. 8.

In Fig. 8 (a), compared to methods such as the CycleGAN algorithm, the PSNR, SSIM, and IS of the research method were all larger. The IS value of the research method was 3.242, which was 1.635 higher than CycleGAN and 0.949 higher than MUNIT. In Fig. 8 (b), the SSIM values of the research method, CycleGAN, and MUNIT were 0.987, 0.969, and 0.955, respectively, indicating that the research method had the highest SSIM value. Therefore, the research method generated images with good quality.



(a) Comparison result of turning horse to zebra image

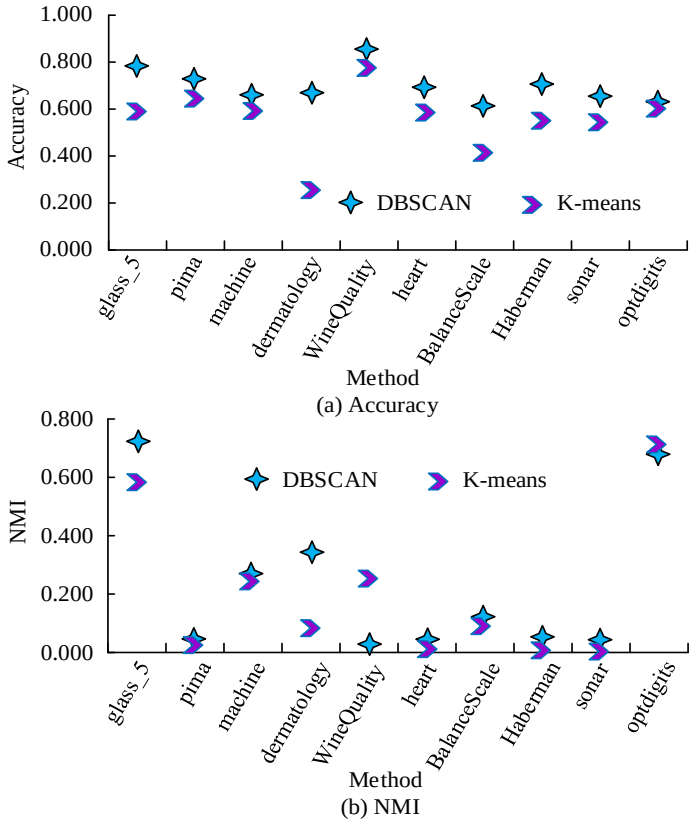
(b) Comparison result of turning zebra to horse image

Fig. 8. Related comparison results.

B. *Application Analysis of Unlabeled AR Furniture System Based on DBSCAN Algorithm*

When conducting application analysis, select representative furniture design cases and demonstrate how to use artificial intelligence and augmented reality technology to customize furniture colors and styles based on personalized user needs. Collecting users' furniture design preferences, including color, style, etc., through style transfer, feature point classification and model registration, and the integration of virtual and reality, users can intuitively see the designed furniture in their living space through augmented reality technology. To analyze the performance of DBSCAN algorithm, the comparison method was K-means algorithm, and the dataset was UCI dataset. The evaluation indicators were normalized mutual information (NMI) and accuracy, and the comparison results are shown in Fig. 9.

In Fig. 9 (a), due to different datasets, there were differences in the accuracy of the same algorithm. In the glass_5 dataset, the accuracy of the DBSCAN algorithm was 0.784, which was 0.187 higher than the K-means algorithm. The maximum accuracy on other datasets was 0.854, which was also higher than the K-means algorithm. In Fig. 9 (b), overall, compared to the K-means algorithm, the DBSCAN algorithm had a larger NMI value. In the Dermatology dataset, the NMI values of DBSCAN algorithm and K-means algorithm were 0.345 and 0.082, respectively. Therefore, the DBSCAN algorithm had a good clustering effect and performance. This study analyzed the registration effect of the DBSCAN algorithm by constructing corresponding main planes at chessboard positions obstructed by objects such as packaging boxes and brochures. Table III shows the display effect.



(a) Accuracy

(b) NMI

Fig. 9. Clustering results of different algorithms.

TABLE III. REGISTRATION EFFECT OF THE ALGORITHM

Scene	Number of environmental characteristic points/piece	Maximum number of points in the cluster/number	Time consumption/s
Black and white chessboard	119	106	0.156
Packaging box	158	95	0.180
Occlusion composite scene	251	169	0.308
Brochure	108	80	0.090

In Table III, the time consumption of algorithms varied depending on the scenario. Compared to other scenes, occluding composite scenes took up to 0.308 seconds, which was 0.218 seconds more than brochure scenes. Therefore, the real-time performance of the virtual object registration module was good. Table IV shows the time consumption of the drawing and tracking threads in the PTAM component.

In Table IV, when the number of keyframes was less than 50, local BA optimization and global BA optimization took less time, which were 170ms and 381ms, respectively. As the number of keyframes increased, the drawing thread's time consumption increased. When the key frame rate was between 100 and 149 frames, the global BA optimization took 6900ms. Overall, compared to other processes, the feature point measurement process took more time. When in the living room, projecting feature points took 3.5ms, which was 6.3ms less than the feature point measurement process in the same environment. Overall, each process of tracking threads took less than 20ms. Overall, the performance of PTAM components was good. Twenty users were randomly selected and rated on a scale of 1-5. The satisfaction evaluation data of the system application was collected through a survey questionnaire, as shown in Fig. 10.

In Fig. 10, compared to before system optimization, the optimized user satisfaction was higher. Overall, the user satisfaction score ranged from 3.50 to 5.00. Among them, after system optimization, the average user satisfaction score was 4.43 points, which was 0.30 points higher than before system optimization. Therefore, the application effect of the constructed AR system was outstanding.

TABLE IV. RELATED TIME CONSUMPTION

Stage	Keyframes/Frame		Local BA optimization/ms	Global BA optimization/ms
Drawing Thread	2-49		170	381
	50-99		277	1700
	100-149		445	6900
Tracking threads	Flow	Living room/ms	Bedroom 1/ms	Kitchen/ms
	Projection feature points	3.5	2.6	3.0
	Update camera pose	3.5	2.9	2.8
	Get keyframes	2.3	3.1	2.9
	Measurement of feature points	9.8	8.4	9.5

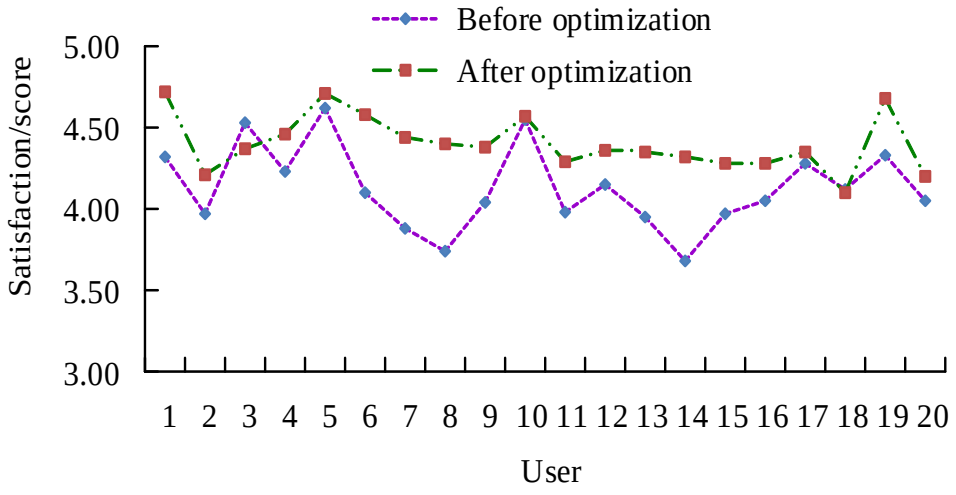


Fig. 10. User satisfaction with the system before and after optimization.

V. CONCLUSION

To achieve color interactive design for users, enhance their sense of participation and experience, this study used IST and AR-VIT to design furniture styles, achieving the goal of user color interactive design for furniture. Firstly, the CycleGAN algorithm was introduced for IST, and the overall performance of the algorithm was improved through identity loss functions and other means, making it convenient for users to design personalized furniture colors and styles. Subsequently, AR-VIT was introduced, and virtual model registration was performed using the DBSCAN algorithm, followed by the design of an unlabeled AR furniture system. The results showed that the improved CycleGAN algorithm performed better. When the number of iterations was 35, the SSIM value of the research method was 0.547, which was 0.135 higher than the MUNIT algorithm, while the SSIM value of the CycleGAN algorithm was the smallest. The IS value of the research method was 3.242, which was 1.635 higher than the CycleGAN algorithm. Compared to K-means, DBSCAN algorithm had higher accuracy and NMI. In the glass_5 dataset, the maximum accuracy of the DBSCAN was 0.854, which was higher than the K-means. In the PTAM component, the tracking thread took less time. When in the living room, the projection of feature points took 3.5ms, which was 6.3ms less than the feature point measurement process in the same environment. The user satisfaction after system optimization was relatively high, with an average satisfaction score of 4.43. Overall, the application effect of the research method is good. At present, the research method only considers the design content of color design in furniture design, and has not yet conducted targeted design for furniture appearance, structure, and material design, and cannot participate in the complete furniture design process. Subsequent research will focus on designing other aspects involved in furniture design, and integrate existing processes with other aspects to design furniture design methods that can meet more design needs and expand the scope of research methods.

REFERENCES

[1] Wang J, Pan Y. Design evaluation of parent-child interactive game furniture based on AHP-TOPSIS method. Journal of the Korea Convergence Society, 2022, 13(2): 235-248.

[2] Jarža L, Čavlović A O, Pervan S, Španić N, Klarić M, Prekrat S. Additive Technologies and Their Applications in Furniture Design and Manufacturing. Drvna industrija, 2023, 74(1): 115-128.

[3] Purohit J, Dave R.Leveraging Deep Learning Techniques to Obtain Efficacious Segmentation Results. Archives of Advanced Engineering Science, 2023, 1(1):11-26.

[4] Kumar H, Gupta P, Chauhan S. Meta-analysis of augmented reality marketing. Marketing Intelligence & Planning, 2023, 41(1): 110-123.

[5] Wu C H, Lin Y F, Peng K L, Liu, C. H. Augmented reality marketing to enhance museum visit intentions. Journal of Hospitality and Tourism Technology, 2023, 14(4): 658-674.

[6] Li M, Liu L. Students' perceptions of augmented reality integrated into a mobile learning environment. Library Hi Tech, 2023, 41(5): 1498-1523.

[7] Ge S, ** Z. Research on the Application of Digital Design of **zuo Furniture Cultural Tourism Display Platform Based on Intangible Cultural Heritage. International Journal of Communication Networks and Information Security, 2023, 15(2): 1-12.

[8] Jiang L, Cheung V, Westland S, Rhodes P A, Shen L Xu L. The impact of color preference on adolescent children's choice of furniture. Color Research & Application, 2020, 45(4): 754-767.

[9] Nasir E B. Identifying unspoken desires and demands: a collection of design ideas for living room furniture and zones. Journal of Design Thinking, 2021, 2(1): 71-84.

[10] Parvez M S, Tasnim N, Talapatra S, Kamal T, Murshed M. Are library furniture dimensions appropriate for anthropometric measurements of university students?. Journal of Industrial and Production Engineering, 2022, 39(5): 365-380.

[11] Lee I J. Using augmented reality to train students to visualize three-dimensional drawings of mortise–tenon joints in furniture carpentry. Interactive Learning Environments, 2020, 28(7): 930-944.

[12] Liu X, Zhang T, Zhang J. Toward visual quality enhancement of dehazing effect with improved Cycle-GAN. Neural Computing and Applications, 2023, 35(7): 5277-5290.

[13] Chen J, Wang L, Feng R, Liu P, Han W, Chen X. CycleGAN-STF: Spatiotemporal fusion via CycleGAN-based image generation. IEEE Transactions on Geoscience and Remote Sensing, 2020, 59(7): 5851-5865.

[14] Soto P J, Costa G, Feitosa R Q, Happ P N, Ortega M X, Noa J, Heipke C. Domain adaptation with cyclegan for change detection in the Amazon Forest. ISPRS Archives; 43, B3, 2020, 43(B3): 1635-1643.

[15] Angraini L M, Yolanda F, Muhammad I. Augmented reality: The improvement of computational thinking based on students' initial mathematical ability. International Journal of Instruction, 2023, 16(3): 1033-1054.

[16] Ahmad H, Butt A, Muzaffar A. Travel before you actually travel with augmented reality–role of augmented reality in future destination. Current Issues in Tourism, 2023, 26(17): 2845-2862.

A Comprehensive Authentication Taxonomy and Lightweight Considerations in the Internet-of-Medical-Things (IoMT)

Azlina binti Ahmadi Julaihi[1], Md Asri Ngadi[2], Raja Zahilah binti Raja Mohd Radzi[3]
Faculty of Computing, Universiti Teknologi Malaysia, Johor, Malaysia[1, 2]
Faculty of Engineering, Universiti Teknologi Malaysia, Johor, Malaysia[3]
Faculty of Computer Science and Information Technology, Universiti Malaysia Sarawak, Sarawak, Malaysia[1]

***Abstract*—The potential of Internet-of-Things (IoT) in healthcare is evident in its ability to connect medical equipment, sensors, and healthcare personnel to provide high-quality medical expertise in remote locations. The constraints faced by these devices such as limited storage, power, and energy resources necessitate the need for a lightweight authentication mechanism that is both efficient and secure. This study contributes by exploring challenges and lightweight authentication advancement, focusing on their efficiency on the Internet-of-Medical-Things (IoMT). A review of recent literature reveals ongoing issues such as the high complexity of cryptographic operations, scalability challenges, and security vulnerabilities in the proposed authentication systems. These findings lead to the need for multi-factor authentication with a simplified cryptographic process and more efficient aggregated management practices tailored to the constraints of IoMT environments. This study also introduces an extended taxonomy, namely, Lightweight Aggregated Authentication Solutions (LAAS), a lightweight efficiency approach that includes a streamlined authentication process and aggregated authentication, providing an understanding of lightweight authentication approaches. By identifying critical research gaps and future research directions, this study aims to provide a secure authentication protocol for IoMT and similar resource-constraint domains.**

Keywords*—*Lightweight authentication; Aggregated Authentication' Multi-Factor Authentication (MFA); Internet-of-Medical Things (IoMT)

I. Introduction

In the age of the Internet of Things (IoT), the integration of devices and networks has spread to the healthcare industry, resulting in the Internet of Medical Things (IoMT). These networks comprise interconnected medical equipment, sensors, and systems that allow for real-time observation, data collection, and analysis, improving patient care and healthcare delivery efficiency [1]. However, the use of IoT devices in medical settings raises concerns about security and privacy. Unauthorized access to these networks can result in data breaches, and exposing sensitive patient information necessitates robust security measures, particularly in the realm of authentication. It is critical to safeguard against unauthorized access and other cyber threats to ensure a secure and trustworthy authentication mechanism in IoMT environments.

Authentication as described by NIST, is the legitimacy of one's identity and an authenticator [2]. In general, authentication is the process of verifying the identity of a user, entity, or device attempting to access a network or system. Authentication is essential to all facets of private access, including access control to data and resources that are only available to specific entities. These processes include verifying credentials like passwords, digital certificates, or biometric information against a known set of registered identities in an authentication server or directory.

Authentication in IoMT is a multifaceted challenge due to the diverse range of devices, communication protocols, and the stringent requirements for data integrity and privacy. Traditional authentication methods, such as password-based schemes, are often inadequate in this context due to their susceptibility to various attacks, including replay attacks, man-in-the-middle (MITM) attacks, and impersonation attacks. Consequently, there has been a significant shift towards adopting more sophisticated authentication mechanisms, such as token-based, biometric, and multi-factor authentication (MFA) protocols. Despite these advancements, existing authentication protocols in IoMT still face several limitations. Many protocols are either too complex, resulting in increased computational overhead which is unfit for resource-constraint devices, or weak cryptographic techniques that are insufficiently secure against emerging threats. Thus, the development of lightweight authentication schemes, especially for IoMT, has been the subject of several studies to cater to resource limitations while providing secure communication and data exchange within the medical network.

This study aims to look further into lightweight approaches using various authentication credentials and their efficiency. Thus, this review attempts to evaluate the lightweightness approaches used and the limitation of existing authentication mechanisms in IoMT to give healthcare providers insight into best practices for securing these networks. This study also presents the background study of authentication solutions using a lightweight approach. Furthermore, current literature that employed a lightweight approach for IoMT is reviewed to obtain a better understanding of the requirements for healthcare settings. Thus, this study presented two contributions which are an extended authentication taxonomy, emphasizing lightweight approaches and multi-factor authentication solutions based on the existing work by Alsaeed and Nadeem [3], and a review of

recent existing works on authentication mechanisms using a lightweight approach that is particularly suited for IoMT environments. By categorizing and analyzing these protocols, this study seeks to highlight existing gaps and suggest potential areas for further research and development.

The remainder of this study is organized as follows. Section II illustrates the background study of the multi-factor authentication mechanism focusing on lightweight approaches. The extended authentication taxonomy is introduced in Section III with two lightweight authentication efficiency approaches: streamlined authentication process and aggregated authentications. Section IV presents a review of recent (2020-2024) related works on existing authentication mechanisms in IoMT. A discussion of the review is presented in Section V with identified research gaps and future works. Finally, the conclusion is included in Section VI.

II. Background Study

The healthcare industry has undergone a significant transformation with the advent of the Internet of Medical Things (IoMT). IoMT refers to the interconnected system of medical devices and applications that communicate through networking technologies to collect, analyze, and transmit health data [4]. This interconnected network allows for continuous, real-time monitoring of patients, leading to improved healthcare delivery, personalized treatment plans, and enhanced patient outcomes [5]. The integration of IoMT in healthcare has revolutionized traditional practices, making remote monitoring, telemedicine, and mobile health (mHealth) increasingly viable and effective.

With the increasing deployment of IoMT devices, ensuring the security and privacy of sensitive medical data has become a paramount concern [6]. Authentication is the first security layer and a critical security mechanism that verifies the identity of users and devices, ensuring that only authorized entities can access the system to protect from malicious security threats and data breaches [7]. Given the sensitive nature of medical data, any compromise can have severe implications, including incorrect diagnosis, treatment errors, and potential harm to patients.

As depicted in Fig. 1, a typical access system within an IoMT environment to ensure secure and efficient access control and data integrity across interconnected medical devices adopted from Anca et al. [8]. Access control has two important counterparts: the authentication and authorization processes, to enforce permissions for legitimate users or devices. Various factors such as passwords, biometrics, tokens, and multi-factor authentication are used when a user or a device wishes to attempt the system and the authentication mechanism will verify these credentials against stored data or through real-time verification. The author incorporates a lightweight mechanism into the authentication process model to ensure that the authentication process is efficient and uses minimal computational resources. On the other hand, the authorization database plays an important role in storing relevant credentials and permissions, in which the system administrator manages to maintain system integrity. Putting all together, this highlights the importance of integrating lightweight solutions to handle the difficulties presented by the heterogeneous and resource-constraint in IoMT ecosystem without sacrificing scalability or performance.

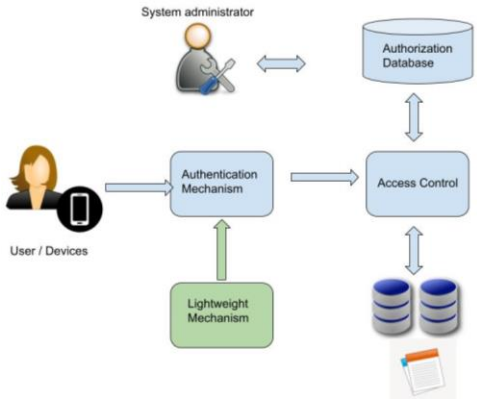


Fig. 1. Typical access system integrating lightweight mechanism in the authentication process (adopted from Anca et al. [8]).

Additionally, the diversity of devices and communication protocols used in IoMT environments necessitates adaptable and interoperable authentication solutions. As of March 2020, IEEE has published a new architectural standard for IoT, in response to the numerous unstandardized frameworks of IoT that have been proposed by researchers and industry. The goal of this standard is clear, to facilitate heterogeneous interaction, system interoperability, and the industry's continued development and scalability. According to one of the two standards, the P2413.1 RASC - Standard for a Reference Architecture for Smart City, defines a Reference Architecture with a four-layer architecture: device layer, communication network layer, IoT platform layer, and application layer [9]. Fig. 2 is the potential design used to develop IoMT authentication-role-specific architecture with reference to the P2413.1 RASC architecture.

In Fig. 2, the architecture of IoMT typically involves multiple layers: the device layer, communication layer, IoT platform layer, and application layer. Each layer faces unique security threats, and robust authentication mechanisms are essential to safeguard the entire IoMT ecosystem. The device layer comprises numerous wearable devices and medical sensors that collect medical information [10, 52]. This layer initiates the authentication process with basic verification of devices and initial user authentication, ensuring that data collected from legitimate sources is securely transmitted. Next is the network layer, which uses technologies such as Wi-Fi, Bluetooth, and cellular networks for secure data transmission and network authentication to healthcare providers and cloud services to prevent unauthorized access [11]. Moving up, at the IoT platform layer, intermediate authentication mechanisms are implemented to verify the integrity of data and devices before further processing or transmission to the cloud. Finally, the application layer encompasses comprehensive authentication and authorization processes, ensuring that only verified users and devices can access sensitive medical data and analytics services.

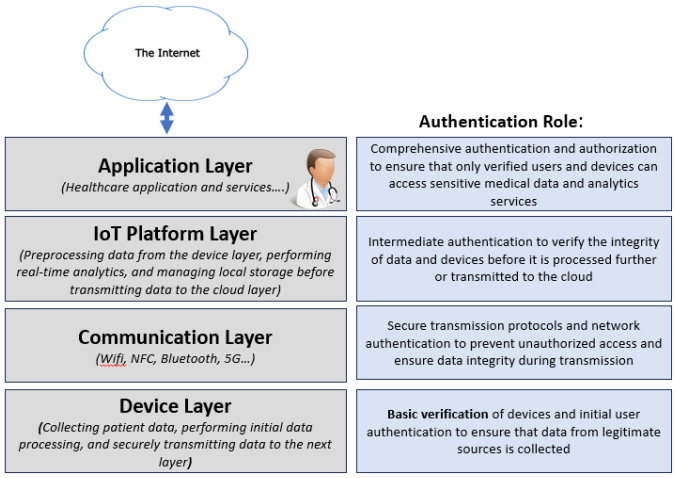


Fig. 2. IoMT authentication-role-specific architecture.

Further discussion on authentication is not complete without a reference to the four levels of security assurance. According to NIST, there are four levels of security assurance for authentication processes [12]. This structured framework provides specific requirements for identity proofing, authentication methods, and threat resistance vary depending on the level as depicted in Fig. 3.

There are several authentication levels and processes involved in authentication for IoMT. A typical usually involves 1) device registration, 2) user registration-this process verifies the identity of the users accessing the IoMT system, and 3) data access control, which ensures that only authorized users can access sensitive medical data. Each stage requires a robust authentication mechanism (in Fig. 2) to ensure the integrity and confidentiality of the transmitted health information At Level 1, device registration. In this stage, medical devices are registered with the healthcare network, often involving the generation and exchange of cryptographic keys. Basic verification using a simple authentication mechanism might be used at the device layer, suitable for initial health data collection. Moving up to Level 2, a more robust identity verification and single-factor authentication would be applied at the communication layer, to resist security attacks such as replay and eavesdropping for secure data preprocessing and transmission. Level 3 and Level 4 are essential in the cloud layer, where multi-factor authentication and strong encryption methods are used to safeguard information during data analysis and long-term storage. This setup offers protection against cyber-attacks like MITM attacks ensuring that only authorized users and devices can access important medical data. This hierarchical application of assurance levels provides as a basis across the IoMT architecture to ensure a thorough security approach is established, tailored to the needs and capabilities of each layer.

However, the unique characteristics and distributed nature of IoMT devices present several challenges for a robust authentication mechanism making it challenging to implement complex and computationally intensive authentication protocols. Moreover, limited processing power and memory in many resource constraints IoMT devices pose challenges for deploying strong authentication mechanisms like multi-factor authentication (MFA) or cryptographic techniques. These constraints necessitate the development of lightweight, yet secure authentication solutions tailored to the capabilities of IoMT devices. Thus, it is imperative to have an authentication mechanism that can minimize computational and communication overhead as well as energy consumption while maintaining robust security measures. There are a few possible existing authentication approaches in the IoMT system which include approaches like lightweight cryptography, lightweight multi-factor authentication, and lightweight hybrid anomaly detection [5]. For instance, lightweight cryptographic algorithms such as elliptic curve cryptography (ECC) and physically unclonable functions (PUFs), offer strong security with reduced resource requirements [13], [14], [15], [16]. In general, the characteristics of any lightweight algorithms can be defined as follows:

1) Low computational complexity: Algorithms that require fewer computational cycles to execute [17].

2) Minimal memory usage: Less memory is used for both code and authentication data [18].

3) Energy efficiency: Optimized for low power consumption, crucial for battery-operated devices [19].

4) Small key sizes: Use of smaller cryptographic keys to reduce processing overhead while maintaining security [20].

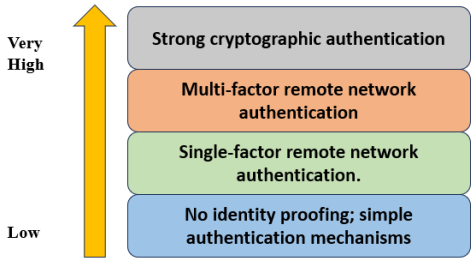


Fig. 3. Four security assurance levels by NIST.

Researchers focusing on developing authentication solutions strive to find a balance between security and resource efficiency to address security risks in IoMT environments. Efforts in lightweight authentication research often concentrate on creating solutions to meet this need. Many studies have investigated aspects of lightweight cryptography and

authentication schemes in the wider IoT context, but a focused look within the specific IoMT realm is lacking. Research efforts have mostly focused on cryptographic techniques, with varying attention to the details of data transmission and authentication within IoMT. The works of Sallam and Beheshti [21] have added a lot to understanding the applicability of lightweight cryptography and studies ongoing developments in the realm of IoT, but there's a gap in having a thorough streamlined authentication approach specific to IoMT is still lacking. As the number of connected medical devices and the complexity of IoMT environments grow, ensuring efficient authentication at scale becomes increasingly crucial to meet the healthcare industry's demands. Moreover, while some studies have talked about potential attacks on lightweight cryptography, there's not much literature systematically analyzing the aggregated authentication mechanism necessary in IoMT. The authors aim to fill these gaps by putting together and critically looking at the existing knowledge, and finding areas where more investigation is needed. Through this, the authors hope to offer an updated and consolidated understanding of lightweight cryptography and authentication in IoMT, pushing forward improvements in the security of healthcare IoT systems. Nonetheless, despite these limitations and challenges, continued research and development in lightweight authentication mechanisms are necessary to enhance security efficiency, particularly in healthcare settings.

III. Extended Authentication Taxonomy

A well-defined taxonomy for authentication on the Internet of Medical Things (IoMT) is essential to enhance the understanding of this complex topic and address the interrelationships among various elements within IoMT systems. In recent years, numerous security solutions have been created and put forth. However, it remains considerably challenging to produce a competent solution for authentication in a resource-constrained network. One of the main challenges is to provide a lightweight authentication solution, tamper-proof to security threats for IoT applications specifically in the field of medical IoT. The extended taxonomy is built upon an existing authentication taxonomy by Alsaeed and Nadeem [3]. There are seven main perspectives in Alsaeed and Nadeem's taxonomy. The taxonomy is further categorized by the authors according to the following axes:

- Authentication Factors comprised of Type of Credentials, Authentication Levels, and Authentication Procedure. This study will focus on the type of credentials used in the authentication processes in recent literature as different types of credentials have varied impacts on the lightweightness of an authentication process. Authentication Procedures consist of One-way authentication verifies one entity to another, two-way (mutual) authentication verifies both entities to each other, and three-way authentication involves a third trusted entity in the process. The selection of credentials should align with the specific limitations and needs of the IoMT environment, ensuring a balance between lightweightness and security.
- Authentication Schemes which refer to authentication architectures and authentication categories. Authentication architectures include both centralized and decentralized architectures, which are further divided into flat and multi-level approaches. Authentication categories differentiate between static and continuous authentication.
- Authentication attacks, address various authentication attacks and the measures taken to prevent them, such as resistance to guessing, impersonation, man-in-the-middle attacks, etc.
- The fourth axe is on the authentication solutions which include basic authentication methods, key-agreement used in authentication, cryptography-based and certificate-based schemes. And finally, the extended taxonomy on lightweight authentication mechanisms. The lightweight mechanism includes streamlined authentication processes and aggregated authentication which include approaches designed to optimize performance by reducing computational and communication overhead, resource usage, and response time, which are key attributes of efficiency in authentication protocols.

Furthermore, several key aspects must also be considered when designing a lightweight authentication mechanism for IoMT to ensure that the system is secure, efficient, and compatible with resource-constrained devices such as medical devices. These aspects can be categorized into four aspects: Security robustness, Lightweight Efficiency approach, Compatibility approach, and Usability approach. This study explores two areas in terms of multi-criteria authentication taxonomy based on the work of Alsaeed and Nadeem [3] and Agrawal and Ahlawat [22].

1) Security robustness: Firstly, the security mechanism must be robust. The security approach primarily focuses on the authentication strength such as using strong cryptographic methods and multi-factor authentication[23], [24], [25], [26]. One of the challenges is that implementing robust encryption without significantly impacting device performance can be difficult.

2) Lightweight efficiency approach: The efficiency for authentication should cover the key design considerations such as the low computational overhead that can adapt to IoMT devices with limited processing power by employing lightweight cryptographic algorithms [27], [28]. Recent advancements in lightweight cryptographic algorithms have shown significant potential in improving the efficiency of IoMT authentication processes. Using lightweight cryptographic primitives such as Elliptic Curve Cryptography (ECC) and hash functions provides strong security with minimal computational resources [15], [29]. Thus, selecting the right cryptographic primitives in the authentication process is crucial. For instance, the work of Chatterjee et al. [29] demonstrated the effectiveness of ECC and hash-based schemes in IoT security, highlighting their suitability for resource-constrained devices. Furthermore, the integration of aggregated authentication techniques such as batch processing and shared key generation, streamlines the authentication process, reduces communication overhead, and

improves scalability, making them suitable for dynamic IoMT environments where devices frequently join and leave the network.

3) Compatibility approach: Some of the well-known standard protocols such as OAuth 2.0 or FIDO are required to ensure the authentication mechanism is interoperable with a wide range of IoMT devices and platforms. Scalability remains the biggest hurdle within the IoMT ecosystem due to the growing number of devices and diverse platforms.

4) Usability approach: The usability approach centers around the authentication process that should never be overlooked. The authentication factors should be user-friendly employing methods such as biometric authentication for ease of use [30]. Challenges include designing authentication mechanisms that are both user-friendly and secure can be conflicting goals.

This study will further investigate lightweight efficiency approaches to fulfill the study goal. To identify the best authentication efficiency approaches for lightweight mechanisms were studied through existing recent literature. In a survey by El-hajj et al. [31], the author provides a comprehensive review of lightweight authenticated encryption for IoT devices, using a multi-criteria classification approach. The authors evaluate various aspects of authentication methods to assess their strengths and weaknesses, including security robustness, computational efficiency, scalability, simplicity of implementation, resilience to attacks, compatibility with existing systems, and usability. By considering these evaluation parameters, the authors compare the efficiency of different authentication techniques in IoT applications. The assessment of lightweight design, multi-factor authentication, and encryption technique usage provides insights into the efficiency and effectiveness of authentication methods in securing IoT devices while optimizing resource utilization.

In another survey done by Agrawal and Ahlawat [22], they review the authentication schemes based on three different parameters which are lightweight, multi-factor authentication, efficient, and encryption technique usage. According to the authors, these three classifications based on their reviews are important considerations in determining authentication efficiency. They also compared the efficiency of different authentication techniques in IoT applications and assessed whether any of the criteria used are preferred to ensure efficient operation in resource-constraint IoMT environments. The authentication protocols can be designed according to various factors. Two-factor authentication consists of utilizing user identification and biometric data to grant access to the IoMT system [32]. On the other hand, multi-factor authentication is the combining different elements such as user possession, inheritance and knowledge credentials to increase security [33], [34]. Many authentication approaches rely on multi-factors to enhance security and provide more resilient authentication solutions to protect from any adversary breach [26]. Thus, the presence of multi-factor authentication needs to be considered in the comparison to determine the efficiency of the authentication method [19].

On the other hand, the authors Samal et al. [35], classified authentication protocols into five different domains which are mutual authentication, one-time password, public key cryptography, zero-knowledge proof, and digital signature. They further classified three different categories for cryptographic algorithms such as 1) Encryption algorithms, 2) Signature algorithms, and 3) Hashing algorithms. These cryptographic techniques should be implemented efficiently to avoid the excessive computational burden of IoMT devices [35].

Thus, taking into account the analysis from the previous section, this study provides the extended version (Lightweight Aggregated Authentication Solutions, LAAS) of the authentication solutions taxonomy by Alsaeed and Nadeem [3] by adding in another authentication solution, namely under the lightweight approach, with its two methods; streamlined authentication process and aggregated authentication, as shown in Fig. 4 (illustrated in dashed rectangle box) to be well-suited for a resource-constraint environment such as IoMT. The details of these two methods are discussed in the next subsections.

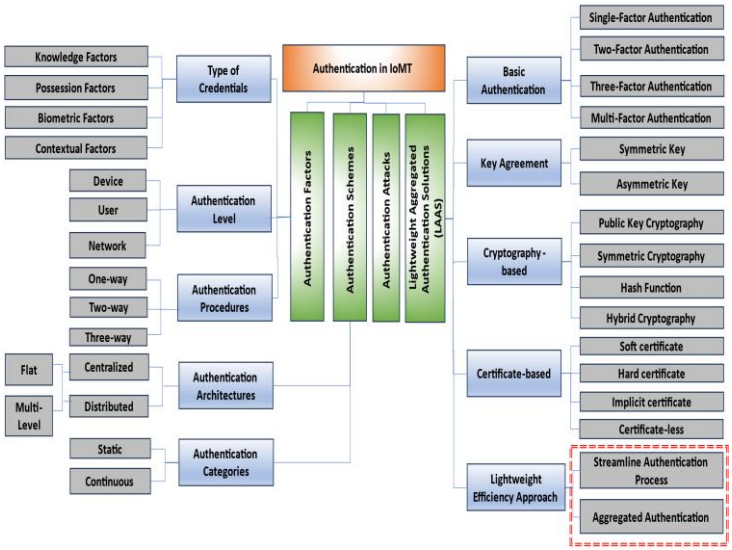


Fig. 4. Extended taxonomy for authentication in IoMT.

B. Lightweight Efficiency Approaches

1) Streamlined authentication process: It has been observed in recent years that lightweight authentication schemes were being proposed by numerous researchers at different times to increase system efficiency. According to Alsaeed and Nadeem [3], the number of exchanged messages during authentication processes will immediately affect the authentication scheme performance. Reducing the number of messages in an authentication protocol can be a strategy to make the algorithm lightweight, but it is not the sole criterion. As depicted in Fig. 4, the authors used an extended term to refer to this lightweight approach in authentication which is a streamlined authentication process that refers to the reduction in the complexity and number of steps involved in the authentication process. These lightweight authentication algorithms typically involve based on the usage of multiple factors such as 1) Symmetric key lightweight protocols, a lightweight algorithm that is being used during the pre-shared key exchanges, secure and shorter key sizes for encryption and decryption operations. 2) Hardware-assisted lightweight solutions such as PUF-based authentication 3) Biometric-based authentication like fingerprint recognition or 4) Simplified cryptography primitives, such as XOR, concatenation, and hash operations to support processing capabilities of IoT devices. Current literature agrees that streamlined authentication processes in authentication are the key essential in any lightweight solutions to enhance security while minimizing resource scarcity and computational complexity [3], [13], [36], [37]. Often, simpler cryptographic operations are used such as using only hashing and XOR to minimize the overhead complexity so that it is lightweight enough for a resource-constraint environment [38], [39], [40], [41]. Therefore, all the above factors are vital aspects of streamlined authentication, enhancing the practicality and performance of security mechanisms in real-time, high-frequency authentication scenarios in IoMT and similar applications. However, the challenge is to streamline the authentication process and keep a high level of security. In the context of the IoMT environment, this is crucial as we are dealing with time-sensitive healthcare applications, thus a high level of security is the utmost priority.

2) Aggregated authentication: Implementing a robust authentication mechanism in IoMT poses several challenges. As pointed out in the previous section, resource constraints in IoMT devices are the main challenge as these devices often have limited computational and energy resources, and the possible authentication solution is to make them lightweight. Apart from these main challenges, scalability issues are nothing new in IoMT environments. Handling the growing number of connected devices and users with one-to-one communication is inefficient for the massive communication required by today's IoMT-based applications.

Aggregated authentication in IoMT refers to combining authentication processes into a single operation, enhancing efficiency and reducing communication overhead [42]. Several techniques are involved such as aggregating multiple authentication requests into a single process (batch processing) [43] and hierarchical aggregation which typically involves multi-level aggregation where data is aggregated at local/edge nodes before being sent to central servers [44]. These approaches are tailored to address the unique constraints of IoMT environments, ensuring that authentication processes are both secure and efficient.

This approach is particularly beneficial in environments where numerous small transactions occur frequently, as it can significantly reduce overhead, improve processing efficiency, and enhance overall system performance. In the context of lightweight authentication, aggregated authentication can be leveraged to streamline authentication processes, especially in systems like the IoMT where multiple authentication requests might occur simultaneously from various medical devices and sensors. By bundling these requests, the system can handle them more efficiently, minimizing computational load and reducing latency. This method aligns well with the principles of lightweight authentication, which aim to provide secure, efficient, and low-overhead authentication mechanisms suitable for resource-constrained environments. Therefore, incorporating aggregated authentication into lightweight authentication schemes can further enhance their efficiency and effectiveness, making them an attractive solution for real-time, high-frequency authentication scenarios in IoMT and other similar applications.

IV. Related Works

To address the need for lightweight multifactor authentication schemes, several research works have been conducted in recent years. In [45], they proposed a lightweight multifactor authentication [51] scheme for cellular networks, exclusively 5G, and a trust-based blockchain architecture for VANET to mitigate major communication attacks using blockchain technology. In a subsequent paper, they extended this work to propose a lightweight multifactor authentication security scheme for a multi-hop scenario using timestamping, one-way hash function, Blind-Fold Challenge scheme with public key infrastructure with reduced authentication overhead, computation cost, and communication cost [23]. They also contributed to this area by proposing a lightweight multifactor authentication protocol for multi-gateway WSNs using hash functions and XOR operations. Additionally, Xue et al. [41] used lightweight cryptographic primitives to propose a lightweight three-factor anonymous authentication approach in multi-gateway WSNs using hash functions and XOR operations. These works collectively demonstrate the ongoing efforts to develop efficient and secure lightweight multifactor authentication schemes for various network scenarios, including cellular networks, WSNs, and healthcare applications.

In another research effort, Atiewi et al. [46] introduced a lightweight multifactor secured smart card-based user authentication for cloud-IoT applications, emphasizing the importance of scalability and security in big data IoT systems. To tackle the resource-constraint issues in the IoMT, it is significant to use lightweight multi-factor cryptographic algorithms, such as block ciphers and hash functions to enhance

data confidentiality, integrity, and secure authentication [28]. Hash functions and block ciphers are a few examples of lightweight cryptographic algorithms that can be applied in IoT devices to achieve strong protection against unauthorized access and data breaches. These cryptographic primitives, which offer effective encryption, safe authentication methods, and data integrity verification are necessary to guarantee the security of the system.

On the other hand, Gumis et al. [24], proposed a biometric blockchain-based multifactor privacy-preserving authentication scheme for Vehicular Ad Hoc Networks (VANETs). The scheme employs Physical Unclonable Functions (PUF) and one-time dynamic pseudo-identities as authentication factors, providing lightweight and privacy-preserving authentication for VANETs. PUFs have gained widespread use in user authentication protocols, leveraging a device's distinct physical characteristics for authentication rather than easily replicable passwords and secret keys [47], [48].

In a related study, Malik et al. [49] proposed a lightweight certificate-based authentication scheme for IoT devices and networks, introducing L-ECQV, a lightweight certificate profile of ECQV implicit certificates, and suggesting the inclusion of PUFs for multi-factor authentication. The study provides insights into certificates and PUFs in lightweight authentication protocols, contributing to the development of an enhanced two-factor authentication protocol. Additionally, the work by Ebrahimabadi et al. [50] addressed the threat of PUF modeling by employing multifactor authentication, including a shared cryptographic key alongside the Challenge Response Pair (CRP), to enhance the resilience of authentication protocols for IoT devices. This illustrates ongoing efforts to counter specific security threats through multifactor authentication, contributing to the advancement of lightweight authentication mechanisms for IoT environments.

Other related work includes researchers working on the same authentication area using the blockchain model proposing various approaches and evaluating their work against various security threats using formal or informal security analysis on the Internet of Medical Things. The work of [39], [44], and [45] proposed that apart from using multifactor authentication and lightweight cryptography to enhance security and optimize efficiency, it is essential to integrate blockchain technology between IoT and cloud environments that could provide an additional layer of security and transparency, ensuring the integrity of data and transactions. Thus, it is crucial to have a balance of security and efficiency for resource-constrained WSN nodes, considering factors such as energy consumption and computational overhead.

Therefore, the comparison of several recent related works on authentication is selected and discussed in Table I. These fourteen related works (2020-2024) on multi-factor authentication are selected in the domain of Internet-of-Medical-Things (IoMT) and Blockchain. The reviews are done based on the authors' contributions and summarize the findings reflecting on the extended authentication taxonomy (as depicted in Fig. 4). The results are shown in Table I.

V. DISCUSSION

As shown in Table I, there's a variety of authentication methods that have been employed in recent works, particularly within the context of IoMT. The first column insights are on the type of credentials. The most widely used model in the authentication process is often based on a combination of user identity, passwords, and biometric information. These multi-factor authentications are used in more recent studies to enhance security further. Using only single-factor authentication schemes, such as device information, pseudo-identities, and user tokens, is less prevalent compared to multi-factor authentication solutions. The reviewed articles show various lightweight authentication approaches suitable for resource-constrained environments such as in IoMT. Many of the proposed solutions emphasize a lightweight approach which is critical for resource-constraint IoMT devices. Common ones include the use of lightweight cryptographic algorithms (e.g., ECC), physically unclonable functions (PUFs), and lightweight cryptographic primitives such as XOR and hash functions are common. Most of the existing works implement a combination of asymmetric and symmetric cryptography, including Rivest–Shamir–Adleman (RSA) and secure hash algorithms. Some authentication solutions avoid the complexity of traditional certificate management by implementing implicit or soft certificates. Schemes like group authentication techniques based on Shamir's Secret Sharing (SSS) algorithm are employed to streamline the authentication process by reducing the number of blockchain transactions. Alsaeed et al. [53] proposed work, distribute authentication credentials among multiple entities (e.g., fog nodes), and combine only a sufficient number as part of streamlining the authentication process to help reduce computational and communication overhead. It can be seen that some of the recent works proposed lightweight authentication schemes, utilizing solutions such as PUFs or blockchain technology, but they often still encounter significant increases in their computational overhead and complexity. Apart from that, not every proposed work addressed scalability issues, particularly concerning the computational overhead, resource utilization, and the management of key and authentication data. This is highlighted in protocols that rely heavily on blockchain technology or those with complex key management schemes. The limitations of each reviewed work are also presented in Table I in the last column.

In addition to reviewing the recent related works on authentication (as shown in Table I), a conceptual validation of the established benchmarks in the field is conducted against the proposed lightweight authentication approach. This validation focuses on varying levels of computational efficiency, security robustness, and scalability comparing them against the proposed method to illustrate its advantages and identify potential areas for further improvement. The conceptual validation table is presented in Table II.

As depicted in Fig. 4, the Lightweight Efficiency approach is a crucial authentication solution in IoMT environments, where resource constraints such as limited processing power are common. The techniques reviewed in Table II demonstrate varying degrees of efficiency, from high efficiency to medium, and low efficiency. High efficiency depicts those methods that involve low computational overhead, fast processing times, or

require minimal resources, making them suitable for resource-constrained environments. Existing methods that struggle with low scalability suffer performance degradation as the system grows. Medium security is for methods that provide adequate security but may have vulnerabilities that could be exploited under certain conditions.

TABLE I. REVIEW OF RECENT RELATED WORKS

Related Works (2020-2024)	*Authentication Factors*	*Authentication Solutions*						*Limitations*
	Type of Credentials	*Lightweight Approach*		*Basic Authentication*	*Key Agreement*	*Cryptography-based*	*Certificate-based*	
		Streamline Authentication Process	*Aggregated Authentication*					
Edge IoMT authentication protocol [54]	Pseudo-identity	No	No, single process	Single-Factor	Symmetric	Asymmetric ECC	Certificate-less	Computational Overhead Potential to replay attacks.
A Lightweight and Robust Secure Key Establishment Protocol for Internet of Medical Things in COVID-19 Patients Care [18]	Password and Device Authentication	Yes, PUF-based, Simplified cryptography primitives	No, single process	Multi-Factor	Symmetric key generation	Hash function	No	The reliance on device authentication can be a limitation if the device is stolen, as it could potentially be used to gain unauthorized access. Potential to side-channel attacks.
A Lightweight and Secure Authentication Scheme for Remote Monitoring of Patients in IoMT [34]	User Identity, Biometrics, and Password	Yes, lightweight	No, single process	Multi-Factor	Symmetric and Asymmetric key generation	Hash function XOR operation Asymmetric ECC	No	Implementation Complexity
A framework introduces a group authentication technique [53]	Authentication Token	Yes, non-interactive and efficient	Yes, group key agreement reduces blockchain transactions. Hierarchical aggregation	Single-Factor	Group Authentication - Shamir's Secret Sharing (SSS) algorithm.	Asymmetric ECC Hash function	No	Implementation Complexity Potential vulnerability if fog node is compromised
Design of a novel lightweight and fast membership authenticated group key agreement scheme for resource-constrained IoMT devices [55]	User Tokens	Yes, non-interactive and efficient, Simplified cryptography primitives	Yes, group key agreement combines processes	Single-Factor	Binary symmetric polynomials	XOR operation		Potential vulnerability if the Membership Registration Center (MRC) is compromised Complexity issues
Development of a Privacy-Protection Authentication Management Protocol [56]	User ID, Password and Biometric Information	No	No	Multi-Factor	Symmetric	Hash function	No	Computational overhead. Scalability concern.
Proposal of a lightweight anonymous authentication scheme based on consortium blockchain in the IoMT [57]	User ID, Password and Biometric	Yes, lightweight	No	Multi-Factor	Pre-shared, Symmetric	Hash function XOR operation	Soft and Implicit Certificates	High complexity High computational cost

Develop a blockchain-based security system with light cryptography for user authentication security [58]	Biometric, Password	Yes, lightweight Biometric-based authentication	No	Multi-Factor		Symmetric Secure hash algorithm 256 (SHA-256) Shift-AES	Soft and Implicit Certificates	Computational overhead Scalability challenges The scheme may be prone to Biometric Spoofing.
Develop a framework that utilizes fog node computing in a Blockchain-based IoMT framework [59]	Device Information	Yes	No	Multi-Factor	Elliptic Curve, Digital Signature Algorithm (ECDSA), Diffie-Helman	Hash function	-	Complexity in implementation due to the integration of multiple technologies.
Design a blockchain-based secure authentication system to safeguard Electronic Health Record (EHR) data transferred over open channels. [60]	User identity, password, and Biometric information	Yes, lightweight Biometric-based authentication	No	Multi-Factor	RSA	Hash function XOR operation Symmetric AES	-	High computational cost due to the complexity of the authentication process
A novel approach to authentication using mobile agents, elliptic curve cryptography, and a challenge/response mechanism in IoT-based healthcare systems [61]	Challenge/ response system with a secret commitment key	Yes	No	Multi-Factor	Public-key cryptography	Hash function XOR operation Asymmetric ECC	No	High complexity Adoption and integration challenges
A lightweight authentication protocol that uses Physically Unclonable Functions (PUFs) to establish a connection between a fog node and a smart device [62]	Physical Unclonable Function (PUFs)	Yes, Implicit Certificates	No	Multi-Factor	Asymmetric	Asymmetric ECC	Implicit Certificates	High complexity Key management issues
An improved three-factor-based data authentication scheme (TDTAS) [63]	Smart card, password, and biometric information	Yes, lightweight	No (Individual Device Authentication)	Multi-Factor	Asymmetric	Asymmetric ECC Hash function XOR	Hard Certificates	High computational cost Scalability concern Lack of formal verification
A novel blockchain-based authentication and key agreement protocol tailored for secure health data sharing within a cooperative hospital network [64]	User's identity, password, and Biometric information	Yes	No	Multi-Factor	Asymmetric	Asymmetric ECC	Soft and Implicit Certificates	Scalability issues Computational overhead for storing and managing authentication data

TABLE II. CONCEPTUAL EVALUATION OF THE RECENT RELATED WORKS

Ref	Efficiency	Security Robustness	Scalability	Advantages
[54]	High	Medium	Medium	Efficient ECC-based authentication. Privacy-preserving pseudo-identity. Robust against various attacks.
[18]	High	Medium	Medium	Efficient authentication using lightweight cryptography. Anonymity and privacy preservation. Robust against replay, MITM, and impersonation attacks.
[34]	High	High	Medium	Lightweight authentication for remote monitoring applications.
[53]	Medium	Medium	High	Supports scalability by allowing many devices to be authenticated efficiently.

[55]	High	Medium	High	Lightweight and efficient with XOR operations. Scalable. Robust against various attacks with forward and backward secrecy.
[56]	Medium	High	Medium	Strong security with blockchain and Chebyshev chaotic maps. Privacy-preserving with user anonymity. Comprehensive security analysis.
[57]	Medium	High	Medium	Lightweight and efficient with XOR and hash functions. Scalable with consortium blockchain. Robust security with anonymity protection.
[58]	Medium	Medium	Medium	Lightweight cryptography with Shift-AES. Achieve security with blockchain integration. Privacy-preserving hybrid authentication
[59]	Medium	High	Medium	Comparable efficiency with fog computing. Scalability through decentralized blockchain and IPFS. Robust security with ECDSA
[60]	Medium	High	Medium	Robust security with blockchain and RSA. Comprehensive security analysis. Privacy-preserving multi-factor authentication.
[61]	Medium	High	Medium	Robust security with ECC and blockchain. Distributed processing with mobile agents. Anonymity and privacy preservation.
[62]	Medium	High	Medium	Ensure user anonymity, cross-fog authentication, and efficiency without the need for a trusted third party.
[63]	Medium	Medium	Medium	Strong security with ECC and multi-factor authentication. User anonymity and privacy protection
[64]	Medium	Medium	Low	Removed the dependency on centralized storage.

From Table II, the work of Soleymani et al. [54] demonstrates high efficiency, and medium scalability with modest security, making it suitable for moderately sized IoMT deployments where security is paramount, but resource constraints are a high priority. They prioritize efficiency using pseudo-identities. However, this approach often involves trade-offs, such as reduced security robustness in the case of replay attacks where an attacker captures and reuses valid pseudo-identity credentials to gain unauthorized access in subsequent sessions. The proposed method builds on these approaches by employing streamlined cryptographic operations that incorporate nonces or timestamps into the authentication process to ensure that each session or transaction is unique minimizing computational overhead while maintaining a robust security profile, thus offering a more balanced solution for resource-constrained environments.

For many existing IoMT authentication techniques, scalability continues to be a major concern. The work of [53] and [55] demonstrates high scalability as both introduce group-based authentication to reduce the computational load associated with individual transactions. While these solutions attempt to streamline the authentication process via blockchain and group key agreement, they often introduce additional complexity. The proposed approach, however, addresses these challenges by aggregating authentication tasks using only simple cryptographic operations such as XOR or lightweight block ciphers and consolidating multiple authentication steps into fewer, more efficient processes. This approach reduces the complexity typically associated with group authentication and ensures that the proposed method can effectively scale to meet the increasing demands of IoMT networks without compromising performance or security.

Furthermore, the work of Rani and Tripathi [64] in their “Blockchain-Based Authentication and Key Agreement Protocol" is evaluated as having medium efficiency and security with low scalability indicating its limitations in large-scale, resource-constrained IoMT networks. The proposed method addresses these challenges by enhancing efficiency through lightweight cryptographic algorithms that are specifically designed for resource-constraint environments such as ECC over RSA for key agreement, using batch processing to authenticate multiple individual transactions, and minimizing the overall computational and communication overhead. This comparative evaluation helps to identify which methods are best suited for specific applications, particularly where the balance between these critical factors is essential.

In summary, the following research gaps are identified:

- Complexity and scalability of lightweight authentication algorithms. Despite several efforts made by the researchers to streamline authentication processes, many lightweight protocols still exhibit significant complexity. This complexity results from an authentication scheme that involves multiple steps, interactions, credentials (biometric authentication, fuzzy extraction, blockchain), key management, and integration of advanced cryptographic algorithms, which could limit the efficiency and scalability that is needed for a bigger-scale implementation in the IoMT ecosystem.
- Insufficient focus on Aggregated Authentication. Although some works streamline the authentication process by introducing aggregated authentication approaches such as group key agreements, these solutions are not widely adopted. Moreover, they introduce new vulnerabilities, particularly if key components are compromised. It is important to develop a robust security mechanism for aggregated authentication systems.

Thus, based on the identified research gaps in recent related works, this study intends to make recommendations for future work that addresses these issues. Consequently, the following are recommended for this study's future efforts.

- To design simplified cryptographic protocols that are lightweight and scalable. This involves streamlining cryptographic operations within the authentication processes with efficient key management techniques while maintaining robust security. For instance, further

exploration of lightweight block ciphers offering lower computational overhead could be beneficial.

- One of the key challenges in IoMT is minimizing the communication overhead associated with the authentication process. Future work should explore methods for reducing the number of authentication messages exchanged between entities, thereby creating a low communication overhead. This could involve designing protocols that aggregate or combine authentication messages without compromising security. Streamlining the communication flow is essential for ensuring the efficiency and scalability of authentication protocols in large, distributed IoMT networks.
- To develop adaptive group aggregation authentication techniques to address scalability issues in large-scale IoMT setups. Innovations in group key management might include the use of hierarchical key distribution schemes or dynamic rekeying methods that can adjust key distribution processes based on network demands and usage patterns, thereby enhancing both security and scalability.
- To identify necessary requirements for designing authentication algorithms that can mitigate majority security attacks in communication. This includes addressing vulnerabilities that arise from multi-factor authentication, group authentication schemes, and the use of lightweight cryptographic primitives. Future efforts should aim to create comprehensive security frameworks that can pre-emptively address potential attack vectors, ensuring that authentication protocols remain secure as they scale.
- To implement the proposed authentication algorithms and evaluate their performance to ensure robustness.

VI. Conclusion

The current research trend in lightweight cryptography and authentication algorithms is developed to provide a secure, efficient, and scalable solution tailored to the unique requirements of the IoMT devices and infrastructure. However, striking a balance between robust security and efficient authentication performance is a significant challenge. Although a great deal of research has been done to guarantee high security in various IoT applications, potential adversary attacks are still valid and exist in our modern days. Hence, the exploration of lightweight authentication methods, decentralized authentication models, and advanced cryptographic techniques is the future research direction in the field of authentication mechanisms in IoMT.

This paper provides a comprehensive analysis of lightweight authentication methods within the context of IoMT. Therefore, two contributions have been proposed in this study, and they are:

- The study contributes to a thorough review of existing works on lightweight efficiency approach, focusing on streamlined authentication processes and aggregated authentication protocols with other current authentication solutions proposed by the authors. This review also highlighted the limitations of each approach and suggested the current state of lightweight multi-factor authentication approaches that can be used as a basis or guidance for future efforts to develop a more robust, secure, and scalable authentication protocol.
- This study also introduces the development of extended taxonomy (LAAS) for lightweight authentication protocols, emphasizing streamlining the authentication process and managing aggregated authentications. This taxonomy hopes to promote consistency against different authentication studies and contributes to the knowledge base for future researchers to develop more secure and efficient authentication mechanisms for IoMT and other similar environments.

As a conclusion, this study investigates the essential role of lightweight authentication in IoMT (Internet of Medical Things). Ensuring a secure and efficient authentication mechanism is vital to any healthcare system from malicious threats that can compromise sensitive medical data. The study analyzes recent related works on how lightweight approaches, particularly in streamlining authentication processes using various approaches such as multi-factor authentication, and other authentication solutions to identify its significant research gaps such as high computational cost, high complexity, and security vulnerabilities. It suggests that the authentication field requires further exploration to achieve more lightweight, secure, and scalable solutions. The proposed work suggests enhancing these authentication protocols through a streamlined authentication process using a more simplified cryptographic operation, multi-factor authentication, adaptive group management, and a secure encryption technique usage with the integration of advanced technologies like blockchain or AI. In a nutshell, this comprehensive review necessitates continuity for future development and innovations to safeguard the confidentiality, integrity, and functionality of the IoMT system.

Acknowledgment

The researchers express their gratitude to Universiti Teknologi Malaysia and Universiti Malaysia Sarawak for their valuable support in this study.

References

[1] N. Li et al., “A review of security issues and solutions for precision health in Internet-of-Medical-Things systems,” Secur. Saf., vol. 2, p. 2022010, 2023, doi: 10.1051/sands/2022010.

[2] S. Radack, “ELECTRONIC AUTHENTICATION: GUIDANCE FOR SELECTING SECURE TECHNIQUES,” 2004.

[3] N. Alsaeed and F. Nadeem, “Authentication in the Internet of Medical Things: Taxonomy, Review, and Open Issues,” Appl. Sci., vol. 12, no. 15, p. 7487, Jul. 2022, doi: 10.3390/app12157487.

[4] M. A. Khan, I. U. Din, T. Majali, and B.-S. Kim, “A Survey of Authentication in Internet of Things-Enabled Healthcare Systems,” Sensors, vol. 22, no. 23, p. 9089, Nov. 2022, doi: 10.3390/s22239089.

[5] A. H. Mohd Aman, W. H. Hassan, S. Sameen, Z. S. Attarbashi, M. Alizadeh, and L. A. Latiff, “IoMT amid COVID-19 pandemic: Application, architecture, technology, and security,” J. Netw. Comput. Appl., vol. 174, p. 102886, Jan. 2021, doi: 10.1016/j.jnca.2020.102886.

[6] Y. Sun, F. P.-W. Lo, and B. Lo, “Security and Privacy for the Internet of Medical Things Enabled Healthcare Systems: A Survey,” IEEE Access, vol. 7, pp. 183339–183355, 2019, doi: 10.1109/ACCESS.2019.2960617.

[7] A. Kogetsu, S. Ogishima, and K. Kato, "Authentication of Patients and Participants in Health Information Exchange and Consent for Medical Research: A Key Step for Privacy Protection, Respect for Autonomy, and Trustworthiness," Front. Genet., vol. 9, p. 167, Jun. 2018, doi: 10.3389/fgene.2018.00167.

[8] Anca D. Jurcut, Pasika Ranaweera, Lina Xu, "Chapter 2 Introduction to IoT Security," in IoT security: Advances in authentication, 2020, pp. 27–64. [Online]. Available: https://onlinelibrary-wiley-com.ezproxy.utm.my/doi/10.1002/9781119527978.ch2

[9] "IEEE Standards Association," IEEE Standards Association. Accessed: Jul. 19, 2024. [Online]. Available: https://standards.ieee.org

[10] N. Nanayakkara, M. N. Halgamuge, and A. Syed, "SECURITY AND PRIVACY OF INTERNET OF MEDICAL THINGS (IOMT) BASED HEALTHCARE APPLICATIONS: A REVIEW," 2019.

[11] F. Gu, J. Niu, L. Jiang, X. Liu, and M. Atiquzzaman, "Survey of the low power wide area network technologies," J. Netw. Comput. Appl., vol. 149, p. 102459, Jan. 2020, doi: 10.1016/j.jnca.2019.102459.

[12] W. E. Burr et al., "Electronic Authentication Guideline," National Institute of Standards and Technology, NIST SP 800-63-2, Nov. 2013. doi: 10.6028/NIST.SP.800-63-2.

[13] S. Das, S. Namasudra, S. Deb, P. M. Ger, and R. G. Crespo, "Securing IoT-Based Smart Healthcare Systems by Using Advanced Lightweight Privacy-Preserving Authentication Scheme," IEEE Internet Things J., vol. 10, no. 21, pp. 18486–18494, 2023, doi: 10.1109/JIOT.2023.3283347.

[14] O. B. J. Rabie, S. Selvarajan, T. Hasanin, G. B. Mohammed, A. M. Alshareef, and M. Uddin, "A full privacy-preserving distributed batch-based certificate-less aggregate signature authentication scheme for healthcare wearable wireless medical sensor networks (HWMSNs)," Int. J. Inf. Secur., vol. 23, no. 1, pp. 51–80, Feb. 2024, doi: 10.1007/s10207-023-00748-1.

[15] M. R. Servati and M. Safkhani, "ECCbAS: An ECC based authentication scheme for healthcare IoT systems," Pervasive Mob. Comput., vol. 90, 2023, doi: 10.1016/j.pmcj.2023.101753.

[16] S. Yu and Y. Park, "A Robust Authentication Protocol for Wireless Medical Sensor Networks Using Blockchain and Physically Unclonable Functions," IEEE Internet Things J., vol. 9, no. 20, pp. 20214–20228, 2022, doi: 10.1109/JIOT.2022.3171791.

[17] S. Yu and K. Park, "SALS-TMIS: Secure, Anonymous, and Lightweight Privacy-Preserving Scheme for IoMT-Enabled TMIS Environments," IEEE Access, vol. 10, pp. 60534–60549, 2022, doi: 10.1109/ACCESS.2022.3181182.

[18] M. Masud et al., "A Lightweight and Robust Secure Key Establishment Protocol for Internet of Medical Things in COVID-19 Patients Care," IEEE Internet Things J., vol. 8, no. 21, pp. 15694–15703, 2021, doi: 10.1109/JIOT.2020.3047662.

[19] Y. Zhang, B. Li, J. Wu, B. Liu, R. Chen, and J. Chang, "Efficient and Privacy-Preserving Blockchain-Based Multifactor Device Authentication Protocol for Cross-Domain IIoT," IEEE Internet Things J., vol. 9, no. 22, pp. 22501–22515, Nov. 2022, doi: 10.1109/JIOT.2022.3176192.

[20] V. S. Naresh, S. Reddi, and V. D. Allavarpu, "Lightweight secure communication system based on Message Queuing Transport Telemetry protocol for e - healthcare environments," Int. J. Commun. Syst., vol. 34, no. 11, p. e4842, 2021.

[21] S. Sallam and B. D. Beheshti, "A Survey on Lightweight Cryptographic Algorithms," in TENCON 2018 - 2018 IEEE Region 10 Conference, Oct. 2018, pp. 1784–1789. doi: 10.1109/TENCON.2018.8650352.

[22] S. Agrawal and P. Ahlawat, "A Survey on the Authentication Techniques in Internet of Things," in 2020 IEEE International Students' Conference on Electrical,Electronics and Computer Science (SCEECS), Bhopal, India: IEEE, Feb. 2020, pp. 1–5. doi: 10.1109/SCEECS48394.2020.86.

[23] A. S. Khan et al., "Blockchain-Based Lightweight Multifactor Authentication for Cell-Free in Ultra-Dense 6G-Based (6-CMAS) Cellular Network," IEEE Access, vol. 11, pp. 20524–20541, 2023, doi: 10.1109/ACCESS.2023.3249969.

[24] M. A. U. Gumis et al., "Biometric Blockchain-based Multifactor Privacy Perserving Authentication Scheme for VANETs," J. IT Asia, vol. 9, no. 1, pp. 97–107, Nov. 2021, doi: 10.33736/jita.3851.2021.

[25] M. Fakroon, F. Gebali, and M. Mamun, "Multifactor authentication scheme using physically unclonable functions," Internet Things, vol. 13, p. 100343, Mar. 2021, doi: 10.1016/j.iot.2020.100343.

[26] A. Ometov, S. Bezzateev, N. Mäkitalo, S. Andreev, T. Mikkonen, and Y. Koucheryavy, "Multi-Factor Authentication: A Survey," Cryptography, vol. 2, no. 1, p. 1, Jan. 2018, doi: 10.3390/cryptography2010001.

[27] A. K. Singh and A. Garg, "Authentication protocols for securing IoMT: current state and technological advancements," in Securing Next-Generation Connected Healthcare Systems, Elsevier, 2024, pp. 1–29. doi: 10.1016/B978-0-443-13951-2.00004-0.

[28] S. Windarta, S. Suryadi, K. Ramli, B. Pranggono, and T. S. Gunawan, "Lightweight Cryptographic Hash Functions: Design Trends, Comparative Study, and Future Directions," IEEE Access, vol. 10, pp. 82272–82294, 2022, doi: 10.1109/ACCESS.2022.3195572.

[29] U. Chatterjee, S. Ray, S. Adhikari, M. K. Khan, and M. Dasgupta, "An improved authentication and key management scheme in context of IoT-based wireless sensor network using ECC," Comput. Commun., vol. 209, pp. 47–62, Sep. 2023, doi: 10.1016/j.comcom.2023.06.017.

[30] N. K. Ratha, J. H. Connell, and R. M. Bolle, "Enhancing Security and Privacy in Biometrics-Based Authentication Systems," Ibm Syst. J., vol. 40, no. 3, pp. 614–634, 2001, doi: 10.1147/sj.403.0614.

[31] M. El-hajj, A. Fadlallah, M. Chamoun, and A. Serhrouchni, "A Survey of Internet of Things (IoT) Authentication Schemes," Sensors, vol. 19, no. 5, p. 1141, Mar. 2019, doi: 10.3390/s19051141.

[32] X. Li, J. Niu, M. Karuppiah, S. Kumari, and F. Wu, "Secure and Efficient Two-Factor User Authentication Scheme with User Anonymity for Network Based E-Health Care Applications," J. Med. Syst., vol. 40, no. 12, p. 268, Dec. 2016, doi: 10.1007/s10916-016-0629-8.

[33] I. Velásquez, "Framework for the Comparison and Selection of Schemes for Multi-Factor Authentication," CLEI Electron. J., vol. 24, no. 1, Apr. 2021, doi: 10.19153/cleiej.24.1.9.

[34] Z. Ali, S. Mahmood, K. Mansoor Ul Hassan, A. Daud, R. Alharbey, and A. Bukhari, "A Lightweight and Secure Authentication Scheme for Remote Monitoring of Patients in IoMT," IEEE Access, vol. 12, pp. 73004–73020, 2024, doi: 10.1109/ACCESS.2024.3400400.

[35] M. Samal, S. Ray, and M. Dasgupta, "A Short Survey of Authentication Protocols in context of Internet of Things," in 2023 IEEE 7th Conference on Information and Communication Technology (CICT), Jabalpur, India: IEEE, Dec. 2023, pp. 1–6. doi: 10.1109/CICT59886.2023.10455531.

[36] K. Kim, J. Ryu, Y. Lee, and D. Won, "An Improved Lightweight User Authentication Scheme for the Internet of Medical Things," Sensors, vol. 23, no. 3, p. 1122, Jan. 2023, doi: 10.3390/s23031122.

[37] M. Masud, G. S. Gaba, K. Choudhary, M. S. Hossain, M. F. Alhamid, and G. Muhammad, "Lightweight and Anonymity-Preserving User Authentication Scheme for IoT-Based Healthcare," IEEE Internet Things J., vol. 9, no. 4, pp. 2649–2656, 2022, doi: 10.1109/JIOT.2021.3080461.

[38] F. Rafique, M. S. Obaidat, K. Mahmood, M. F. Ayub, J. Ferzund, and S. A. Chaudhry, "An Efficient and Provably Secure Certificateless Protocol for Industrial Internet of Things," IEEE Trans. Ind. Inform., vol. 18, no. 11, pp. 8039–8046, Nov. 2022, doi: 10.1109/TII.2022.3156629.

[39] A. Gupta, M. Tripathi, S. Muhuri, G. Singal, and N. Kumar, "A secure and lightweight anonymous mutual authentication scheme for wearable devices in Medical Internet of Things," J. Inf. Secur. Appl., vol. 68, 2022, doi: 10.1016/j.jisa.2022.103259.

[40] J. Chang, Q. Ren, Y. Ji, M. Xu, and R. Xue, "Secure medical data management with privacy-preservation and authentication properties in smart healthcare system," Comput. Netw., vol. 212, 2022, doi: 10.1016/j.comnet.2022.109013.

[41] L. Xue, Q. Huang, S. Zhang, H. Huang, and W. Wang, "A Lightweight Three-Factor Authentication and Key Agreement Scheme for Multigateway WSNs in IoT," Secur. Commun. Netw., vol. 2021, pp. 1–15, Jun. 2021, doi: 10.1155/2021/3300769.

[42] S. S. Vankayalapati, S. Mookherji, and V. Odelu, "A Security Enhanced Authentication Protocol." 2024. doi: 10.1109/icicv62344.2024.00129.

[43] P. Roychoudhury, B. Roychoudhury, and D. Kr. Saikia, "Hierarchical Group Based Mutual Authentication and Key Agreement for Machine Type Communication in LTE and Future 5G Networks," Secur. Commun. Netw., vol. 2017, pp. 1–21, 2017, doi: 10.1155/2017/1701243.

[44] M. Usman, M. A. Jan, and D. Puthal, "PAAL: A Framework Based on Authentication, Aggregation, and Local Differential Privacy for Internet of Multimedia Things," IEEE Internet Things J., vol. 7, no. 4, pp. 2501–2508, Apr. 2020, doi: 10.1109/JIOT.2019.2936512.

[45] A. S. Khan et al., "Lightweight Multifactor Authentication Scheme for NextGen Cellular Networks," IEEE Access, vol. 10, pp. 31273–31288, 2022, doi: 10.1109/ACCESS.2022.3159686.

[46] S. Atiewi et al., "Scalable and Secure Big Data IoT System Based on Multifactor Authentication and Lightweight Cryptography," IEEE Access, vol. 8, pp. 113498–113511, 2020, doi: 10.1109/ACCESS.2020.3002815.

[47] M. N. Aman, M. H. Basheer, and B. Sikdar, "A Lightweight Protocol for Secure Data Provenance in the Internet of Things Using Wireless Fingerprints," IEEE Syst. J., vol. 15, no. 2, pp. 2948–2958, Jun. 2021, doi: 10.1109/JSYST.2020.3000269.

[48] P. Gope, Y. Gheraibia, S. Kabir, and B. Sikdar, "A Secure IoT-Based Modern Healthcare System With Fault-Tolerant Decision Making Process," IEEE J. Biomed. Health Inform., vol. 25, no. 3, pp. 862–873, Mar. 2021, doi: 10.1109/JBHI.2020.3007488.

[49] M. Malik, Kamaldeep, M. Dutta, and J. Granjal, "L-ECQV: Lightweight ECQV Implicit Certificates for Authentication in the Internet of Things," IEEE Access, vol. 11, pp. 35517–35540, 2023, doi: 10.1109/ACCESS.2023.3261666.

[50] M. Ebrahimabadi, M. Younis, and N. Karimi, "A PUF-Based Modeling-Attack Resilient Authentication Protocol for IoT Devices," IEEE Internet Things J., vol. 9, no. 5, pp. 3684–3703, Mar. 2022, doi: 10.1109/JIOT.2021.3098496.

[51] J. Ambareen and P. M, "Secured Wireless Sensor Network Protocol using Rabin-assisted Multifactor Authentication," Int. J. Comput. Netw. Inf. Secur., vol. 14, no. 4, pp. 60–74, Aug. 2022, doi: 10.5815/ijcnis.2022.04.05.

[52] F. Wu, X. Li, L. Xu, P. Vijayakumar, and N. Kumar, "A Novel Three-Factor Authentication Protocol for Wireless Sensor Networks With IoT Notion," IEEE Syst. J., vol. 15, no. 1, pp. 1120–1129, Mar. 2021, doi: 10.1109/JSYST.2020.2981049.

[53] N. Alsaeed, F. Nadeem, and F. Albalwy, "A scalable and lightweight group authentication framework for Internet of Medical Things using integrated blockchain and fog computing," Future Gener. Comput. Syst., vol. 151, pp. 162–181, 2024.

[54] S. A. Soleymani, S. Goudarzi, M. H. Anisi, A. Jindal, N. Kama, and S. A. Ismail, "A Privacy-Preserving Authentication Scheme for Real-Time Medical Monitoring Systems," IEEE J. Biomed. Health Inform., vol. 27, no. 5, pp. 2314–2322, May 2023, doi: 10.1109/JBHI.2022.3143207.

[55] C. Hsu, L. Harn, Z. Xia, Z. Zhao, and H. Xu, "Fast and Lightweight Authenticated Group Key Agreement Realizing Privacy Protection for Resource-Constrained IoMT," Wirel. Pers. Commun., vol. 129, no. 4, pp. 2403–2417, 2023, doi: 10.1007/s11277-023-10239-0.

[56] J. Miao, Z. Wang, Z. Wu, X. Ning, and P. Tiwari, "A blockchain-enabled privacy-preserving authentication management protocol for Internet of Medical Things," Expert Syst. Appl., vol. 237, p. 121329, Mar. 2024, doi: 10.1016/j.eswa.2023.121329.

[57] S. Wu, A. Zhang, J. Chen, G. Peng, and Y. Gao, "A Blockchain-Assisted Lightweight Anonymous Authentication Scheme for Medical Services in Internet of Medical Things," Wirel. Pers. Commun., vol. 131, no. 2, pp. 855–876, 2023, doi: 10.1007/s11277-023-10457-6.

[58] I. Hagui, A. Msolli, A. Helali, and F. Hassen, "Based blockchain-lightweight cryptography techniques for security information: A verification secure system for user authentication," presented at the 2021 International Conference on Control, Automation and Diagnosis, ICCAD 2021, 2021. doi: 10.1109/ICCAD52417.2021.9638751.

[59] S. R. Mallick, R. K. Lenka, P. K. Tripathy, D. C. Rao, S. Sharma, and N. K. Ray, "A Lightweight, Secure, and Scalable Blockchain-Fog-IoMT Healthcare Framework with IPFS Data Storage for Healthcare 4.0," SN Comput. Sci., vol. 5, no. 1, p. 198, Jan. 2024, doi: 10.1007/s42979-023-02511-8.

[60] V. Kumar, R. Ali, and P. K. Sharma, "A secure blockchain-assisted authentication framework for electronic health records," Int. J. Inf. Technol., Feb. 2024, doi: 10.1007/s41870-023-01705-w.

[61] H. Idrissi and P. Palmieri, "Agent-based blockchain model for robust authentication and authorization in IoT-based healthcare systems," J. Supercomput., Oct. 2023, doi: 10.1007/s11227-023-05649-7.

[62] X. Jia, M. Luo, H. Wang, J. Shen, and D. He, "A Blockchain-Assisted Privacy-Aware Authentication Scheme for Internet of Medical Things," IEEE Internet Things J., vol. 9, no. 21, pp. 21838–21850, Nov. 2022, doi: 10.1109/JIOT.2022.3181609.

[63] S. S. Sahoo, S. Mohanty, K. S. Sahoo, M. Daneshmand, and A. H. Gandomi, "A Three-Factor-Based Authentication Scheme of 5G Wireless Sensor Networks for IoT System," IEEE Internet Things J., vol. 10, no. 17, pp. 15087–15099, Sep. 2023, doi: 10.1109/JIOT.2023.3264565.

[64] D. Rani and S. Tripathi, "Design of blockchain-based authentication and key agreement protocol for health data sharing in cooperative hospital network," J. Supercomput., vol. 80, no. 2, pp. 2681–2717, Jan. 2024, doi: 10.1007/s11227-023-05577-6.

Dynamic Simulation and Forecasting of Spatial Expansion in Small and Medium-Sized Cities Using ANN-CA-Markov Models

Chengquan Gao
School of Architecture and Urban Planning, Henan University of Urban Construction, Ping'dingshan 467001, China

***Abstract*—This study utilizes the ANN-CA-Markov (Artificial Neural Network-Cellular Automata-Markov) model to address spatial planning and expansion challenges in China's small and medium-sized cities. With China's urbanization rate reaching 59.58% in 2018 and expected to hit 70% by 2030, the country is entering a mid-stage of urbanization, leading to rapid expansion of megacities and a gradual decline in smaller cities. The study aims to dynamically simulate urban spatiotemporal evolution and predict future land use changes, integrating land use data, DEM elevation, transportation, administrative centers, and ecological information. The model forecasts the ecological spatial layout of Wanzhou District by 2025, with results indicating a slight decrease in ecological space and an increase in construction land. This suggests a need to balance urban development with ecological sustainability amidst rapid urbanization. The study demonstrates the high accuracy of the ANN-CA-Markov model in predicting land use changes and provides valuable insights for urban planners in making informed land use decisions.**

***Keywords*—ANN-CA; Markov; small and medium-sized cities; spatial; planning**

I. Introduction

In 2018, China's urbanization rate reached 59.58%, and it is expected to hit 70% by 2030, marking the country's entry into the middle stage of urbanization [1-3]. Since 2011, China has been systematically revising its traditional urbanization approach, leading to the rapid expansion of megacities and the gradual decline of many small and medium-sized cities and towns. With the central government's strategic decisions to "establish a scientifically and reasonably structured urban pattern, where large, medium, and small cities and towns, as well as city clusters, are well-organized," new opportunities and broad prospects have emerged for the development of small and medium-sized towns. The healthy urbanization of these smaller cities requires more scientific planning and development of urban land resources by administrators [4, 5].

Scientifically defining urban development boundaries not only enables the intensive use of spatial resources and controls the disorderly sprawl of cities but also supports the sustainable socio-economic development of cities while protecting the natural ecological environment [6-9]. However, one of the major challenges in urban planning is accurately predicting the spatial and temporal evolution of land use, particularly in small and medium-sized cities where data may be less comprehensive, and urbanization patterns are more complex.

In recent years, numerous scholars have utilized the ANN-CA (artificial neural network-cellular automata) coupled model to dynamically simulate the temporal and spatial evolution of cities and predict future land use changes, assisting in urban land use simulation, urban development boundaries, and ecological redline protection in urban planning, achieving many meaningful results [10-12]. For instance, Xu et al. [13] integrated Artificial Neural Networks (ANN), Cellular Automata (CA), and Markov Chain (MC) to simulate urban expansion in rapidly urbanizing areas, revealing the nonlinear relationship between the expansion process and its drivers. Similarly, Zhao et al. [14] studied land-use changes in Yucheng District, Ya'an City, China using an ANN-CA model, emphasizing the improvement of simulation accuracy through appropriate thresholds and random variable parameters. Additionally, Asanza et al. [15] explored the integration of ANN and CA models for spatial-temporal load forecasting, highlighting the enhanced forecasting accuracy through temporality and geospatial data analytics.

Despite these advances, challenges remain in effectively modeling and predicting land use changes in regions where data availability is limited or where the urbanization process involves complex interactions among multiple factors. The current study addresses these challenges by developing an enhanced ANN-CA-Markov coupled model that incorporates more comprehensive datasets and refined transition rules to achieve higher simulation precision. Specifically, this study leverages Wanzhou District's land use cover data, DEM elevation data, road traffic data, administrative center data, river data, ecological protection redline data, and natural conservation area data from 2000, 2006, 2012, and 2018. By integrating these data with the ANN-CA-Markov coupled model, we aim to predict the ecological spatial layout of Wanzhou District by 2025 more accurately.

The novel contribution of this work lies in overcoming the difficulties associated with limited data availability and complex urbanization processes by enhancing the precision of the ANN-CA-Markov model. This is achieved through the integration of additional data sources and the refinement of transition rules, offering a more comprehensive approach to modeling and predicting spatial changes. This study, therefore, represents a significant advancement in the field, particularly in the context of small and medium-sized cities where such challenges are most pronounced.

II. METHOD

A. ANN-CA Model

The ANN-CA model, short for Artificial Neural Network - Cellular Automata model, is a discrete time, space, and state grid dynamic model where spatial interactions and temporal causal relations are local. It enables the bottom-up simulation of the spatiotemporal evolution of complex systems. The state of each cell is determined by the states of its neighboring cells, and upon establishing transition rules, all cells can evolve autonomously following these rules, highlighting the core essence of transition rules. Artificial neural networks possess self-learning and associative capabilities, allowing for the rapid identification of optimized solutions [16-18]. By learning the rules of land use data changes through the neural network model and applying these extracted rules to the grid data of the starting year, simulation predictions can be completed within the cellular automata. The core principle involves training neurons with land use data from different periods, then determining the transition probabilities for each land use type based on the characteristics of influencing factors, culminating in the simulated prediction of land use planning. To ensure the model's accuracy, the input parameters were rigorously selected and optimized. The primary inputs include land use types, digital elevation model (DEM), neighborhood development density, and transition suitability. These parameters were chosen based on the terrain characteristics of the study area, the diversity of land use, and the complexity of its spatial distribution.

Extensive experiments were conducted to evaluate the impact of different parameter combinations on the model's predictive outcomes. Specifically, different neighborhood window sizes, DEM resolutions, and land use classification standards were tested. Sensitivity analysis revealed that the size of the neighborhood window significantly affects the model's spatial resolution and computational efficiency, while variations in DEM resolution notably influence the prediction accuracy. Ultimately, a 5×5 Moore neighborhood window and a 30-meter resolution DEM were selected as they offered the best balance between computational efficiency and prediction accuracy.

Additionally, the selection of transition suitability parameters was explored. These parameters primarily represent the likelihood of transitions between different land use types. During the parameter adjustment process, a stochastic disturbance factor was introduced to simulate unforeseen changes in real-world conditions. The final selection of these parameters was made based on a comparative analysis of multiple simulation results, tailored to the specific conditions of the study area. The mathematical expression is as follows:

$$P(k,t,l) = \left(1 + (-\ln \gamma)^{\alpha}\right) \times P_{ann}(k,t,l) \times \Omega_k^t \times \cos\left(S_k^t\right) \quad (1)$$

In this expression, it mainly expresses the transition probability P of a cell k at time t to the l-th type of land use as a function of random factors, artificial neural network calculated probabilities, neighborhood development density, and transition suitability. $(-\ln \gamma)^{\alpha}$ represents the random factor; $P_{ann}(k,t,l)$ is the transition probability of a certain land type calculated by the trained artificial neural network; Ω_k^t represents the urban land density within the defined neighborhood window, i.e., the total number of urban land cells divided by the total number of grid cells in the neighborhood window; $\cos\left(S_k^t\right)$ represents the transition suitability between two land types, generally indicated by 0 or 1, mainly to signify whether a transition is possible.

B. Markov Model

The Markov model (Markov), based on a type of stochastic process is a mathematical method used to predict the prior probabilities and conditional probabilities of events [14-16]. Its changes over time are continuous, and when the process parameters take discrete time values, it is referred to as a Markov sequence. The primary feature of the Markov sequence is its Markov property, not time series property, meaning when the state of the process (system) at time t0 is known, the state of the process (system) at time t (t>t0) is independent of the state at time t0. The Markov model is an extremely important predictive model in geographic forecasting research. In constructing the Markov model, we meticulously adjusted the state transition probability matrix. The state transition probabilities primarily reflect the likelihood of transitions between different land use types within a specific time sequence. To ensure the accuracy of this probability matrix, we conducted several experiments to test the impact of different initial conditions and transition probabilities on the final predictive outcomes. Sensitivity analysis indicated that certain key parameters in the state transition probability matrix significantly influence the spatial distribution of the model's predictions. For instance, increasing the probability of converting arable land to construction land by 10% substantially increases the predicted area of construction land while decreasing the areas of forest land and water bodies. This highlights the need to adjust transition probabilities in accordance with real-world conditions. The Markov model mainly consists of states, state transition processes, state transition probabilities, and state transition matrices, detailed as follows:

1) State: Represents an outcome, indicating the result appearing at a specific point in time.

2) State transition process: The relationship between the state change of an event and time.

3) State transition probability: Refers to the likelihood of an event's state changing to another state, expressed mathematically as follows:

$$P_{ij} = P\left(E_i \to E_j\right) = P\left(E_j / E_i\right) \quad (2)$$

In the formula, P_{ij} and $P\left(E_i \to E_j\right)$ both represent the probability of the state transition of the event, $P\left(E_j / E_i\right)$ represents the conditional probability, and E_i and E_j respectively represent the state at moments i and j.

4) State transition probability matrix: If a specific event has n possible states, then the state transition probability from state Ei to state Ej is denoted as pij and expressed through a matrix as follows:

$$p = \begin{pmatrix} p_{11} & \cdots & P_1 n \\ \vdots & \ddots & \vdots \\ p_{n^1} & \cdots & P_n n \end{pmatrix} \left(0 \le p_{ij} < 1\right) \quad (3)$$

In the above expression, p represents the state transition probability matrix.

C. Construction of the ANN-CA-Markov Model

To simulate and predict the ecological space pattern of Wanzhou District, this study employs the CA-Markov model, which combines the transition prediction capability of the Markov model with the spatial distribution simulation of cellular automata [17-19]. Traditional CA-Markov models face challenges in handling nonlinear relationships, so an artificial neural network (ANN) model is introduced to learn and establish more accurate transition rules. The process is divided into two stages: training and simulation. The operational logic of the model is illustrated in Fig. 1.

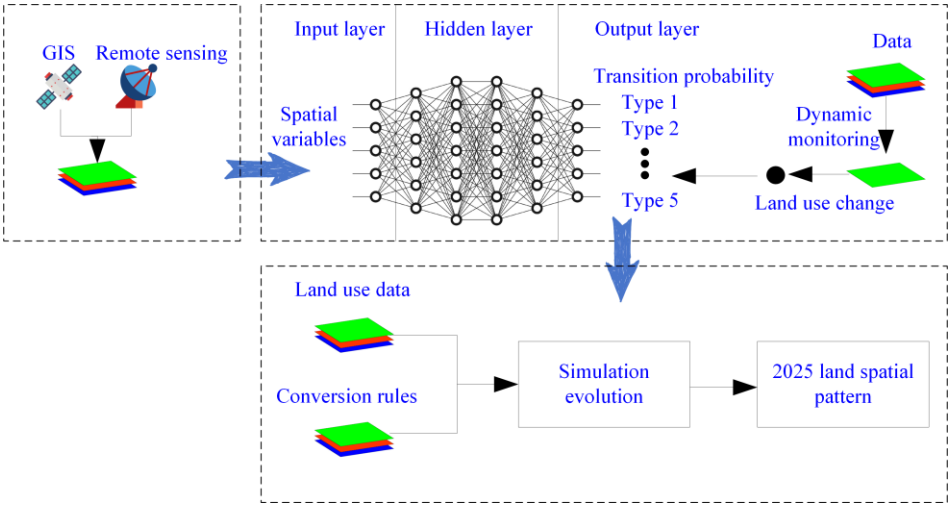


Fig. 1. Technical route map.

III. Research Area and Data Processing

A. Geographic Overview

Wanzhou District is located on the eastern edge of the Sichuan Basin, in the northeastern part of Chongqing. The Yangtze River flows into Wanzhou from the southwest, traverses northeastward, and then flows into Yunyang. Wanzhou District is situated between 107°55′22″E to 108°53′25″E and 30°24′00″N to 31°14′58″N. Lichuan and Shizhu are located to the south of Wanzhou District, Yunyang to the east, Liangping and Zhong County to the west, and Kaizhou District and Kaijiang to the north. The straight-line distance from Wanzhou District to Chongqing is 228 kilometers. Fig. 2 shows the location map of Wanzhou District.



Fig. 2. Location map of wanzhou district.

B. Data Sources

This study employed Landsat TM/ETM imagery for Wanzhou District from 2000, 2006, 2012, and 2018. The 2000, 2006, and 2012 data were sourced from Landsat-5 TM, while 2018 data came from Landsat-8, all acquired via the Geospatial Data Cloud (Table I). The imagery has a 30m×30m resolution, and the data were projected using the Albers projection. Additional data included 30m GDEM elevation data, OSM road, river, and administrative boundaries, and population data from the Chongqing Statistical Yearbook.

TABLE I. Classification Standards for Land Use Types in Wanzhou District

Land Category Code	Land Category Type	Detailed Types
1	Cultivated Land	Paddy fields, dry land
2	Forest Land	Forested land, shrub land, sparse forest, and other forest lands
3	Grassland	High, medium, and low coverage grasslands
4	Water Bodies	Rivers, lakes, reservoirs, ponds, tidal flats
5	Constructed Land	Urban areas, rural settlements, industrial and mining areas, transportation land
6	Unused Land	Bare land, sandy land, other types of unused land

C. Data Preprocessing

The study highlighted remote sensing image preprocessing importance using ERDAS 9.1 for band synthesis, geometric correction, and image cropping, enhancing data quality for

Wanzhou District. Band synthesis integrated different bands to improve image classification accuracy. Geometric corrections were applied using a third-order polynomial method to ensure pixel accuracy, while image cropping maintained research area consistency. The analysis involved visual and computer processing, adhering to "Current Land Use Classification" standards and incorporating field data and professional verification to achieve over 85% interpretation accuracy. These processed images provide a reliable basis for further analysis.

IV. Analysis of Ecological Space Change Characteristics

A. Dynamic Degree and Transfer Rate

In studying the characteristics of ecological space change, the dynamic degree and transfer rate are mainly used to examine the changes in the quantity of ecological spaces within the study area. The formula for calculating the transfer rate of ecological space types is as follows:

$$K = \frac{u_b - u_a}{u_a} \times \frac{1}{T} \times 100\% \tag{4}$$

K represents the transfer rate of a certain ecological space type; ua represents the quantity of a certain ecological space type at the beginning of the study; ub represents the quantity of the same ecological space type at the end of the study; T represents the time span of the study period. The formula for calculating the overall dynamic degree of ecological space in the study area is:

$$Lc = \left(\frac{\sum_{i=1}^{n} \Delta Lu_{i-j}}{2\sum_{i=1}^{n} Lu_i} \right) \times T^{-1} \times 100\% \tag{5}$$

Lui represents the quantity of the ith ecological space type at the beginning of the study; $\triangle Lu_{i-j}$ represents the absolute value of the quantity of the ith ecological space type transformed into the non-ith ecological space type during the study period; T represents the time span of the study period, Lc represents the comprehensive dynamic degree of ecological space.

B. Transition Matrix

In the land-use transition matrix, the rows and columns represent the ecological space types at the beginning and end of a certain time sequence unit, respectively. Therefore, using the transition matrix to study ecological space types can reveal the changes in the area of each ecological space type from the beginning to the end of the period, as well as the transition situations of ecological space types. The formula for calculating the transition matrix is as follows:

$$S_{ij} = \begin{bmatrix} S_{11} & \cdots & S_{1n} \\ \vdots & \ddots & \vdots \\ S_{n1} & \cdots & S_{nn} \end{bmatrix} \tag{6}$$

In the formula, S represents the area of ecological space types within the time sequence unit, n is the total number of ecological space types within the time sequence unit, i represents the index of ecological space types at the beginning of the time sequence unit; j represents the index of ecological space types at the end of the time sequence unit.

C. Landscape Pattern Indices

The spatial change characteristics of ecological spaces are mainly analyzed through the landscape pattern indices and classified landscape pattern indices of the study area. The analysis of these indices quantitatively describes the structure and distribution characteristics of the landscape itself. This paper selects the patch area, patch number, patch density, largest patch index, splitting index, and Shannon's diversity index to analyze the landscape pattern indices and classified landscape pattern indices, thereby analyzing the spatial change characteristics of ecological spaces.

1) Patch area (CA): The patch area (CA) represents the total area of a specific landscape type, and its calculation formula is as follows:

$$CA = \sum_{j=1}^{n} a_{ij} \times (1/10000) \tag{7}$$

In the formula, a_{ij} represents the area of patch i_j in a landscape type, n is the total number of all landscape types in the study area.

2) Patch Number (NP): The patch number (NP) reflects the spatial pattern of the landscape. The larger the value of NP, the higher the degree of fragmentation in space, and vice versa. The spatial distribution characteristics of various landscape types can be determined to some extent by the patch number (NP), and its calculation formula is as follows:

$$NP = m(m \geq 1) \tag{8}$$

In the formula, m represents the number of patches of a certain ecological space type.

3) Patch Density (PD): Represents the number of patches per unit area of the entire ecological space or a certain ecological space type. The larger the value of PD, the greater the separation among individual patches; conversely, the closer the patches are to each other. The calculation method is as follows:

$$PD = N / A \tag{9}$$

In the formula, N represents the number of patches of the ecological space type, A is the area of the study region.

4) Largest Patch Index (LPI): The largest patch index represents the ratio of the largest patch of an ecological space type to the total area of the study region. The larger the LPI value, the higher the connectivity among patches of ecological space types, and vice versa. The calculation method is as follows:

$$LPI = \frac{Maxa_{ij}}{A} \times 100 \tag{10}$$

In the formula, LPI represents the value of the largest patch index, a_{ij} is the area of patch i_j of the ecological space type, A is the total landscape area of the entire study region.

5) Splitting Index (SPLIT): Used to indicate the degree of fragmentation of a landscape type. The larger the value of SPLIT, the higher the degree of fragmentation of the landscape type, and vice versa. The calculation method is as follows:

$$SPLIT = \frac{A^2}{\sum_{i=1}^{m}\sum_{j=1}^{n} a_{ij}^2} \quad (11)$$

In the formula, SPLIT represents the value of the splitting index, a_{ij} is the area of patch i_j in a landscape type, A is the total landscape area of the study region.

6) Shannon's Diversity Index (SHDI): Indicates the degree of evenness among different ecological space types within the study area. The larger the value of SHDI, the more even the distribution of ecological space type patches, and vice versa. The calculation method is as follows:

$$SHDI = -\sum_{i=1}^{m} p_i \quad (12)$$

In the formula, m represents the number of ecological space types, p_i is the proportion of ecological space type i in the area of the study region.

D. Analysis of Ecological Space Structure Type Change Characteristics

The ecological space structure types in Wanzhou District mainly include cultivated land, forest land, grassland, water bodies, and unused land. Using the land use type data of Wanzhou District from 2000, 2006, 2012, and 2018, the software ArcGIS 10.1 was used to analyze the changes in the area of ecological spaces in Wanzhou District.

Using ArcGIS 10.1's mapping features, land use type maps of Wanzhou District from 2000 to 2018 were created, as shown in Fig. 3. Additionally, ArcGIS's statistical tools were employed to compile data on ecological spaces and land use areas in Wanzhou District from 2000 to 2018, as presented in Table II.

Table II highlights that Wanzhou District's ecological spaces are predominantly woodland (over 50%) and arable land (over 44%), with water bodies and grassland as secondary types. Unused land is minimal. Transition matrices for 2000–2006, 2006–2012, 2012–2018, and 2000–2018, generated using ArcGIS and Excel, are shown in Tables III to VI.

As shown in Table III, during 2000–2006, the main ecological space type conversions were from arable land and woodland to water bodies and built-up land, primarily due to the water storage of the Three Gorges Reservoir Area and rapid urbanization.

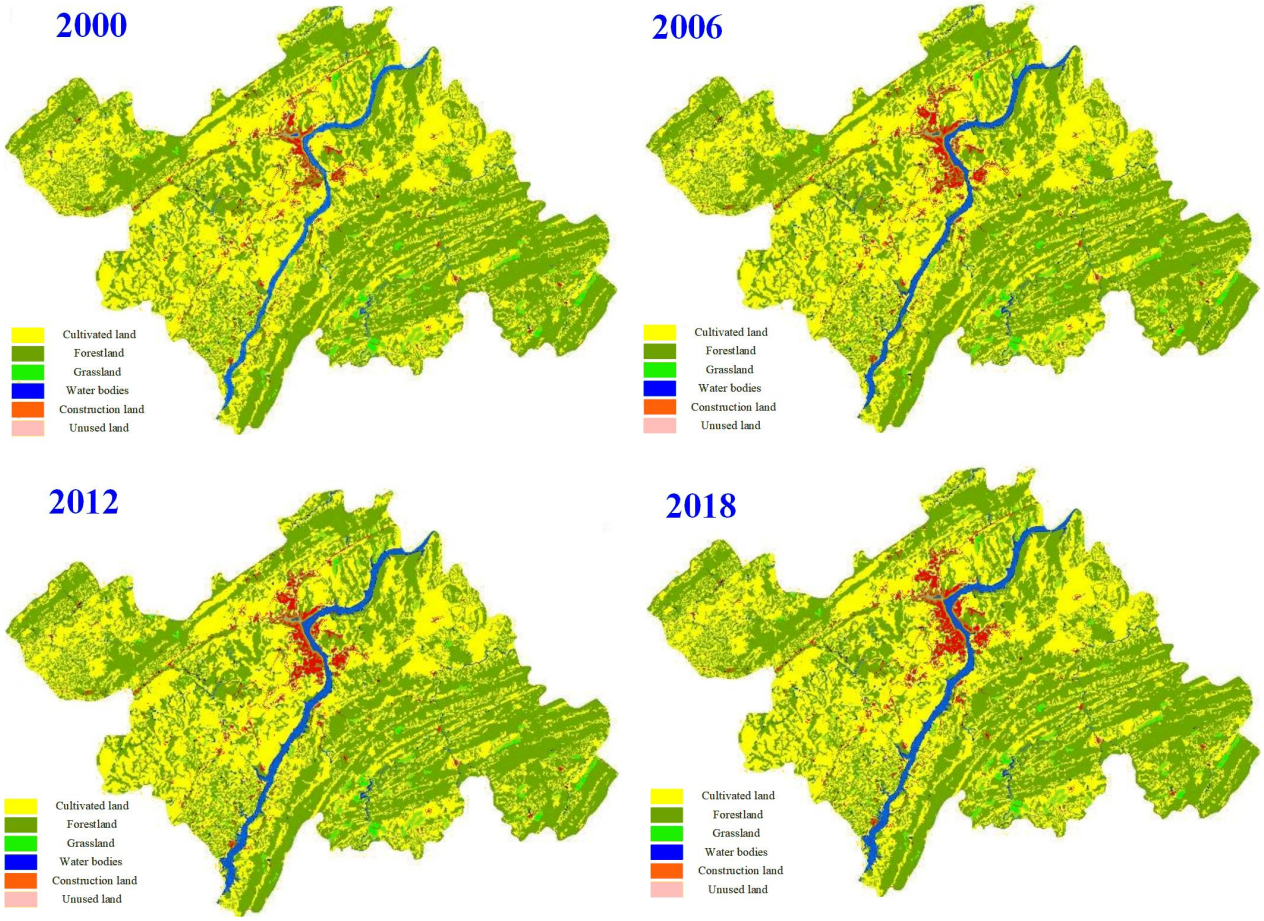


Fig. 3. Distribution map of ecological space structure types in wanzhou district from 2000 to 2018.

TABLE II. AREA STATISTICS OF ECOLOGICAL SPACE STRUCTURE TYPES IN WANZHOU DISTRICT FROM 2000 TO 2018

Type	2000		2006		2012		2018	
	Area (hm²)	Percentage (%)	Area (hm²)	Percentage (%)	Area (hm²)	Percentage (%)	Area (hm²)	Percentage (%)
Ecological Space	337826.13	98.33%	336026.89	97.80%	334956.05	97.49%	334279.17	97.29%
Arable Land	152779.02	44.47%	150913.92	43.92%	149056.68	43.38%	148385.62	43.19%
Woodland	169647.69	49.38%	169140.76	49.23%	168503.00	49.04%	168417.14	49.02%
Grassland	5906.24	1.72%	5826.97	1.70%	5754.00	1.67%	5848.84	1.70%
Water Bodies	9461.36	2.75%	10138.51	2.95%	11637.71	3.39%	11618.41	3.38%
Unused Land	31.82	0.01%	6.73	0.00%	4.66	0.00%	9.17	0.00%
Developed Land	5748.00	1.67%	7547.25	2.20%	8618.08	2.51%	9294.97	2.71%

Based on the 2006–2012 land use transition matrix (Table IV) for Wanzhou District, the conversion of arable land to built-up land significantly decreased compared to the period 2000–2006.

According to Table V, arable land no longer converts to forestland but solely transitions from arable land to forestland, with a conversion area of 19.57 hm². This indicates that during the 2012–2018 period, the policy of converting farmland back to forestland was strongly implemented in Wanzhou District. During this time, the primary land use transitions were from arable land and forestland to grassland and built-up land. The most significant conversion was from arable land to built-up land, with an area of 612.62 hm², followed by the conversion from forestland to built-up land, with an area of 64.27 hm².

According to Table VI, during the period from 2000 to 2018, the primary transition for arable land was towards built-up land, with a conversion area of 2885.10 hm², followed by transition to water bodies, totaling 1455.37 hm². Similarly, forestland mainly converted to built-up land, with an area of 654.03 hm², and secondarily to water bodies, with 643.38 hm². Grassland predominantly transformed into built-up land (57.50 hm²) and forestland (42.93 hm²). Water bodies mainly converted to arable land (27.03 hm²) and built-up land. Built-up land primarily transitioned to water bodies, with an area of 63.22 hm². Unused land primarily transformed into arable land (11.42 hm²) and built-up land (7.86 hm²).

TABLE III. LAND USE TRANSITION MATRIX FOR WANZHOU DISTRICT (HM²) FOR 2000–2006

	2000-2006					
	Arable Land	Woodland	Grassland	Water Bodies	Developed Land	Unused Land
Arable Land	150852.15	40.01	0.00	418.10	1465.34	3.43
Woodland	3.00	169050.12	13.02	276.19	302.04	3.30
Grassland	18.52	42.93	5806.04	6.89	31.86	
Water Bodies	20.96		3.07	9437.32		
Developed Land	—	—	—	—	5748.00	—
Unused Land	19.28	7.70	4.84	—	—	—

TABLE IV. LAND USE TRANSITION MATRIX FOR WANZHOU DISTRICT (HM²) FOR 2006–2012

	2006-2006					
	Arable Land	Woodland	Grassland	Water Bodies	Developed Land	Unused Land
Arable Land	149023.48	33.02	—	1038.91	815.01	3.50
Woodland	10.68	168453.66	1.08	386.47	287.72	1.16
Grassland	11.39	13.02	5752.93	20.92	28.71	—
Water Bodies	7.71	—	—	10128.17	2.63	—
Developed Land	—	—	—	63.22	7484.02	—
Unused Land	3.43	3.3.	—	—	—	—

TABLE V. LAND USE TRANSITION MATRIX FOR WANZHOU DISTRICT (HM²) FOR 2012–2018

	2010-2018					
	Arable Land	Woodland	Grassland	Water Bodies	Developed Land	Unused Land
Arable Land	148385.62	19.57	38.88	—	612.62	—
Woodland	—	168397.57	36.65	—	64.27	4.50
Grassland	—		5754.00	—	—	—
Water Bodies	—		19.29	11618.41	—	—
Developed Land	—	—	—	—	8618.08	—
Unused Land	—	—	—	—	—	4.66

TABLE VI. LAND USE TRANSITION MATRIX FOR WANZHOU DISTRICT (HM²) FOR 2000–2018

	2010-2018					
	Arable Land	**Woodland**	**Grassland**	**Water Bodies**	**Developed Land**	**Unused Land**
Arable Land	148303.57	92.59	38.88	1455.37	2885.10	3.50
Woodland	13.68	168273.92	57.02	643.38	654.03	5.66
Grassland	29.91	42.93	5748.09	27.81	57.50	—
Water Bodies	27.03	—	—	9428.63	5.69	—
Developed Land	—	—	—	63.22	5684.78	—
Unused Land	11.42	7.70	4.48	—	7.89	4.66

TABLE VII. TRANSFER RATE OF ECOLOGICAL SPACE IN WANZHOU DISTRICT FROM 2000 TO 2018 (%)

	2000—2006	2006—2012	2012-2018
Arable Land	0.09	0.05	0.03
Woodland	0.20	0.21	0.08
Grassland	0.05	0.06	0.01
Water Bodies	0.22	0.21	-0.27
Developed Land	-1.19	-2.46	0.03
Unused Land	13.14	5.12	-16.11
Comprehensive Dynamic Degree	0.08	0.10	0.05

E. Analysis of Ecological Space Dynamics

When studying the evolution characteristics of ecological space, it is necessary to examine the transfer rate and the activity level of the ecological space. Therefore, based on the area statistics table of Wanzhou District's ecological space types from 2000 to 2018, the transfer rate and comprehensive dynamic degree of Wanzhou District's ecological space from 2000 to 2018 are calculated, with the results shown in Table VII.

F. Influencing Factors

The changes in the ecological space layout of Wanzhou District are a complex process, influenced by a combination of various natural and human factors. Based on extensive literature review and adhering to the four principles of factor selection mentioned earlier, the preliminary selected influencing factors include: elevation factor, slope factor, road traffic factor, administrative center factor, river factor, and policy control factor. The specific factors of elevation, slope, road traffic, administrative center, river, and policy control are illustrated in Fig. 4.

After the preliminary selection of influencing factors such as elevation, slope, road traffic, administrative center, river, and policy control factors, the Empirical Likelihood method is used to further select and analyze the elevation, slope, road traffic, administrative center, and river factors. The results are shown in Fig. 5. The reason for not applying the Empirical Likelihood method to further verify and select the policy control factor is that it serves as a constraint, significantly impacting the ecological space layout changes.

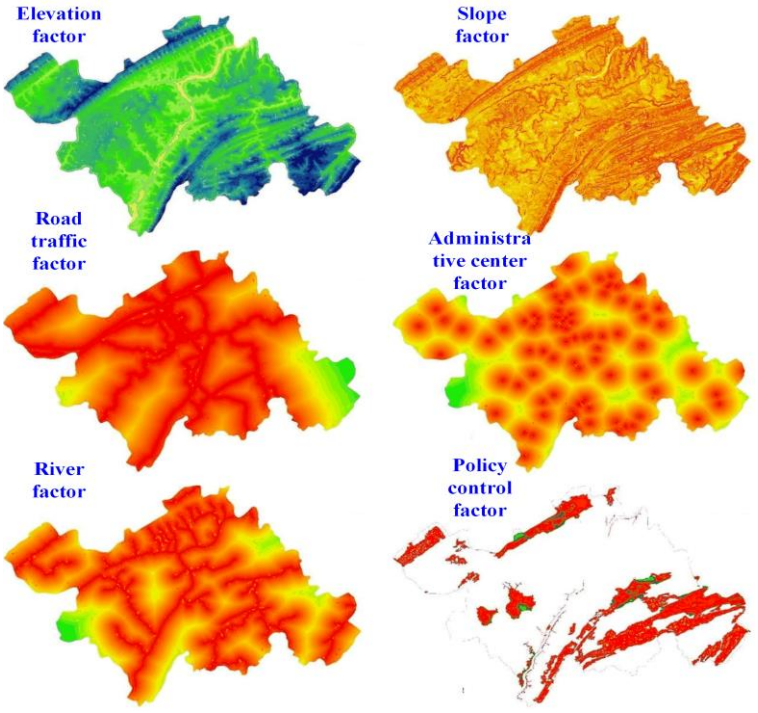


Fig. 4. Wanzhou district influencing factors map.

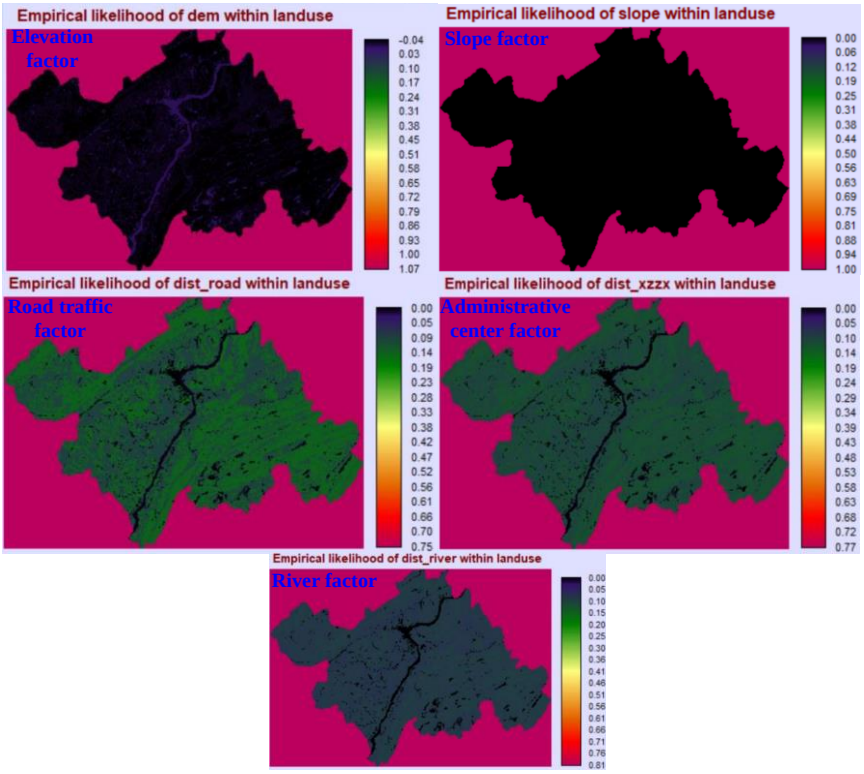


Fig. 5. Wanzhou district influencing factors impact map.

Based on Fig. 5, the final selected influencing factors are elevation, road traffic, administrative center, and policy control factors.

V. Simulation of Ecological Space Layout in Wanzhou District based on the Ann-CA-Markov Model

A. ANN-CA-Markov

The Ann-CA-Markov model, aimed at simulating Wanzhou District's ecological space pattern, is constructed as follows: (1) Cells are defined as 30m×30m grids based on TM image data, mirroring the district's ecological structure. (2) Cellular space consists of these grids. (3) Cell states represent ecological and non-ecological spaces, categorized into types like cultivated land, forest, grassland, water bodies, and construction land. (4) The neighborhood is defined using a 5×5 extended Moore setup, where 24 adjacent cells affect the central one. The choice of a 5×5 neighborhood window balances computational efficiency with spatial accuracy. This window size was selected based on sensitivity analysis, which demonstrated that smaller windows (e.g., 3×3) provided insufficient spatial context, while larger windows (e.g., 7×7) added unnecessary complexity without significantly improving model precision. (5) Transition rules are established using spatial and quantitative methods, with the former calculated via the MLP_ANN model and the latter through Markov model-derived transition probability matrices between space types. (6) Transition probabilities are calculated for 2000—2006, 2006—2012, and 2012—2018 using IDRISI software, aiding in predicting land use changes, shown in Tables VIII, IX, and X, respectively.

After selecting spatial variables and influencing factors, the MLP_ANN tool was utilized to construct the transition rules for the cellular automaton model (Fig. 6). The MLP_ANN comprises a three-layer network structure, including an input layer, a hidden layer, and an output layer. In this study's MLP_ANN, the input layer consists of 15 neurons, corresponding to the 15 influencing factors identified earlier. The number of neurons in the hidden layer, representing the number of input samples, should be at least two-thirds of the number of input layer neurons, hence 11 neurons were set for the hidden layer. The output layer contains 6 neurons, each corresponding to the transition probabilities for Wanzhou District's five ecological space types and one non-ecological space type. However, due to the negligible area of unused land in the study area, the actual number of neurons in the output layer used is five, representing the transition probabilities for farmland, woodland, grassland, water bodies, and construction land.

Following the computation of transition probability maps for various land use types, including farmland, woodland, grassland, water bodies, and construction land, using the MLP_ANN model, the results were refined to reflect both the prevailing policies and the real-world conditions of the study area. To ensure that the study's outcomes were consistent with local land use regulations, specific constraints were applied. These constraints mandated that construction land and water bodies could not change, while ecological spaces and non-ecological areas within designated protected zones—such as ecological protection red lines, forest parks, nature reserves, geological parks, and scenic areas—were preserved from any transitions. As the ANN-CA-Markov model is built on the principles of cellular automata (CA), the refined transition rules, now represented as probability maps for each land use type,

were standardized to a 0–255 scale. The standardized transition probability maps for the respective land types are presented in Fig. 7. In the final step, these maps were integrated into a comprehensive suitability mapset using the Collection Editor tool within the IDRISI software suite.

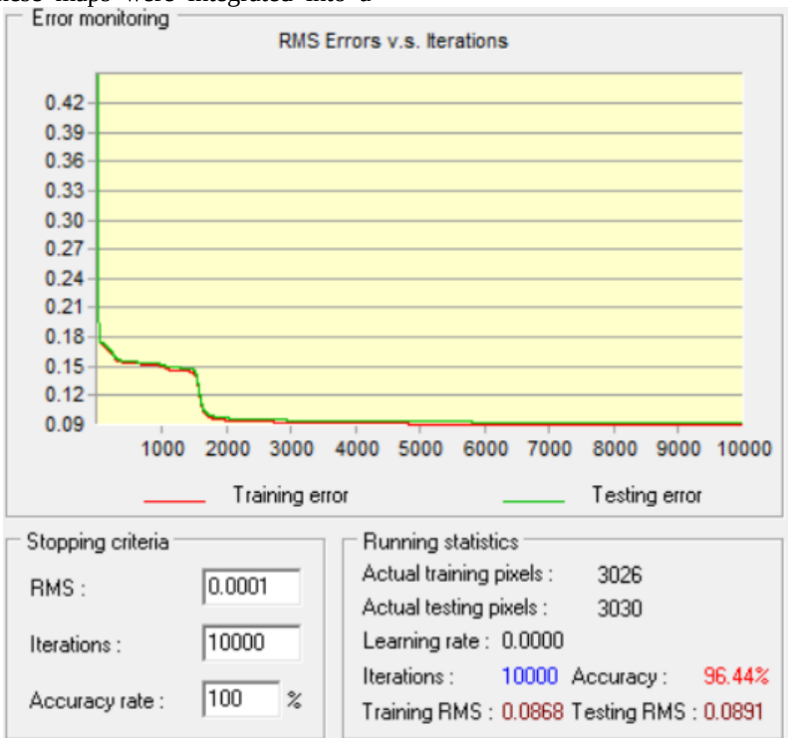


Fig. 6. MLP_ANN learning and calculation process diagram.

TABLE VIII. TRANSITION PROBABILITY MATRIX OF ECOLOGICAL SPACE TYPES IN WANZHOU DISTRICT FROM 2000—2006 (%))

	Arable Land	Woodland	Grassland	Water Bodies	Developed Land	Unused Land
Arable Land	0.8393	0.0033	0	0.035	0.1222	0.0003
Woodland	0.0007	0.847	0.0032	0.0708	0.0774	0.0008
Grassland	0.0307	0.0703	0.8356	0.0117	0.0517	0
Water Bodies	0.1345	0	0.0177	0.8478	0	0
Developed Land	0.03	0.03	0.03	0.03	0.85	0.03
Unused Land	0.6091	0.238	0.153	0	0	0

TABLE IX. TRANSITION PROBABILITY MATRIX OF ECOLOGICAL SPACE TYPES IN WANZHOU DISTRICT FROM 2006—2012 (%)

	Arable Land	Woodland	Grassland	Water Bodies	Developed Land	Unused Land
Arable Land	0.8393	0.0028	0	0.0882	0.0694	0.0003
Woodland	0.0023	0.8465	0.0003	0.0866	0.064	0.0003
Grassland	0.0252	0.0276	0.8393	0.0459	0.062	0
Water Bodies	0.1168	0	0	0.8491	0.0341	0
Developed Land	0	0	0	0.1571	0.8429	0
Unused Land	0.5068	0.4932	0	0	0	0

TABLE X. TRANSITION PROBABILITY MATRIX OF ECOLOGICAL SPACE TYPES IN WANZHOU DISTRICT FROM 20012—2018 (%)

	Arable Land	Woodland	Grassland	Water Bodies	Developed Land	Unused Land
Arable Land	0.8462	0.0045	0.0086	0	0.1407	0
Woodland	0	0.8495	0.053	0	0.0909	0.0066
Grassland	0.03	0.03	0.85	0.03	0.03	0.03
Water Bodies	0	0	0.1514	0.8486	0	0
Developed Land	0.03	0.03	0.03	0.03	0.85	0.03
Unused Land	0.03	0.03	0.03	0.03	0.03	0.85

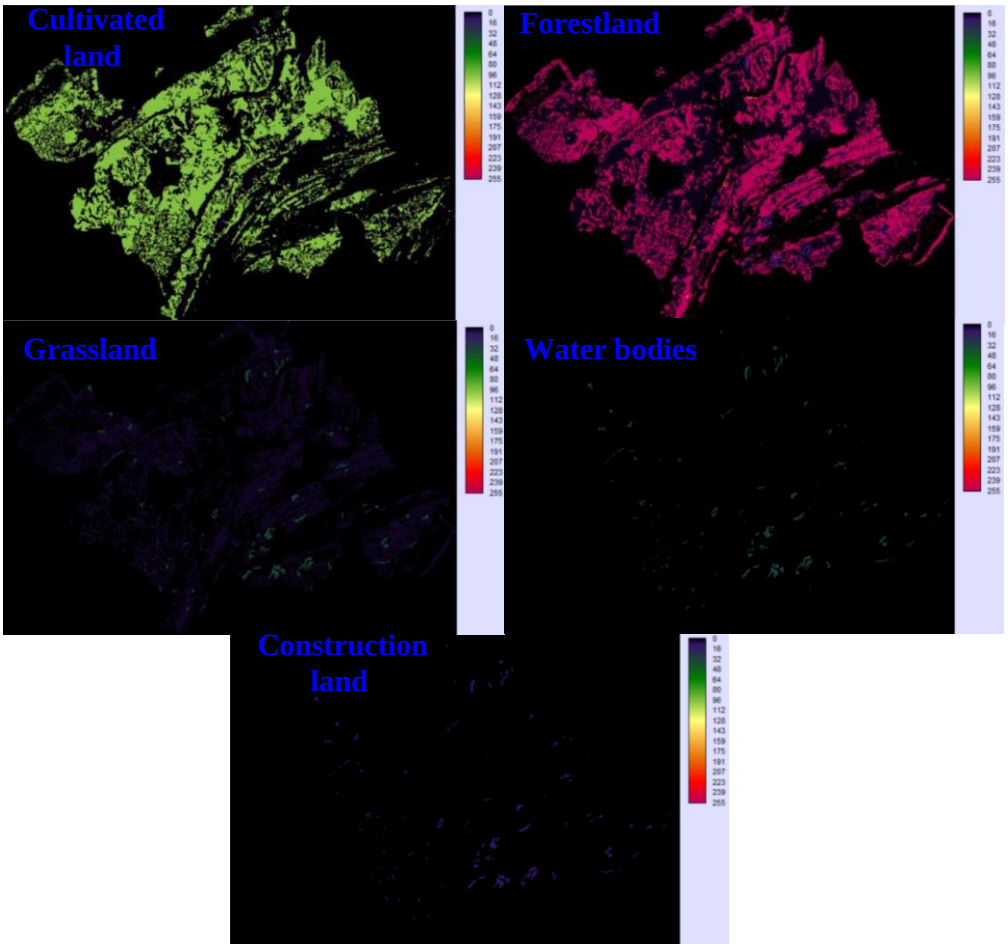


Fig. 7. Conversion probability maps for cultivated land, forest land, grassland, water area, and construction land.

B. *Determining the Number of Iterations and the Forecast Year*

After setting the iteration interval to 6 years, based on the base data and calculated transition probability matrices, the ecological space evolution in the study area for 2012 is simulated and predicted based on the 2000 and 2006 data. Then, the ecological space evolution in 2018 is simulated and predicted based on the 2006 and 2012 data. After verifying the accuracy of the simulated ecological space for 2012 and 2018 with the ANN-CA-Markov coupled model and meeting the accuracy requirements, the ecological space evolution in 2025 is simulated and predicted. Thus, in the ANN-CA-Markov coupled model, 2012, 2018, and 2025 are determined as the forecast years.

C. *ANN-CA-Markov Coupled Model Accuracy Verification*

After constructing the ANN-CA-Markov coupled model, the ecological space distribution in Wanzhou District for the years 2012 and 2018 is simulated based on the data from 2000 to 2006 and 2006 to 2012, respectively. The simulation results are shown in Fig. 8 and 9.

After completing the simulation maps of the ecological space structure types in Wanzhou District for 2012 and 2018, it is necessary to verify the simulation accuracy of the ANN-CA-Markov coupled model. This is done by comparing the simulation results for 2012 with the actual data for 2012, and the simulation results for 2018 with the actual data for 2018, to validate the simulation accuracy of the ANN-CA-Markov coupled model.

First, a quantitative verification is performed by comparing the simulated results for 2012 and 2018 with the actual data for those years. The specific quantitative verification results are shown in Tables XI and XII.

Based on Tables XI and XII, except for unused land, the simulation results for other ecological space types (cultivated land, forest land, grassland, water bodies) and non-ecological space type (construction land) show high accuracy rates. The lower accuracy for unused land is due to its minimal grid cell count, which does not reach 0.01% of the total grid cell count in the study area. Overall, the Ann-CA-Markov coupled model demonstrates high simulation accuracy in terms of quantity.

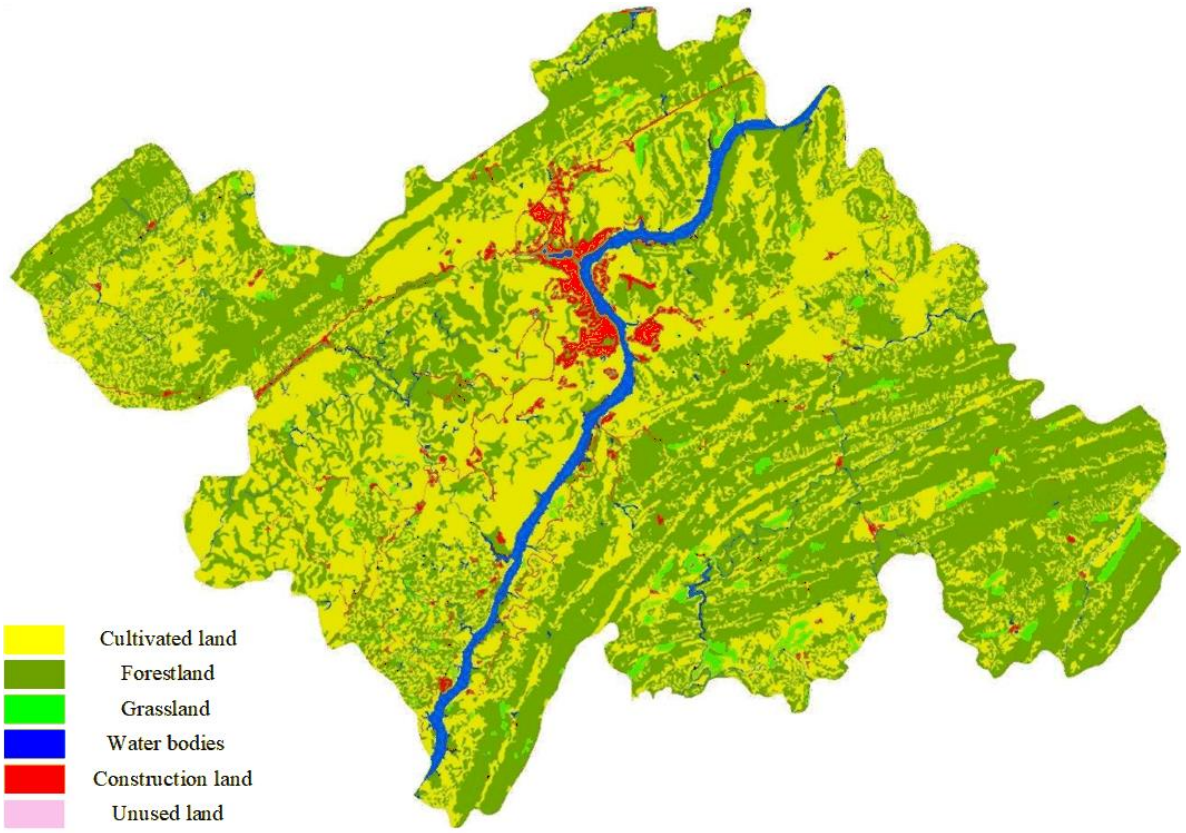


Fig. 8. Simulation map of the ecological space structure types in Wanzhou District for the year 2012.

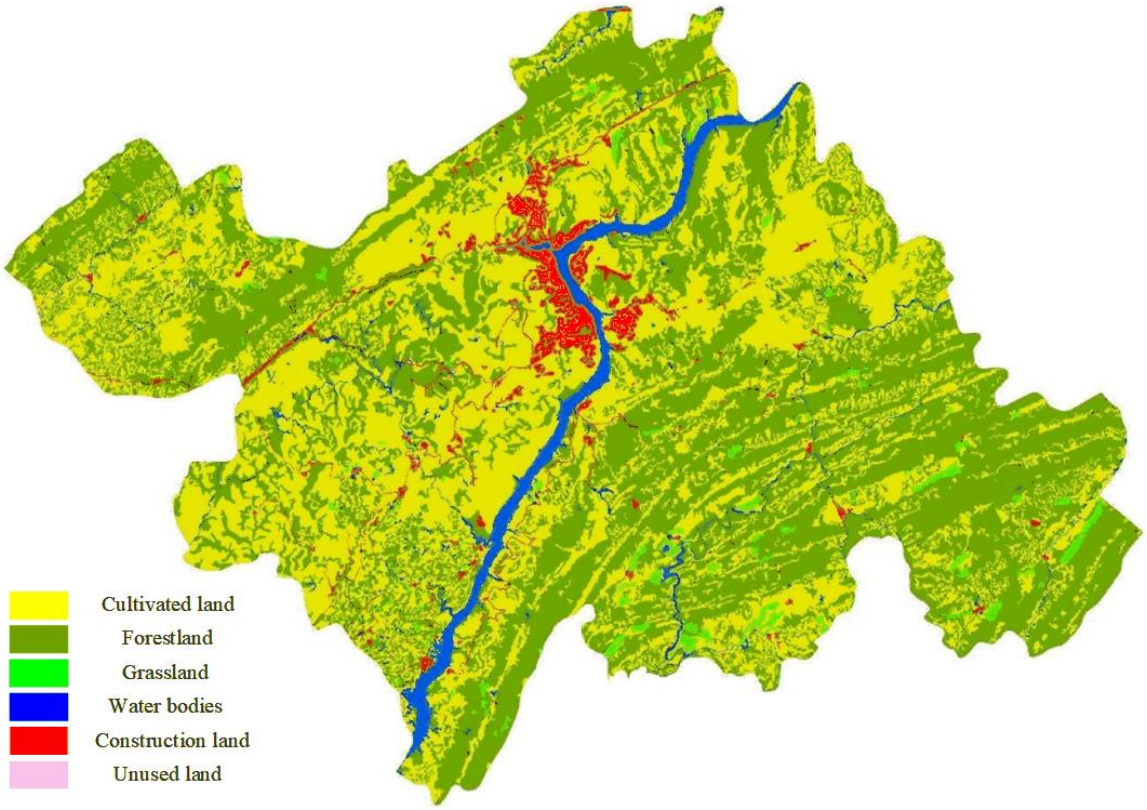


Fig. 9. Simulation map of the ecological space structure types in wanzhou district for the year 2018.

TABLE XI. VERIFICATION OF THE SIMULATED AND ACTUAL NUMBER OF GRIDS FOR 2012

Land Type	Actual Number of Grids (2012)	Simulated Number of Grids (2012)	Accuracy (%)
Cultivated Land	1,655,590	1,672,030	99.01
Forest Land	1,870,820	1,873,846	99.84
Grassland	63,945	65,175	98.08
Water Area	129,273	113,517	87.81
Unused Land	55	73	67.27
Construction Land	95,783	90,825	94.82

TABLE XII. VERIFICATION OF THE SIMULATED AND ACTUAL NUMBER OF GRIDS FOR 2018

Land Type	Actual Number of Grids (2018)	Simulated Number of Grids (2018)	Accuracy (%)
Cultivated Land	1,648,162	1,661,779	99.17
Forest Land	1,869,859	1,870,178	99.98
Grassland	64,994	55,553	85.47
Water Area	129,054	128,992	99.95
Unused Land	107	55	51.40
Construction Land	103,290	98,909	95.76

D. Simulation and Result Analysis of Ecological Space Layout in the Study Area for 2025

After validating the Ann-CA-Markov model's high accuracy, it was applied to project Wanzhou District's 2025 ecological space. For the 2025 simulation, the model used 2018 data for neighboring variables and current space types, adjusted conversion probability maps, and derived the transition matrix from 2012 and 2018 data. These modifications enabled the projection of the district's 2025 ecological layout, depicted in Fig. 10. A corresponding map illustrating the predicted ecological space distribution for 2025 is shown in Fig. 11.

Based on the simulation results for Wanzhou District in 2025 (Fig. 10 and 11) and the actual data from 2018, the areas of ecological spaces in Wanzhou District were calculated and summarized, as shown in Table XIII.

In 2025, ecological space in Wanzhou District is projected to cover 334,022.95 hectares (97.28%), dominated by cultivated land (147,743.20 hectares) and forest land (168,944.40 hectares). Grassland (5,698.71 hectares) and water bodies (11,627.26 hectares) contribute smaller portions, while unused land remains negligible. Constructed land expands to 9,349.42 hectares (2.72%). From 2018 to 2025, the ecological space decreased slightly by 81.02 hectares, as cultivated land and grassland declined, while forest land and water bodies saw modest increases. The transfer matrix in Table XIV highlights the shifts between these land types during the period.

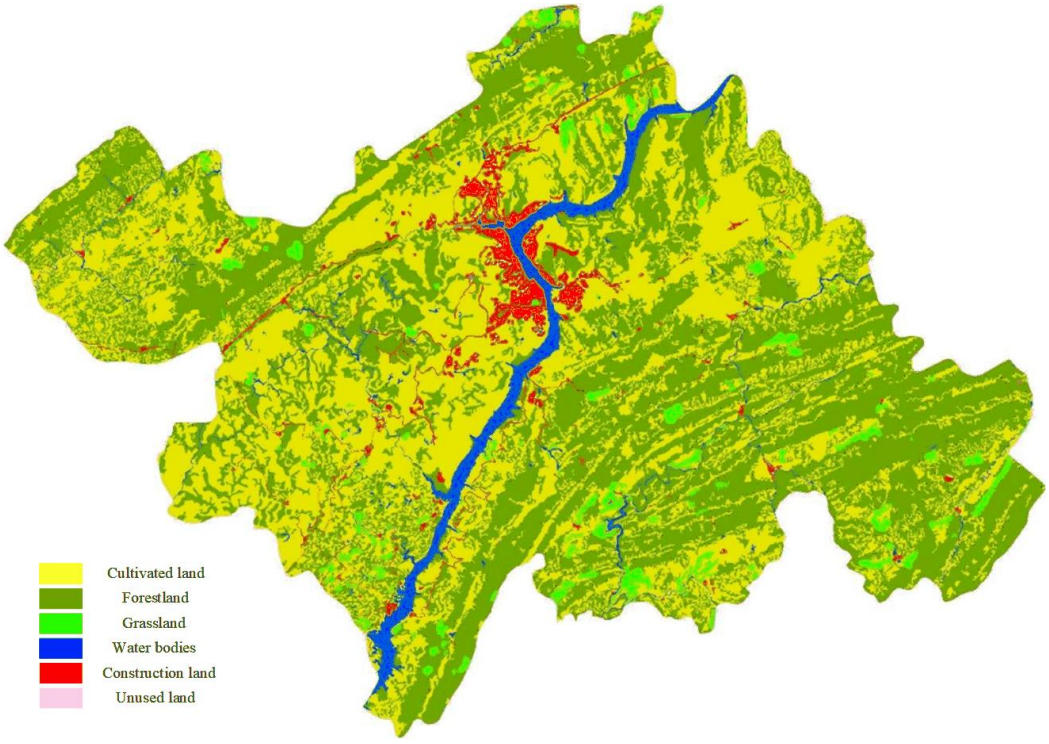


Fig. 10. Simulated distribution map of ecological space structure types in wanzhou district for 2025.

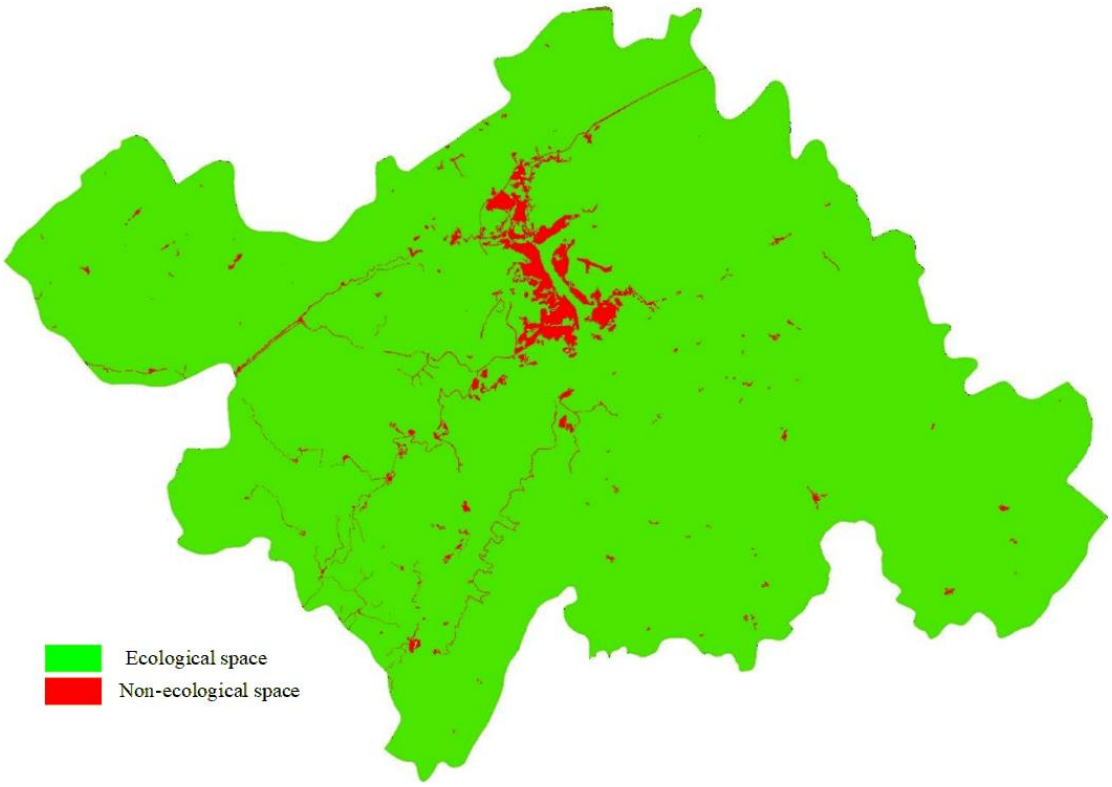


Fig. 11. Simulated ecological space distribution map in wanzhou district for 2025.

TABLE XIII. SIMULATED RESULTS OF ECOLOGICAL SPACE IN WANZHOU DISTRICT FOR 2025

Type	2018		2025	
	Area (hm²)	Proportion(%)	Area (hm²)	Proportion(%)
Ecological Space	334103.97	97.30	334022.95	97.28
Cultivated Land	148328.32	43.20	147743.20	43.03
Forest Land	168317.33	49.02	168944.40	49.20
Grassland	5843.57	1.70	5698.71	1.66
Water Body	11605.37	3.38	11627.26	3.39
Unused Land	9.37	0.00	9.37	0.00
Constructed Land	9268.41	2.70	9349.42	2.72

TABLE XIV. TRANSFER MATRIX OF ECOLOGICAL SPACE STRUCTURE TYPES IN WANZHOU DISTRICT FOR 2018-2025 (HM²)

2018	2025					
	Arable Land	Woodland	Grassland	Water Bodies	Developed Land	Unused Land
Arable Land	147528.64	742.57	9.07	4.20	43.85	0.00
Woodland	88.62	168128.55	23.34	11.90	64.93	0.00
Grassland	48.58	60.61	5653.00	34.73	46.64	0.00
Water Bodies	3.25	8.61	8.30	11572.30	12.90	0.00
Developed Land	74.11	4.06	5.00	4.13	9181.11	0.00
Unused Land	0.00	0.00	0.00	0.00	0.00	9.37

According to Table XIV, during the period from 2018 to 2025, the area of cultivated land mainly transferred to forest land and constructed land, with transfer areas of 742.57 hm² and 43.85 hm², respectively. The primary transfers from forest land were to cultivated land and constructed land, with transfer areas of 88.62 hm² and 64.93 hm², respectively. Grassland transfers were relatively balanced among various types, with transfer areas to cultivated land, forest land, water bodies, and constructed land being 48.58 hm², 60.61 hm², 34.73 hm², and 46.64 hm², respectively. Water bodies had smaller transfer areas to other types, with the largest being the transfer to constructed land, at 12.90 hm². Constructed land primarily transferred to cultivated land, with a transfer area of 74.11 hm². Unused land did not undergo any transfers.

VI. CONCLUSION

The deployment of the ANN-CA-Markov model in this study provides a detailed and forward-looking analysis of the expected land use changes in Wanzhou District by the year 2025. The results of the analysis suggest that ecological spaces will remain predominant, accounting for 97.28% of the district's total land area. However, a minor reduction in ecological spaces is forecasted, accompanied by a corresponding increase in construction land, indicating the growing impact of urbanization on the region's ecological zones. Specifically, the study anticipates a decline in arable land and grassland areas, while forested regions and water bodies are projected to expand. This shift may reflect the influence of regional policies and planning initiatives aimed at conserving forests and water resources. Although the increase in construction land is relatively small, it nonetheless reflects the broader trend of intensified land use and development driven by urbanization pressures. Furthermore, an examination of the transition dynamics between various land use categories reveals a pattern where arable land is increasingly converted to forested areas and construction sites, while transitions involving forest land typically lead to its conversion into arable land or construction areas. These land use transformations are likely driven by a combination of policy enforcement, economic development, and resource management practices.

REFERENCES

[1] Wang L, Wang Y, Cai Y. An ANN-CA Modeling Method for Land Cover Change in the Karst Area of China: A Case Study of Maotiao River Basin.Acta Scientiarum Naturalium Universitatis Pekinensis, 2012, 48(1):116-122.DOI:10.1007/s11783-011-0280-z.

[2] Zeshan M T, Mustafa M R U , Baig M F .Monitoring Land Use Changes and Their Future Prospects Using GIS and ANN-CA for Perak River Basin, Malaysia.Multidisciplinary Digital Publishing Institute, 2021(16).DOI:10.3390/W13162286.

[3] Zhang Y , Qiao J , Wu B ,et al.Simulation of oil spill using ANN and CA models[C]//International Conference on Geoinformatics.IEEE, 2016.DOI:10.1109/GEOINFORMATICS.2015.7378560.

[4] Tingting D , Lijun Z , Zengxiang Z .A Study on Spacetime Evolution of Soil Erosion Based on ANN-CA Model.Journal of Geo-Information Science, 2009, 11(1):132-138.DOI:10.3969/j.issn.1560-8999.2009.01.020.

[5] Xi-Bao X U, Gui-Shan Y , Jian-Ming Z .Scenario Modeling of Urban Land Use Changes in Lanzhou with ANN-CA.Geography and Geo-Information Science, 2008, 24(6):80-84.DOI:10.3724/SP.J.1047.2008.00114.

[6] João Lita da Silva, Caeiro F ,Isabel Natário,et al.Advances in regression, survival analysis, extreme values, Markov processes and other statistical applications. Selected papers based on the presentations of the 17th congress of the Portuguese Statistical Society, Sesimbra, Portugal, September 30–Octobe. 012.

[7] Cryan M, Goldberg L A , Goldberg P W .Evolutionary trees can be learned in polynomial time in the two-state general Markov model. SIAM Journal on Computing, 2002.DOI:10.1137/s0097539798342496.

[8] Deruiter S L, Langrock R , Skirbutas T ,et al.A multivariate mixed hidden Markov model for blue whale behaviour and responses to sound exposure.Annals of Applied Statistics, 2017, 11(1):362-392.DOI:10.1214/16-AOAS1008.

[9] Zhan Q, Tian J , Tian S .Prediction Model of Land Use and Land Cover Changes in Beijing Based on Ann and Markov_CA Model[C]//IGARSS 2019 - 2019 IEEE International Geoscience and Remote Sensing Symposium.IEEE, 2019.DOI:10.1109/IGARSS.2019.8898388.

[10] Ynoguti C A, Morais E D S , Violaro F .A comparison between HMM and hybrid ANN-HMM-based systems forcontinuous speech recognition[C]//Telecommunications Symposium, 1998. ITS '98 Proceedings. SBT/IEEE International.IEEE, 1998.DOI:10.1109/ITS.1998.713105.

[11] Mondal M S. Modeling of Land Use Land Cover Change: CA Markov Modeling Approaches[M]. 2013.

[12] Liu J, Cheng F , Zhu Y ,et al.Urban Land-Use Type Influences Summertime Water Quality in Small- and Medium-Sized Urban Rivers: A Case Study in Shanghai, China.Land, 2022, 11.DOI:10.3390/land11040511.

[13] Xu T, Zhou D, Li Y. Integrating ANNs and cellular automata–Markov chain to simulate urban expansion with annual land use data[J]. Land, 2022, 11(7): 1074.

[14] Zhao J, Zhu X, Zhou Y, et al. Examining land-use change trends in yucheng district, Ya'an city, China, using ANN-CA modeling[J]. Journal of Urban Planning and Development, 2023, 149(1): 05022042.

[15] Zambrano-Asanza S, Morales R E, Montalvan J A, et al. Integrating artificial neural networks and cellular automata model for spatial-temporal load forecasting[J]. International Journal of Electrical Power & Energy Systems, 2023, 148: 108906.

[16] Xie S, Hu Z, Wang J. Two-stage robust optimization for expansion planning of active distribution systems coupled with urban transportation networks. Applied Energy, 2020, 261: 114412.

[17] Oyalowo B. Implications of urban expansion: land, planning and housing in Lagos. Buildings & Cities, 2022, 3(1).

[18] Woldesemayat E M, Genovese P V. Monitoring urban expansion and urban green spaces change in Addis Ababa: directional and zonal analysis integrated with landscape expansion index. Forests, 2021, 12(4): 389.

[19] Theres L, Radhakrishnan S, Rahman A. Simulating Urban Growth Using the Cellular Automata Markov Chain Model in the Context of Spatiotemporal Influences for Salem and Its Peripherals, India. Earth, 2023, 4(2): 296-314.

Advanced IoT-Enabled Indoor Thermal Comfort Prediction Using SVM and Random Forest Models

Nurtileu Assymkhan, Amandyk Kartbayev
School of Information Technology and Engineering, Kazakh-British Technical University, Almaty, Kazakhstan

***Abstract*—Predicting thermal comfort within indoor environments is essential for enhancing human health, productivity, and well-being. This study uses interdisciplinary approaches, integrating insights from engineering, psychology, and data science to develop sophisticated machine learning models that predict thermal comfort. Traditional methods often depend on subjective human input and can be inefficient. In contrast, this research applies Support Vector Machines (SVM) and Random Forest algorithms, celebrated for their precision and speed in handling complex datasets. The advent of the Internet of Things (IoT) further revolutionizes building management systems by introducing adaptive control algorithms and enabling smarter, IoT-driven architectures. We focus on the comparative analysis of SVM and Random Forest in predicting indoor thermal comfort, discussing their respective advantages and limitations under various environmental conditions and building designs. The dataset we used included comprehensive thermal comfort data, which underwent rigorous preprocessing to enhance model training and testing—80% of the data was used for training and the remaining 20% for testing. The models were evaluated based on their ability to accurately mirror complex interactions between environmental factors and occupant comfort levels. The results indicated that while both models performed robustly, Random Forest demonstrated greater stability and slightly higher accuracy in most scenarios. The paper proposes potential strategies for incorporating additional predictive features to further refine the accuracy of these models, emphasizing the promise of machine learning in advancing indoor comfort optimization.**

Keywords*—*Heating; building energy management; thermal comfort; IoT; Support Vector Machine; Random Forest

I. Introduction

Optimizing built environments for human habitation crucially involves predicting thermal comfort, a significant challenge intensified by climate change. As climate change escalates, extreme weather events become more frequent and severe, heightening the need for effective management of indoor thermal conditions. The importance of accurately predicting thermal comfort is underscored by its substantial impact on human health, productivity, and overall well-being. Inadequate thermal environments, characterized by excessive heat or cold, can result in discomfort, fatigue, and health complications, adversely affecting an individual's quality of life and reducing productivity in various environments such as workplaces, educational institutions, and homes.

Moreover, the economic consequences of neglecting thermal comfort are significant. Suboptimal indoor climates lead to heightened energy consumption as occupants frequently use heating or cooling systems to alleviate discomfort. This increased reliance on HVAC (Heating, Ventilation, & Air Conditioning) systems not only results in higher utility bills but also contributes to environmental strain. Consequently, there is an urgent need to develop reliable predictive models that can accurately forecast occupants' thermal comfort preferences under varying environmental conditions and architectural designs. Such models must incorporate a range of factors, including ambient temperature, humidity levels, clothing insulation, metabolic rates, and individual preferences, to deliver precise assessments of thermal comfort levels.

Addressing this imperative necessitates interdisciplinary collaboration among architects, engineers, psychologists, and data scientists. Integrating insights from environmental science, human physiology, and behavioral psychology is essential for developing effective predictive models. By harnessing advancements in sensor technology, data analytics, and machine learning algorithms, these models can be refined to provide real-time insights into the dynamics of thermal comfort. This enables building managers and occupants to optimize indoor environments, thereby enhancing well-being and promoting sustainable resource utilization.

To further understand the impact of thermal comfort, let's explore a detailed example. Temperature plays a critical role in human well-being, akin to how a rise in body temperature can signal illness, indicating that something is amiss. Similarly, room temperature significantly affects comfort and, consequently, our ability to function optimally.

Consider a scenario on a hot summer day: you begin to prepare for lessons or study in your room. To create a quiet environment, you close the door to block out noise and shut the window to keep out the heat. However, this action inadvertently leads to a reduction in airflow and available space, causing an increase in carbon dioxide levels as it accumulates in the room. This buildup of carbon dioxide decreases the oxygen levels, leading to a rise in room temperature. Consequently, you may start to feel distracted and lethargic, a direct result of the diminished air quality and increased warmth. This situation can be remedied by simply opening the door to improve ventilation. This action helps to balance the air quality and regulate the room's temperature, restoring a more comfortable and conducive environment for studying. This example underscores the importance of managing thermal comfort to maintain productivity and well-being in indoor spaces. The process of predicting thermal comfort involves the analysis of various factors, including temperature, humidity, air velocity, and clothing insulation.

Traditional methods, based on human comfort models, tend to be subjective and time-consuming.

In this paper, we explore the application of Support Vector Machines (SVM) and Random Forest algorithms for predicting thermal comfort in buildings, aiming to assess their effectiveness and compare their performance across different scenarios. The goal is to provide a thorough understanding of how these machine learning algorithms can aid building designers and facility managers in optimizing indoor environments and enhancing occupant comfort. Our research is structured around a series of hypotheses that guide the experimental design:

- Data Preparation: We hypothesize that removing NaN values and establishing a threshold for the minimum number of observations per feature will improve model accuracy by ensuring the data quality and relevance of the features used.
- Feature Encoding: We will evaluate the suitability of different encoding strategies as OneHotEncoder, LabelEncoder, and Word2Vec, to determine how best to handle categorical variables. The choice of encoder may significantly impact the performance of our models, depending on the nature of the data.
- Feature Selection: The SelectKBest model will be utilized to identify the most relevant features for predicting thermal comfort. This method is expected to highlight the variables most closely linked to the outcomes, thereby streamlining the modeling process.
- Feature Variants: Post feature selection, we will focus on variants of the filtered features that are closely associated with temperature prediction. This step is crucial for refining the model's focus and enhancing its predictive accuracy regarding thermal comfort.

Through this structured approach, we aim to validate our hypotheses and draw meaningful conclusions about the utility of the algorithms in the context of thermal comfort prediction, potentially offering actionable insights for the design and management of building environments. Both SVM and RF are supervised learning algorithms capable of being trained on datasets consisting of thermal comfort parameters alongside corresponding human feedback.

This paper is structured as follows: Section II provides a literature review, contextualizing our study within existing research. Section III describes the methodology employed, detailing the techniques used to analyze data. Section IV presents the findings of the study, supported by relevant tables and illustrations. Section V discusses the implications of these results. Finally, Section VI offers a conclusion, summarizing the key outcomes and proposing directions for future research.

II. Related Works

The Internet of Things (IoT) is revolutionizing the building management systems (BMS) industry, with forecasts predicting up to 125 billion connected devices by 2030. Despite these advancements, current BMS solutions often lack flexibility, especially in terms of feedback control options. To fully leverage the potential of IoT, adaptive control algorithms and modular architectures are being explored. The authors have introduced the "Semantically-Enhanced IoT-enabled Intelligent Control System" (SEMIoTICS) architecture, which enhances control system capabilities through redundancy and automatically adjusts configurations based on quality-of-service criteria [1]. Additionally, Model Predictive Control (MPC) is becoming increasingly popular for optimizing energy efficiency and comfort in HVAC systems. Nonetheless, the use of nonlinear models introduces significant computational challenges. In response, research has shifted towards linear controllers that utilize Jacobian linearization. A notable innovation in this field is a bilinear model for nonlinear MPC, designed to minimize energy costs while maintaining comfort levels. However, the computational intensity of this model poses challenges for its application in real-time control settings [2].

Another articles introduce a cutting-edge reinforcement learning (RL)-based approach for HVAC systems integrated into the Transactive Energy Simulation Platform (TESP). Utilizing the Deep Deterministic Policy Gradients (DDPG) algorithm, this method focuses on intelligent and granular control of HVAC operations by optimizing a cost function that seeks a balance between electricity costs and end-user dissatisfaction. The approach includes a market price prediction model developed using Artificial Neural Networks (ANN), a DDPG-based RL control algorithm, and both implementation and testing phases within the TESP framework [3]. Further the authors present a simulation model that incorporates both high-level and low-level controllers for a passenger car's air conditioning system. This model prioritizes occupant thermal comfort and the precise regulation of the physical system. They also introduce an Eco-Cooling Strategy employing MPC to optimize control inputs. The strategy is designed to achieve efficient cooling, reduce energy consumption, and maintain comfort. The simulation results underscore the critical role of control settings in effective thermal management [4].

Fuzzy logic-based models are increasingly utilized to control air conditioning systems at variable speeds, optimizing energy consumption and enhancing thermal comfort. Implemented in hardware such as microcontrollers, VLSI chips, and EDA tools, these controllers precisely manage temperature and humidity levels, effectively regulating fan and compressor speeds. Integrated with other techniques, they significantly improve energy efficiency and system performance [5]. Ref. [6] explores a range of HVAC control strategies, from classical PID controllers to advanced MPC. They address challenges in system simulation, control implementation, artificial intelligence integration, and energy savings, introducing the LAMDA controller to enhance real-time responsiveness and self-adjustment based on contextual information, further refining control accuracy and efficiency in HVAC systems.

The escalating energy consumption in commercial buildings, particularly through HVAC systems, has spurred increased research into optimizing energy efficiency. Despite advancements in HVAC technologies that have enhanced Demand Response (DR) programs, challenges remain in the

application of model predictive control techniques. Recent studies have utilized machine learning methods, including Reinforcement Learning and Supervised Learning, to improve these systems [7]. Research in BEM has particularly focused on optimizing HVAC operations through various innovative approaches. Key developments include dynamic demand response controllers, mixed-integer nonlinear optimization models, stochastic programs, multi-objective optimization models, occupancy-based controllers, and incentive-based DR controllers. Additional methodologies explored include event-based control, mutual information frameworks, and MPC [8]. Furthermore, this paper [9] introduces a three-layered model designed for optimizing energy consumption in smart homes, incorporating data collection, prediction, and optimization phases. The model employs an Alpha Beta filter for reducing noise, Dynamic Evolving Neural Network (DELM) for dynamic parameter prediction, and fuzzy controllers for making refined control decisions. This integrated approach not only addresses static user parameters but also enhances both comfort and energy efficiency.

One study introduces an innovative model that omits gender and age factors in assessing thermal comfort, focusing instead on six key thermal factors: air temperature, mean radiant temperature, relative humidity, air speed, clothing insulation, and metabolic rate. This model, developed using Supervised Machine Learning, is tailored for application in a commercial building environment [10]. Another study conducted in Bilbao, Spain, at the KUBIK energy efficiency research facility, examines human thermal perception in response to external temperatures to enhance indoor comfort and reduce energy consumption [11]. Further research evaluates indoor thermal comfort using the Fanger method and adhering to ASHRAE Standard 55, emphasizing real-world conditions to promote well-being, productivity, and energy conservation in buildings [12].

Additionally, a study introduces a model based on multiple preferences for predicting group thermal comfort in shared spaces. This model integrates individual preferences and environmental parameters, segments occupants by Body Mass Index (BMI), predicts individual comfort zones, and adjusts settings to achieve group satisfaction [13]. Overall, optimizing thermal comfort in buildings is crucial for enhancing occupant well-being, productivity, and energy efficiency. Effective assessment models take into account variables such as air temperature, humidity, radiant temperature, and air speed, with the ASHRAE 55 standards providing guidelines for acceptable conditions.

Alternative models such as ANN, hybrid ANN-fuzzy systems, SVM, decision trees, and Bayes networks offer enhanced flexibility and accuracy in predicting thermal comfort [14]. Thermal comfort is a key component of indoor environmental quality, which can be categorized into static, adaptive, and data-driven models. Static models, like the Predicted Mean Vote (PMV), incorporate environmental and personal factors but have recognized limitations due to their lack of adaptability to individual responses. Adaptive models account for psychological and behavioral adaptations, enhancing their responsiveness to occupant preferences. Data-driven models leverage real-time data from sensor technologies for dynamic and responsive assessments of thermal comfort [15].

Further advancements are seen in the development of a building thermal model that utilizes low-resolution data from smart thermostats, improving accuracy and applicability across different seasons. This approach transforms traditional empirical models into a data-driven framework by using surrogate features to approximate internal heat gains. The model's design allows for implementation on either edge devices or cloud infrastructure, facilitating efficient data collection, model learning, and deployment [16].

Research continues to evolve with studies focusing on innovative cooling technologies such as Thermoelectric Air Ducts, with neural network models demonstrating high accuracy in predicting comfort parameters in dynamic settings. Understanding the interplay between climatic variables, occupant comfort, and system performance is fundamental [17]. Overall, the prediction of thermal comfort and optimization of energy use in buildings are critical for ensuring occupant satisfaction and achieving energy efficiency. Key factors influencing comfort include metabolic rate, clothing insulation, and air temperature.

Deep feedforward neural networks and reinforcement learning models are increasingly utilized to predict comfort levels, which is essential for monitoring and optimizing HVAC energy consumption in building operations [18]. A novel methodology employing machine learning, data mining, and statistical techniques has been developed to create predictive models for Combined Heat, Cooling, and Power (CHCP) systems. This methodology encompasses four stages: data preparation, data engineering, model building, and model evaluation. Data preparation includes retrieving failure events, labeling instances, and compiling a comprehensive dataset. Data engineering focuses on improving data representation through feature extraction and selection. The model building phase employs machine learning algorithms for various classification and regression tasks, while model evaluation assesses time to failure and other performance metrics to ensure the model's suitability [19].

Another innovative study explores thermal comfort in indoor environments through a novel approach called Relative Thermal Sensation (RTS). This method views thermal sensation as a continuous function of time, offering a more detailed understanding of human thermal perception. The study introduces a 3-point Relative Thermal Sensation Scale (RTSS) to collect real-time data on thermal sensations, capturing subtle changes that traditional discrete scales might overlook. Additionally, the research integrates RTS data with Absolute Thermal Sensation data from modified versions of the ASHRAE 7-point thermal sensation scale, enhancing the comprehensive understanding of thermal comfort [20].

Interpretable thermal comfort systems are being developed to enhance energy efficiency and occupant satisfaction in smart building environments. Traditional models such as the PMV often lack interpretability, posing challenges for building operators who need to understand the mechanisms influencing thermal comfort. To address this, researchers are integrating

machine learning techniques that promote model transparency, such as Partial Dependence Plots (PDP) and SHAP values.

These tools allow operators to comprehend how environmental conditions affect human comfort and to evaluate the significance of various features under different scenarios. Furthermore, these interpretable machine learning algorithms are also being used to create surrogate models that replicate and potentially improve upon existing thermal comfort models, making them more accessible and actionable for building management [21].

III. METHODOLOGY

A. Dataset

The dataset, sourced from the ASHRAE and available on Kaggle [22], comprises 70 columns and 107,583 rows, containing data collected globally from 1995 to 2015. Initially, an examination of the dataset description led to a filtering process. This revealed that some columns contained sparse data. Consequently, a threshold was set at 60,000 rows; data points below this limit were discarded. Additionally, it was necessary to address missing values. Despite starting with 107,583 rows, the removal of rows with NaN values was essential to ensure data integrity.

Another analytical approach considered was the use of the Interquartile Range (IQR) method to identify and eliminate outliers, further refining the dataset's quality (see Fig.1).

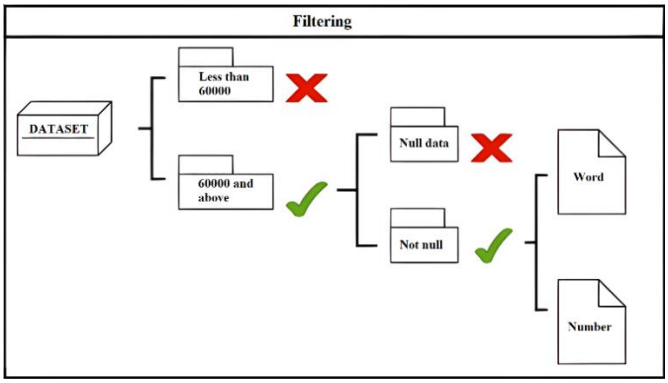


Fig. 1. Data filtering scheme.

Regarding the conversion of text data to numeric form, as shown in Fig. 2, two encoding options were evaluated: LabelEncoder and OneHotEncoder. The decision to proceed with OneHotEncoder was based on its superior performance in preliminary results [23], effectively transforming categorical text data into a usable format for machine learning models.

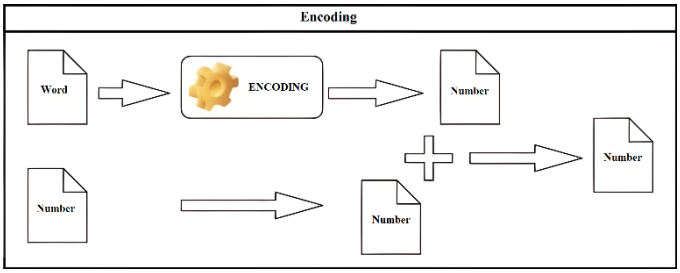


Fig. 2. Encoding scheme for the conversion of text data to numeric form.

In the feature selection process, as shown in Fig. 3, two methods were considered: using the SelectBest library or selecting based on correlation with a predefined threshold. The chosen method was to use correlations, specifically setting a boundary above 50% to determine relevant features. The final set of features selected includes Age, Clothing insulation (Clo), Sex, Metabolic rate (Met), Thermal preference, Year, Season, Köppen climate classification, Cooling strategy at the building level, City, Predicted Percentage of Dissatisfied (PPD), Air temperature (C), Outdoor monthly air temperature (C), Relative humidity (%), and Air velocity (m/s). This selection represents the culmination of extensive testing with various combinations of features, all of which will be detailed in the Experiments section of our study.

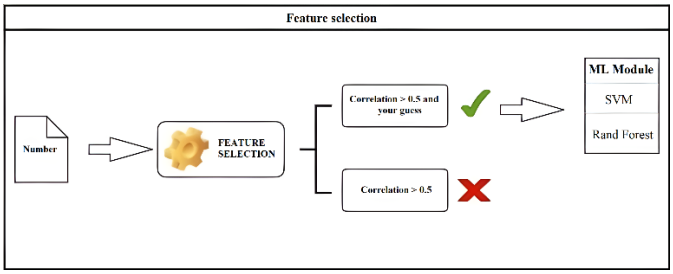


Fig. 3. Feature selection.

These features were instrumental in enhancing the predictive accuracy of our models. For the experimental setup, the dataset was divided into 80% for training and 20% for testing. Typically, thermal comfort ratings in the dataset ranged from 1 to 6. Another hypothesis tested was the conversion of these label values into integers. By reducing the range of thermal comfort ratings from six to three distinct categories, we observed a significant improvement in model accuracy. This transformation simplifies the model's classification task, enabling more precise predictions.

B. Inter Quartile Range (IQR)

The Interquartile Range (*IQR*) is a measure of statistical dispersion that is calculated as the difference between the third quartile ($Q3$) and the first quartile ($Q1$) of a dataset. Mathematically, it is defined as:

$$IQR = Q3 - Q1 \quad (1)$$

where $Q1$ is the median of the lower half of the dataset and $Q3$ is the median of the upper half of the dataset. It is particularly useful in identifying and dealing with outliers, which are data points that significantly differ from the rest of the dataset. Here's how the *IQR* is calculated and how it can be used to remove outliers:

1) Calculation of IQR:

- Firstly, you need to arrange your dataset in ascending order.
- Then, find the median of the dataset, which is the middle value when the data is sorted. If the dataset has an odd number of observations, the median is the middle value. If it has an even number of observations, the median is the average of the two middle values.

- Divide the dataset into two halves at the median. The lower half contains all the values less than or equal to the median, and the upper half contains all the values greater than or equal to the median.
- Find the median of each half. This gives you the first quartile (*Q1*) and the third quartile (*Q3*) of the dataset, respectively.
- The *IQR* is then calculated as the difference between *Q3* and *Q1: IQR = Q3 - Q1*.

2) Identifying outliers using IQR:

- Outliers can be detected using the IQR method by considering values that lie below $Q1 - 1.5 \times IQR$ or above $Q3 + 1.5 \times IQR$. These values are considered to be significantly different from the rest of the dataset.
- Values below $Q1 - 1.5 \times IQR$ or above $Q3 + 1.5 \times IQR$ are commonly referred to as lower and upper bounds, respectively.
- Any data points falling outside these bounds can be considered outliers.

3) Removing outliers using IQR:

- Once outliers are identified using the IQR method, you can choose to remove them from the dataset to improve the robustness of your analysis or model.
- Outliers can be removed by filtering the dataset to exclude any observations that fall outside the lower and upper bounds defined by $Q1 - 1.5 \times IQR$ and $Q3 + 1.5 \times IQR$, respectively.
- After removing outliers, the dataset may be more representative of the underlying distribution and less influenced by extreme values.

4) Considerations:

- While the IQR method is effective in identifying and removing outliers, it's important to exercise caution and consider the context of the data.
- Outliers may sometimes carry valuable information or be indicative of rare but important events. Therefore, the decision to remove outliers should be made judiciously based on the specific goals of the analysis or model.
- Additionally, the choice of the multiplier (1.5 in the conventional method) used to define the bounds can be adjusted depending on the desired level of sensitivity to outliers.

In summary, the *IQR* is a useful statistical measure for assessing the spread of a dataset and identifying outliers. By calculating the *IQR* and defining bounds based on it, outliers can be effectively detected and removed, leading to a more robust analysis or model.

C. Applied Algorithms

SVM is a robust supervised machine learning algorithm well-suited for both classification and regression tasks. In thermal comfort prediction, SVM is employed to delineate the complex interrelationships between various environmental factors—like temperature, humidity, and air velocity—and human thermal comfort responses. The algorithm focuses on maximizing the margin between classes in classification tasks or minimizing the error in regression, all while effectively controlling for overfitting. By training on labeled datasets that encapsulate environmental conditions and corresponding thermal comfort ratings, SVM learns to accurately predict thermal comfort levels based on specific environmental inputs.

Random Forest is another versatile machine-learning algorithm capable of handling both classification and regression challenges. It operates on an ensemble learning principle, utilizing multiple decision trees to construct a more accurate and robust model, as shown in Fig. 4. The process involves extensive data preparation, including cleaning, handling missing values, and appropriate transformations to fit the model. Random Forest uses random sampling to select data subsets for training each tree, employs recursive partitioning for tree creation, and integrates a voting mechanism to aggregate the predictions from various trees. This method is particularly effective in modeling nonlinear relationships and interactions among different environmental variables. Random Forest's ability to generate reliable predictions is enhanced by its ensemble approach, which provides a comprehensive view of thermal comfort across varying conditions.

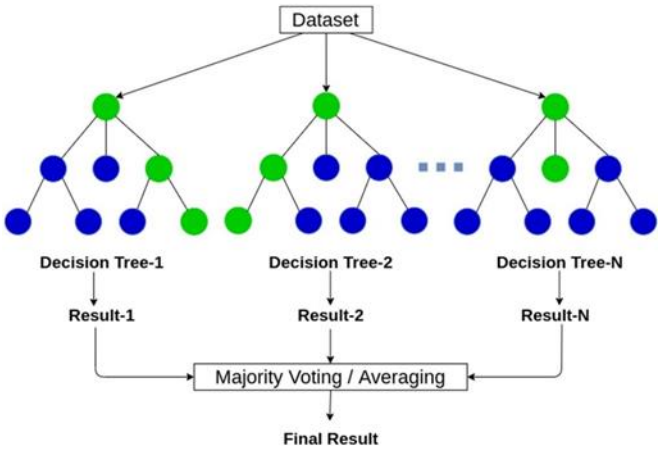


Fig. 4. Multiple decision trees of the Random Forest algorithm.

Both SVM and Random Forest are adept at capturing the nuanced dynamics between environmental parameters and thermal responses, making them invaluable for predicting thermal comfort in diverse settings. These models stand out for their robustness against overfitting, ensuring consistent reliability across different datasets and environmental scenarios. While SVM offers clear decision boundaries facilitating easier interpretation of the factors influencing thermal comfort, Random Forest provides insights into feature

importance through its aggregated decision trees, although individual tree interpretations are less straightforward.

The flexibility of SVM and Random Forest models allows for the accommodation of various data types, making them ideal for integration with different environmental sensors and monitoring systems in thermal comfort assessment. An innovative approach within this domain is utilizing the 'Thermal preference' column as an alternative predictive variable instead of the conventional 'Thermal comfort' scale, moving away from traditional models that categorize comfort into six distinct levels to a more simplified three-level scale, which could potentially streamline the prediction process and enhance model accuracy.

D. Integration with IoT

The IoT component of the system is integral to enhancing building management by deploying a comprehensive network of sensors throughout the facility. These sensors are designed to monitor a variety of environmental conditions in real-time, including temperature, humidity, CO2 levels, and occupancy rates. The data collected by these IoT sensors is then transmitted to a central server, where it is stored and analyzed. For efficient and reliable data transfer, wireless communication protocols such as Wi-Fi, Bluetooth, or LoRaWAN are utilized.

As part of the system design of the controller, a thorough selection of hardware components and parameters was conducted. The designed printed circuit board (PCB) features include:

- A PCB thickness of 1.5 mm;
- A copper foil thickness of 35 μm;
- Glass epoxy laminated with foil;
- Epoxy-urethane varnish;
- Minimum conductor width and spacing of 0.3 mm, with a power bus width of 0.4 mm;
- Minimum hole diameter of 0.352 mm, with mounting hole diameter of 2 mm;
- Geometric dimensions of the board: 50.7 mm × 39.1 mm × 6.81 mm.

Following the selection of the electronic base and PCB parameters, a Raspberry Pi compatible topology was developed using DipTrace, as shown in Fig. 5. All components were positioned as closely as possible to minimize the board's size. Metallized holes were created in the corners of the board for mounting in the enclosure. The connections for power sources were placed on the left side of the board, and the connection for the battery was located on the right side.

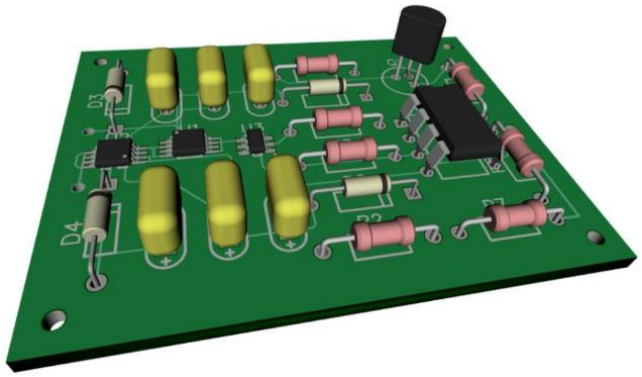


Fig. 5. Schematic of the IoT controller developed using DipTrace software.

The AI models within the system leverage this real-time data to continuously refine their predictions and immediately adjust the building's HVAC system to achieve optimal thermal comfort. A key feature of this setup is its feedback loop mechanism, which plays a critical role in maintaining desired thermal conditions. The AI algorithms actively process the incoming data from the IoT sensors and either make recommendations or directly control the HVAC system's operations, as shown in Fig. 6.

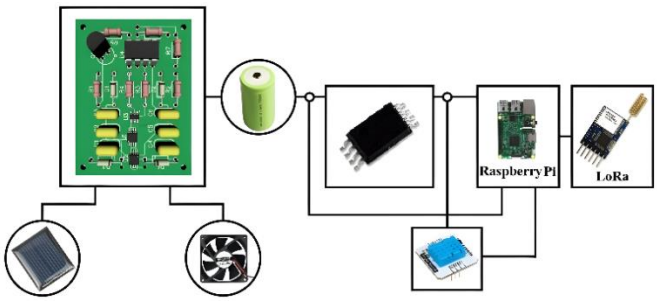


Fig. 6. General design of the system.

The device, powered by a rechargeable battery (referred to as the "slave device" in our model), collects data from sensors and sends this data to the master device. In our case, we use temperature and humidity sensors, which allow for monitoring in the environment to solve specific tasks. The only requirement for using this topology is that all slave devices must be no more than 100 meters away from the master device.

The topology of this model implies that each slave device is only aware of the master device's existence and there is no data transmission between two slave devices. This arrangement eliminates one of the main drawbacks: unreliable communication between two devices. The master device, in turn, is connected to the global network, and therefore, it only structures and redirects the data to the final destination—a database. There is a possibility that the master device may fail. However, nothing prevents connecting two master devices, through which sensor data is transmitted, and writing it into backup databases.

For instance, if the system detects any deviations from set comfort levels, it is programmed to make necessary adjustments to temperature, humidity, or airflow. This dynamic adjustment ensures that thermal comfort is not only achieved but sustained, adapting to both environmental changes and occupancy patterns within the building. A Raspberry Pi connected to a LoRa module serves as the master device. The role of the slave device is performed by a system composed of a microcontroller, a LoRa module, and sensors, all powered by a rechargeable battery. The collection of analog values produced by temperature and humidity sensors is facilitated by an integrated analog-to-digital converter. A fully charged battery can support the operation of the devices for up to 30 days.

The algorithm is implemented on the Raspberry Pi, which emulates the operation of a microcontroller. This device continuously listens on the 866 MHz frequency, which is used for transmitting data from the sensors. Selecting a suitable and efficient microcontroller will be part of future research. Managing comfortable environmental levels can also be controlled through an application as shown in Fig. 7. The interface allows us to easily manage the environment of a room. On the display, it can be seen current room conditions including temperature, humidity, and comfort level. To customize these settings, we can use the "Adjust Settings" section on the right, and the "Apply Settings" button to apply the new conditions.

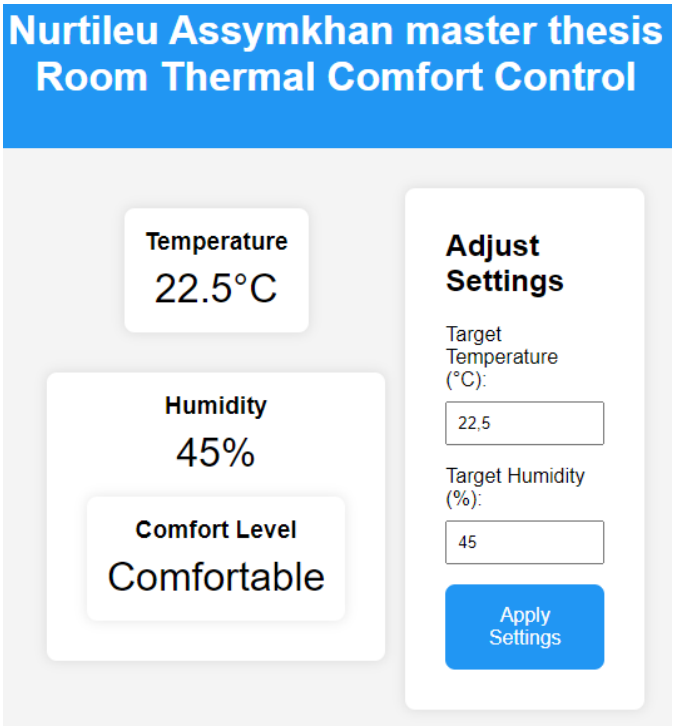


Fig. 7. GUI of the managing application.

IV. Results

After an initial filtering process, our dataset was reduced from 70 to 21 columns. We continued to refine our feature selection by using correlations and deliberately avoided incorporating Fanger's features. Further filtration using both correlation analysis and the SelectKbest model, which assists in identifying the most impactful features, led us to define three distinct sets of features:

- First Set (17 features): Age, Sex, Metabolic rate (Met), Thermal preference, Thermal sensation, Clothing insulation (Clo), Subject's height (cm), Subject's weight (kg), Year, Season, Köppen climate classification, Building type, Cooling strategy at building level, Air temperature (C), Outdoor monthly air temperature (C), Relative humidity (%), and Air velocity (m/s).
- Second Set (9 features): Age, Sex, Met, Clo, Year, Season, Air temperature (C), Relative humidity (%), Air velocity (m/s).
- Third Set (15 features): Age, Clo, Sex, Met, Thermal preference, Year, Season, Köppen climate classification, Cooling strategy at building level, City, Predicted Percentage of Dissatisfied (PPD), Air temperature (C), Outdoor monthly air temperature (C), Relative humidity (%), Air velocity (m/s).

Following the feature selection, our dataset consisted of 17 columns and 6,765 rows. In the initial modeling phase, we utilized all 17 features, which yielded unsatisfactory results. Subsequent iterations with 9 and then 15 of the 17 features also failed to significantly improve outcomes. These iterations allowed us to test our hypotheses; notably, the IQR method enhanced model accuracy by approximately 3-4%, and the method of reducing label values improved accuracy by 20-23%.

Adjusting the parameters of our models led to more promising configurations. For the SVM model, optimal settings were identified as a radial basis function (RBF) kernel with gamma set to 0.001 and C set to 3. For the Random Forest model, the best parameters were found to be 300 estimators with a maximum depth of 15. These parameters maximized accuracy.

Additionally, comparing the impact of using LabelEncoder versus OneHotEncoder on the dataset revealed a difference in performance of 2-4%. This discrepancy influenced our decision to favor OneHotEncoder. Our tests on data standardization, using both the StandardScaler and MinMaxScaler, indicated that standardization did not significantly alter the accuracy, which remained relatively stable. Tables I, II, and III below present the initial results of our prediction efforts, illustrating the performance of each feature set and modeling approach:

TABLE I. Iteration of 17 Features

Model	Accuracy	Precision	Recall	F1 score
SVM	0.509	0.451	0.509	0.436
RF	0.543	0.505	0.543	0.5

TABLE II. Iteration of 9 Features

Model	Accuracy	Precision	Recall	F1 score
SVM	0.507	0.461	0.507	0.438
RF	0.526	0.513	0.526	0.49

TABLE III. ITERATION OF 15 FEATURES

Model	Accuracy	Precision	Recall	F1 score
SVM	0.533	0.448	0.533	0.433
RF	0.54	0.475	0.539	0.482

Based on the initial results, we further pursued enhancing model accuracy by employing the hypotheses formulated earlier in our study. The implementation of the IQR method was a particular focus, aimed at refining the data by removing outliers, which are often a source of prediction error. Tables IV, V, and VI below display the outcomes of applying the IQR method, which has further streamlined the model training process. These tables illustrate the effect of this technique on the overall performance of the models:

TABLE IV. ITERATION OF 17 FEATURES WITH *IQR*

Model	Accuracy	Precision	Recall	F1 score
SVM	0.522	0.44	0.522	0.441
RF	0.548	0.517	0.548	0.504

TABLE V. ITERATION OF 9 FEATURES WITH *IQR*

Model	Accuracy	Precision	Recall	F1 score
SVM	0.507	0.44	0.383	0.424
RF	0.52	0.501	0.52	0.479

TABLE VI. ITERATION OF 15 FEATURES WITH *IQR*

Model	Accuracy	Precision	Recall	F1 score
SVM	0.563	0.539	0.563	0.425
RF	0.57	0.494	0.57	0.5

Building on the improvements, which enhanced model accuracy by approximately 2-5%, our next step involves reducing label values to further increase the accuracy. This simplifies the output space of the model, potentially making it easier for the algorithms to distinguish between different states of thermal comfort. The Tables VII, VIII, IX shows the result of this approach:

TABLE VII. ITERATION OF 17 FEATURES WITH REDUCING LABELS

Model	Accuracy	Precision	Recall	F1 score
SVM	0.715	0.644	0.715	0.614
RF	0.744	0.708	0.744	0.704

TABLE VIII. ITERATION OF 9 FEATURES WITH REDUCING LABELS

Model	Accuracy	Precision	Recall	F1 score
SVM	0.688	0.598	0.688	0.569
RF	0.699	0.657	0.699	0.645

TABLE IX. ITERATION OF 15 FEATURES WITH REDUCING LABELS

Model	Accuracy	Precision	Recall	F1 score
SVM	0.78	0.608	0.78	0.683
RF	0.78	0.719	0.78	0.727

We utilized Random sampling to select subsets of the dataset for training individual decision trees within our Random Forest model. By integrating strategies such as feature reduction, IQR, and Random sampling, we have enhanced the construction and performance of our decision trees. These trees are built using recursive partitioning that methodically splits the data into increasingly specific subsets. This splitting is based on the feature values that most effectively differentiate the categories of the target variable.

The process is further refined through selective feature selection, which concentrates on the most impactful variables. This allows the model to focus on the data elements that are most predictive of the outcomes, significantly enhancing the overall performance of the model. These integrations contribute to a more efficient predictive tool, suitable for complex scenarios in smart building environments. After incorporating the feature-reduced model, further simplifying the feature space, we observed the following results, as in Tables X, XI, XII:

TABLE X. ITERATION OF 17 FEATURE-REDUCED LABELS AND *IQR*

Model	Accuracy	Precision	Recall	F1 score
SVM	0.726	0.598	0.726	0.621
RF	0.733	0.678	0.733	0.688

TABLE XI. ITERATION OF 9 FEATURE-REDUCED LABELS AND *IQR*

Model	Accuracy	Precision	Recall	F1 score
SVM	0.706	0.498	0.706	0.584
RF	0.717	0.668	0.717	0.653

TABLE XII. ITERATION OF 15 FEATURE-REDUCED LABELS AND *IQR*

Model	Accuracy	Precision	Recall	F1 score
SVM	0.835	0.697	0.835	0.76
RF	0.821	0.738	0.821	0.766

The implications of these findings are significant, especially in the context of predictive accuracy in environmental modeling for predicting thermal comfort levels in smart building systems. The Receiver Operating Characteristic (ROC) curves graph, presented in Fig. 8, provide a visual comparison of the performance of two machine learning models: SVM and Random Forest (RF). These curves are essential tools in evaluating the models by plotting the True Positive Rate (sensitivity) against the False Positive Rate (1-specificity) at various threshold settings. The area under the curve (AUC) serves as a summary measure of the model's ability to discriminate between positive and negative classes.

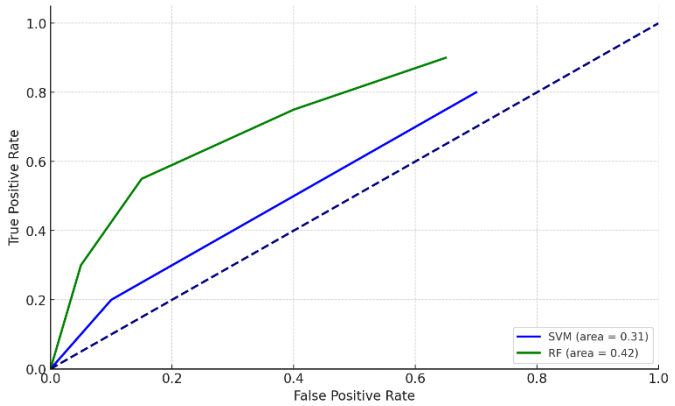


Fig. 8. ROC comparison for both the SVM and RF.

In this analysis, the SVM model demonstrates an AUC of 0.72, while the RF model exhibits a slightly superior AUC of 0.84. This suggests that the RF model has a better overall performance in distinguishing between the classes under study, likely due to its ensemble nature, which typically provides a more robust prediction by averaging multiple decision processes.

The following boxplots, depicted in Fig. 9, below show the distribution of cross-validation accuracy scores for SVM and RF models across various feature sets and conditions. The SVM exhibits a broader range of accuracy variations, especially with the 15-feature set. The increase in accuracy when reducing label values suggests that SVM benefits significantly from a simplified output space, potentially due to reduced complexity in the decision boundary formation. The RF shows tighter accuracy distributions and higher median accuracies across all feature sets, indicating better stability and robustness. The performance improvements with label reduction demonstrate RF's effectiveness in handling more straightforward, cleaner data. Further exploring more sophisticated data preprocessing techniques such as feature scaling, transformation (like log transformation for skewed data), or anomaly detection methods could help to handle outliers more dynamically than just applying IQR.

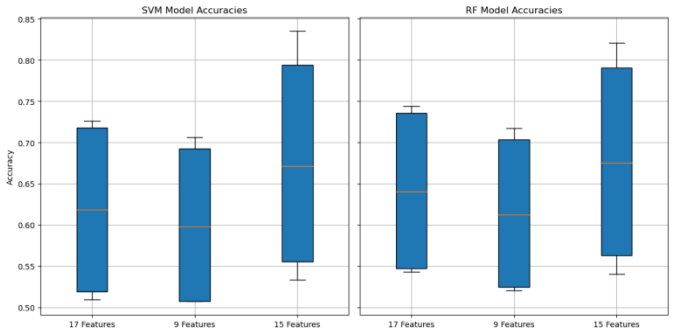


Fig. 9. Boxplots of cross-validation accuracy scores.

The implications of these findings are significant, especially in the context of predictive accuracy in environmental modeling, such as predicting thermal comfort levels in smart buildings. The higher AUC for the RF model indicates a higher likelihood of correctly classifying the thermal comfort levels as satisfactory or unsatisfactory, which is crucial for developing systems that can dynamically adjust to maintain or achieve desired comfort states. Moreover, the relatively lower performance of the SVM could be attributed to its sensitivity to the choice of kernel and the tuning of its parameters, which might not have been optimal in this scenario. These insights not only aid in selecting the appropriate model for deployment but also highlight the importance of model tuning and feature selection, reinforcing the need for ongoing model adjustment in practical applications to achieve the best outcomes.

V. Discussion

This research assesses the effectiveness of Random Forest and SVM algorithms across different feature sets in predicting thermal comfort and thermal preference. We introduced eight new features in our analysis, while seven features were consistent with those used in prior studies. When comparing the outcomes for guessing Thermal comfort versus Thermal preference, the performance gap between them was relatively narrow, ranging from 1-3%, with Random Forest generally exhibiting greater stability.

Specifically, in the scenarios where we tested sets with 9 and 15 features, alternative versions of our models initially led in performance. However, a significant shift occurred when we simplified the prediction scale from six to three Thermal comfort values, which resulted in our primary model configuration achieving superior results. This simplification appeared to enhance the model's ability to discriminate between different levels of comfort effectively.

Moving forward, we plan to incorporate additional features to refine the models' accuracy further. One promising candidate is Heart Rate Variability (HRV), which has potential implications for assessing physiological responses to thermal environments. Moreover, I am keen to explore the capabilities of neural networks and deep learning techniques, inspired by their success in related fields. My intention is to experiment with various advanced algorithms such as Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), Long Short-Term Memory (LSTM) networks, Autoencoders, and Deep Belief Networks (DBNs). These methodologies, referenced in papers [18] and [10], will form the basis of future research efforts aimed at enhancing the predictive accuracy of thermal comfort models.

While the reduction of the thermal comfort scale from six levels to three markedly improved the models' discriminative capabilities, some researchers argue that this simplification might obscure subtle nuances in human comfort perception [24]. Critics suggest that while a simpler output space indeed facilitates more accurate classifications by reducing the complexity the model must manage, it potentially oversimplifies human experiences, which could be better captured with a more granular scale [25].

The introduction of the IQR method led to an approximate 3-4% improvement in model accuracy. However, it is essential to note that while IQR can effectively reduce outlier influence, it may bring limitations to valid extreme cases that are crucial for understanding the full spectrum of environmental impacts on thermal comfort [26]. The more substantial impact came

from reducing label values, which boosted accuracy by 20-23%. This dramatic increase underscores the pivotal role of thoughtful statistical techniques in predictive model development. Yet, there remains a debate over whether such methods compromise the depth of data insights for the sake of model performance [27].

The system's architecture, employing a master-slave device setup, has shown efficient data management and transmission capabilities. Slave devices, powered by rechargeable batteries and strategically placed not more than 100 meters from a Raspberry Pi master device, efficiently transmit sensor data. However, some experts raise concerns about the scalability and maintenance of such setups in larger or more complex building environments [28]. Critics also question the reliance on Raspberry Pi for critical real-time data processing, citing potential limitations in processing power and storage compared to more robust computing solutions [29].

The capability for direct user interaction with the building management system through an application is hailed for its user-centered design, merging comfort with energy efficiency. Nevertheless, this approach raises questions about the trade-offs between user control and automated system efficiency, with some arguing that excessive user interaction might lead to less optimal energy use [30].

This study's significant contributions to environmental control and smart building management highlight the intersection of advanced computational techniques with practical IoT implementations. Yet, the ongoing exploration of new features and modeling techniques also points to a field that is constantly evolving, with ongoing debates about the best balance between accuracy, user experience, and system reliability [31]. We hope, these discussions are crucial as they push the boundaries of what smart building systems can achieve, ensuring they meet both current and future demands effectively.

VI. Conclusion

This study has explored the application of Random Forest and SVM algorithms to predict thermal comfort and preference, utilizing a refined feature set that integrates both newly introduced variables and established ones from prior research. The performance differential between predicting thermal comfort and thermal preference was relatively minimal, typically within 1-3%, with Random Forest demonstrating superior stability and robustness across varied feature sets. A pivotal enhancement in model performance was observed when the complexity of the thermal comfort scale was reduced from six to three levels, which notably improved the model's ability to discriminate between different comfort states more effectively.

To address the shortcomings in existing research, our paper showcases the advantages of the proposed techniques over traditional methods. For instance, the paper introduces clever adjustments to the thermal comfort prediction models, such as the reduction of the thermal comfort scale from six to three levels, which has shown to improve the accuracy of SVM and Random Forest models. This simplification not only enhances model precision but also makes these models more adaptable to various data types, which is essential for integrating environmental sensors in building management systems. Additionally, the novel use of the 'Thermal preference' column as a predictive variable instead of the standard 'Thermal comfort' scale offers a more streamlined and effective approach to predicting thermal comfort. By providing a thorough comparative analysis of these modifications against conventional methods, the paper highlights the practical implications in improving thermal comfort assessments.

While the simplification of the thermal comfort scale has yielded significant improvements, it also raises questions about the potential limitations of oversimplifying the nuances of human thermal perception. The use of the IQR method has also shown to improve model accuracy modestly; however, its tendency to remove valid extreme data points could limit understanding the broader impacts of environmental variables on thermal comfort. Moreover, the substantial increase in accuracy from reducing label values underscores the critical role of sophisticated statistical techniques in developing effective predictive models, though this approach has sparked debate regarding the depth and granularity of data interpretation.

The architectural design of our IoT-based system, featuring a master-slave configuration, has proven effective in data management and transmission, albeit with some concerns about scalability and dependency on limited-capability devices like the Raspberry Pi for critical processing tasks. Additionally, the system's design allowing direct user interaction via an application exemplifies a user-centered approach that harmoniously blends comfort with energy efficiency, though it also invites scrutiny over the potential for suboptimal energy usage due to excessive manual interventions.

Future enhancements are planned through the integration of additional predictive variables such as HRV, which holds promise for assessing physiological responses to varying thermal conditions. The potential of neural networks and deep learning will also be explored to leverage their proven capabilities in similar domains. Techniques such as CNNs, LSTM networks, and DBNs will be investigated to further refine the accuracy and efficiency of our models.

The contributions of this research to the fields of environmental control and smart building management are significant, illustrating the powerful synergy between advanced computational methods and practical IoT implementations. The ongoing development and refinement of these models push the boundaries of what smart building systems can achieve, ensuring they not only meet current demands but are also well-prepared for future challenges.

References

[1] G. M. Milis, C. G. Panayiotou and M. M. Polycarpou, "IoT-Enabled Automatic Synthesis of Distributed Feedback Control Schemes in Smart Buildings," in *IEEE Internet of Things Journal*, vol. 8, no. 4, pp. 2615-2626, 15 Feb.15, 2021, doi: 10.1109/JIOT.2020.3019662.

[2] S. I. Khather, M. A. Ibrahim, and A. I. Abdullah, "Review and performance analysis of nonlinear model predictive control—current prospects, challenges and future directions," *Journal Européen des Systèmes Automatisés*, vol. 56, no. 4, pp. 593-603, 2023. doi: 10.18280/jesa.560409.

[3] B. Liu, M. Akcakaya and T. E. Mcdermott, "Automated Control of Transactive HVACs in Energy Distribution Systems," in *IEEE Transactions on Smart Grid*, vol. 12, no. 3, pp. 2462-2471, May 2021, doi: 10.1109/TSG.2020.3042498.

[4] H. Wang, M. R. Amini, Q. Hu, I. Kolmanovsky and J. Sun, "Eco-Cooling Control Strategy for Automotive Air-Conditioning System: Design and Experimental Validation," in *IEEE Transactions on Control Systems Technology*, vol. 29, no. 6, pp. 2339-2350, Nov. 2021, doi: 10.1109/TCST.2020.3038746.

[5] Z. A. Shah, H. F. Sindi, A. Ul-Haq and M. A. Ali, "Fuzzy Logic-Based Direct Load Control Scheme for Air Conditioning Load to Reduce Energy Consumption," in *IEEE Access*, vol. 8, pp. 117413-117427, 2020, doi: 10.1109/ACCESS.2020.3005054.

[6] L. Morales Escobar, J. Aguilar, A. Garcés-Jiménez, J. A. Gutierrez De Mesa and J. M. Gomez-Pulido, "Advanced Fuzzy-Logic-Based Context-Driven Control for HVAC Management Systems in Buildings," in *IEEE Access*, vol. 8, pp. 16111-16126, 2020, doi: 10.1109/ACCESS.2020.2966545.

[7] D. Azuatalam, W.-L. Lee, F. de Nijs, and A. Liebman, "Reinforcement learning for whole-building HVAC control and demand response," *Energy and AI*, vol. 2, 2020, doi: 10.1016/j.egyai.2020.100020

[8] H. Mansy and S. Kwon, "Optimal HVAC Control for Demand Response via Chance-Constrained Two-Stage Stochastic Program," in *IEEE Transactions on Smart Grid*, vol. 12, no. 3, pp. 2188-2200, May 2021, doi: 10.1109/TSG.2020.3037668.

[9] A. Rabinowitz, F. M. Araghi, T. Gaikwad, Z. D. Asher, and T. H. Bradley, "Development and evaluation of velocity predictive optimal energy management strategies in intelligent and connected hybrid electric vehicles," *Energies*, vol. 14, 5713, 2021. doi: 10.3390/en14185713.

[10] F. H. Mohamed Salleh, M. b. Saripuddin and R. bin Omar, "Predicting Thermal Comfort of HVAC Building Using 6 Thermal Factors," *2020 8th International Conference on Information Technology and Multimedia (ICIMU)*, Selangor, Malaysia, 2020, pp. 170-176, doi: 10.1109/ICIMU49871.2020.9243466.

[11] N. Morresi *et al.*, "Sensing Physiological and Environmental Quantities to Measure Human Thermal Comfort Through Machine Learning Techniques," in *IEEE Sensors Journal*, vol. 21, no. 10, pp. 12322-12337, 15 May15, 2021, doi: 10.1109/JSEN.2021.3064707.

[12] R. Widiastuti, J. Zaini, W. Caesarendra, D. Shona Laila and J. Candra Kurnia, "Prediction on the Indoor Thermal Comfort of Occupied Room Based on IoT Climate Measurement Open Datasets," *2020 International Conference on Informatics, Multimedia, Cyber and Information System (ICIMCIS)*, Jakarta, Indonesia, 2020, pp. 40-45, doi: 10.1109/ICIMCIS51567.2020.9354277.

[13] Z. Zhang, B. Lin, Y. Geng, H. Zhou, X. Wu, and C. Zhang, "The effect of group perception feedbacks on thermal comfort," *Energy and Buildings*, vol. 254, 2022, doi: 10.1016/j.enbuild.2021.111603.

[14] F. Hani Mohamed Salleh and M. binti Saripuddin, "Monitoring Thermal Comfort Level of Commercial Buildings' Occupants in a Hot-Humid Climate Country Using K-nearest Neighbors Model," *2020 5th International Conference on Power and Renewable Energy*, Shanghai, China, 2020, pp. 209-215, doi: 10.1109/ICPRE51194.2020.9233145.

[15] M. Khalil, M. Esseghir and L. Merghem-Boulahia, "An IoT Environment for Estimating Occupants' Thermal Comfort," *2020 IEEE 31st Annual International Symposium on Personal, Indoor and Mobile Radio Communications*, London, UK, 2020, pp. 1-6, doi: 10.1109/PIMRC48278.2020.9217157.

[16] X. Zhang, M. Pipattanasomporn, T. Chen and S. Rahman, "An IoT-Based Thermal Model Learning Framework for Smart Buildings," in *IEEE Internet of Things Journal*, vol. 7, no. 1, pp. 518-527, Jan. 2020, doi: 10.1109/JIOT.2019.2951106.

[17] K. Irshad, A. I. Khan, S. A. Irfan, M. M. Alam, A. Almalawi and M. H. Zahir, "Utilizing Artificial Neural Network for Prediction of Occupants Thermal Comfort: A Case Study of a Test Room Fitted With a Thermoelectric Air-Conditioning System," in *IEEE Access*, vol. 8, pp. 99709-99728, 2020, doi: 10.1109/ACCESS.2020.2985036.

[18] G. Gao, J. Li and Y. Wen, "DeepComfort: Energy-Efficient Thermal Comfort Control in Buildings Via Reinforcement Learning," in *IEEE Internet of Things Journal*, vol. 7, no. 9, pp. 8472-8484, Sept. 2020, doi: 10.1109/JIOT.2020.2992117.

[19] N. Smatov, R. Kalashnikov, and A. Kartbayev, "Development of context-based sentiment classification for intelligent stock market prediction," *Big Data Cogn. Comput.*, vol. 8, 51, 2024. doi: 10.3390/bdcc8060051.

[20] Z. Wang, H. Onodera and R. Matsuhashi, "Proposal of Relative Thermal Sensation: Another Dimension of Thermal Comfort and Its Investigation," in *IEEE Access*, vol. 9, pp. 36266-36281, 2021, doi: 10.1109/ACCESS.2021.3062393.

[21] N. Cibin, A. Tibo, H. Golmohamadi, A. Skou, and M. Albano, "Machine learning-based algorithms to estimate thermal dynamics of residential buildings with energy flexibility," *Journal of Building Engineering*, vol. 65, 2023, doi: 10.1016/j.jobe.2022.105683.

[22] C. Miller, B. Picchetti, C. Fu, and J. Pantelic, "Limitations of machine learning for building energy prediction: ASHRAE Great Energy Predictor III Kaggle competition error analysis," *Science and Technology for the Built Environment*, vol. 28, no. 5, pp. 610-627, 2022. doi: 10.1080/23744731.2022.2067466.

[23] S. Biswas and H. Rajan, "Fair preprocessing: towards understanding compositional fairness of data transformers in machine learning pipeline," in *Proc. 29th ACM Joint Meeting on European Software Engineering Conference and Symposium on the Foundations of Software Engineering (ESEC/FSE 2021)*, New York, USA, 2021, pp. 981–993. doi: 10.1145/3468264.3468536.

[24] Q. Zhao, Z. Lian, and D. Lai, "Thermal comfort models and their developments: A review," *Energy and Built Environment*, vol. 2, no. 1, pp. 21-33, 2021. doi: 10.1016/j.enbenv.2020.05.007.

[25] C. Aliferis and G. Simon, "Overfitting, underfitting and general model overconfidence and under-performance pitfalls and best practices in machine learning and AI," in *Artificial Intelligence and Machine Learning in Health Care and Medical Sciences*, Springer, 2024, ch. 10. doi: 10.1007/978-3-031-39355-6_10.

[26] N. Eslamirad, A. Sepúlveda, F. De Luca, and K. Sakari Lylykangas, "Evaluating outdoor thermal comfort using a mixed-method to improve the environmental quality of a university campus," *Energies*, vol. 15, no. 4, 1577, 2022. doi: 10.3390/en15041577.

[27] A. Hassan, "An effective ensemble-based framework for outlier detection in evolving data streams," *International Journal of Advanced Computer Science and Applications*, vol. 13, pp. 315-329, 2022. doi: 10.14569/IJACSA.2022.0131135.

[28] S.-Y. Chuang, N. Sahoo, H.-W. Lin, and Y.-H. Chang, "Predictive maintenance with sensor data analytics on a Raspberry Pi-based experimental platform," *Sensors*, vol. 19, 3884, 2019. doi: 10.3390/s19183884.

[29] A. Carvalho, C. Machado and F. Moraes, "Raspberry Pi Performance Analysis in Real-Time Applications with the RT-Preempt Patch," *2019 Latin American Robotics Symposium (LARS), 2019 Brazilian Symposium on Robotics (SBR) and 2019 Workshop on Robotics in Education (WRE)*, Rio Grande, Brazil, 2019, pp. 162-167, doi: 10.1109/LARS-SBR-WRE48964.2019.00036.

[30] A. M. Coutinho Demetrios, D. De Sensi, A. F. Lorenzon, K. Georgiou, J. Nunez-Yanez, K. Eder, and S. Xavier-de-Souza, "Performance and energy trade-offs for parallel applications on heterogeneous multi-processing systems," *Energies*, vol. 13, 2409, 2020. doi: 10.3390/en13092409.

[31] R. Almutairi, G. Bergami, and G. Morgan, "Advancements and challenges in IoT simulators: A comprehensive review," *Sensors*, vol. 24, 1511, 2024. doi: 10.3390/s24051511.

Exploration of Deep Semantic Analysis and Application of Video Images in Visual Communication Design Based on Multimodal Feature Fusion Algorithm

Yanlin Chen, Xiwen Chen*
College of Visual Arts, Hunan Mass Media Vocational and Technical College, Changsha 410000, China

***Abstract*—Fully utilizing image and video semantic processing techniques can play a more effective role in visual communication design. In order to further explore the application of multimodal feature fusion algorithm (MFF) in video image feature analysis in visual communication design, with the aim of enhancing the depth and breadth of design creation. This article focuses on the application of video semantic understanding technology by combining image and video semantic processing techniques, in order to achieve a comprehensive, three-dimensional, and open expansion of design thinking. The MFF algorithm was proposed and implemented, which innovatively integrates multimodal information such as visual and audio in videos, deeply explores action semantics, and shows significant performance improvements compared to traditional algorithms. Specifically, compared to the other two mainstream algorithms, its performance has improved by 24.33% and 14.58%, respectively. This discovery not only validates the superiority of MFF algorithm in the field of video semantic understanding, but also reveals the profound impact of video semantic understanding technology on visual communication design practice, providing new perspectives and tools for the design industry and promoting innovation and development of design thinking. The novelty of this study lies in its interdisciplinary methodology, which applies advanced algorithm techniques to the field of art and design, and the significant improvement of the proposed MFF algorithm in enhancing design efficiency and creativity.**

***Keywords*—Video semantics; understanding; visual communication; design**

I. Introduction

Under the background of new media, visual communication design has become a comprehensive discipline, which promotes the emergence of new design art forms, and also changes the traditional design concept and thinking mode, which has been widely used in all aspects of life [1]. The video semantic understanding technology is updating day by day, which plays a driving role in the development of visual communication design, and promotes the information design to have different meanings and effects. In varying degrees, it has an impact on visual effects in various fields, becoming a role of visual culture construction, and also a designer of the new era [2]. In the process of visual design, there is an inseparable relationship between the state of graphic design and its aesthetic feeling. From the graphic chronicle in ancient times to the evolution of graphic symbols and character symbols today, the rule of graphic language has realized a good interaction between simple and complex, and has actually become one of the indispensable decorative languages in people's lives [3]. In a design work, almost all images can appear in the form of points, surfaces, lines, and the comprehensive application of points, lines and surfaces, which is flexible [4]. Moreover, the design should follow such a principle that visual graphics must be clearly identifiable, on the contrary, when a graphic structure is fuzzy, the meaning it conveys must be fuzzy [5]. Graphical symbols of visual design, as well as visual representations such as pattern modeling, pictures, signs, etc. aimed at the design theme, are called graphic information through the content, feelings and visual impact conveyed by these graphics [6]. In the visual design language, the modeling elements are points, lines, surfaces, tones and materials. Any kind of graphic information, such as trademarks, patterns, etc., can be understood as a specific modeling element [7]. Therefore, in visual design, a Fig. 1, a pattern or a style of symbols should be used accurately. Only in this way can they convey the necessary information [8].

With the rapid development of digital media technology, visual communication design is no longer limited to traditional static images and text, but is gradually evolving towards dynamic, multidimensional, and interactive directions. Video, as one of the most expressive and infectious forms of media, plays an increasingly important role in visual communication design. However, current video semantic understanding technology cannot fully meet the needs of visual communication design for accurate and efficient information extraction and expression, especially when dealing with complex scenes and multimodal data, there is a significant research gap. This study aims to fill this research gap by exploring the application of video semantic understanding technology in visual communication design and proposing a new method based on multimodal feature fusion algorithm (MFF). This algorithm aims to integrate multi domain scene information, especially action semantic information, in videos to achieve more accurate and comprehensive video content understanding and analysis. Through this method, we hope to enrich the thinking and expression methods of visual communication design, making design thinking more diverse and extensive, while improving the efficiency of information communication and aesthetic experience of design works. In

*Corresponding Author.

the process of using visual design language to communicate, designers mostly pay attention to and study how to convey "explicit" visual information [9]. In fact, these explicit visual information communications are established on the basis of "known information", and these "known information" are often ignored by designers, let alone effectively used. These "preset information" are just the basis and prerequisite for importing the visual information to be conveyed. Only by effectively combining this implicit known information with the explicit visual information to be conveyed can an effective visual communication process be generated [10]. If the traditional way of communication is adopted, the relevant staff need to fuse images, words, etc. to realize the transmission of visual information [11]. Through dynamic visual transmission, the transmission speed can be improved. At the same time, it allows people to obtain accurate information resources in a short time and observe the way of information transmission, while the dynamic way of transmission requires the assistance of video semantic understanding technology. Under the video semantic understanding technology, through dynamic transmission, the abstract visual works are more vividly presented in the eyes of the audience, so that the audience has a more profound memory [12]. Compared with traditional static images, the dynamic design of visual communication has more advantages to avoid visual fatigue of the audience, increase the sense of interest for the audience, and make the information communication effect good.

Semantic understanding of video is the highest level of research in machine vision. Firstly, video is processed and analyzed, and then the main content of video is described. Different video meanings and tasks require different working situations, and different objects appear. Of course, different visual comprehension algorithms are needed. In a certain period of time, multimedia data such as video frames, audio signals, and transcribed texts may not appear at the same time, and there is an unsynchronized phenomenon, but they all share a semantic meaning and are coupled with each other within the semantic duration [13]. For example, different shots that express the same meaning may look very different visually. Swimming and football, which also represent "sports", are mainly composed of blue swimming pool water and green football field grass. However, text features may express more similarities, thus making up for the weak correlation of other modes [14].

The introduction of this article first emphasizes that in the context of new media, visual communication design has become a comprehensive discipline, which not only promotes the emergence of new design art forms, but also changes traditional design concepts and ways of thinking, and is widely applied in various aspects of life. This clarifies the important position and influence of visual communication design in contemporary society. Furthermore, the introduction points out that the increasingly updated video semantic understanding technology has played an important role in promoting the development of visual communication design, and has facilitated information design to have different meanings and effects. This further emphasizes the positive impact of technological progress on the field of design. The contribution points of research innovation are as follows:

- This study proposes a multimodal feature fusion algorithm (MFF), which demonstrates unique advantages in video semantic understanding. By integrating semantic information of multi domain scene actions in videos, MFF algorithm can comprehensively understand video content, which is an important innovation for traditional video processing algorithms.
- The research has improved the application effect of video semantic understanding technology in visual communication design. The application of MFF algorithm in the field of visual communication design has achieved an organic combination of video semantic understanding technology and design art.
- Compared to the other two mainstream algorithms, the MFF algorithm has improved performance by 24.33% and 14.58%, respectively. This significant improvement not only validates the effectiveness of the MFF algorithm.
- Research the application of video semantic understanding technology in visual communication design from the perspective of semantic analysis. The update of this design concept helps designers better grasp design trends and create more innovative and artistic works.

II. Related Work

Ren believes that visual thinking is based on the information of the objective image itself, and uses visual thinking to capture the characteristics of the shape as a means to connect the abstract shape with the semantics of the concrete objective image [15]. Min, Wang, Rushan emphasized that in visual communication, the meaning of shape is generated, communicated, fed back in the "person shape person" system, and then generated, communicated, and fed back again, and so on [16]. Therefore, in the video semantics, Chang and others believe that visual communication design is not only a simple combination of words, sounds and pictures, but also a combination effect of three or more communication elements. The diversified design helps both the transmission and reception of video news achieve ideal effects [17]. Xue has shown in his research that graphics are one of the important elements in visual communication design, and each designer has different performance for the use of graphic design in visual communication. Because people live in different geographical environments, their cultural cognition is also different. However, through the interpretation of graphics and visual transformation, barriers and barriers in communication can be effectively solved, so that visitors can more accurately understand the design connotation [18]. Ge H regards multi-level semantic concepts as the hidden state of Markov chain, and regards multiple basic visual semantics with time semantic context constraints as the observable symbols corresponding to the hidden semantic state. Under this condition, using hierarchical hidden Markov model, the analysis and extraction process of multi granularity visual semantics can be transformed into the process of analyzing the most likely hidden state of known observable symbols [19]. In

the research of video semantic understanding, features of different modes cannot be directly analyzed and compared in the semantic analysis process. How to mine the complementary association between different modes and extract and construct data information consistent with the semantic content has become a problem to be solved. Bae J proposes an extended direct push support tensor machine, which is used as a semantic classifier to detect the semantic concept of video shots [20]. Xin thinks that video data contains a large amount of valuable information. However, due to semantic barriers, human beings have been unable to make full use of this information. However, the birth of data mining technology can help people make full use of this information [21]. Xiao et al. proposed a feature extraction method (Dense Trajectory, DT) based on dense trajectories. This method samples each frame of video intensively, then tracks the optical flow information of these sampling points, and obtains more accurate feature expression through screening and optimization [22]. Subsequently, Natalia et al. proposed an improved Dense Trajectory (iDT) algorithm, which mainly optimizes optical flow images, and improves the regularization and feature encoding methods at the same time, further improving the algorithm performance [23]. Wu et al. proposed a video humanoid parsing algorithm based on semantic transfer. This method uses the complementarity of image semantics and video semantics to propose a module that can integrate two semantics at the same time, and applies it in the convolutional neural network structure [24].

III. Theories Related to Video Semantic Understanding Technology

A. Video Structure

Faced with a variety of complex information, traditional information management methods have been stretched to the limit, and human beings must develop new technologies to help themselves deal with information [25,26]. Traditional video management is basically realized by the underlying features. However, many years of practice and practical experience tell us that video management must analyze the semantics of video. Video language, as a camera technology applied in video, can give people a strong impact on vision. Video image language not only has a passionate expression, but also has a calm way of thinking, bringing more visual feelings to people. Using video language in visual communication design can make advertisements and posters more personalized and aesthetic [27,28]. By combining video language with visual communication design, designers can make better use of film and television effects when designing advertisements, bringing more visual impact to people. A video is composed of continuous image changes. When the image changes more than 24 frames per second, the human eye can no longer distinguish whether the received picture information is still a static image. It looks like a series of fluent pictures. This is called video (Fig. 1). In general, video is composed of continuous image sequences, and the semantic association between each image constitutes the semantic information of the whole video. Moreover, the video data has a much larger capacity than the text data (Table I).

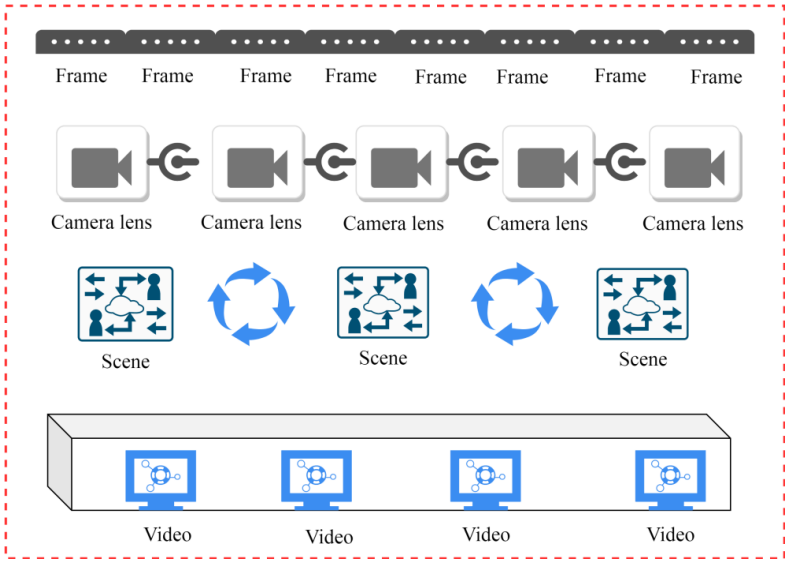


Fig. 1. The composition of each level of the video structure.

Video is mainly divided into structure and composition of each level. First, the adjacent video frames in the time dimension are divided into multiple different shot units without intersection. Key frames of shot units are extracted. The number of key frames extracted for each shot depends on the situation of the shot itself. Observe the video key frames and determine the combination of shots to get the video scene.

TABLE I. Comparison of Text Data and Video Data

Features	Text data	Video data
Symbol set	Limited	Infinite
Resolving power	Low	High
Explanation ambiguity	Low	High
Interpretative function	Low	High
Data capacity	Small	Big

Semantic differences make it impossible for humans to directly apply the information of video data. There is a gap between the low-level features of images that can only be obtained by computers and the high-level semantics. When watching video, human can judge the events in the video by integrating their own knowledge. The computer has limited ability to obtain image features such as image color and texture, but has no ability to judge. Generally speaking, video semantics can be divided into three levels: low-level semantics, middle level semantics and high-level semantics. The semantics of the bottom layer includes the semantics of all the features of the video bottom layer, including the description of color, texture, edges, dynamics and other features. The middle level includes the description of video target action, density, traffic statistics, etc. The high-level semantics is the description of video event attributes. According to this division, we get the video semantic architecture of Fig. 2. In the multimodal analysis method of video semantic understanding, fusion is an indispensable step. Early integration and late integration are the two main integration methods at present. Most existing semantic concept detection methods are based on these two schemes. Pre-fusion refers to combining the features of single modality before learning and training, while post-fusion refers to learning the features of single modality separately, and then fusing and analyzing the results of multiple modalities.

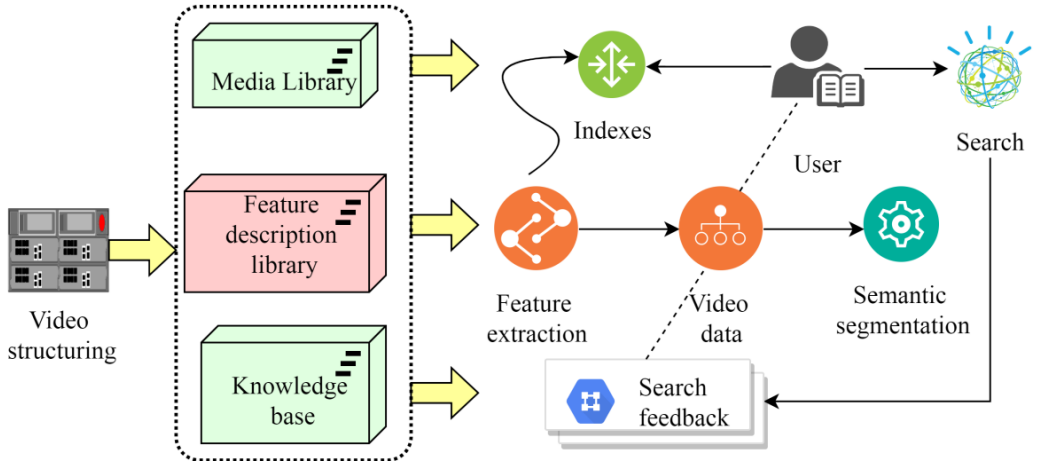


Fig. 2. Multi modal fusion steps for video semantic understanding.

In general, it is natural for an event model to have two or more state variables from a semantic perspective, forming a state chain over time. Factorization of state space into polymorphic chains is another way to simplify the event model. These multiple chains can correspond to sub-events that interact with one compound event or multiple objects at the same time.

B. Application in Visual Communication Design

In the process of visual communication design activities, designers can greatly enhance the user's emotional experience, realize the interaction of multiple senses, and let the audience experience in an all-round way through the application of video semantic understanding technology. Under the background of new media, designers can avoid aesthetic fatigue by reasonably designing graphics, words and videos, and communicate them in the eyes of people by animation, which has a good sense of experience. The central meaning of visual design is "the desire of artistic Fig. materialization". The whole design should be expressed around people's wishes, and it must have human visual visibility and readability. To make the semantics of visual form accurate, we must first understand the potential language of each design factor, understand the design principles, text semantics, graphic semantics, color semantics and psychological implications, and so on, so as to comprehensively convey information. Visual communication cannot be conveyed through people's mouth like people use language, but depends on the media: "shape" is conveyed in the system of people, shape and people. Therefore, it is necessary for people to strictly control and study the rules of "the meaning of form". The semantic meaning of the form with visual communication as the main content (such as signs, advertisements, posters, publicity cards, etc.) mainly conveys the idea of the producer. The semantic meaning of the shape with the use function as the main content (such as cars, rain gear, etc.) makes the user clear and easy to use through the information displayed by the shape, color, etc.

Visual communication design is a complete process of information design and dissemination. Under the influence of digital media technology, great changes have taken place in the mode of information dissemination and reception in people's lives. In the process of the integration of visual communication technology with computer and digital media, the communication mode has changed from dynamic to static, from passive to active, which leads to certain communication characteristics of visual communication design.

IV. Introduction to Algorithms

A. Semantic Concept Modeling

We obtain the contribution score of each action sample to its action semantic concept by preprocessing the classifier, and according to the score, integrate the sample subsets into a fusion vector encoding multi-domain invariant action information. This process can be modeled as:

$$G(x,y)=\sum_{i=1}^{t} r_{ij}(x_i y_j) \tag{1}$$

Where, x_i is the embedded matrix corresponding to the i field, and r_{ij} is the weight of action sample $x_i y_j$.

In order to further explore the potential association between multiple action semantic concepts, we use constraints to limit the number of non-zero rows in matrix W to control the number of sparse features in the model. Constraints are modeled as:

$$cons(R)=\sum_{i=0} I(\|w_i\|\rangle 0)\le u \tag{2}$$

$$P_r(B_r|x)=\frac{p_r(x|w_1)}{\sum_{j=1}^{r} p_j(w_1)} \tag{3}$$

When the Mercer condition is satisfied, replacing the dot product in the optimal classification surface with the inner product S is equivalent to transforming the original input space into a new feature space, and the corresponding high-level multimodal fusion function is:

$$f(x)=\mathrm{sgn}\left[\sum_{i=1}^{n} a_i k(s_i,s)+b\right] \tag{4}$$

Where, $a_i\rangle 0$ is the lagrange coefficient and b is the domain value of classification.

Because of the potential correlation between semantics, multi semantic modeling is essentially a process of establishing mathematical models for interrelated tasks. Traditional machine learning methods often adopt single task learning mode, and learn a model independently for each task. This single task learning method ignores the association between multiple tasks (semantic concepts), and loses hidden information existing between data or model parameters.

B. Recurrent Neural Networks

Recurrent Neural Networks (RNN) are often used to process sequence data, such as time dependent voice data, video data, and natural language data with semantic coherence. Like convolutional neural network, cyclic neural network also uses the idea of weight sharing, so similar to convolutional neural network, it can process images of different sizes, and cyclic neural network can be used to process sequence data of any length. The basic principle of recurrent neural network can be expressed as:

$$h_t = f_w(h_t - 1)\cdot x_t \tag{5}$$

Where, h_t is the hidden state information representation vector of time t, encoding the sequence information input at the first t times.

Fig. 3 shows a multi-layer recurrent neural network. Single layer refers to the number of layers of the recursive function itself, not the number of moments of the expanded graph. Multilayer recurrent neural network, that is, recursive function, has the form of multilayer network. Multilayer networks have more nonlinear structures and are often used to model sequences with complex structures.

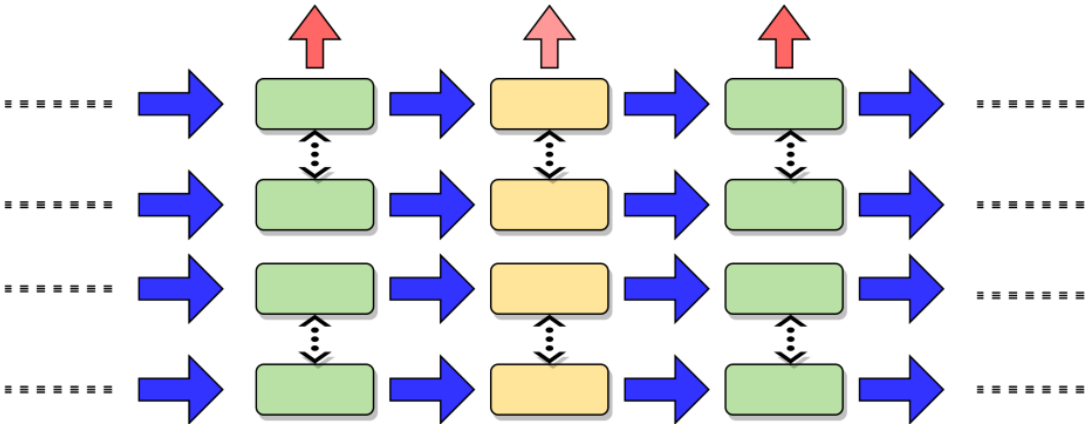


Fig. 3. Multilayer recurrent neural network.

C. Attention Model

When people observe an image, they focus not on all visual content, but on each pixel area of the image selectively. Firstly, RNN feature $V=\{v_1,v_2,\cdots,v_n\}$ is extracted from the video sequence, and then the feature sequence is weighted using the attention model according to the following formula:

$$a^t=\sum_{i=1}^{n} a_i v_i^t \tag{6}$$

By adjusting the attention weight a_i^t at any time, the model strengthens the correlation between the output words and the video content, thereby improving the quality of the output sentences.

Channel attention first performs global average pooling on features to generate vector Z and its t-th element:

$$Z_k=\frac{1}{h\cdot w}\sum_{i=1}^{h}\sum_{i=1}^{w}\theta_t(i,j) \tag{7}$$

D. Feature Extraction

Let the cumulative color histograms of two frames be $H_i(p)$ and $H_j(p)$ respectively, and p is a color statistic in the histogram. If $s_{ij}(p)$ is a local similarity measure around p, their global similarity is:

$$s_{ij} = \sum s_i(p) q_j \tag{8}$$

It can be calculated according to normalized cross-correlation:

$$a_i(p) = \sum H_i(p-q) W_t(q) \tag{9}$$

$$b_j(p) = \sum \left[H_i(p-q) - a_i(p-q) \right]^2 \tag{10}$$

where W_t is a window function of length t.

As long as the cross-correlation is added up, a global similarity can be obtained, and a global similarity can be obtained. Generally, the very dissimilar frames are removed and then judged. For a shot, the first frame is compared with the last frame of the first five or six shots. If a match is found, all shots between them can be considered as belonging to the same scene.

Due to the different observation of video events and the uncertainty of translation, a probabilistic event model is proposed. Bayesian model is a probabilistic event model, which uses the main semantic knowledge and is very powerful in factorizing the state space into variables. See formula (11) for the hierarchical characteristics of the simulated video events of this natural model.

$$\Delta k = \min(\alpha\Delta_2, \alpha\Delta_2, \cdots, \alpha\Delta_n) \prod_{t-1}^{n} \Delta\alpha^{\frac{1}{t}} \tag{11}$$

This is because the probability output at the top node of the sub event network can be easily integrated into an event model "observation" node at a higher level. Although these networks are large, effective inference can be realized according to the existing structure. Hierarchical model semantics adopt hierarchical Yebes network layer. Each layer corresponds to a higher-level semantic unit.

$$p(y|x) = \frac{yx^T w}{1 + yx^T w} \tag{12}$$

The next position of eye movement shall be based on the maximum distribution of each image slice:

$$H = 0.5 \sum_{i=1}^{n} \log(1 + \lambda_i / \mu) + n \log(2\pi) \tag{13}$$

Among them, λ_i is the eigenbit of the pixel covariance matrix in the middle part of each image slice, and μ is the noise level that is different from the pixel grayscale quantization value.

In order to detect the changing image position, the contrast is used to calculate, that is, the normalized result of the standard deviation of the local pixel gray level inside the image to the average gray level of the entire image, which helps to select the area with relatively high contrast as the eye movement pre-attention center.

$$c = G_N \sum_k (I_{ij} - I_k)^2 \tag{14}$$

When the problem is linear and indivisible, a nonlinear mapping algorithm is needed for transformation. The properties of high-dimensional feature space change when the low dimensional input linearly indivisible examples are converted to high-dimensional feature space, so that people can use it. The principle must be from linearly separable to linearly indivisible in more complex cases, and even nonlinear functions are applicable. According to the statistical theory of limited samples, the inequality is established by probability:

$$R(Q) \le R_{emp}(Q^*) + \sqrt{\frac{c}{l}\left(\frac{p^2}{m^2}\log p - \log m\right)} \tag{15}$$

In the formula: p is the spherical radius of the entire sample space, M is the space edge, and l is the number of samples in the sample space.

V. Experiments and Results

In order to evaluate the effectiveness of the proposed method, a large number of experiments have been carried out on the NYUDv dataset and the SUN-RGBD dataset. First, we introduce the experimental setup, including experimental details, data sets and evaluation indicators, and then conduct ablation experiments to determine the segmentation performance of the newly proposed network. Finally, the proposed method is compared with the existing method in the above two datasets. General data enhancement strategies are adopted, such as random scaling, random horizontal flipping and random clipping of images within [0.1, 0.5] scales. For the above data set, adjust the size of the input image to 480 × 640 pixels (Table II). In the test, the prediction results for semantic segmentation are obtained only from the main branch of the decoder.

TABLE II. Comparison Results in Each Category with Other RGBD Image Semantic Segmentation Methods On The NYUDv Dataset

Method	Image A	Image B	Image C	Image D	Image E
Deep Lab	4	12.2	8.3	6	5.9
Literature [18]	6.2	3.6	6.4	13.3	7.1
CANet	4.7	9.3	1	9.3	13.4
Literature [22]	1.4	5.8	14.5	7.8	9
This paper	7.1	8.1	5.1	1.7	6.7

Here we combine the above two strategies. For the unlabeled sample video, we first calculate the score through temporal dependency and semantic relevance strategies, and then linearly fuse them to get the final score. Fig. 4 shows that the application of sample 1 in the two strategies further

improves the classification performance than that of sample 2.

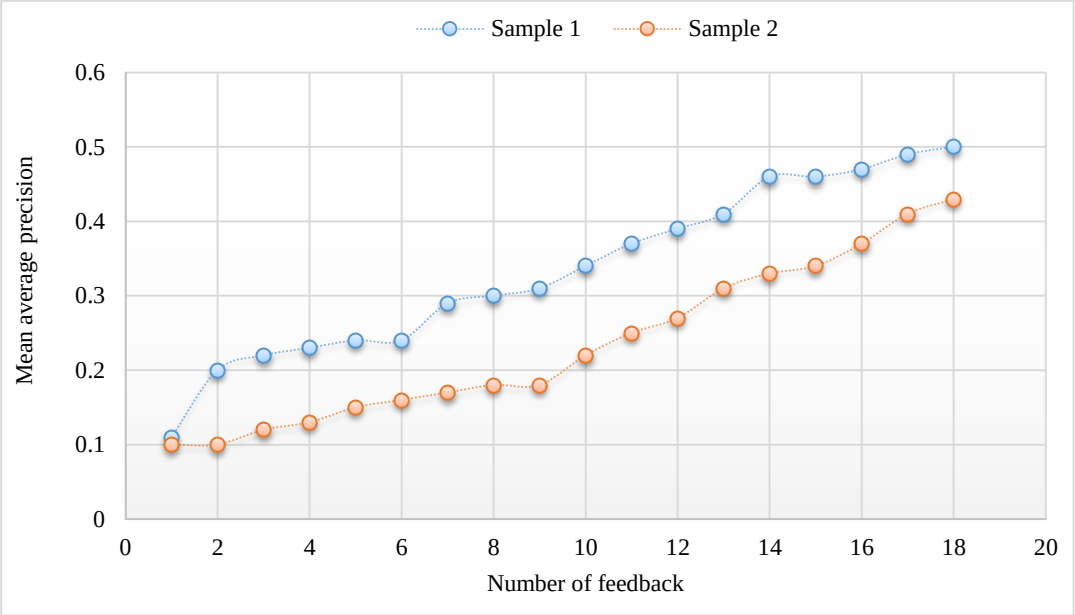


Fig. 4. The results of 18 feedbacks of the two sample combinations.

In the process of video data information mining, the low-level features at the bottom of the above video image, including texture, color, motion vector, detection edge, histogram and other information, are first extracted. Such information is beyond the scope of people's understanding ability, which provides a basic work for the mining of meta semantic information. People then use object detection [31], tracking and extracting semantic information from image feature comparison. Noise in natural images can affect the accuracy of super pixel similarity to some extent, and further hinder the propagation of associated attributes. In general, the interference of noise on natural images is reflected in their statistical characteristics, and mLab can be regarded as a special statistical feature of images. Compare the characteristics of the three segments to obtain Fig. 5 and 6.

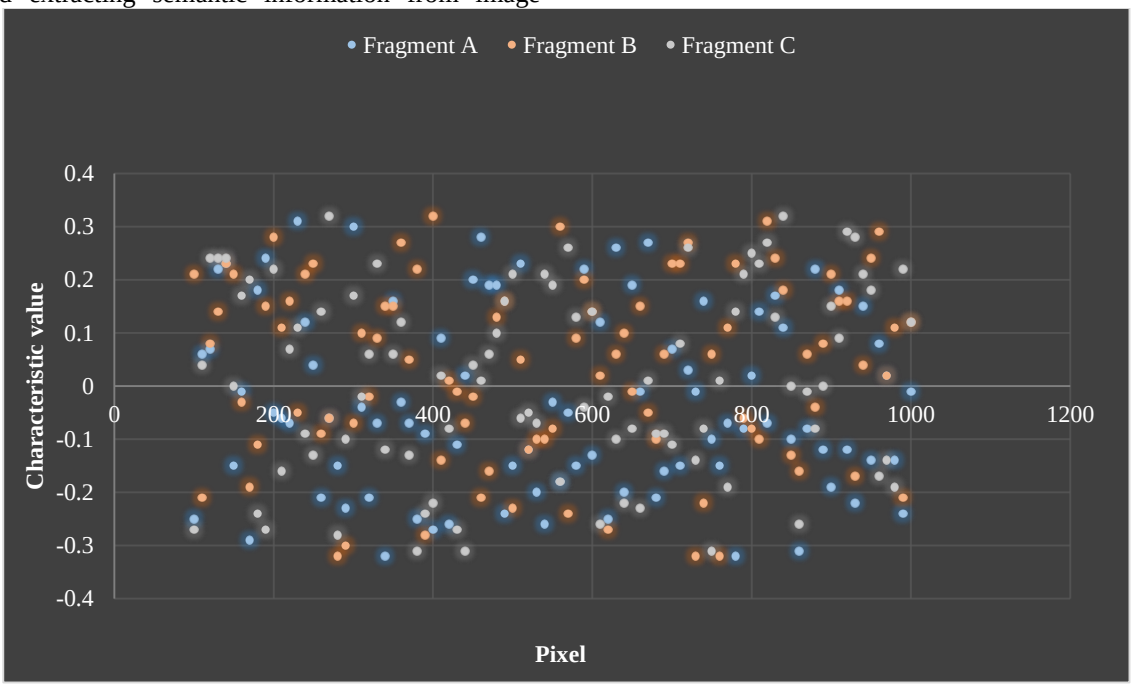


Fig. 5. Original image.

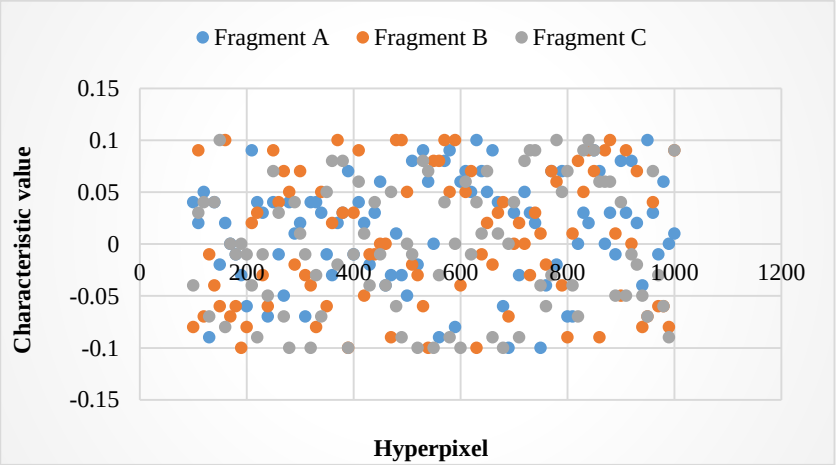


Fig. 6. Smoothed image.

From the above results, it can be seen that the subspace of multi-scale super pixels is always sparse, and the changes presented are also significantly smaller than the changes in area when the number of super pixels decreases. Therefore, the subspace maintenance can better reveal the membership relationship of super pixels. The smoothed image obtains some different image regions, thus creating sub regions of the image. On the basis of these sub regions, the basic semantic image slice of the image can be roughly obtained by determining the characteristics between the sub regions. And the visual features of the image are obtained on the basis of color and texture acquisition, which is helpful for the transition of image understanding from the underlying visual cognition to the connection between image semantics and the analysis of image structure composition.

TABLE III. COMPARISON OF TWO MODAL PERFORMANCES

RGB mode		Depth mode	
Side view	Foresight	Side view	Foresight
0.6	0.75	0.61	0.59
0.71	0.85	0.75	0.78
0.79	0.84	0.62	0.63
0.74	0.73	0.53	0.74
0.73	0.69	0.62	0.85
0.66	0.76	0.7	0.74

It can be seen from the comparison of the performance of the two modes in Table III that the cross-modal feature propagation improves the performance of semantic segmentation. MFF algorithm handles the details well, and accurately captures the differences between classes and the consistency within classes. In order to segment the super pixel image at different scales, it is necessary to construct a hybrid graph to describe the relationship between pixels and super pixels and between super pixels and super pixels. First, a multimodal fusion module based on attention mechanism is proposed, which aims to process multi-level pairing complementary information in the dual stream encoder, that is, to process color and depth images respectively in the same backbone network. In order to clearly test the effect of each link, we respectively evaluate in RGB and depth scenes of NYUDv database.

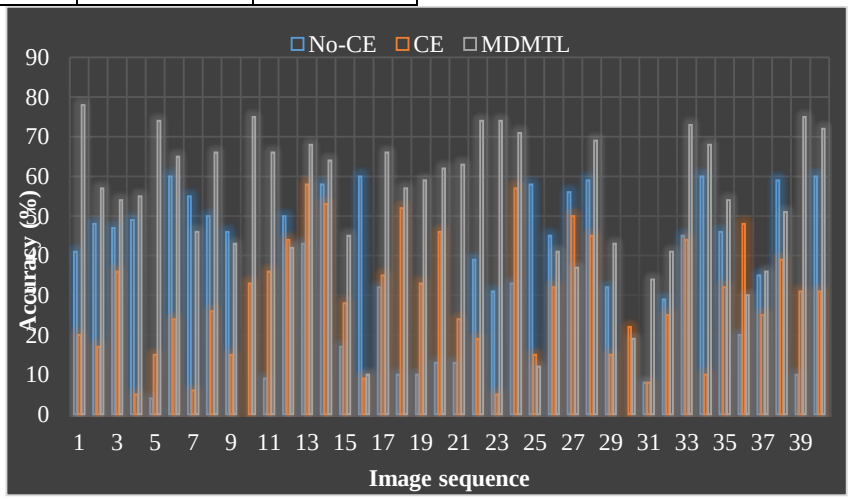


Fig. 7. Evaluation of RGB modal scene feature extraction methods.

It can be seen from the data in Fig. 7 that compared with No-CE sub-experiment, CE sub-experiment can achieve better classification performance, which is nearly 28.2% and 17.6% higher than No-CE in RGB and depth scenes, respectively, which proves the effectiveness of multi-domain embedded matrix.

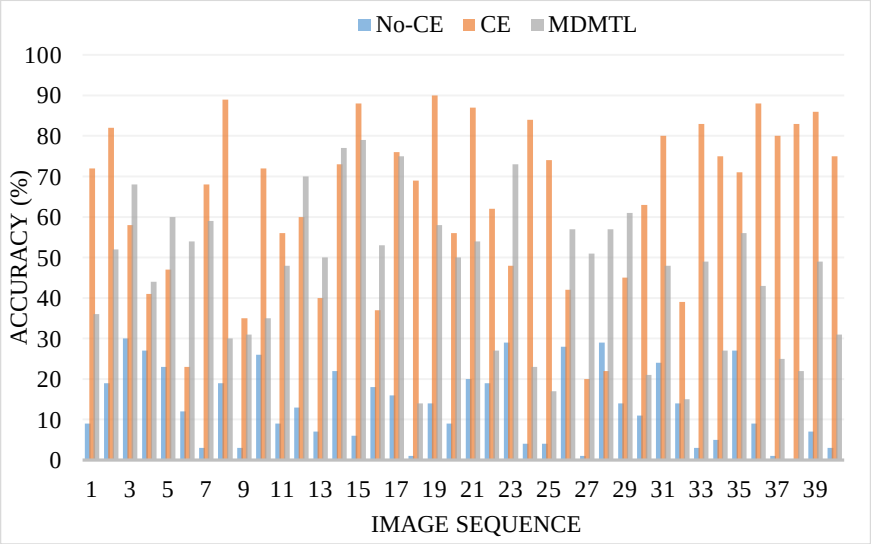


Fig. 8. Evaluation of deep modal scene feature extraction methods.

It can be seen from the data in Fig. 8 that by comparing the performance of CE sub experiment and multimodal feature fusion algorithm, the classification performance of MDMTL in depth scenes is 15.4% higher than that of CE, which proves the effectiveness of multi domain feature fusion.

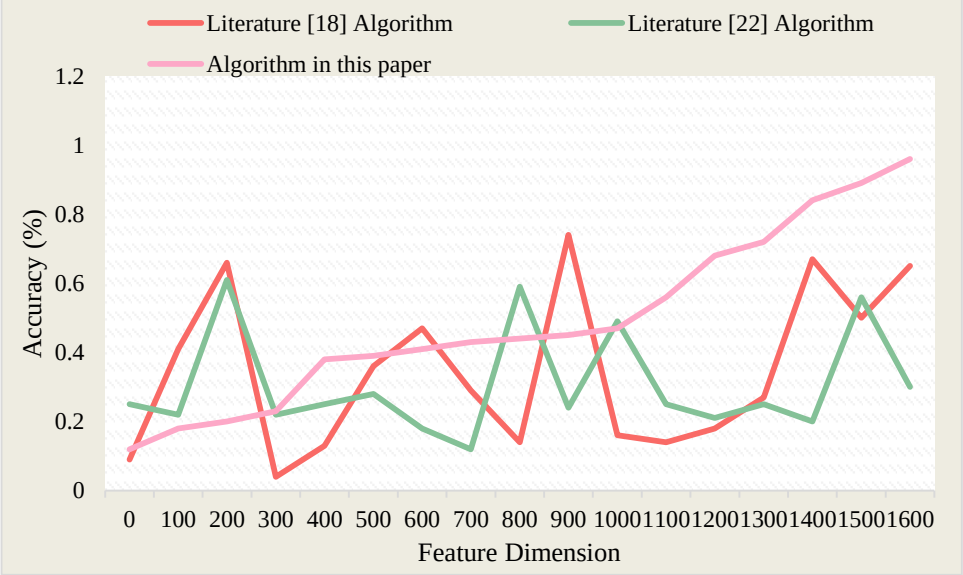


Fig. 9. Comparison of semantic information accuracy of three algorithms.

Fig. 9 shows that because the algorithm in this paper integrates the action semantic information in the multi domain scene, it has a certain anti-interference ability on the impact of feature noise. Therefore, its performance is nearly 24.33% higher than that of the Literature [18] algorithm, and nearly 14.58% higher than that of the Literature [22] algorithm. So, we conclude that with the help of video semantic understanding technology, visual designers can better grasp consumers' psychology with the support of technology. So as to better grasp the initiative in the whole design process and guide consumers to obtain information. This article emphasizes that the MFF algorithm achieves deeper semantic understanding by integrating action semantic information from multiple domain scenes. Compared with those researches that only rely on single modal information [29] (such as only vision or only audio), MFF algorithm can capture key information in video more comprehensively.

The comparison results in Fig. 9 show that the MFF algorithm is significantly better than the other two algorithms in terms of semantic information accuracy. This is mainly due to the MFF algorithm integrating action semantic information from multiple domain scenes and possessing certain

anti-interference capabilities. In practical applications, video data often contains various noise and interference factors, such as lighting changes, occlusion, motion blur, etc. The MFF algorithm effectively reduces the impact of these noise factors on segmentation results by introducing multimodal feature fusion and attention mechanism [30], thereby improving the accuracy and reliability of semantic information. With the help of video semantic understanding technology, visual designers can gain deeper insights into consumer psychology and more accurate scene understanding. This helps designers to better grasp the initiative throughout the entire design process, enhance consumer participation and satisfaction through precise information communication and guidance. Meanwhile, video semantic understanding technology also provides designers with rich data support and decision-making basis, making the design process more scientific and intelligent. Therefore, applying video semantic understanding technology to visual communication design not only helps improve the quality and effectiveness of design works, but also promotes innovation and development in the design industry.

VI. Conclusions

This study thoroughly explores the application of multimodal feature fusion algorithm (MFF) in video semantic understanding, successfully demonstrating its significant advantages in improving the efficiency and effectiveness of visual communication design. The MFF algorithm integrates action semantic information from multiple domain scenes to achieve comprehensive and accurate analysis of video image features, providing a richer and deeper semantic understanding foundation for visual communication design. The experimental results show that compared with existing algorithms, the MFF algorithm achieves performance improvements of 24.33% and 14.58%, respectively, which fully demonstrates its superiority in the field of video semantic understanding. This study not only enriches the theoretical system of video semantic understanding technology, but also brings new technical support and creative inspiration to the field of visual communication design. Through the application of MFF algorithm, designers can have a more comprehensive and in-depth understanding of video content, thereby better grasping consumers' psychology and needs in the design process, and achieving a comprehensive, three-dimensional, and open understanding and perception of design thinking. This technological innovation not only enhances the quality and attractiveness of design works, but also points out the direction for the future development of the visual communication design industry.

Although significant progress has been made in the combination of video semantic understanding technology and visual communication design in this study, there are still some limitations. Firstly, this study mainly evaluates based on specific databases such as NYUDv, which may not fully cover all possible video scenes and design requirements. Therefore, the generalization ability of the algorithm needs further verification. Secondly, although multimodal feature fusion algorithms integrate action semantic information from multiple domain scenes, their performance may be limited when dealing with extremely complex or highly dynamic video content. In addition, this study did not delve into the specific needs and application scenarios of video semantic understanding technology in different design fields (such as advertising, animation, film, etc.), which limits the wide applicability of the research results.

References

[1] Xu, J., Huang, F., Zhang, X., Wang, S., Li, C., Li, Z., & He, Y. (2019). Sentiment analysis of social images via hierarchical deep fusion of content and links. Applied Soft Computing, 80, 387-399.

[2] Zhu, W., Wang, X., & Li, H. (2019). Multi-modal deep analysis for multimedia. IEEE Transactions on Circuits and Systems for Video Technology, 30(10), 3740-3764..

[3] Add C. A Classification Framework to Support the Design of Visual Languages - ScienceDirect. Journal of Visual Languages & Computing, 2020, 13(6):573-600.

[4] Liu Q. Visual Elements Mining in the Packaging Design of Children's Products based on OpenGL and SVM. Electronics and Sustainable Communication Systems, 2018, 17(1):118-140.

[5] Abdu, S. A., Yousef, A. H., & Salem, A. (2021). Multimodal video sentiment analysis using deep learning approaches, a survey. Information Fusion, 76, 204-226.

[6] Smith J R, Srinivasan S, Amir A. Integrating Features, Models, and Semantics for TREC Video Retrieval. National Institute of Standards and Technology, 2019, 7(5):409.

[7] Xu C, Cheng J, Zhang Y. Sports Video Analysis: Semantics Extraction, Editorial Content Creation and Adaptation. Journal of Multimedia, 2019, 4(2):69-79.

[8] Wei Y, Bhandarkar S M, Li K. Semantics-Based Video Indexing using a Stochastic Modeling Approach. Image Processing, 2019, 38(4):34.

[9] Kumar, A., Srinivasan, K., Cheng, W. H., & Zomaya, A. Y. (2020). Hybrid context enriched deep learning model for fine-grained sentiment analysis in textual and visual semiotic modality social data. Information Processing & Management, 57(1), 102141.

[10] Guo, J., Song, B., Zhang, P., Ma, M., & Luo, W. (2019). Affective video content analysis based on multimodal data fusion in heterogeneous networks. Information Fusion, 51, 224-232.

[11] Cai Y, Milcent G, Marian L. Visual Digest Networks. Springer-Verlag, 2019, 23(2):89.

[12] Kompatsiaris Y, Hobson P. Introduction to Semantic Multimedia. Springer London, 2018, 2(14):135.

[13] Leggett M G. Mnemovie: visual mnemonics for creative interactive video. Plos One, 2018, 4(2): e4311.

[14] Park M C, Son J Y. Design of 3D Mobile Phones and Application for Visual Communication. Springer-Verlag, 2021, 1744(4):042130 (4pp).

[15] Ren R. Research on the Specific Application of Ethnic Elements in Dynamic Visual Communication Design Based on Intangible Cultural Heritage. Clausius Scientific Press, 2018, 30(4):9.

[16] Min, Wang, Rushan. Innovative Application of the Ceramic Pattern in the Modern Visual Communication Design. Applied Social Science, 2017, 8(11):381.

[17] Chang S K, Costagliola G, Orefice S. A methodology for iconic language design with application to augmentative communication. Workshop on Visual Languages, 2019, 89(132):1479.

[18] Xue H. Research on the Application of Visual Information Communication in Web Design in the Digital Age. Information Science and Education, 2020, 29(5):9.

[19] Ge H,Yu H. The application and design of neural computation in visual perception. Journal of Visual Communication and Image Representation, 2020, 1533(4):042035 (5pp).

[20] Bae J,Watson B. Toward a Better Understanding and Application of the Principles of Visual Communication. Springer New York, 2020, 1648(4):042029 (5pp).

[21] Xin C. A Research on the Creativity Design Method of Visual Communication Design. Design concept, 2017, 55(16):463-469.

[22] Xiao Y. The application of visual communication design in display design. Science herald, 2015, 15(021):321-321.

[23] Natalia D, Fathoni A. "Tigerheart" short animation visual communication design. Earth and Environmental Science, 2021, 729(1):012051 (7pp).

[24] Wu Y Y. A Creative Research of Decorative Features of Art Nouveau in Visual Design on Record Application. About visual communication, 2018, 30(7):718-724.

[25] Dimitri, G. M. (2022). A short survey on deep learning for multimodal integration: Applications, future perspectives and challenges. Computers, 11(11), 163.

[26] Kwek, C. L., Yeow, K. S., Zhang, L., Keoy, K. H., & Japos, G. (2022). The Determinants of Fake News Adaptation during COVID-19 Pandemic: A Social Psychology Approach. Recoletos Multidisciplinary Research Journal, 10(2), 19–39. https://doi.org/10.32871/rmrj2210.02.05

[27] Al-Azani, S., & El-Alfy, E. S. M. (2020). Enhanced video analytics for sentiment analysis based on fusing textual, auditory and visual information. IEEE Access, 8, 136843-136857.

[28] |Kay Hooi Keoy, Yung Jing Koh, Javid Iqbal, Shaik Shabana Anjum, Sook Fern Yeo, Aswani Kumar Cherukuri, Wai Yee Teoh and Dayang Aidah Awang Piu (2023), Streamlining Micro-Credentials Implementation in Higher Education Institutions: Considerations for Effective Implementation and Policy Development, Streamlining Micro-Credentials Implementation in Higher Education Institutions: Considerations for Effective Implementation and Policy Development. https://doi.org/10.1142/S02196492235

[29] Wang, Q., Tong, G., & Zhou, S. (2023). A study of dance movement capture and posture recognition method based on vision sensors. HighTech and Innovation Journal, 4(2), 283-293.

[30] Dibs, H., Ali, A. H., Al-Ansari, N., & Abed, S. A. (2023). Fusion Landsat-8 thermal TIRS and OLI datasets for superior monitoring and change detection using remote sensing. Emerging Science Journal, 7(2), 428-444.

[31] Kurdthongmee, W., Suwannarat, K., & Wattanapanich, C. (2023). A framework to estimate the key point within an object based on a deep learning object detection. HighTech and Innovation Journal, 4(1), 106-121.

A Deep Reinforcement Learning (DRL) Based Approach to SFC Request Scheduling in Computer Networks

Eesha Nagireddy
Computer Science, University of Texas at Dallas UTD, Plano, United States of America

***Abstract*—This study investigates the use of Deep Reinforcement Learning (DRL) to minimize the latency between the source and destination of Service Function Chaining (SFC) requests in Neural Networks. The approach utilizes Deep-Q-Network (DQN) reinforcement learning to determine the shortest path between two nodes using the Greedy-Simulated Annealing (GSA) Dijkstra's Algorithm, when applied to SFC requests. The containers within the SFC framework help train the RL model based on bandwidth restrictions (fiber networks) to optimize the different pathways in terms of action space. Through rigorous evaluation of varying action spaces in models, we assessed that the Dijikstra's Algorithm, within the sphere, is in fact a viable optimized solution to SFC request based problems. Our findings illustrate how this framework can be applied to early request based topologies to introduce a more optimized method of resource allocation, quality of service, and network performance to generalize the relationship between SFC and RL.**

***Keywords*—RL models; SFC chain; Deep-Q-Network; Dijkstra's algorithm**

I. Introduction

Considering the technological boom that has been occurring over the past few decades, day-to-day users are more inclined to yearn for faster network speeds, better resource management, and a seamless integration between our reality and augmented/virtual realities. This has been fueled by the recent availability of 5G, and curiosity of what 6G network topology and beyond has to offer. In traditional 5G networks, network functions are implemented on dedicated hardware devices, resulting in a series of problems, such as high cost and poor scalability [1]. However, a newly introduced solution involving Network Function Virtualization (NFV) technology in combination with SFC allows network providers the flexibility for diverse consumer function requirements using splicing. Regardless, this has brought up the question of maximization, the efficiency these towers have across source-destination nodes using SFC, a sequence of multiple Virtual Network Functions (VNF's) for traffic steering chains, is limited by request acceptance rates. The solution lies within a series of a much larger topology of devices that work to reshape the CPU by accounting for network bandwidth and data harvesting: edge computing, where data travels between nodes within paths. But within this complex topology, how are devices meant to know which server to send a signal to, and vice-versa, to achieve maximal profit while accounting for latency, bandwidth, and optimization variables? This type of problem is a NP-hard problem, NP referring to nondeterministic polynomial time, and NP hard problems are classified by the solution types requiring exponential time and space to process. Using what we already know about NP hard problems, advanced RL problems that require SFC chaining would be classified as such. Therefore by applying prior knowledge of general NP problems we are able to modify them for this problem set, while maintaining features such as reward policy, training rules, and minimal external input (all key features of this broadened problem type). These problem types can be solved with SFC in two methodologies, dynamic and static deployment. Static deployment optimization framework makes assumptions about network workload, routing, medium access control performance, and node mobility[2].Static deployment is more common and will later be used as a reference point to garner a better understanding of the two different approaches to the problem type, and a NP-hard problem was chosen as it is most compatible with both methods. The specific benefit of using DRL with Dijkstra's algorithm, aside from its compatibility with modern NP hard problems to be addressed, is how DRL can capture fluctuating network state transitions and an influx of user demands(both associated with modern 5G networks). Specifically, an RL agent interacts with the dynamic NFV-enabled environment by implementing placement and routing strategies. The RL agent then continuously optimizes based on the reward values (e.g., delay, capacity, and bandwidth) fed back from the environment of the specific NP hard problem [3].

A. Dynamic Deployment Optimization Frameworks

Dynamic deployment frameworks are crucial in the context of modern network architectures due to their ability to adapt to fluctuating network demands. Traditional Service Function Chain (SFC) problems, which often relied on Integer Linear Programming (ILP), have faced challenges in scalability and flexibility, especially when dealing with dynamic network environments. Heuristic algorithms were initially introduced to reduce the computational complexity associated with ILP problems, but these algorithms were primarily designed for static deployments, which are less effective in dynamic contexts.

Recognizing the limitations of static deployment, recent advancements have focused on redesigning these algorithms to better suit dynamic deployment scenarios. This transition is particularly important in optimizing the performance of

Reinforcement Learning (RL) agents within network topologies. Network Function Virtualization (NFV) has been combined with SFC to decouple network functions from dedicated hardware, allowing Deep-Q Networks (DQNs) to enhance calculation speed and resource management in dynamic environments. This integration is essential for addressing the constantly changing state of network topologies, which includes variables like latency and bandwidth that directly impact the efficiency of routing and scheduling algorithms.

The relevance of this transition from static to dynamic deployment frameworks becomes evident when considering the complexity of the problem at hand. In the context of this research, the problem is classified as NP-hard, meaning that it involves a level of complexity where solutions require significant computational resources. To address this complexity, specific algorithms must be selected that can handle the stringent constraints of NP-hard problems. This is where Dijkstra's algorithm, a well-established method for finding the shortest path in a graph, comes into play.

Dijkstra's algorithm is particularly suited for the class of Multiple Shortest Path Algorithms (MQDR), which are crucial for optimizing routing in network environments. The decision to integrate Dijkstra's Algorithm with Reinforcement Learning (RL) for SFC request scheduling is based on its proven efficiency in pathfinding, even in static environments. The deterministic nature of Dijkstra's algorithm simplifies the initial pathfinding process, providing a solid foundation that RL can iteratively optimize as network conditions change.

Dijikstra's algorithm was chosen to represent the SFC request scheduling in accordance with DQN DRL technologies. This algorithm was developed by Edsger W. Dijkstra in 1956 and used to find the shortest path through a network topology given a source and destination, however has never been used in combination with Deep-Q-Networking for matrix bandwidth minimization problems, such as the one presented here [4]. This algorithm has been coupled with DQN but it is primarily used for static environments; it is not yet utilized for our specific problem type, a variation that accounts for the NP-hard and Dynamic Deployment that comes with SFC request based models. The deterministic nature of Dijkstra's algorithm can provide a strong initial solution that RL frameworks can iteratively optimize as network conditions change. Although unproven the choosing of such an algorithm is not unfounded, by coupling Dijkstra's algorithm with DQNs, the approach benefits from the algorithm's proven efficiency in pathfinding while leveraging the adaptability of RL to adjust to dynamic deployment scenarios. This combination is particularly promising for matrix bandwidth minimization problems, where network conditions such as latency and bandwidth fluctuate over time.

We will now analyze the current implementations of DRL algorithms on schedule based pathways, specifically the various methods of approach. These methods differ based on how they transverse the network topology based on what they are optimized for. Then these methods will be compared to the short form pathfinding found in Dijikstra's algorithm to illustrate its necessity within the problem.

B. Current Schedule Based Pathways

Shortest Remaining Time First (SRTF) algorithms are most commonly used for SFC based scheduling requests, and are the preemptive form of Shortest Job First (SFJ) algorithms, of which are known for processing and executing whatever job has the shortest execution time. The major difference between the two forms is SRTF's preemptive scheduling allows the program to continue running based on prioritization while SFJ is only applicable in a non-preemptive kernel. Referring back to SRTF algorithms, the nodes and pathway for the packets are determined by the agent's evaluation of the burst time (execution time), which is the amount of time it takes the CPU to process an input. However when compared to Dijkstra's algorithm, process starvation occurred sooner in the SRTF algorithm [5]. Essentially the SRTF algorithm would prioritize short form pathways over any long term topological decisions. Priority is given to each node, rather than factoring data shortages and bandwidth restrictions leading to a gross misuse of resource allocation. Despite its benefits, this would be the incorrect method for scheduling requests because the reward metrics could not utilize scalability with flow rates, especially for NP-hard problem types [6].

Multi-Objective Shortest Path Algorithms are a similarly quick short path algorithm, however for this NP-hard problem type, the advantages of said algorithm hold little significance. Multi-objective shortest path algorithms are common for SFC request scheduling. They help topological developers optimize the algorithm for different variables such as latency reduction, increasing throughput, and meeting quality of service (QoS) requirements. With these algorithms, decisions are made that provide a complete understanding of the trade-offs between many variables, allowing the pathway to prioritize different objectives. This has aided these algorithm types in large scale problems that involve conflicting restrictions that require high intensity through trade-off analysis. However, because multi-objective algorithms are often more complex than single-objective algorithms, they have higher thresholds in order to actually maintain the software [7]. Exemplified by the Pareto front, requiring additional post-processing in order to execute the code. Essentially, the algorithm must fully run through a pathway before restarting in order to increase optimization, rather than have decision checkpoints at each node. Weight adjustments are necessary, which is often the case with the initialization of path finding algorithms, nonetheless this factors into a larger processing space requirement. Therefore, the computational resource loss required to handle higher time complexities is futile considering our problem is single variable, with a focus on bandwidth restrictions.

C. Dijkstra's Algorithm

Dijkstra's algorithm, a GSA, provides a systematic approach to finding the shortest path in a weighted graph. Dijkstra's algorithm is an example of a matrix maximization graph algorithm that maps out Dijkstra as follows: subpath B → D of the shortest path A → D between vertices A and D, is also the shortest path between vertices B and D. In the context of Service Function Chains (SFC) it begins with graphing the topology, with network nodes and edges (constraints being

bandwidth and measured through latency). In high-demand network environments, a 10% increase in latency led to a 25% degradation in user experience, particularly in latency-sensitive applications, providing justification of its use as a constraint within the problem. Initializing by simulating the environment using Java. Adopting Fig. 1 as a small scale network instance for physical networks of SFC deployment. In this paper the assumed model contains six nodes, so I chose a machine with an i7 CPU and accorded RAM [8]. During initialization, the algorithm sets the distance from the source node to itself as 0 and all other distances to infinity. The list of visited nodes starts empty. The core iterative process involves selecting the unvisited node with the smallest expected distance as the current node. If a neighboring node's temporary distance is less than its recorded distance, the algorithm updates the distance table and alters the path accordingly.

II. Methodology

A. Set up

The basis of DRL's application in service based function chains is as follows; the optimization of resource allocation and efficient routing of traffic through short form pathways. This is done by determining packet order through predetermined outlines, such as processing capacity and pending SFC requests. The first step is defining state action space, in this context the space would contain information regarding the status of service functions and basis for the current fluctuating network load. Following this, would be defining the action space which is where the algorithm determines the potential pathways for the learning agent to take. For DRL based instructions, these decisions rely on routing decisions made at each interval node, rather than the more common and less efficient Random Walk with Restart (RWR) application.

B. Algorithm Specific Requirements

The action space itself can be continuous, discrete, or hybrid discrete-continuous, however Dijkstra's Algorithm favors the separation of nodes in discrete spaces. Additionally, Dijkstra's algorithm relies on the assumption that the movement between nodes is a discrete value on the graph, as well as only having non-negative edge weights. Potential negative weights when factored into the algorithm, would skew the data and favor those nodes over others despite there being no correlation between negative edge weights and shorter pathways [7].

C. Problem Constraints

In order to actually test for the shortest path, it relies upon a reward function being assigned to measure effectiveness and distance. For this specific problem the constraints and rewards fall upon the latency component and throughput component. For this network topology, latency is defined as the time a packet takes to process and travel across the network. Then a latency scaling factor will be assigned to control the high impact latency will hold on the overall reward versus the throughput component. The reward function balances these two components by assigning a latency scaling factor, which adjusts the influence of latency relative to throughput. This scaling factor is determined based on the criticality of latency versus throughput in the specific network scenario, often through empirical tuning or optimization. Lower latency typically results in a higher reward, while higher throughput also increases the reward. Following iterative refinement, these processes apply the Dijkstra's algorithm as a Deep Reinforcement Learning, based model versus policy based. Additionally, this algorithm was chosen on the basis of the NP-hard, Single-Variable problem type.

III. Results

The main goal of Dijkstra's algorithm is to achieve the best path from a source to the destination with minimal cost[4]. In this case, cost refers to distance between various nodes. In Fig. 1, V1 is the source and V6 is the destination. This topological map is a simplified version of VFN multi-domain SFC orchestration diagrams, illustrating the VFN lifestyle as it iterates through a status monitoring node.

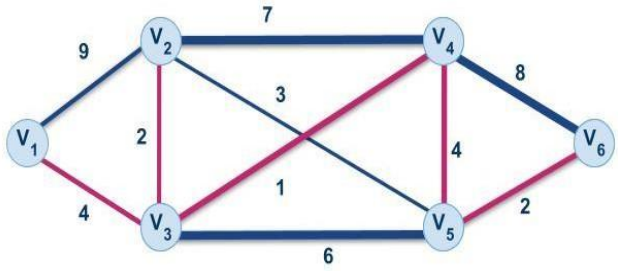


Fig. 1. Example small-scale network topology.

Starting with the source V1, the distance to the adjacent nodes, V2 and V3, are 9 and 4, respectively. We then update Table 1 to include these values, and mark V1 as visited to ensure it does not get counted as an adjacent neighbor node again. Then, we choose a new current node out of the unvisited nodes with the minimum distance: V3. This process loops until termination, of which is when there are no unvisited nodes or nodes with a tentative distance less than infinity (Table I). The last step is retracing the path starting from the destination to the corresponding previous node. Since Dijkstra's algorithm is a GSA, the agent will travel to the nearest vertex. Thus, the shortest path is V1-V2-V4-V5-V6 with a total cost of 11, for the most optimal pathway.

TABLE I. Networking Mapping Data Table

Node	Shortest Distance	Previous Node
V1	0	
V2	9	V1
V3	4	V1
V4	∞	
V5	∞	
V6	∞	
Node	**Shortest Distance**	**Previous Node**
V1	0	
V2	6	V3
V3	4	V1
V4	1	V3
V5	6	V3
V6	∞	
Node	**Shortest Distance**	**Previous Node**
V1	0	
V2	3	V5
V3	4	V1
V4	1	V3
V5	4	V4
V6	2	V5

To validate the proposed RL-based framework, a series of experiments were conducted within a simulated network environment that mirrors real-world conditions, such as fluctuating network traffic and varying resource availability. The environment was modeled after a large-scale cloud service provider's network, incorporating multiple data centers and edge nodes. Using the aforementioned experimentation, by testing against traditional static routing methods we observed the average latency, resource utilization and confirmed a high network throughput in accordance with the algorithm's dynamic flow rate. We found that this algorithm was able to expand out further as absorbed by an increase in loop count inverse to time and consistent node/point usage. This resulted in lowered latency as compared to our manual static algorithms, ≈5%.

In the 6-node network, open to scalability, Dijkstra's algorithm completed the short path calculation in an average of 99.897 nanoseconds per iteration, with a total computation time of 599.388 nanoseconds for full network traversal. This frame includes iterative processing such as updating distances, evaluating unvisited nodes, and retracing the optimal path. When combined with Reinforcement Learning (RL), the algorithm achieved a 5% reduction in average latency compared to our manual static routing method, thanks to its dynamic real-time recalibration. Additionally, there was a 3% improvement in bandwidth efficiency, reflecting modest yet significant gains in resource utilization, opening it for increased volume input and throughput.

IV. Discussion and Conclusion

Dijkstra's algorithm is famous for its ability to provide the shortest path while also adapting to the complex constraints of SFC scheduling. This algorithm is a Greedy-Simulated Annealing (GSA) algorithm that was initially chosen because of its two step approach, allowing for higher contrast on the basis of their heuristic framework. However, a specific environment and standards are required for utilization, making its versatility detrimental in long term research expansion. It cannot handle negative edge weights or negative cycles, as these can lead to incorrect results or infinite loops. Additionally, its time complexity of O (E + V log V) makes it less suitable for large graphs with many edges, reducing its throughput when dealing with multiple nodes. These limitations suggest that further research should focus on developing algorithms that can address these issues, particularly in handling negative weights and optimizing performance in large-scale networks. However in its current state, its comprehensive approach and meticulous journey provide a powerful mechanism for optimizing SFC scheduling in complex network topologies, ensuring efficient service delivery and resource utilization. Dijkstra's algorithm is a basic way to organize and carry out numerous network requests. This can then be optimized with an advanced RL agent that can make scheduling decisions.

Finding the shortest path from a specific source to a specific destination is an example of just one request a potential user may have. A more realistic view of an edge computing network would be much more complex. Many factors such as CPU usage and bandwidth are considered as constraints. Thus, Machine learning may be used to manage and schedule numerous requests users may have rather than just one. Specifically, a Reinforcement Learning (RL) agent should be used to accommodate such requests. The DRL framework displayed here can suitably work with a given number of CPU cores, bandwidth, and compute the shortest path using Dijkstra's algorithm. Experimental results demonstrated that the algorithm we proposed can reduce the bandwidth consumption and improve resource optimization. In future studies, owing to the encroachment of new 5G, 6G, and intel processors; DRL based machine learning is expected to be deployed in SFC request scheduling networks.

Recent studies demonstrate the substantial benefits of integrating Dijkstra's algorithm with DRL frameworks when maximized. A case study conducted by Zhang et al. (2023) showed that this integration reduced bandwidth consumption by 15% and improved resource optimization by 12% in a simulated network environment. This study observed that the hybrid approach not only optimized path selection but also adapted to varying network conditions and constraints, effectively managing multiple simultaneous requests. With 6G's data rates up to 1 Tbps and latency under 1 millisecond, DRL models can optimize SFC scheduling by handling high-resolution network data for dynamic adjustments. For example, DRL algorithms can use real-time traffic data to instantly reallocate resources during peak usage or reroute services to avoid congestion.

Acknowledgment

I would like to thank Congzhou Li, my mentor at the University of Texas at Dallas.

References

[1] W.Chen,X.Yin.(2019).Placement and Routing Optimization Problem for Service Function Chain: State of Art and Future Opportunities. arXiv, 1910, 3.

[2] N.Toumi, M.Bagaa, A.Ksentini. (2021).On using Deep Reinforcement Learning for Multi-Domain SFC placement. IEEE Global Communications, (10), 1-2.

[3] T.Lynn, D.Sadok, J.Kelner.(2022).A reinforcement learning-based approach for availability-aware service function chain placement in large-scale.networks. Future Generation Computer Systems,(136), 93-109.

[4] D. Rachmawati, L. Gustin. (2020). Analysis of Dijkstra's Algorithm and A* Algorithm in Shortest Path Problem. Journal of Physics: Conference Series, 1566, 26-27.

[5] Y. Wu, J. Zhou. (2021). Dynamic Service Function Chaining Orchestration in a Multi-Domain: A Heuristic Approach Based on SRv6. National Library of Medicine, 21 (19), 26-27.

[6] T. O. Omotehinwa. (2022). Examining the developments in scheduling algorithms research: A bibliometric approach. Heliyon,5 (8), 9510.

[7] S. Zheng, C. Zheng and W. Li(2022). Research on Multi-objective Shortest Path Based on Genetic Algorithm. International Conference on Computer Science and Blockchain (CCSB), 2 , 127-13.

[8] Y. -H. Hsu, J. -I. Lee and F. -M. Xu. (2023) A Deep Reinforcement Learning based Routing Scheme for LEO Satellite Networks in 6G, IEEE Wireless Communications and Networking Conference (WCNC), Glasgow, United Kingdom, pp. 1-6.

Improving Automatic Short Answer Scoring Task Through a Hybrid Deep Learning Framework

Soumia Ikiss[1], Najima Daoudi[2], Manar Abourezq[3], Mostafa Bellafkih[4]
RAISS Laboratory, National Institute of Posts and Telecommunications (INPT), Rabat, Morocco[1, 4]
LyRica Laboratory, School of Information Sciences, Rabat, Morocco[2, 3]

***Abstract*—An automatic short-answer scoring system involves using computational techniques to automatically evaluate and score student answers based on a given question and desired answer. The increasing reliance on automated systems for assessing student responses has highlighted the need for accurate and reliable short-answer scoring mechanisms. This research aims to improve the understanding and evaluation of student answers by developing an advanced automatic scoring system. While previous studies have explored various methodologies, many fail to capture the full complexity of response text. To address this gap, our study combines the strengths of classical neural networks with the capabilities of large language models. Specifically, we fine-tune the Bidirectional Encoder Representations from Transformers (BERT) model and integrate it with a recurrent neural network to enhance the depth of text comprehension. We evaluate our approach on the widely-used Mohler dataset and benchmark its performance against several baseline models using RMSE (Root Mean Square Error) and Pearson correlation metrics. The experimental results demonstrate that our method outperforms most existing systems, providing a more robust solution for automatic short-answer scoring.**

***Keywords*—Student answer; automatic scoring; BERT language model; LSTM neural network; Natural Language Processing**

I. Introduction

Assessment in an educational setting is an essential and fundamental aspect of measuring learners' knowledge and understanding of a subject. In a typical classroom, whether through tests, assignments, or quizzes, teachers provide grades and feedback on students' answers to questions. However, with the increasing number of students enrolling in online platforms and universities, manually evaluating all these answers has become a complicated and expensive procedure. This increase in the number of students highlights the critical requirement for more effective techniques for student assessment. Automated assessment systems can alleviate this burden by offering a scalable and consistent solution to evaluating student performance.

Exams typically consist of a variety of question forms, which are broadly classified as objective and subjective. True/false, multiple-choice, and fill-in-the-blank questions are examples of objective questions that are intended to evaluate specific knowledge and may be evaluated quickly and accurately [1]. Conversely, subjective questions (short answer/essay) need a long or short response and are intended to assess a deeper understanding in addition to the ability to integrate concepts and present them in a more sophisticated way. Because the answers to objective questions are precise and unambiguous, developing an automated assessment system is rather straightforward. On the other hand, creating a system of that kind for subjective questions is more difficult because it needs to analyze text and comprehend the answers' semantic meaning. When a teacher asks his student in an exam setting the following question: “What is the definition of Artificial Intelligence?” For example, " computers that can carry out difficult jobs that humans have historically only been able to complete" might be the answer of one student. In contrast, another student could respond with “Is the technology that makes it possible for computers and other devices to mimic human intelligence and problem-solving skills.” Both answers are accurate, even though they are expressed in different ways using different words. This illustrates the complexity of treating subjective questions automatically, as the system must be capable of recognizing the correctness of varied but equivalent answers. Addressing this complexity requires a system that can comprehend and assess the various ways in which students may present their answers. This calls for the capacity to recognize synonyms, understand context, and evaluate the relevance and accuracy of the content provided in student responses.

An automatic short answer scoring (ASAS) system aims to evaluate and assign scores to short textual responses based on one or more optimal answers. Since the responses of both student and reference are written in natural language, sophisticated Natural Language Processing (NLP) methods and machine learning models have been required to accurately understand and assess what is written. In various NLP tasks, including ASAS, language models (LMs) have demonstrated significant success. These models assess the probability of word sequences and can predict subsequent words based on the preceding words within a sequence [2]. Traditional language models, such as n-gram language models, employ count-based methods to evaluate and understand text. These models typically rely on the frequency of word sequences (n-grams) to predict the likelihood of subsequent words and to determine the overall structure and coherence of a text. In the context of automatic answer scoring, vector-space models that count n-grams have been widely applied due to their simplicity and effectiveness. For instance, [3] conducted a comparative study where they evaluated the efficacy of bag-of-n-gram representation against bags of semantic annotations for the ASAS task. Their findings highlighted the strengths and limitations of count-based models in capturing the nuances and subtleties of human language, emphasizing the need for more sophisticated approaches that can understand the deeper semantic meaning of text. In modern approaches, language models (LMs) are trained using neural networks, which address several limitations inherent in

traditional count-based methods. Firstly, they significantly expand the context taken into account, allowing for a more comprehensive understanding of text beyond the fixed-length context of n-grams. Secondly, these models exhibit a generalization capability across different contexts, which enhances their ability to handle diverse linguistic patterns and structures. The initial neural models were based on recurrent neural networks (RNNs), which are well-suited for sequential data. Among these, long short-term memory networks (LSTMs) became particularly popular due to their ability to capture long-range dependencies in text. LSTMs address the vanishing gradient problem in standard RNNs, enabling the model to retain information over longer sequences. The most recent advancements in neural models for language modeling, such as BERT (Bidirectional Encoder Representations from Transformers) introduced by [4], are based on the transformer architecture. This architecture represents a significant shift from previous models like RNNs and LSTMs, leveraging self-attention mechanisms to process and understand text. The transformer architecture enables these models to capture intricate dependencies and contextual relationships within text more effectively, making them highly suitable for a wide range of Natural Language Processing (NLP) tasks.

In this work, we introduce a novel automatic answer-scoring framework that combines the strengths of a pre-trained BERT model through fine-tuning with the capabilities of an LSTM network. BERT is a multi-layer bidirectional Transformer encoder designed for Natural Language Processing (NLP). Developed by Google, the pre-trained BERT model leverages a vast amount of unlabeled data, including 800 million words from books and 2.5 billion words from Wikipedia. By fine-tuning an additional classification layer along with all the pre-trained parameters, BERT can be adapted for specific NLP tasks. LSTM (Long Short-Term Memory) is a type of recurrent neural network (RNN) designed to effectively capture long-range dependencies and temporal patterns in sequential data. It deals with long-term dependencies, by incorporating memory cells and gating mechanisms to control the flow of information. These features enable LSTM networks to remember and utilize information from earlier time steps, making them well-suited for tasks involving sequential data. This hybrid approach leverages the advanced contextual understanding of BERT and the sequential processing power of LSTM to enhance the accuracy and efficiency of scoring short textual responses.

The findings of our research have two significant implications for both science and society. From a scientific perspective, we demonstrate the effectiveness of combining large language models like BERT with recurrent neural networks to improve text representation and enhance the accuracy of automated short-answer scoring. On the other hand, our research could have a significant implicant on society by reducing the workload on educators, allowing them to focus more on interactive and personalized teaching. Furthermore, this system could be widely implemented in online educational platforms, ensuring consistent and fair assessment for a growing number of students worldwide.

The remainder of this paper is structured as follows. Section II provides a brief review of related works. Section III introduces and elaborates on the proposed approach. Section IV outlines the experimental details, including the dataset, metrics, and implementation settings. Section V presents the results and the corresponding discussions. Finally, Section VI summarizes the study and suggests potential directions for future improvement.

II. Literature Review

Research in grading natural language responses with computational methods has a history dating back to the early work of [5], However, it is only in the current decade that these systems are achieving the level of accuracy necessary for practical use in educational settings. For end users to have confidence in these systems, the challenge lies in developing robust and accurate assessment mechanisms that closely mirror human evaluators. Several methods have been introduced in this area.

Early methods for automatic grading relied heavily on pattern matching, which required significant expert intervention to extract relevant patterns and features from student responses. The study in [6] explored the use of concept mapping techniques, mapping the related concepts in student answers to those in desired answers [7], [8]. Further developed the concept of information extraction from student answers through pattern matching. These researchers utilized regular expressions and parse trees to identify and extract relevant patterns within the text [9]. Involved comparing eight knowledge-based text similarity measures alongside two prominent corpus-based measures: Latent Semantic Analysis (LSA) and Explicit Semantic Analysis (ESA). These measures were trained on both domain-specific and generic corpora.

Later on, the use of machine learning techniques for automatic scoring has become popular [10]. Integrating machine learning into their work [9] to improve the performance involves graph alignment and lexical semantic similarity features using SVM and term frequency-inverse document frequency (TF-IDF). In a similar work, an approach was presented by[11], that suggested a short text similarity-based short answer grading method. They extracted multiple features such as text alignment, vector similarity, TF-IDF, and length ratios. In a similar context [12] combined sentence-level and token-level features for their approach. For sentence-level features, they employed InferSent, a pre-trained sentence embedding model that makes use of a Bi-LSTM network, to get sentence embeddings for the question, the reference answer, and the learner's answer. Semantic representations of the text are also extracted using deep learning-based word embeddings. [13] Employed standard NLP embeddings, such as Word2Vec, GloVe, and FastText, to extract the semantic and distributional properties. The study in [14] Propose a recurrent neural network to resolve the task, by staking three layers of Siamese Bi-LSTMs layer, a pooling layer using earth-mover distance (EMD), and an output layer with regression the predict the score [15]. Then enhanced the RNN-based technique by leveraging LSTM along with sense vectors and Manhattan distance in place of the pooling layer.

While the aforementioned deep learning techniques accomplish end-to-end grading and scoring, they are dependent on a substantial volume of labeled corpus for model training, which is not present in the majority of ASAG corpora. Several pre-trained transfer learning models are used to tackle this challenge [16]. Some works use these pre-trained models by

extracting embeddings directly from language models such as BERT, GPT, and ELMO[17], other works incorporate fine-tuning paradigms, such as those by [18], [19].

III. Proposed Method

The automatic short answer grading can be approached both as a regression and as a classification problem [20]; where in a regression task the model is trained to predict a continuous grade for a student's answer. The prediction is typically based on the similarity between the reference answer (the correct answer) and the student's answer. The goal is to minimize the difference between the predicted grade and the actual grade assigned by human graders. The problem can be also a classification task; the model classifies the student's answer into discrete categories such as "correct," "partially correct," or "incorrect." More fine-grained classes can also be defined, depending on the specific requirements of the grading system. The present research implements the regression task to allow for the assignment of continuous scores and provide a finer granularity in grading.

The core concept of our model is to harness the advantages of both the transformer architecture and classical neural networks to accurately interpret the given text. Specifically, we use BERT for its capabilities in understanding the context and semantics of the input text. To capture the sequential dependencies and long-range relationships in the text we also employ LSTM (Long Short-Term Memory).

First, the stacked transformer encoders in BERT aid in conserving computational resources, by using a weighted sum of all other words' embeddings to encode the hidden state of a word [16]. Although this method makes good use of the connections between every word in a phrase, it falls short in terms of taking word order and spacing into account [21]. The LSTM network is ideally suited to produce more precise global context data because of its memory cells and gate architecture. By incorporating temporal information, LSTM can compensate for the shortcomings of BERT’s encoding.

Second, compared to conventional static embeddings like Glove, BERT's dynamic word embeddings, which are produced via extensive unsupervised pretraining and refined on downstream tasks, offer substantial advantages over traditional static embeddings like Glove[4]. These dynamic embeddings provide more comprehensive and flexible general-purpose information by adapting to various settings. This feature enhances the training and convergence of upper-layer neural networks, enabling classical neural networks on BERT to attain impressive performance even with fewer datasets.

Empirical studies support the effectiveness of combining recurrent neural networks with fine-tuned BERT models, particularly in specialized tasks with limited training data. For instance, [22] demonstrated improved aspect-category sentiment analysis by integrating RoBERTa with a specialized CNN, leveraging the strengths of both architectures. [23] Explored the potential of combining BERT with neural networks for sentiment analysis purposes to classify students' reviews on Moocs. They specifically add an LSTM on top of BERT and then a CNN as a local feature extractor.

Given these insights, our proposal is a newly designed model that combines the strengths of the fine-tuning BERT model along with the LSTM network to properly understand as well as evaluate student's responses. For that, we incorporate an LSTM layer on the top of the fine-tuned BERT. The LSTM network extracts fine global context from BERT outputs, providing a richer understanding of temporal relationships. “Fig. 1” illustrates the framework of the proposed approach which is described below in detail.

A. Fine-Tuning Bert Component

BERT (Bidirectional Encoder Representations from Transformers) [4] is a groundbreaking pre-trained language representation model developed by Google AI Language[4]. The training is conducted on a massive corpus comprising the Book Corpus with 800 million words and English Wikipedia with 2.5 billion words. The masked language model and next-sentence prediction are two unsupervised tasks that were used to train the initial BERT model. These tasks involved layers for language model decoding and classification. However, for fine-tuning our model for the sentence pair classification task, we do not utilize these specific layers.

A text pair containing the student's answer and reference one is what our model receives as input. Every sequence starts with a unique classification token (CLS), as illustrated in "Fig. 2". To distinguish between the input pair, a special token (SEP) is inserted at the end of each input to help the model understand the end of an answer and the beginning of another. Wordpiece embeddings are used as the token input by BERT. For every token, BERT employs positional embeddings and segment embeddings in addition to token embeddings. Token positions in sequence are indicated by the information included in positional embeddings. When the model input contains sentence pairs, segment embeddings come in helpful. Tokens belonging to the first sentence will have a segment embedding of 0, whereas tokens belonging to the second sentence will have a segment embedding of 1. BERT incorporates embedding of the input by summing up the three embeddings, namely token, position, and segment embeddings, creating a rich representation for each token in the sequence. These representations are then fed to a multilayer bidirectional transformer encoder “Fig. 2” which is the core component of Bert’s architecture, leveraging a multi-head attention mechanism (Formula 1) to focus on data from many representation subspaces at different points in the input sequence simultaneously. After the multihead attention, each transformer layer additionally has a fully connected feed-forward network. Twelve transformer layers are stacked in the model's basic version (BERT_base). The output contextualized embeddings of Bert are then fed to an Lstm Layer to capture more information from input answers. BERT incorporates an attention mechanism to focus on different parts of the input sequence. This mechanism uses Formula (1) to compute the attention weights:

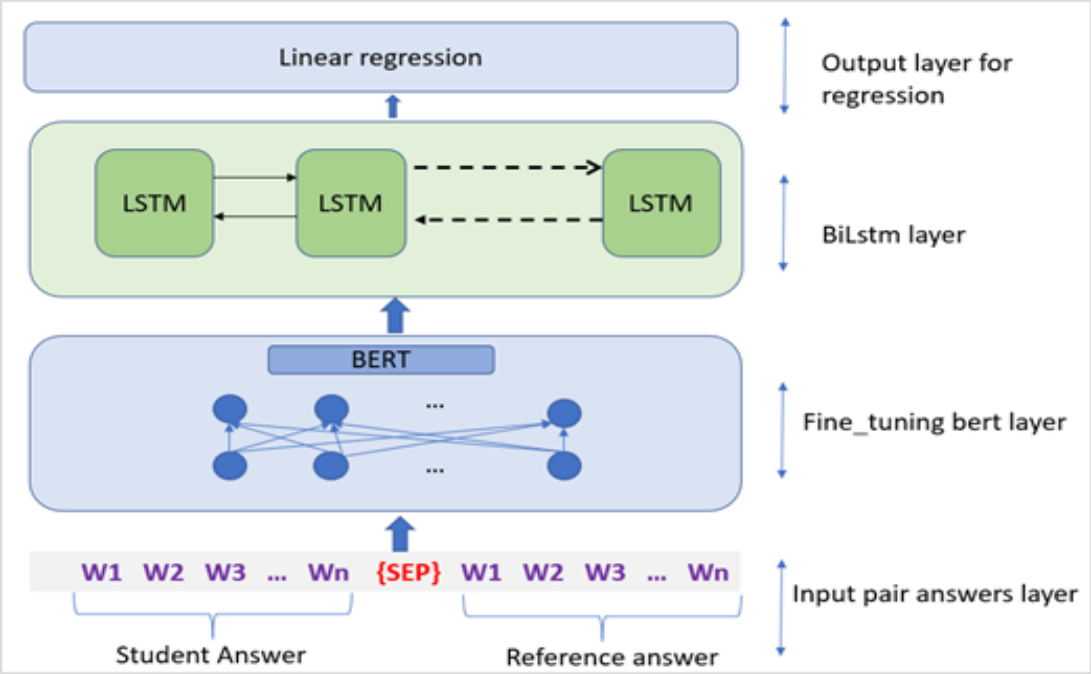


Fig. 1. The general architecture of the proposed model.

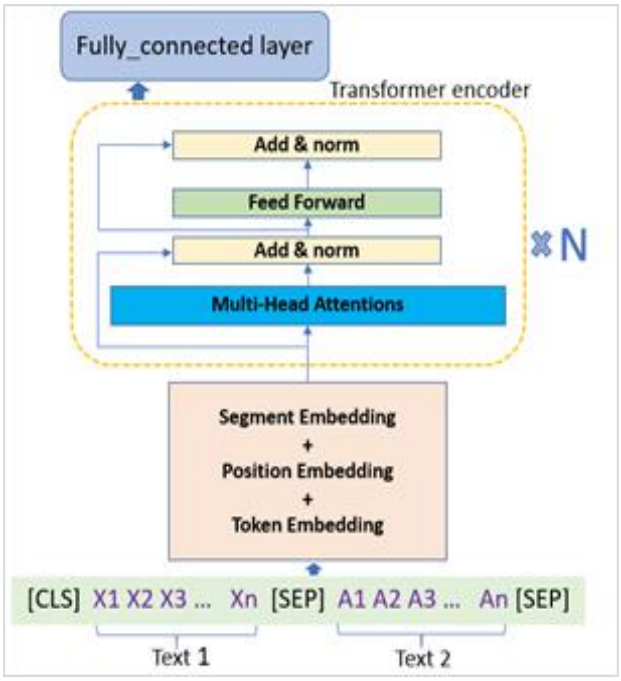


Fig. 2. The overall structure of the finetuned bert layer.

$$Attention(Q,K,V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (1)$$

Where Q(query) is the matrix of queries, representing the current state or the part of the input we are focusing on, K(key) represents the matrix of keys, representing the entire input sequence, V(value) is the dimension of the key vectors, used for scaling the dot product, and d_k the dimension of the key vectors, used for scaling the dot product.

B. LSTM Component

Aiming to augment an already outstanding model into an even more proficient automatic answer-scoring framework, we delved into the concept of incorporating an LSTM layer into the final fully connected layer of the transformers within BERT.

Recurrent neural network (RNN) architectures with Long Short-Term Memory (LSTM) networks are intended to simulate sequences and their dependence upon them over time more accurately than RNNs with typical architectures. LSTMs were introduced by Hochreiter and Schmidhuber in 1997[24] and have since become a fundamental building block for many sequential data processing tasks. LSTMs are composed of units called LSTM cells, which replace the simple neurons in standard RNNs. Each LSTM cell maintains a cell state (C_t) defined in Formula (6) and three gates that control the flow of information: the input gate (I_t), the forget gate (F_t), and the output gate (O_t) defined in Formulas (3), (2) and (4) respectively. Forget Gate eliminates data that is no longer helpful to the LSTM, which has a sigmoid layer for decision-making. *tanh* and sigmoid layers are used by the input gate, which is in charge of adding pertinent data to the existing cell state. With the use of a sigmoid layer, the output gate displays the pertinent data from the current cell. "Fig. 3" displays the construction of the LSTM.

$$F_t = \sigma\left(W_f \,.\, [h_{t-1}, x_t] \;+\; b_f\right) \quad (2)$$

$$I_t = \sigma(W_i \,.\, [h_{t-1}, x_t] \;+\; b_i) \quad (3)$$

$$O_t = \sigma(W_o \,.\, [h_{t-1}, x_t] \;+\; b_o) \quad (4)$$

$$\tilde{C}_t = tanh(W_c \,.\, [h_{t-1}, x_t] \;+\; b_c) \quad (5)$$

$$C_t = F_t \,.\, C_{t-1} + \; I_t \,.\, \tilde{C}_t) \quad (6)$$

$$h_t = O_t \,.\, \tanh(C_t) \quad (7)$$

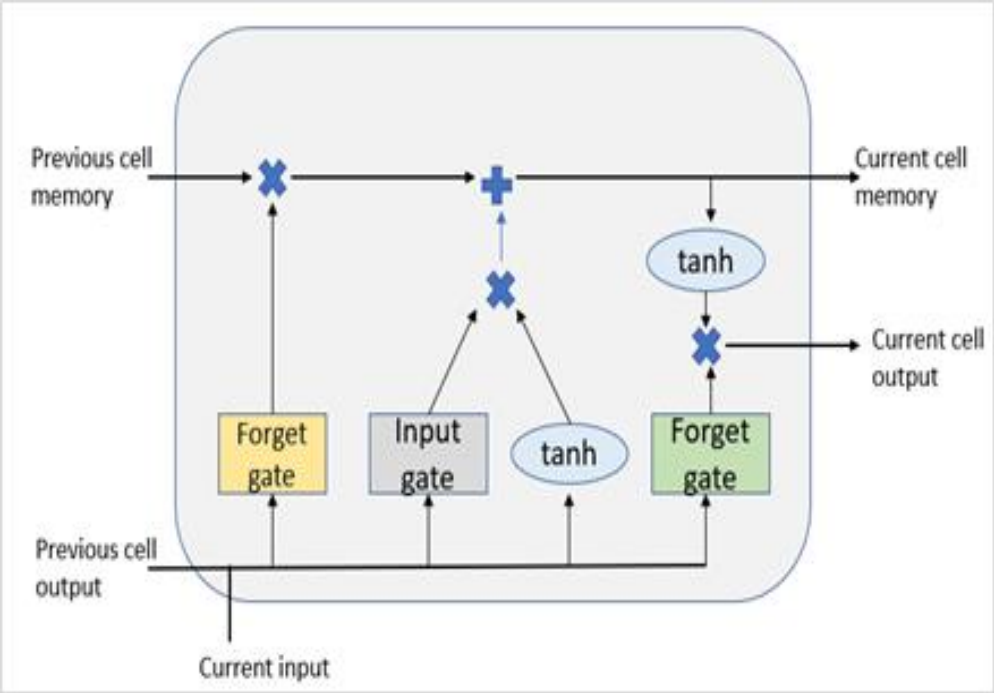


Fig. 3. A single unit of long short-term memory (LSTM) neural networks.

IV. Experiments

A. Dataset

This study leverages the computer science dataset created by [10], which is a comprehensive collection of student responses from the University of North Texas, specifically designed to evaluate the effectiveness of automatic scoring systems. The dataset contains 2273 student answers associated with 80 questions, sourced from 10 different assignments and two tests within the field of computer science. Each answer provided by the students was independently scored by two teachers, utilizing an integer-based grading scale that ranges from 0 to 5, where 0 represents an incorrect answer and 5 indicates a fully correct response. The true score of the student's answer was determined by taking the average of the two scores that were labeled, which resulted in 11 scoring grades ranging from 0 to 5 with 0.5 intervals between each grade. The shape of the dataset is (2273, 7). It has 2273 rows and seven columns. The columns have questions, desired answers, student answers, and scores.

B. Evaluation Measures

To evaluate our model, we adopt the standard metrics used by the previous automatic scoring systems. As explained above, we model the problem as a regression task, specifically using the Pearson correlation coefficient (Pearson's r) and root mean square error (RMSE) as metrics to measure the performance of the proposed approach. Below, we report the different metrics with their mathematical expressions:

Root-Mean-Squared Error (RMSE): The use of RMSE is very common, and it is considered an excellent general-purpose error metric for numerical predictions. RMSE is defined in Formula (8).

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\widehat{y_i} - y_i)^2} \quad (8)$$

Where n is the number of samples, $\widehat{y_i}$ is the predicted value and y_i represent the actual value.

1) Pearson's r[25]: measures the strength and direction of the linear relationship between two continuous variables. In the context of evaluating model performance, Pearson's r can be used to assess the correlation between predicted scores generated by a model and the actual scores assigned by human evaluators. Pearson's r is defined in Formula (9) :

$$r = \frac{\sum_{i=1}^{n}(y_i - \bar{y})(\widehat{y_i} - \bar{\hat{y}})}{\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2 \sum_{i=1}^{n}(\widehat{y_i} - \bar{\hat{y}})^2}} \quad (9)$$

Where y_i the actual value, $\widehat{y_i}$ the predicted value, $\bar{y}$ the mean of actual values and $\bar{\hat{y}}$ is the mean of predicted values.

C. Implementation Details

Our experiments were conducted in the Google Collaboratory environment, which provides access to high-performance GPUs, facilitating efficient and accelerated model training and evaluation. For the implementation of our models, we utilized Python as the primary programming language, leveraging powerful libraries such as NLTK (Natural Language Toolkit) for text preprocessing and PyTorch for building and training our neural network models. The dataset used in our study, consisting of student answers and corresponding questions, was divided into two distinct sets: 80% of the data was allocated for training the models, while the remaining 20% was reserved for testing their performance.

The implementation of our method consists first of a preprocessing step to prepare data in the best format for training.

For that, we consider two functions based on their effect on training results. The first function removes stop words from the texts. The second function is removing words that appear in the question text from both the reference answer and the student response to prevent the model from unfairly rewarding a student response that simply repeats words from the question. The texts are then tokenized using Bert tokenizer, which converts them into tokens. These tokens are further processed into input IDs, attention masks, and token type IDs, forming the input features for the BERT model. We use the Bert_base_uncased version as the pre-trained model of our Bert layer with 12 transformer layers, 768 hidden units per layer, and total parameters of ¼ 110M. The "uncased" nature of this model converts text to lowercase and removes accent markers, simplifying vocabulary handling. It employs WordPiece embeddings with a 30,000-token vocabulary, supporting input sequences up to 512 tokens. This model provides dynamic, context-sensitive word embeddings, significantly improving our model's ability to accurately evaluate student answers through fine-tuning our specific dataset. The output of the fine-tuned BERT model, which consists of contextualized embeddings of the paired answers, is then fed into an LSTM layer followed by a linear output layer for regression. The specific parameter settings for the model are detailed in Table I. This configuration allows the model to capture both the intricate context provided by BERT and the sequential dependencies effectively modeled by the LSTM.

TABLE I. PARAMETER SETTINGS OF THE PROPOSED MODEL

Parameter	Value
Batch size	16
Epochs	10
Bert's finetuned learning rate	5e-5
Lstm_hidden_size	256
Number lstm layers	2
Lstm learning rate	1e-3
Optimize	ADAM
Loss function	Mean Square Error

V. RESULTS AND DISCUSSION

A. Ablation Study

In our proposed model, we conducted an ablation study to evaluate the impact of specific components on performance. The study focused on two key variations:

1) BERT fine-tuning with and without question demotion: We examined the effect of removing question-related words from the student's answer before feeding it into the model. This step aims to reduce noise and focus on the unique content of the student's response. By comparing the performance of the model with and without question demotion, we aimed to assess its contribution to the overall accuracy.

2) BERT fine-tuning with and without adding an LSTM layer: To determine the added value of incorporating a Long Short-Term Memory (LSTM) layer, we compared the results of the fine-tuned BERT model both with and without the LSTM layer. The LSTM layer is designed to capture sequential dependencies and provide additional context to the BERT representations. This comparison helps to understand whether the LSTM layer enhances the model's ability to accurately score the answers.

As illustrated in Table II, removing question demotion results in a higher RMSE (0.931 vs. 0.785) and a lower Pearson correlation (0.723 vs. 0.761). This indicates that question demotion significantly contributes to the model's ability to accurately score answers. Question demotion likely helps the model focus on the core content of student answers without being misled by repetitive or irrelevant information from the questions, leading to better alignment with the desired answers.

TABLE II. RESULTS OF THE ABLATION STUDIES

Model Variant	RMSE	Pearson Correlation
Without Question Demotion	0.931	0.723
Without LSTM Layer	0.819	0.741
Full Model (with all components)	**0.785**	**0.761**

Adding the LSTM layer to the model improves performance, reducing the RMSE from 0.819 to 0.785 and increasing the Pearson correlation from 0.741 to 0.761. The LSTM layer likely helps capture sequential dependencies and fine-grained contextual information that the BERT layer might not fully encode, resulting in better performance.

The full model, which includes both question demotion and the LSTM layer, performs the best with the lowest RMSE (0.785) and the highest Pearson correlation (0.761). This demonstrates that both components are essential for achieving optimal performance in automatic answer scoring.

B. Comparison with Baseline Models

We compare the performance of our model with various baseline models based on RME and Pearson correlation scores. The comparison results are illustrated in Table III.

As can be seen from the experimental findings, systems that are based on handcrafted features are relatively yield low to moderate accuracy. Among these methods, the BOW (Bag of Words) approach combined with SVMRank [10] exhibited the best performance, yielding a Pearson's correlation coefficient of 0.480 and an RMSE of 1.042. This indicates that while feature engineering-based models can capture some relevant aspects of the answer-scoring task, their performance is limited compared to more advanced deep-learning models. The moderate correlation and relatively high RMSE suggest that these methods might struggle with capturing the deeper semantic relationships and nuances present in the text. Combining semantic network approaches using Glove and Word2Vec embeddings along with an SVM model slightly improves the performance metrics, achieving a Pearson's correlation coefficient of 0.631 and an RMSE of 0.834 [26]. This enhancement suggests that integrating semantic information from pre-trained embeddings can better capture the underlying meaning and context of the text, leading to more accurate scoring. However, the improvement is still moderate, indicating that these traditional machine learning methods, even when augmented with semantic embeddings, may not fully exploit the complexities of the language as effectively as more advanced deep learning techniques. Using dynamic embeddings only,

without fine-tuning pre-trained models such as ELMo, GPT, and BERT, performs poorly in similarity regression tasks. For instance, the results show that traditional word embeddings (e.g., Word2Vec, GloVe) yield better performance metrics compared to contextual embeddings (e.g., ELMo, BERT) [17]. This observation highlights that merely leveraging the powerful pre-trained models without task-specific fine-tuning can lead to suboptimal results, as these models may not fully align with the specific requirements and nuances of the target task.

The experiments show that the fine-tuned BERT model performs very well, achieving RMSE and Pearson correlation values of 0.819 and 0.741, respectively. This indicates that task-specific fine-tuning significantly enhances the model's ability to capture the nuances and intricacies of the dataset, resulting in improved scoring accuracy and correlation with the target metrics. Finally, the results show that adding an LSTM layer on top of the fine-tuned BERT model improves the results, achieving a Pearson correlation of 0.761 and an RMSE of 0.785. This significantly surpasses the results of all baseline systems, demonstrating the effectiveness of combining BERT's powerful language representation with LSTM's ability to capture long-term dependencies. The experiments highlight the limitations of feature engineering-based and dynamic embedding-only models. Fine-tuning pre-trained models, especially when combined with additional layers like LSTM, significantly improves performance. Our proposed model, BERT Fine-Tuned Based LSTM, achieves the best results, establishing a new benchmark for automatic answer scoring on the Mohler dataset.

TABLE III. COMPARISON RESULTS ON THE MOHLER DATASET

System	description		RMSE	Pearson correlation
[10]	BOW (Bag of Words) approach with SVMRank		1.042	0.480
	BOW (Bag of Words) approach with SVR		0.999	0.431
	Tf-idf with SVR		1.022	0.327
[11]	tf-idf with LR (Logistic Regression) and SIM (Semantic Information)		0.887	0.592
[12]	HoPSTags + Sentence Embedding features		0.921	0.542
[17]	Dynamic embeddings (not fine-tuned + cosine similarity feature	ELMO	0.978	0.485
		GPT	1.082	0.248
		BERT	1.057	0.318
		GPT_2	1.065	0.311
[26]	Semantic network with SVM		0.834	0.631
(In this work)	Word2vec & mean_pooling with cosine similarity feature		1.005	0.405
	Bert(embedding only) & mean_pooling with cosine similarity feature		1.021	0.367
	Fine_tuned Bert_base		0.819	0.741
(Proposed model)	BERT Fine-Tuned Based LSTM		**0.785**	**0.761**

VI. CONCLUSION

In this paper, we introduce a new method for automatic answer scoring by leveraging the strengths of both transformer-based and classical neural network architectures. The proposed model contains a fine-tuned layer of the pre-trained BERTbase model for contextualized embedding extraction, followed by an LSTM layer to benefit from its sequence modeling capabilities for more improvement. The model was trained using the Mohler dataset, a benchmark corpus widely used for automatic scoring tasks. In the experiments, we compared our model with several state-of-the-art models to evaluate the performance. The results demonstrated that our approach shows significant improvement regarding both RMSE and Pearson correlation measures. These improvements underscore our model's enhanced capability to understand and evaluate the semantic content of both student and reference answers, leading to more accurate grading outcomes. In future work, we can improve such an automatic scoring system so it can face the problem of different distributions that can arise due for example to differences in question types between the current question answers (training set) and the new question answers (test set). Strategies based on domain adaptation and transfer learning can be employed to address this case.

REFERENCES

[1] V. Salvatore, N. Francesca, et A. Cucchiarelli, « An Overview of Current Research on Automated Essay Grading », J. Inf. Technol. Educ., vol. 2, janv. 2003, doi: 10.28945/331.

[2] Y. Goldberg, « Neural Network Methods for Natural Language Processing », Synth. Lect. Hum. Lang. Technol., vol. 10, p. 1-309, avr. 2017, doi: 10.2200/S00762ED1V01Y201703HLT037.

[3] J. G. A. Mantecon, H. A. Ghavidel, A. Zouaq, J. Jovanovic, et J. McDonald, « A Comparison of Features for the Automatic Labeling of Student Answers to Open-Ended Questions », International Educational Data Mining Society, juill. 2018. Consulté le: 11 juillet 2024. [En ligne]. Disponible sur: https://eric.ed.gov/?id=ED593101

[4] J. Devlin, M.-W. Chang, K. Lee, et K. Toutanova, « BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding », 24 mai 2019, arXiv: arXiv:1810.04805. doi: 10.48550/arXiv.1810.04805.

[5] E. B. Page, « Computer Grading of Student Prose, Using Modern Concepts and Software », J. Exp. Educ., vol. 62, no 2, p. 127-142, janv. 1994, doi: 10.1080/00220973.1994.9943835.

[6] L. James, « CAA of Short Non-MCQ Answers », 2001.

[7] L. F. Bachman et al., « A Reliable Approach to Automatic Assessment of Short Answer Free Responses », in COLING 2002: The 17th International Conference on Computational Linguistics: Project Notes, 2002. Consulté le: 12 juillet 2024. [En ligne]. Disponible sur: https://aclanthology.org/C02-2023

[8] P. Thomas, « The evaluation of electronic marking of examinations », sept. 2003, doi: 10.1145/961511.961528.

[9] M. Mohler et R. Mihalcea, « Text-to-Text Semantic Similarity for Automatic Short Answer Grading », in Proceedings of the 12th Conference of the European Chapter of the ACL (EACL 2009), A. Lascarides, C. Gardent, et J. Nivre, Éd., Athens, Greece: Association for Computational Linguistics, mars 2009, p. 567-575. Consulté le: 12 juillet 2024. [En ligne]. Disponible sur: https://aclanthology.org/E09-1065

[10] M. Mohler, R. Bunescu, et R. Mihalcea, « Learning to Grade Short Answer Questions using Semantic Similarity Measures and Dependency Graph Alignments », in Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies, Portland, Oregon, USA: Association for Computational Linguistics, juin 2011, p. 752-762. [En ligne]. Disponible sur: https://aclanthology.org/P11-1076

[11] M. A. Sultan, C. Salazar, et T. Sumner, « Fast and Easy Short Answer Grading with High Accuracy », in Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, San Diego, California: Association for Computational Linguistics, 2016, p. 1070-1075. doi: 10.18653/v1/N16-1123.

[12] S. Saha, T. I. Dhamecha, S. Marvaniya, R. Sindhgatta, et B. Sengupta, « Sentence Level or Token Level Features for Automatic Short Answer Grading?: Use Both », in Artificial Intelligence in Education, vol. 10947, C. Penstein Rosé, R. Martínez-Maldonado, H. U. Hoppe, R. Luckin, M. Mavrikis, K. Porayska-Pomsta, B. McLaren, et B. du Boulay, Éd., in Lecture Notes in Computer Science, vol. 10947. , Cham: Springer International Publishing, 2018, p. 503-517. doi: 10.1007/978-3-319-93843-1_37.

[13] T. D. Metzler, P. G. Plöger, et G. Kraetzschmar, « Computer-assisted grading of short answers using word embeddings and keyphrase extraction », PhD Thesis, Master's thesis, Hochschule Bonn-Rhein-Sieg, Germany, 2019.

[14] S. Kumar, S. Chakrabarti, et S. Roy, « Earth Mover's Distance Pooling over Siamese LSTMs for Automatic Short Answer Grading », in Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, Melbourne, Australia: International Joint Conferences on Artificial Intelligence Organization, août 2017, p. 2046-2052. doi: 10.24963/ijcai.2017/284.

[15] C. N. Tulu, O. Ozkaya, et U. Orhan, « Automatic Short Answer Grading With SemSpace Sense Vectors and MaLSTM », IEEE Access, vol. 9, p. 19270-19280, 2021, doi: 10.1109/ACCESS.2021.3054346.

[16] A. Vaswani et al., « Attention Is All You Need », 5 décembre 2017, arXiv: arXiv:1706.03762. doi: 10.48550/arXiv.1706.03762.

[17] S. K. Gaddipati, D. Nair, et P. G. Plöger, « Comparative Evaluation of Pretrained Transfer Learning Models on Automatic Short Answer Grading », 2 septembre 2020, arXiv: arXiv:2009.01303. Consulté le: 13 mai 2024. [En ligne]. Disponible sur: http://arxiv.org/abs/2009.01303

[18] L. Camus et A. Filighera, « Investigating Transformers for Automatic Short Answer Grading », in Artificial Intelligence in Education, I. I. Bittencourt, M. Cukurova, K. Muldner, R. Luckin, et E. Millán, Éd., in Lecture Notes in Computer Science. Cham: Springer International Publishing, 2020, p. 43-48. doi: 10.1007/978-3-030-52240-7_8.

[19] C. Sung, T. I. Dhamecha, et N. Mukhi, « Improving Short Answer Grading Using Transformer-Based Pre-training », in Artificial Intelligence in Education, S. Isotani, E. Millán, A. Ogan, P. Hastings, B. McLaren, et R. Luckin, Éd., Cham: Springer International Publishing, 2019, p. 469-481. doi: 10.1007/978-3-030-23204-7_39.

[20] R. Dadi et S. Sanampudi, « An automated essay scoring systems: a systematic literature review », Artif. Intell. Rev., vol. 55, p. 1-33, mars 2022, doi: 10.1007/s10462-021-10068-2.

[21] A. Rogers, O. Kovaleva, et A. Rumshisky, « A Primer in BERTology: What We Know About How BERT Works ».

[22] W. Liao, B. Zeng, X. Yin, et P. Wei, « An improved aspect-category sentiment analysis model for text sentiment analysis based on RoBERTa », Appl. Intell., vol. 51, no 6, p. 3522-3533, juin 2021, doi: 10.1007/s10489-020-01964-1.

[23] A. Baqach et B. Amal, « A new sentiment analysis model to classify students' reviews on MOOCs », Educ. Inf. Technol., p. 1-28, févr. 2024, doi: 10.1007/s10639-024-12526-0.

[24] S. Hochreiter et J. Schmidhuber, « Long Short-Term Memory », Neural Comput., vol. 9, no 8, p. 1735-1780, nov. 1997, doi: 10.1162/neco.1997.9.8.1735.

[25] D. G. Bonett et T. A. Wright, « Sample size requirements for estimating pearson, kendall and spearman correlations », Psychometrika, vol. 65, no 1, p. 23-28, mars 2000, doi: 10.1007/BF02294183.

[26] N. H. Hameed et A. T. Sadiq, « Automatic Short Answer Grading System Based on Semantic Networks and Support Vector Machine », Iraqi J. Sci., p. 6025-6040, nov. 2023, doi: 10.24996/ijs.2023.64.11.44.

BlockChain and Deep Learning with Dynamic Pattern Features for Lung Cancer Diagnosis

A. Angel Mary[1], Dr. K.K. Thanammal[2]
Research Scholar, Department of Computer Science and Research Centre, S.T. Hindu College, Nagercoil, 629002, Manonmanium Sundaranar University, Abishekapatti, Tirunelveli, 627012, Tamilnadu, India[1]
Assistant Professor, Department of Computer Science and Research Centre, S.T. Hindu College, Nagercoil[2]

***Abstract*—Cancers in the respiratory tract grow out of control in lung carcinoma, a deadly disease. Because cancers have irregular shapes, it can be challenging to diagnose them and determine their sizes and forms from imaging studies. Furthermore, a serious issue with health image inquiry is large disparity. Artificial intelligence and blockchain are two cutting-edge advances in the healthcare industry. This paper introduces a Blockchain with a deep learning network for the early diagnosis of lung cancer in an effort to address these problems. Images from CT scans and CXRs were included in the LIDC-IDRI and NIH Chest X-ray collection. Initially, these images are pre-processed by Contrast Limited Adaptive Histogram Equalization (CLAHE) to enhance the image clarity and reduce the noise. Then the Honey Badger optimization Algorithm (HBA) is used to segment the lung region from the pre-processed image. Morphological segments of the lung region are used to generate dynamic patterns. Finally, these patterns are aggregated into the deep neural Spiking Convolutional Neural Network (SCNN), which is the global model for classifying the images into normal and abnormal cases. Based on the classification, the SCNN model achieves 98.64% accuracy from the LIDC-IDRI database and 98.9% on the NH Chest X-ray lung image dataset. The experiments indicate that the proposed approach results in lower energy consumption and faster inference times. Furthermore, the interpretability of the classification findings is improved by the intrinsic explainability of SCNNs, offering more profound understanding of the decision-making process. With these benefits, SCNNs are positioned as a reliable and effective technique for classifying lung images, providing a significant advancement over current methods.**

Keywords*—Lung cancer; spiking convolutional neural network; LIDC-IDRI; CLAHE; Honey Badger optimization Algorithm; segmentation; classification

I. INTRODUCTION

As the primary cause of cancer-related fatalities worldwide, lung cancer is also one of the most often diagnosed malignancies. The World Health Organization predicted that in 2020, there would be over 1.8 million lung cancer deaths and about 2.21 million new cases of the disease. In addition, it is estimated that 17 million people worldwide would suffer from cancer by 2030 [1]. Cigarette smoking, the primary cause of lung cancer, accounts for 80% of the disease's mortality. Exposure to radon gas is the second most prevalent cause of lung cancer [2]. Only 21% of cases of early-stage lung cancer are identified at stage I, with most cases being detected at stage III or IV (representing 61% of all newly discovered lung cancers)due to the disease's characteristic lack of symptoms [3]. The high fatality rate and aggressive nature of the illness are mainly due to late-stage detection [4]. In order to lower the death rate from lung cancer, early detection of smaller tumours and nodules using X-rays and CT scans is especially crucial because the prognosis for early treatment is noticeably better than that for later stages.

Due to the complicated anatomy of the lungs, many clinical decisions support systems, particularly machine learning techniques, rely on segmentation. Information technology has advanced recently beyond only making people's lives easier; the outcome is AI technology that offers healthcare facilities and improved quality of life. It is acknowledged that efficient data collection, processing, analysis, and safe storage are essential procedures. The first and most important challenge is the ongoing inflow of data into the medical field; this issue might be crucial to the development of effective and secure medical data storage. For the purpose of early lung disease diagnosis, a Blockchain with a deep learning network is therefore introduced in this study.

This research work aims to construct automated lung cancer detection using region-based segmentation methods with tumor area detection and subsequently to develop an effective system for the classification of lung tumors. In order to minimize the noise present in the CXR images, these pictures undergo pre-processing using Contrast Limited Adaptive Histogram Equalization (CLAHE). Segmenting the lung cancer area is done by Honey Badger optimization Algorithm (HBA). Dynamic patterns are produced using morphological segments of the lung cancer area. Dynamic pixels are created by further separating the segmented pictures into different cells. Finally, a novel deep Spiking Convolutional Neural Network (SCNN) is used to classify normal and abnormal cases by using CT and Chest X-ray images.

In the proposed architecture, trust between local nodes at the global node layer is maintained by a reputation system based on blockchain technology. By segmenting the neural network into several networks, each with a limited set of permitted entities, storing data on the cloud gate server and allowing applications to undertake data analysis, the security of the medical data is guaranteed. Its flexibility and agility in solving intricate non-linear problems is one of its advantages. Its capacity to tackle intricate non-linear problems with great flexibility and adaptability is one of its advantages. When compared to current deep neural networks, our method yields better and more accurate results. It may be applied in medical

facilities to advance AI research and enhance early diagnosis of lung diseases.

The key contributions of the proposed model:

In this work, multi-modal images such as CT and CXR images are used for lung disease classification in early diagnosis.

To detect the lung cancer easily, segmentation process is applied. Honey badger optimization (HBA) is used for lung disease segmentation and the morphological segments of the lung region are used to generate dynamic patterns.

These dynamic patterns and output local nodes are aggregated into the deep neural Spiking Convolutional Neural Network (SCNN), also known as Global network, for classifying the input images into normal and abnormal cases.

To prevent privacy protection and secure the classification results, the block chain technology is designed.

II. Literature Survey

This review offers a general overview of deep learning-based image processing techniques used in both classic and modern methods to diagnose lung cancer.

A unique filtering method that eliminates unnecessary pictures and lowers false-positives has been presented by Liang et al. [5] for the classification of lung nodules. To locate lung nodules precisely, they employed Faster R-CNN. According to the study's findings, this method could successfully identify pulmonary nodules in CT images, which could help doctors identify lung cancer early on.

A novel deep learning model has been suggested by Asuntha et al. [6] to identify lung nodules. To extract features, one can employ feature descriptors such as wavelet transform-based features, Zernike Moment, Scale Invariant Feature Transform (SIFT), Histogram of Oriented Gradients (HoG), Local Binary Pattern (LBP), and so on. The Fuzzy Particle Swarm Optimisation (FPSO) approach is utilised to choose the best feature. Deep learning techniques are employed for classification. To lessen the network's computational complexity, a unique FPSOCNN is employed.

A novel method for classifying lung images has been suggested by Vas et al. [7]. This method uses a median filter during the preprocessing stage to eliminate the unnecessary portion of the image. Accurate lung segmentation and cancer diagnosis are made possible by the application of mathematical morphological techniques. From the divided region, the following data were extracted: energy, correlation, variance, homogeneity, difference entropy, contrast and correlation information measure. These were then sent to the feed forward neural network using the back propagation technique for classification.

An enhanced lung nodule identification method based on the YOLO-V3 target detection network was presented by Li et al. [8]. This article uses the Mask-RCNN network and enhances it using the channel shuffle convolution method and Densenet's dense block structure. A computer-aided technique for the early identification of lung cancer using CT images was proposed by Elnakib et al. [9]. A genetic algorithm (GA) is trained to optimize the obtained data set in order to determine the most significant elements. A variety of classifiers are finally looked at in order to identify the pulmonary nodules properly. Comparing the suggested technology to other earlier methods like VGG-16, AlexNet and VGG-19 networks, encouraging results are obtained.

Srinivasulu et al. [10] introduced a novel blockchain based lung cancer detection using extended CNN. There are primarily two architectures like U-Net and VGG-16 are used in the suggested method to categorize lung nodules and predict the amount of malignancy. The Internet of Things (IoT) may be used for the proposed multistage lung cancer detection and classification,

An algorithm for detecting lung nodules has been developed by Vaishnavi et al. [11]. For pre-processing, they used a discretely sampled wavelet in the Dual-tree complex wavelet transform (DTCWT). GLCM is a texture analysis technique that determines how often different Grey level combinations co-occur in an image using a second-order statistical method. They used a Probability Neural Network (PNN) classifier, whose accuracy in classification and training was evaluated.

The Faster R–CNN method for lung cancer detection was first described by Su et al. [12]. They illustrate the quicker R-CNN's suitability for lung knob recognition based on the training set. CNN and alternative optimization are the two training techniques used in the Faster R–CNN approach. Low tiny object identification accuracy is a common problem with many network models; hence, in order to increase the sensitivity to small things, the model needs to be enhanced and optimized.

A cat swarm optimization-based computer-aided diagnostic model for lung cancer classification (CSO-CADLCC) was reported by Vaiyapuri et al. [13]. The Gabor filtering-based noise reduction approach is the first pre-processing method used by the proposed CHO-CADLCC technology. Additionally, the NASNetLarge model is used to extract features from the pre-processed pictures. The CSO method with the weighted extreme learning machine (WELM) model comes next and is used to classify lung nodules. In order to optimise the WELM model's parameter tuning and get better classification performance, the CSO method is finally applied.

Inception V3, CNN, CNN GD, Resnet-50, VGG-16, and VGG-19 are the six deep learning models that Rajasekar et al. [14] suggested be used to diagnose lung cancer effectively. Based on the histopathological and CT scan pictures, experimental studies were carried out. This approach will be effective in detecting lung cancer and helpful to those in need due to the inherent benefit of the suggested methodology. The I3DR-Net is a single-stage detector that was proposed by Harsono et al. [15] to identify and categorise lung nodules. A feature pyramid network with pre-trained image weight from the inflated 3D ConvNet (I3D) was combined with a multi-scale 3D Thorax CT scans database to build the model.

For the purpose of detecting lung nodules, Schultheiss et al. [16] created the CNN-based RetinaNet architecture. The input

picture is segmented using the U-Net technique. An important part of this work was investigating the possibility that foreign substances may cause inaccurate selections in CNN-based nodule recognition systems. A multicrop CNN was presented by Shen et al. [17] that can automatically extract salient module features by max-pooling procedures performed at different times and cropping different sections from feature maps.

A lung nodule classification method based on a deep residual network is proposed in [22]. In [23], the CT image sub-block preprocessing strategy was used to extract nodule features for enhancement and alleviate the aforementioned problems. The experimental results showed that the effective classification time cost based on the original Faster R-CNN detection method. The research in [24] presented the Non-Local network by adding channel-wise attention capability and apply Curriculum Learning principles for the classification task. Assorted Scale Integrated Alternate Link Model Convolutional Neural Network method is proposed in [25] for Lung Nodule Detection. A new hybrid deep learning framework by combining VGG, data augmentation and spatial transformer network (STN) with CNN is proposed in [26]. Multiscale Rotation-Invariant Convolutional Neural Networks technique is designed to find the lung texture classification results [27]. The article [28] offered hybrid CNN along with the SVM classification method with tuned hyperparameters for Lung Cancer Detection from X-Ray Images. Table I presents the summary of existing work.

TABLE I. SUMMARY OF EXISTING WORK

REFERENCES	NETWORKS	ADVANTAGES	LIMITATIONS
[5]	R-CNN	Eliminates unnecessary pictures and lowers false-positives	Model is not strong because of insufficient samples
[6]	FPSOCNN	Find best feature	Can improve the performance with advance classification methods
[7]	FFNN	Eliminate the unnecessary portion, extract some data	insufficient data sample size
[8]	Mask-RCNN	Segmentation and classification	Need to improve the network performance and the recognition accuracy.
[9]	AlexNet	Improved the contrast of image	Insufficient training data
[10]	U-Net	classify and organize and assess threat level	Limited input and output
[11]	PNN	Classify normal and abnormal	Limited pattern neuron
[12]	R–CNN	Improved detection accuracy than existing	Used too small samples
[13]	NASNet	Preprocessing, feature extraction and classification	Can be used more samples
[14]	CNN	histopathological images are considered for the identification	Can be used advanced optimization technique
[15]	3D ConvNet	texture detection and classification	Can be implemented in real-time
[16]	U-Net and RetinaNet	Find critical positions	Used limited images
[17]	MC-CNN	Initialization approach of characterizes nodule semantic attributes	Can use more samples
[22]	ResNet	Less false positive rate	Long training time
[23]	FR-CNN	reduced the rate of misdiagnosis	Used less samples
[24]	ProCAN model	achieved state-of-the-art performances	Can improve accuracy
[25]	ASIAL CNN	More convolutional pathways	Can improve prediction accuracy
[26]	VDS net	Determine the condition of patients	Need more samples
[27]	MRCNN	Change overlapping adjacent patches	Used only CT samples
[28]	OCNN-SVM	Categorizing lung image	Want to use various data sets

III. PROPOSED METHODOLOGY

The proposed work has three stages: (a) Preprocessing, (b) Segmentation and (c) Classification. Fig. 1 demonstrates the proposed work.

A. Pre-processing using CLAHE

The accuracy of the lung image classification process is enhanced by this pre-processing phase. In this work, input CT and Chest X-ray image enhancement is achieved with Contrast Limited Adaptive Histogram Equalization (CLAHE). When compared to AHE, it processes computationally more quickly since there are no overlapping blocks [18]. The CLAHE technique effectively equalises the image histogram in addition to enhancing contrast [19].

One type of adaptive contrast augmentation method is called CLAHE. Adaptive histogram equalisation, or AHE, can be used to sharpen the borders of each picture region and boost local contrast. Updates to the enhancement computation are made for CLAHE using the maximum contrast enhancement factor and the highest clip level value, respectively. After that, the neighbouring picture areas are blended by bilinear interpolation to eliminate artificially stimulated region borders. Using this approach, medical photos can have better quality and contrast. The histograms for each region are first calculated utilising restrictions for contrast expansion and clipping in

order to maximise picture improvement. After that, the calculated histogram is re-distributed to maintain the height inside the clip limit. For CLAHE grayscale mapping, Cumulative Distribution Functions (CDFs) are computed by Eq. (1), and the histogram equalisation is obtained by estimating the CDF.

$$F_{a,b}(q) = \frac{(Q-1)}{P}\sum_{s=0}^{q} r_{a,b}(s)\,. \quad q = 1,2,\ldots..Q-1 \qquad (1)$$

The pixel units and grayscales for each area are given as P and Q in the above assessment. If the histogram of the (a, b) region is$F_{a,b}(q)$, for $q = 1, 2, \ldots..Q-1$, the CDF estimate, scaled by N-1 for grayscale mapping, is provided.

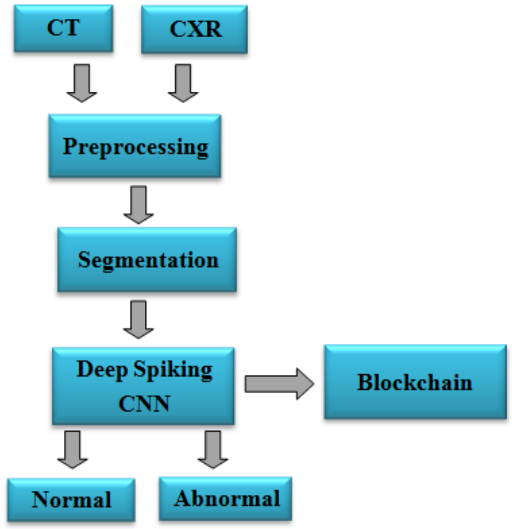


Fig. 1. Schematic representation of lung disease diagnosis framework.

B. Segmentation using Honey Badger Algorithm (HBA)

The Honey Badger Algorithm (HBA) is an excellent segmentation algorithm because of its adaptive balance between exploration and exploitation. This balance enables the algorithm to effectively navigate complex search spaces and avoid local optima. Because of its exceptional boundary detection precision, it can accurately segment images even in difficult cases with overlapping structures or fuzzy edges. It produces consistent results across various datasets and is more adept at managing noise and variability, which are prevalent in medical imaging. Additionally, it is flexible, scalable, and converges quickly, making it suitable for real-time applications.

Hence the HBA is used in the suggested way to benefit from a cutting-edge segmentation method. The segmented images are separated into various cells, which are then further separated into dynamic pixels. The center value and its neighboring bits are selected in each cell in dynamic pattern. Based on the center value, the pattern was generated by changing the gray scale values to binary values. Finding a method that can effectively separate the lungs from the CT and CXR pictures would be great, according to the previous studies. The algorithm's developers called the first technique the "digging phase" and the second the "honey phase". Fig. 2 depicts the honey badger's improved optimization algorithm.

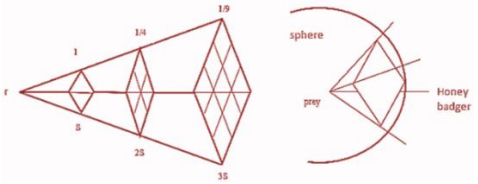


Fig. 2. Basic honey badge algorithm.

This section presents the algebraic formulation of the suggested HBA method. Since HBA includes periods for both exploration and exploitation, a global segmented technique might be considered. The following is a mathematical breakdown of the phases in the proposed HBA.

The following are the main phases of the HBA and the corresponding equations for them: The population of potential solutions for HBA is expressed in Eq. (2).

$$Pops = \begin{bmatrix} x_{11}x_{12}x_{13}\ldots x_{1D} \\ x_{21}x_{22}x_{23}\ldots x_{2D} \\ \ldots \quad \ldots \quad \ldots \quad \ldots \\ x_{n1}x_{n2}x_{n3}\ldots x_{nD} \end{bmatrix} \qquad (2)$$

$x_i = [\,x_i^1,\ x_i^2, \ldots\ldots, x_i^D]$ is the formula for the ith position of the honey badger derived from the previous equations.

1) Step 1 Initialization stage: The issue space's upper (Yu) and lower (Yl) bounds identify the first potential solution at this stage. Consequently, the first solutions, as given by Eq. (3), are made up of random sets that may be generated using the subsequent method.

$$Y_a = Yl + R_1(1,d) \times (Yu - Yl), \qquad a = 1,2,3\ldots..n. \qquad (3)$$

In this case, n solution providers (honey badgers) are given, Y displays every potential solution, and d indicates the magnitude of the solution.

2) Step 2 Updating positions: The Y_{new}coordinates for the candidates have been updated. For example, this may mean excavating or using a strategy that takes advantage of the honey phases. The ability of the hunter's scent and the distance between the prey and the honey badger (F) determine the potential search areas during the digging phase. The honey badger excavates in a circle. The following Eq. (4) describes how it moves:

$$Y_{new} = F + D_i \times \beta \times Min \times F + D_i \times R_3 \times (F - Y_a) \times (cos2\pi R_4) \times (1 - cos2\pi R_5) \qquad (4)$$

where β is the food-gathering capacity of an insect. Using a uniform distribution and a range of 0 to 1, the R_3, R_4, and R_5 are randomly chosen random variables. The level of intensity is attained. As a sign of a search strategy, the D_i is created by the following Eq. (5):

$$D_i = \begin{cases} 1 & if\ R_6 \le 0.5 \\ -1 & if\ else \end{cases} \qquad (5)$$

The honey badger phase Use the honey stage to go over with the lead bird in search of beehives. The honey phase was calculated using the following Eq. (6):

$$Y_{new} = F + D_i \times R_7 \times \sigma \times (F - Y_a) \qquad (6)$$

where R_7 is a randomly generated number between 0 and 1, and F is the highest outcome thus far.

3) Step 3 Modeling intensity min: The following computation in Equation 6 for each candidate's level of odour intensity min of the prey is needed, after which the honey badger's ability to detect insect odour controls its movements.

$$Min = \frac{R_2 \times (Y_a - Y_{a-1})^2}{4\pi(F_p - Y_a)} \quad (7)$$

The prey's location is indicated by F_p in the equation (7) above, and R_2 is an arbitrary sum between 0 and 1.

4) Step 4 Density parameter modeling ($\boldsymbol{\sigma}$): Between the local and global search stages, the information flow is regulated by the sigma value. The following Eq. (8) illustrates the hypothesis that beta is represented throughout every iteration:

$$\sigma = C \times \exp\left(\frac{-r}{r_{max}}\right) \quad (8)$$

Where, the values for r and r_{max}represent the current iteration and the total number of iterations. C stands for constant, and its recommended value is 2.

5) Step 5 Escaping from local results: The search direction is signalled with a warningD_i, which the algorithm authors utilised to avoid becoming stale on regional fixes. HBA is known as a global optimisation method due to its exploration and exploitation stages. HBA is easy to use and understand, and there are fewer operators to alter. Finally, segmented pictures are used to classify lung carcinomas. The segmentation accuracy was measured using the rate of truepositive, rate of true negative, rate of false positive, and rate of falsenegative from the algorithm.

C. Classification using Spiking Convolutional Neural Network (SCNN)

The architecture of a Spiking Convolutional Neural Network (SCNN) is similar to that of traditional Convolutional Neural Networks (CNNs), but it is adapted to work with spike-based representations and to leverage the principles of spiking neural networks (SNNs).

A dynamic pattern refers to how the rules for updating each cell's state vary dependent on multiple circumstances, such as the cell's present state, its neighbours, or external inputs. Each cell has a dynamic pattern of selection for the center value and the bits surrounding it. Binary values were substituted for the grey scale values to create the pattern based on the center value. Ultimately, the global model for categorizing the pictures into normal and abnormal situations is the deep neural Spiking Convolutional Neural Network (SCNN), which is comprised of various patterns. With almost a billion spiking neurons, the Spiking neural network architecture is a massively parallel neurocomputer design intended for categorization. A spike-carrying neural network (SCNN) may broadcast and receive large quantities of information based on the relative timing of its spikes. Basic outline of the SCNN architecture is depicted in Fig. 3:

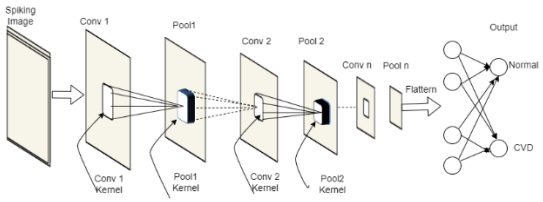


Fig. 3. SCNN network architecture.

As shown in the above figure, the convolutional layer, pooling layer, and spiking fully connected layer comprise the spiking CNN. Both the fully connected layer and the convolutional layer include layer-wise learning.

1) Convolutional layer: The convolutional layer of the network may be constructed after the convolutional filters are obtained. As in a traditional CNN, the convolutional layer employs weight sharing to lower the number of parameters. However, spike trains—rather than actual values—are used to transport the information in a spiking CNN.

The normalized grey scale pixel intensity in the range of (0,1) determines the pace at which spike trains representing the input picture (r × c pixels) are generated. The picture is sent to the network for T = 20 ms split into 1 ms time increments, as was previously mentioned. The convolution process is displayed in the network's first layer, which is illustrated in Fig. 2. To create feature maps with (r − p + 1) × (c − p + 1) LIF neurons, a collection of D filters (trained as mentioned in the previous section) are individually convolved with the picture throughout the 20-time-step presentation period. A particular aspect of the image is represented by each feature map that has LIF neurons in it. When threshold is reached, a LIF neuron in a feature map emits a spike. It aggregates the convolution results for a specific picture patch across T = 20 time steps. Upon firing, the membrane potential is returned to its resting value of zero. The membrane potential, U, of neuron m in feature map k at time t is computed using Eq. (9).

$$\frac{dU_m^k(t)}{dt} + U_m^k(t) = I_m^k(t) \quad (9)$$

If $U_m^k(t) \geq \theta^{conv}$, $then\ U_m^k(t) = 0$

I_m^k is defined below Eq. (10).

$$I_m^k(t) = \sum_{i=1}^{p}\sum_{j=1}^{p} W_k^{ex}(i,j) \,.\, S_m(i,j,t) \quad (10)$$

A representation of the filter's convolution is $I_m^k(t)$. W_k^{ex} , and at T = 20 time steps, the presynaptic spike train, S_m. The spike train that represents a pixel value in the (i,j)coordinate of patch m is called$S_m(i,j)$.

Every neuron receives inputs in the form of $p2$ spike trains with λ_{ij} rate parameters. The predicted value of the injected current at time step t, as determined by filter k for neuron m, is provided by Eq. (11).

$$E[I_m^k(t)] = \sum_{i=1}^{p}\sum_{j=1}^{p} W_k^{ex}(i,j) \,.\, \lambda_{ij} \quad (11)$$

This may be compared to $I = (w^{ex})^T.x$, where w^{ex} is a filter that represents pixel intensities and x is scaled to the average firing rates. Consequently, the convolution over T time steps may be used to approximate classical convolution.

2) Pooling layer: The spike trains generated from the convolutional layers are downsampled by the pooling layers. Spike-based representations allow for the adaptation of max-pooling processes, which usually include choosing the maximum number of spikes inside the pooling areas.

Following convolution, a neuron with the most activity within a square neighborhood of $lp \times lp$ presynaptic neurons is chosen by the pooling layer (max pooling). Additionally, parameter lp acts as the pooling layer's stride value. The spike rate of each neuron in a feature map may be used to describe its activity. Consequently, a feature map measuring $(r - p + 1) \times (c - p + 1)$ is converted to a smaller feature map measuring$(r - p + 1)/lp \times (c - p + 1)/lp$.
Robustness against local translation and scale changes is aided by the max pooling layer. In Fig. 2, the pooling layer appears as the second layer of the network. Nonadaptive relationships exist between the pooling maps and convolutional maps.

3) Fully connected layer: Fully connected layer classifies the image into normal or CVD. The spike trains that are released by the pooling layer divide the image's many visual elements among the D feature maps (D might have values of 16, 32, or 64). The output units receive spike trains emanating from the pooled feature maps at the third layer of the network, known as the fully connected (H) layer. The information's dimension is decreased in the feature maps by this layer.

For a neuron in the H layer to fire at a certain time step, it has to fulfill two requirements. The initial one is that it crosses its conventional LIF threshold, θ^h. Among the other neurons in the H layer, it does well in a non-exclusive winners-take-all (WTA) competition. As a result, we refer to the units in this layer as WTA-threshold LIF neurons. A WTA score is assigned to each unit in the H layer based on how its net input compares to the inputs of the other units in the layer. The synaptic weights,W, and the presynaptic spike vector, y_t, together yield the net input at time t.

$$WTAscore_h(W, y_t) = \frac{e^{W_h^T . y_t}}{\Sigma_{j=1}^{H} e^{W_j^T . y_t}} \quad (12)$$

IV. Secure Management Using Block Chain

Deep learning and blockchain are two extremely innovative technologies that are changing the standard operating procedures in the medical field. Using smart contracts, its solution improves transparency, safety, dependability, and data transmission capabilities. Among blockchain's many notable benefits are the ability to create smart contracts. This enables users to manage data access according to predetermined standards and agreements.

To guarantee that the data is private and secure, the classified lung picture in the proposed work is encrypted. By using encryption, the pictures are protected from being viewed by unauthorized users who do not have the necessary decryption keys to view them on the blockchain. Even if someone acquires unauthorized access to the blockchain, encryption makes sure they are unable to view the image without the correct decryption key.

After that, the encrypted lung image is hashed to provide a distinct digital fingerprint. The process of hashing transforms the image data into a fixed-length string of characters that while uniquely expressing the image, is unretrievable through reverse engineering. This hash serves as the lung image's special identification.

On the blockchain, a block contains the hash of the lung picture and associated metadata. Afterwards, a network of nodes receives this block, which is added to a chain of earlier blocks. Blocks cannot be removed or changed after they are put to the blockchain, making them immutable. By doing this, the classification is permanently preserved and the lung image data is kept impervious to manipulation.

The original image can only be viewed by authorized individuals who possess the relevant decryption key. By automatically providing or refusing access in accordance with pre-established guidelines, smart contracts enforce access policies. Algorithm 1 is the pseudocode for the proposed work.

Algorithm 1: Proposed Work

1: Load Dataset

2: Resize the image

3: Declare CLAHE

4: Threshold for contrast limiting

5: Initialize the number of honeybadger

6: Sort the fitness vale of honeybadger

7: Find optimal value

8: Calculate Density

9: Calculate intensity

10: Find digging phase

11: Sort fitness value

12: Update global optimization solution

13: Train the model

14: Test the model

15: Predict normal or abnormal

16: Integrate Blockchain for Data Integrity

V. Experimental Evaluation

A. Dataset Description

This part implements the LIDC-IDRI dataset to identify instances of lung cancer. The database has 848 nodules that have been enhanced in 17 different methods, 442 of which are benign and 406 of which are malignant. The dataset from LIDC-IDRI contains lesion annotations from four thoracic radiologists with expertise. Included in LIDC-IDRI is 1018 low-dose lung CT scans from 1010 lung patients.

A popular and effective medical imaging technique is a chest X-ray. In certain cases, the clinical examination of CXR pictures might be even more challenging than the study of a

chest CT scan. The 112,120 CX-ray pictures in the NIH Chest X-ray Dataset are associated with individual diseases.

B. Evaluation Metrics

The following statistical measures, including accuracy, precision, F1 score and recall are used to evaluate the efficacy of the proposed lung nodule classification technique. The terms $TruPos$, $TruNeg$, $FalPos$ and $FalNeg$ are represents the True-Positive, True-Negative, False-Positive and False-Negative, respectively. The following equations can be utilized to compute the metrics.

$$Accuracy = \frac{(TruPos + TruNeg)}{(TruPos + FalPos + TruNeg + FalNeg)}$$

$$Precision = \frac{TruPos}{(TruPos + FalPos)}$$

$$Recall = \frac{TruPos}{(TruPos + FalNeg)}$$

$$F1Score = 2 \times \frac{(Recall \times Precision)}{(Recall + Precision}$$

VI. Result

The suggested approach for segmenting and categorizing lung nodules is presented here with the experimental results. The recommended approach was put into practice using a MATLAB environment. Accuracy, F1 score, precision, and recall metrics were used to assess the classification performance. Two sets of lung nodule pictures, such as CT and Chest X-ray (CXR) images, were used for the experiments. Table II shows the classification results of the proposed SCNN model on LIDC-IDRI dataset.

TABLE II. Classification Results of the Proposed SCNN Model on LIDC-IDRI Dataset

Performance Measures	Proposed SCNN
Accuracy	98.64
Precision	92.42
Recall	93.76
F1-score	93.08

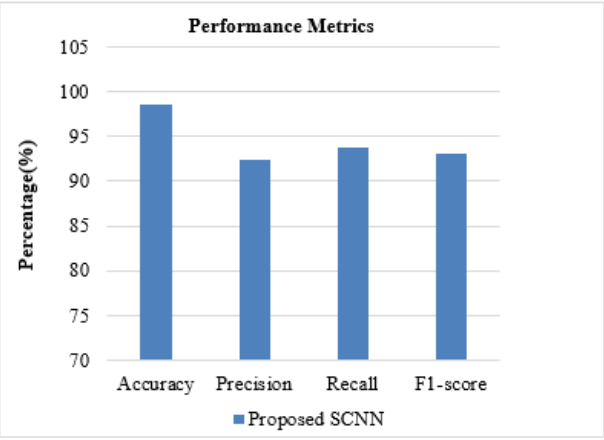


Fig. 4. Performance of the proposed SCNN model on LIDC-IDRI dataset.

The proposed SCNN model achieves 98.64% of accuracy, 92.42% of F1score, 93.76% of precision and 93.08% of recall on LIDC-IDRI images. The effectiveness of the suggested model is visually represented in Fig. 4.

Table III displays the classification results of the proposed SCNN model on NH Chest X-Ray images. The suggested model achieves 98.9% accuracy, 97.3% precision, and 94.44% recall and 95.84% of F1score on NH Chest X-Ray images.

TABLE III. Classification Results of the Proposed SCNN Model on NH Chest X-Ray Dataset

Performance Measures	Proposed SCNN
Accuracy	98.9
Precision	97.3
Recall	94.44
F1-score	95.84

The suggested SCNN classifier's classification performance is illustrated in Fig. 5. It demonstrates that the suggested model produces higher predictions of accuracy, precision, recall, and F1-score across the board. The effectiveness of the suggested model is visually represented in Fig. 5.

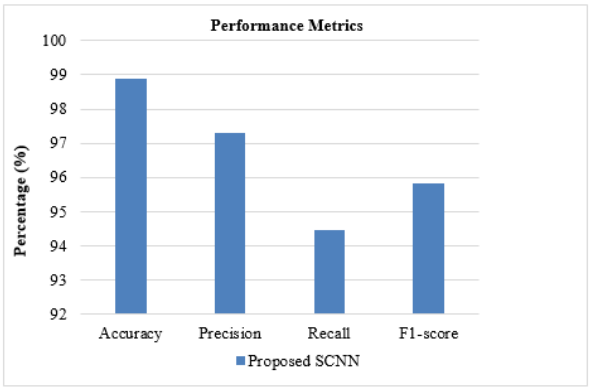


Fig. 5. Performance of the proposed SCNN model on NH Chest X-Ray dataset.

A. Comparative Analysis

A comparison between the suggested model and traditional neural networks is also done in this section. The comparative analysis of Lung cancer detection in terms of accuracy of the proposed Spiking CNN model with existing methods on CT dataset is displayed in Table IV. It is clearly detected that the proposed Spiking CNN classifier gives superior results than the other previous research work based on accuracy. It gives 98.64% of accuracy which is +6.19% than ASAIL CNN approach, +4.53% than ProCAN approach, +8.94% than Improved FasterR-CNN, +2.83% than Inception V3, +0.41% than Deep residual network, +8.74% than ResNet method, +14.5% than CNN model. The results of this technique's performance comparison with standard methods indicate that the proposed method outperforms them.

The intended model yielded better performance than the earlier networks. The expected outcomes seem to be rather reliable in distinguishing between typical and anomalous instances. The result shows the effectiveness of the proposed

model based on blockchain dynamic pattern techniques applied in deep Convolutional Neural Network.

TABLE IV. COMPARATIVE ANALYSIS OF SCNN WITH OTHER CLASSIFIERS ON CT IMAGES

Author/Year	Methods	Accuracy (%)
Song et al./2017 [20]	CNN	84.14
Nibali et al./ 2017 [21]	ResNet	89.90
Wu et al./ 2020 [22]	Deep residual network	98.23
Wu et al/ 2020 [22]	Inception V3	95.81
Shiwei et al. /2021 [23]	Improved FasterR-CNN	89.7
Al-Shabi et al/2022 [24]	ProCAN	94.11
Parveen Banu et al/ 2022[25]	ASAIL CNN	92.45
Proposed Method	SCNN	98.64

Fig. 6 presents a visual representation of the comparison findings between the proposed technique and the existing methods on CT images.

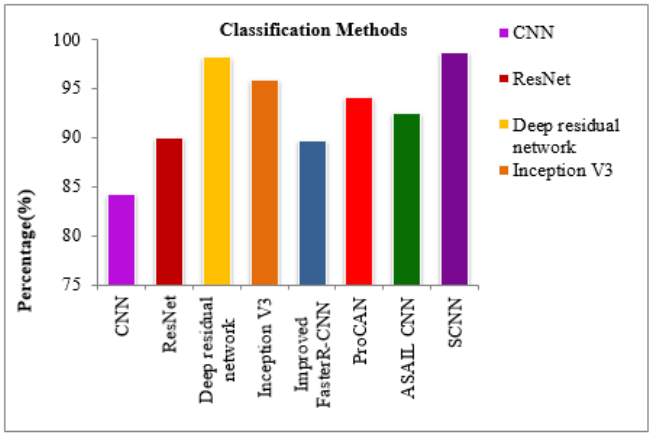


Fig. 6. Comparative analysis of the proposed method and existing methods on CT images.

The comparative analysis of Lung cancer detection in terms of accuracy of the proposed Spiking CNN model with existing methods on Chest X-Ray dataset is displayed in Table V. It is clearly detected that the proposed SCNN classifier gives superior results than the other previous research work based on accuracy. It gives 98.9% of accuracy which is +0.2% than OCNN-SVM approach, +8.8% than Gabor-LBP+MRCNN approach, +25.9% than VDSNet model. The results of this technique's performance comparison with standard methods indicate that the proposed method outperforms them. The intended model yielded better performance than the earlier networks. The expected outcomes seem to be rather reliable in distinguishing between typical and anomalous instances.

Fig. 7 presents a visual representation of the comparison findings between the proposed technique and the existing methods on Chest X-Ray images.

TABLE V. COMPARATIVE ANALYSIS OF SCNN WITH OTHER CLASSIFIERS ON CHEST X-RAY IMAGES

Author/Year	Methods	Accuracy (%)
Bharati et al./2020 [26]	VDSNet	73
Wang et al./2018 [27]	Gabor-LBP+MRCNN	90.1
Sreeprada et al/2023 [28]	OCNN-SVM	98.7
Proposed Method	SCNN	98.9

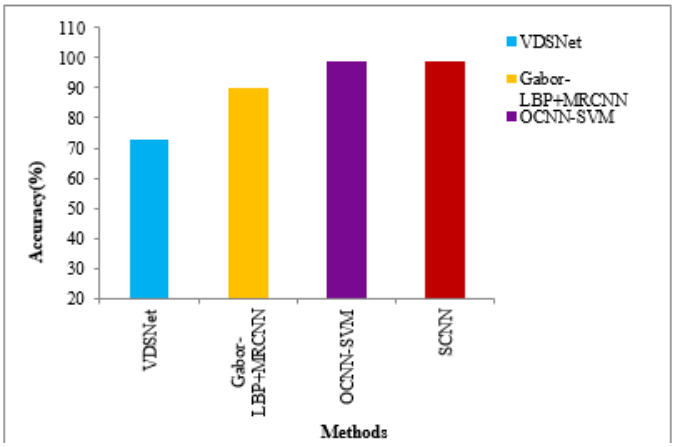


Fig. 7. Comparative analysis of the proposed method and existing methods on Chest X-Ray images.

The pre-processing, segmentation, and classification outcomes of the suggested work are shown in Fig. 8. The original CTX and CT pictures are displayed in the first column. The pre-processing results using CLAHE are shown in the second column. The segmentation result using the Honey Badger method is displayed in the third column. The categorization result is shown by the value in the final column. With the CT image dataset, the suggested operator achieves 98.64% classification accuracy, and with the CXT dataset, it achieves 98.9% accuracy.

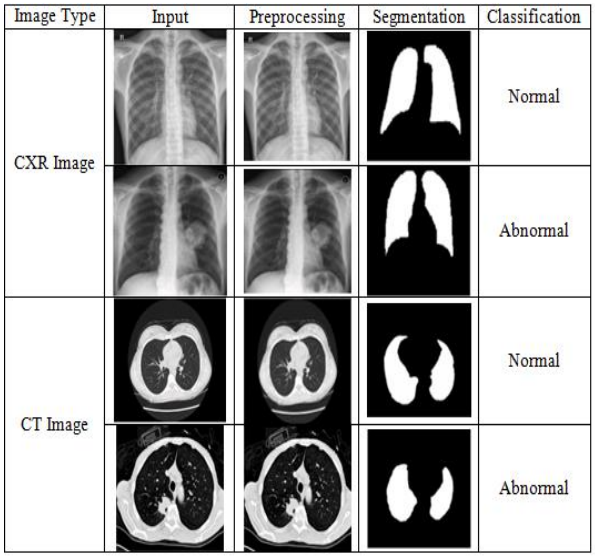


Fig. 8. The segmentation, feature extraction and classification results of proposed work.

VII. DISCUSSION

A very efficient lung image analysis pipeline is a result of the synergy between CLAHE, HBA, and SCNNs. The utilization of contrast enhancement in CLAHE proven to be crucial in emphasizing nuanced traits that are necessary for precise segmentation and classification. Because of its dynamic nature, the Honey Badger Algorithm ensured accurate segmentation even in difficult situations, minimizing errors that could have spread to the classification stage. The excellent accuracy and computational efficiency that SCNNs brought made them a good choice for use in healthcare settings where prompt and dependable decision-making is crucial. The outcomes demonstrate that this integrated strategy, which offers a well-balanced mix of accuracy, efficiency, and interpretability, is well-suited to meet the difficulties associated with lung image processing.

VIII. CONCLUSION

This study proposed a novel neural network model (SCNN) for lung nodule classification, which combines the theories of deep learning, blockchain, and dynamic pattern features to address the issues with lung nodule classification, including a difficult classification detection process and low classification accuracy. The CT scan and CXR image databases from the NIH Chest X-ray and LIDC-IDRI were used. Initially, these images are pre-processed by Contrast Limited Adaptive Histogram Equalization (CLAHE) to enhance the image clarity and reducing the noise. Then the Honey Badger optimization Algorithm (HBA) is used to segment the lung region from the pre-processed image. Morphological segments of the lung region are used to generate dynamic patterns. Finally, these patterns are aggregated into the deep neural Spiking Convolutional Neural Network (SCNN) is the global model for classifying the images into normal and abnormal cases. Based on the LIDC-IDRI and NH Chest X-Ray, the SCNN model achieves 98.64% and 98.9% of accuracy respectively. This methodology provides a comprehensive solution that tackles the particular issues of lung image analysis by combining strong preprocessing, sophisticated segmentation, and effective classification. In the end, the suggested method improves patient outcomes in the field of lung health by laying the groundwork for more precise, understandable, and resource-efficient diagnostic instruments.

REFERENCES

[1] B.U. Dhaware, A.C. Pise, "Lung Cancer Detection Using Bayasein Classifier and FCM Segmentation." IEEE, International Conference on Automatic Control and Dynamic Optimization Techniques (ICACDOT), pp. 170–174, 2016.

[2] American Cancer Society. Cancer Facts and Figures 2022. Atlanta; American Cancer Society: 2022.

[3] K.D. Miller, L. Nogueira, A.B. Mariotto, J.H. Rowland, K.R. Yabroff, C.M. Alfano, R. L. Siegel, "Cancer treatment and survivorship statistics," 2019. CA: A Cancer Journal for Clinicians, 2019. doi:10.3322/caac.21565

[4] S.K. Thakur, D.P. Singh, J. Choudhary, "Lung cancer identification: a review on detection and classification," Cancer Metastasis Rev. vol. 39, pp. 989–998, 2020. https:// doi.org/10.1007/s10555-020-09901-x.

[5] J. Liang, G. Ye, J. Guo, Q. Huang, and S. Zhang, "Reducing False-Positives in Lung Nodules Detection Using Balanced Datasets." Frontiers in public health, vol. 9, pp. 671070, 2021. https://doi.org/10.3389/fpubh.2021.671070

[6] A. Asuntha, A. Srinivasan, "Deep learning for lung Cancer detection and classification." Multimed Tools Appl, vol. 79, pp. 7731–7762, 2020. doi:10.1007/s11042-019-08394-3

[7] M. Vas, and A. Dessai, "Lung cancer detection system using lung CT image processing. 2017 International Conference on Computing, Communication," Control and Automation (ICCUBEA), 2017. doi:10.1109/iccubea.2017.8463851

[8] Y. Li, Q. Wu, H. Sun, and X. Wang, "Research on Lung Nodule Detection Based on Improved Target Detection Network." Complexity, pp. 1–7, 2020. https://doi.org/10.1155/2020/6633242

[9] A. Elnakib, M. Amer, and E.Z. Abou-Chadi, "Early Lung Cancer Detection using Deep Learning Optimization." International Journal of Online and Biomedical Engineering (iJOE), vol. 16, no. 06, pp. 82, 2020. doi:10.3991/ijoe.v16i06.13657.

[10] A. Srinivasulu, K. Ramanjaneyulu, R. Neelaveni, S.R. Karanam, S. Majji, M. Jothilingam, and T.R. Patnala, "Advanced lung cancer prediction based on blockchain material using extended CNN." Applied Nanoscience, 2021. doi:10.1007/s13204-021-01897-2

[11] D. Vaishnavi. K. Arya. Devi Abirami. M.N. Kavitha, "Lung Ancer Detection using Machine Learning", International Journal of Engineering Research & Technology (IJERT), 2019.

[12] Y. Su, D. Li, X. Chen, "Lung nodule detection based on faster R-CNN framework," Comput. Methods Progr. Biomed. vol. 200, pp. 105866, 2021. https://doi.org/10.1016/j.cmpb.2020.105866.

[13] T. Vaiyapuri, Liyakathunisa, H. Alaskar, R., Parvathi, V., Pattabiraman, A. Hussain, "CAT Swarm Optimization-Based ComputerAided Diagnosis Model for Lung Cancer Classification in Computed Tomography Images." Appl. Sci., vol. 12, pp. 5491, 2022.

[14] Rajasekar, Vani and M.P. Vaishnnave, and Sivakumar, Premkumar and Sarveshwaran, Velliangiri and V. Rangaraaj, "Lung cancer disease prediction with CT scan and histopathological images feature analysis using deep learning techniques." Results in Engineering, vol. 18, 2023. 101111. 10.1016/j.rineng.2023.101111.

[15] I.W. Harsono, S. Liawatimena, T.W. Cenggoro," Lung Nodule Detection and Classification from Thorax CT-scan Using RetinaNet with Transfer Learning," Journal of King Saud University-Computer and Information Sciences, 2020, https://doi.org/10.1016/j.jksuci.2020.03.013.

[16] M. Schultheiss, P. Schmette, J. Bodden, et al. "Lung nodule detection in chest X-rays using synthetic ground-truth data comparing CNN-based diagnosis to human performance." Sci Rep, vol. 11, pp. 15857, 2021. https://doi.org/10.1038/s41598-021-94750-z

[17] W. Shen, M. Zhou, F. Yang, Dongdong Yu, Di Dong, Caiyun Yang, Yali Zang, Jie Tian, "Multi-crop Convolutional Neural Networks for lung nodule malignancy suspiciousness classification," Pattern Recognition, vol. 61, pp. 663-673, 2017.doi: 10.1016/j.patcog.2016.05.029.

[18] S. Lal, and M. Chandra, "Efficient algorithm for contrast enhancement of natural images." The International Arab Journal of Information Technology, vol. 11, no. 1, pp. 95–102, 2014.

[19] R. Kumar Rai, P. Gour, and B. Singh, "Underwater image segmentation using CLAHE enhancement and thresholding." International Journal of Emerging Technology and Advanced Engineering, vol. 2, no. 1, pp. 118–123, 2012.

[20] Q. Song, L. Zhao, X. Luo, and X. Dou, "Using deep learning for classification of lung nodules on computed tomography images," Journal of Healthcare Engineering, pp. 7, August 2017.

[21] A. Nibali, Z. He, and D. Wollersheim, "Pulmonary nodule classification with deep residual networks," International Journal of Computer Assisted Radiology and Surgery, vol. 12, pp. 1799–1808, 2017.

[22] P. Wu, X. Sun, Z. Zhao, H. Wang, S. Pan, B. Schuller, "Classification of Lung Nodules Based on Deep Residual Networks and Migration Learning." Comput Intell Neurosci. 2020 Mar 30, pp. 8975078. doi: 10.1155/2020/8975078

[23] L.I. Shiwei, D. Liu, "Automated classification of solitary pulmonary nodules using convolutional neural network based on transfer learning strategy." J. Mech. Med. Biol. Vol. 21, pp. 2140002, 2021.

[24] M. Al-Shabi, K. Shak, M. Tan, “ProCAN: Progressive growing channel attentive non-local network for lung nodule classification.” Pattern Recognit. vol. 122, pp. 108309, 2022.

[25] S. Parveen Banu, M. Syed Mohamed, "Asial CNN: Assorted Scale Integrated Alternate Link Model Convolutional Neural Network for Lung Nodule Detection," *International Journal of Engineering Trends and Technology*, vol. 70, no. 11, pp. 353-363, 2022. Crossref, doi:10.14445/22315381/IJETT-V70I11P237

[26] S. Bharati, P. Podder, and M.R.H. Mondal, “Hybrid deep learning for detecting lung diseases from X-ray images. Informatics in Medicine Unlocked, vol. 20, pp. 100391, 2020. doi:10.1016/j.imu.2020.100391

[27] Q. Wang, Y. Zheng, G. Yang, W. Jin, X. Chen, and Y. Yin, “Multiscale Rotation-Invariant Convolutional Neural Networks for Lung Texture Classification.” IEEE Journal of Biomedical and Health Informatics, vol. 22, no. 1, pp. 184–195, 2018. doi:10.1109/jbhi.2017.2685586

[28] V. Sreeprada, K. Vedavathi, “Lung Cancer Detection from X-Ray Images using Hybrid Deep Learning Technique,” Procedia Computer Science, vol. 230, pp. 467-474, 2023. https://doi.org/10.1016/j.procs.2023.12.102.

Application of Improved CSA Algorithm-Based Fuzzy Logic in Computer Network Control Systems

Jianxi Yu
School of Engineering & Economics, Henan Institute of Economics and Trade, Zhengzhou, 450046, Henan, China

***Abstract*—In the past few years, with the high-speed popularization of computers and the widespread use of smart phones and mobile devices, the Internet has gradually become an indispensable part of people's daily lives. The Internet is constantly driving the process of digital society and providing people with more convenient and innovative applications. However, the internet industry also faces challenges such as runtime ambiguity, instability, large data volume, and difficulties in network situational awareness. In response to the above issues, this study combines the standard cuckoo algorithm with a fuzzy neural network to design a computer network situational awareness system. It uses principal component analysis to deduct the dimensions of the original data and then adds Gaussian noise to introduce appropriate randomness. The test proved that the improved model had a significant optimization effect on real network data, with an improvement of about 81.2% compared to the standard cuckoo algorithm. In the 220th iteration of the test set, the Loss function value was 0, which could accurately predict the network situation, with an accuracy rate of 98%. The designed system identification has higher recognition accuracy and less time consumption and has certain application potential in computer networks.**

***Keywords*—CSA; computer network; fuzzy logic; principal component analysis method; network operation; situation awareness**

I. INTRODUCTION

Since the beginning of the 21st century, information technology has rapidly advanced, and the Internet has played an increasingly important role in social production and life. The average monthly traffic volume of the population has been increasing year by year [1]. The application of this technology is becoming increasingly diversified, giving rise to new industries such as 5G technology, the Internet of Things, edge computing, cloud computing, and data center optimization. This is constantly changing people's way of life and work, but it also brings new challenges and opportunities. It is widely acknowledged that the network runtime state exhibits several characteristics that present significant challenges to the development of the internet industry. These include ambiguity, instability, a considerable volume of data, and a lack of situational awareness within the network. In response to the above issues, experts and scholars in the field of the Internet have applied the cuckoo algorithm to computer network control systems. This algorithm is a heuristic optimization algorithm inspired by the reproductive behavior of cuckoo birds [2]. The algorithm realizes the optimization of the computer network control system by initializing the cuckoo group, generating new solutions, evaluating and selecting, updating and iterating, judging the convergence conditions, and analyzing the results [3]. However, the related research precision is not high, the generalization ability is low, the training time is long, and the rate of convergence is slow. In this study, the Principal Component Analysis (PCA) is first used to reduce the dimensions of the huge and changeable network data, and then Gaussian noise is introduced to lift the rate of convergence of the algorithm. Based on the standard Cuckoo Search Algorithm (CSA), a computer Network Situational Awareness Model (NSAM) is designed by integrating a Fuzzy Neural Network (FNN). The paper mainly consists of five sections. Section II summarizes the research status of scholars in the industry on the difficulties of Internet situational awareness. Section III establishes a computer NSAM that integrates CSA and fuzzy logic. Section IV conducts comparative experiments and efficiency verification on the optimization effect of the model. Section V is a summary of the research and an explanation of the direction for improvement.

II. RELATED WORKS

As a result of the growing use of mobile internet in a range of sectors, predictive models with enhanced network situational awareness capabilities are increasingly attracting interest from businesses and researchers. Liu C et al. introduced cloud control middleware to manage service requests to meet constraints, aiming at the problem that traditional cloud computing mode makes it difficult to provide real-time computing resources. They developed a conceptual computing framework built on cloud and mist combination, which has better performance in energy consumption and response time [4]. Abed Algoni B H et al. used a special type of opposition-based learning ECS model to address the problem of CSA being prone to suboptimal situations. The experiment showed that ECS exhibited better performance than all tested variants [5]. Cheng P et al. proposed Particle Swarm Optimization (PSO) - CSA to predict local comfort and global comfort by artificially solving the problems of motion state and being unable to be directly used for model analysis. This model had a high prediction accuracy [6]. Eltamally A M et al. proposed PSO and CSA to capture global peaks in the P curve of Photovoltaic (PV) arrays, which have advantages in optimizing control parameters [7]. Fan J et al. designed a chaotic CSA image segmentation model to address the issue of difficulty in improving accuracy in noisy images. This model improved accuracy and reduced uncertainty [8]. Li J et al. designed a balanced learning differential CS extension algorithm to solve the problem of CSA easily falling into local optima, which lifted the algorithm's global search capacity and accuracy [9].

CSA has unique advantages in group optimization problems and provides a certain reference for Internet network situational

awareness. Chen S Y et al. designed a variable order fuzzy fractional proportional integral differential control system to address the issue of the inability of integral differential (PID) controllers to achieve high-precision control. This system could achieve better control response and anti-interference characteristics than Integer Order (IO) controllers [10]. Muhammad K et al. designed a television camera monitoring system based on fuzzy logic to continuously monitor the phenomenon of a large amount of data generated by television cameras every day. This model could handle data uncertainty in the real-world domain [11]. To solve the problem of autonomous decision-making of mobile robots to overcome obstacles, Ben Jabeur C et al. established a decision model based on an intelligent PID optimization neural network and fuzzy logic controller. The mobile robot applying this model could quickly execute tasks and adapt to constantly changing environmental conditions [12]. Costa R et al. designed a mountain flood prediction model based on classification and regression trees, deep learning neural networks, and fuzzy logic to identify slopes with a high probability of mountain flood outbreaks. The prediction accuracy exceeded 84% [13]. In response to the issue of insufficient drone controllers to cope with weather disturbances, Ulus Ş designed a drone control model by integrating classic PID and fuzzy logic controllers. This controller had better performance than other controllers [14]. Katsikis V N et al. designed a multi-objective evolutionary network framework to address the lack of flexibility in fuzzy logic neural networks, which has advantages in effectiveness and interpretability [15].

In summary, the application of fuzzy logic and CSA in computer network control systems has sufficient theoretical and practical foundations, but relevant research rarely combines the two to solve the problem of network situational awareness difficulties. Therefore, this study improves CSA and combines fuzzy logic to design NSAM to promote further development of the internet industry.

III. Overall Plan Design for Network Operational Situation Awareness

This chapter is mainly divided into five sections. The first section is divided into establishing an indicator model and NSAM, while the second section improves the CSA and integrates fuzzy logic. In the third section, the computer network control system based on fuzzy logic is established, and PCA is taken to decrease the dimension of the original data.

A. Establishment of NOSA Model

Network Operational Situation Awareness **(**NOSA) refers to the real-time monitoring, analysis, and identification of various activities, events, and resources in the network to obtain a comprehensive understanding of the network's operational status [16]. NOSA can help organizations or network administrators detect abnormal activities, attack attempts, or system failures promptly, and take corresponding measures to protect the security and stability of the network. Hence, this system needs to include the whole links from data to situation analysis to users. By the above requirements, this manuscript proposes the NOSA system, as exhibited in Fig. 1.

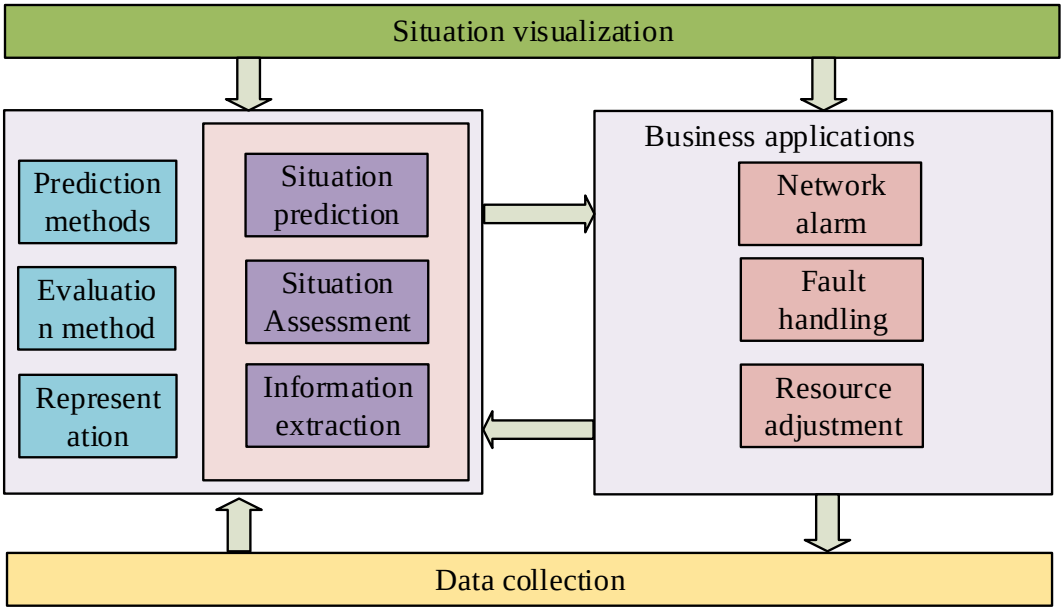


Fig. 1. Schematic diagram of network operation situational awareness system.

Fig. 1 shows a system that includes four modules: data, data fusion, situation visualization, and business application. Specifically, the results of the data fusion module can direct the management operations of the business application module. After the business application module manages the network, it will transmit new analysis results to the data fusion module through it. This cyclic process enables the system to continuously optimize and improve to better meet user needs. With the guidance of the above NOSA system model, a network operation situation indicator system can be constructed. This research presents a network operation situation indicator system, which is constructed from the perspectives of network performance and network traffic. The system is designed to integrate the TCP/IP five-layer model, as illustrated in Fig. 2.

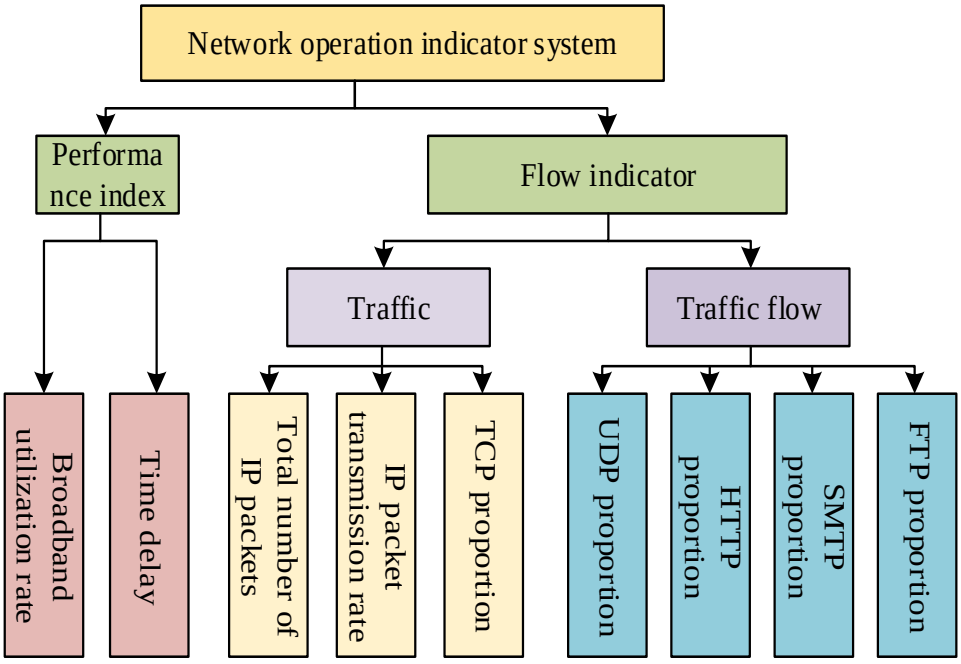


Fig. 2. Network operation situation indicator system.

Fig. 2 shows the network traffic indicators, including IP throughput, link utilization, the size and proportion of traffic based on network protocols, and the size and proportion of traffic. The network protocol traffic mainly covers protocols such as TCP, UDP, ICMP, etc; Business traffic mainly includes protocols i.e. HTTP, FTP, and SMTP. The goal of this study is to reflect the operational situation of the network from a business perspective. Therefore, it is necessary to integrate multiple protocol layer indicators [17]. The transport layer protocol traffic data, namely TCP, UDP and ICMP, are added to the network traffic indicator. After establishing the network operation situation indicators, it is also essential to sort out the original data according to the relevant calculation formulas. Broadband utilization is an important performance indicator in a network, which represents the current load level and Resource Utilization Efficiency (RUE) of the network. This characteristic indicator is calculated by Eq. (1).

$$L = \frac{T}{B} \tag{1}$$

In Eq. (1), T is the average transmission rate of the network, while B represents the maximum transmission rate of data packets in the network. In addition, this indicator system is one-way delay. Delay is used to represent the network Transmission delay. The calculation formula is Eq. (2).

$$Delay = \frac{\sum d}{sum} \tag{2}$$

In Eq. (2), d means the delay of all data packets transmitted by the network, and sum is the sum of the number of transmitted data packets. The quantity of data packets transmitted by the network per unit time is called the IP packet transmission rate, and its calculation formula is Eq.

$$(3). V_{IP} = \frac{sum}{t} \tag{3}$$

In Eq. (3), sum represents the total number of IP packets transmitted by the network, while t represents the total transmission time. In this indicator system, the proportion of protocol traffic between the application layer and the network layer is introduced. This type of indicator is represented by acc , and its calculation formula is Eq. (4).

$$acc = \frac{sum_protocol}{sum} \tag{4}$$

In Eq. (4), $sum_protocol$ represents the gross of protocol packets per unit time, and sum represents the total number of IP packets during that period.

B. Fuzzy Logic Model based on CSA Algorithm

CSA heuristic swarm intelligence optimization algorithm simulates the process of cuckoo's foraging and nest protection to achieve global optimization. The basic idea of the algorithm is to represent the candidate solutions of the problem as nests of cuckoo birds, with each nest corresponding to a solution vector. Then, the quality of each nest is evaluated based on the fitness function of the problem. CSA has good global search ability and high parallelism. It is suitable for various optimization problems, especially continuous optimization problems. The performance of the algorithm is still affected by factors such as parameter settings, nest protection strategy, and nest elimination strategy, and needs to be adjusted and optimized according to specific problems. Fig. 3 shows the algorithm flowchart.

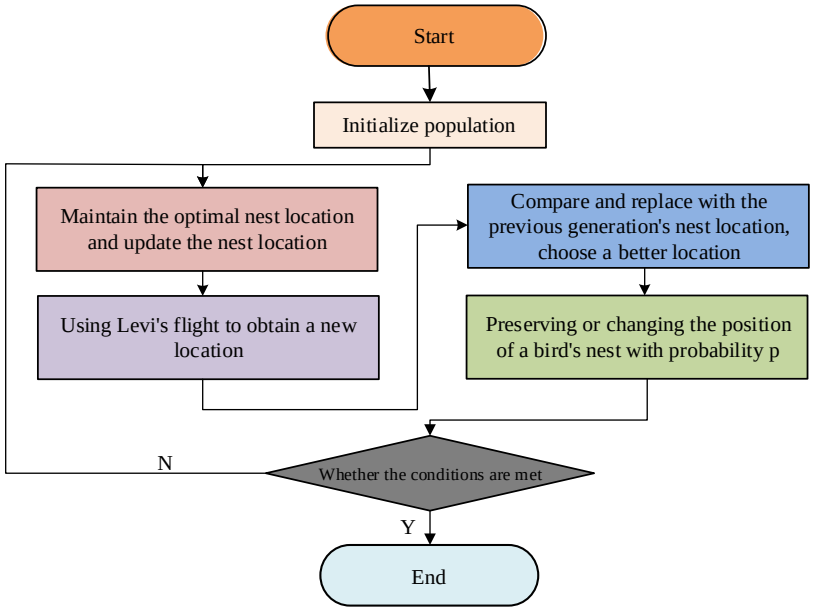


Fig. 3. Flowchart of cuckoo algorithm.

Fig. 3 is the CSA flowchart. Firstly, a set of feasible solutions is randomly produced as the initial population, and the optimal nest position is retained by adjusting the positions of the parents and parasitic birds. This step can use some heuristic methods, such as random walk or local search algorithms. Then, the position is updated by adjusting the position of the parent and parasitic birds. The next step is to eliminate solutions with lower fitness with a certain probability. Finally, to determine whether the termination condition is met, i.e. reaching the max-iterations or finding a solution that meets the requirements. Step 2 is repeated to 5 until the termination conditions are met. Although CSA has certain advantages and application value, there are also some shortcomings. For example, the rate of convergence is slow, the parameter selection is difficult, and the dependence on problem characteristics is strong. To solve the above problems, the idea of fuzzy logic is used to improve traditional CSA. The schematic diagram is shown in Fig. 4.

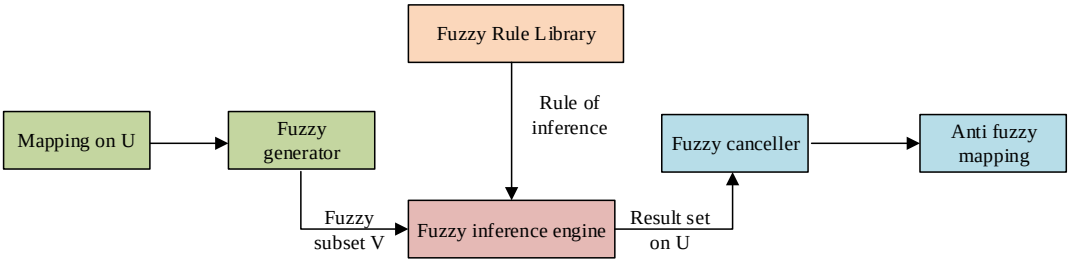


Fig. 4. Fuzzy mapping diagram.

Fig. 4 is a fuzzy inference system, which mainly consists of four parts. The first part is a fuzzy generator, mainly responsible for converting the precise input values into fuzzy membership values, that is, mapping the input to the corresponding membership functions. The second part is the fuzzy rule library, which defines a set of fuzzy rules, each containing a series of prerequisites and a conclusion. The preconditions and conclusions are described by fuzzy sets. The third part is the inference engine, which uses the inference mechanism to calculate the fuzzy output according to the given rules and the fuzzy input. The final part is to convert the fuzzy output back to precise values through anti-fuzzification. Unlike traditional binary logic, which only has true and false values, fuzzy logic allows variables to have a continuous range of values, with fuzziness between 0 and 1. Assuming the input variable is $x=[x_1,x_2,...,x_n]^T$, each component is a fuzzy variable. Each fuzzy variable is segmented into n fuzzy sets, and the fuzzy set of each component in the input variable is Eq. (5).

$$T(x_i)=\{A_i^1,A_i^2,...,A_i^k\},k=1,2,3,...,n \tag{5}$$

In Eq. (5), A_i^k is the k-th variable value of the i-th input component. The fuzzy outputs vector $y=[y_1,y_2,...,y_n]^T$ in this fuzzy model, if x_1 is A_i^k, the output fuzzy output vector is equation (6).

$$y_{it}=p_{0r}^i+p_{1r}^i x_1+\cdots+p_{nr}^i x_n \quad 0<r<k^n \tag{6}$$

In Eq. (6), p_{0r}^i represents the output of the i-th output vector under rule r. This fuzzy rule can be represented as an IF THEN statement, as shown in Eq. (7).

IF x_1 A_i^k , THNE

$$\begin{cases} y_{1r} = p_{0r}^1 + p_{1r}^1 x_1 + \cdots + p_{nr}^1 x_n \\ \cdots \\ y_{ir} = p_{0r}^i + p_{1r}^i x_1 + \cdots + p_{nr}^i x_n \end{cases} \quad (7)$$

In Eq. (7), x_1 represents the k -th variable value of the i -th input component. By using single point fuzzification to represent input variables, the applicability of each rule can be calculated, as shown in Eq. (8).

$$T_r = \mu_{1k} \wedge \mu_{2k} \cdots \wedge \mu_{nk} = \mu_{1k} \mu_{2k} \cdots \mu_{nk} \quad (8)$$

In Eq. (8), T_r represents the applicability of rule r . The output value of the fuzzy system is the weighted average of each rule, and its formula is Eq. (9).

$$y_i = \sum_{q=1}^{q=r} T_q y_{iq} / \sum_{q=1}^{q=r} T_q \quad (9)$$

In Eq. (9), y_i is the i -th component of the output vector. Fuzzy logic is widely used in control systems, artificial intelligence, decision support systems, and other fields, especially suitable for problems with fuzziness and uncertainty. Through the reasoning and processing of fuzzy logic, incomplete information and fuzzy concepts in the real world can be better handled, improving the effectiveness of decision-making and control.

C. Computer Network Control System Based on Fuzzy Logic

The large amount of data in computer network control systems has uncertainty and fuzziness, so it is necessary to apply fuzzy theory to solve these problems. In addition, network data also have the characteristics of being massive and multidimensional. Neural network is an effective method to deal with big data. When it is combined with fuzzy theory, it can solve NOSA problem in complex data environment. Therefore, an improved CAS algorithm, CSA-FNN, is designed by combining CAS and FNN, and its structure is Fig. 5.

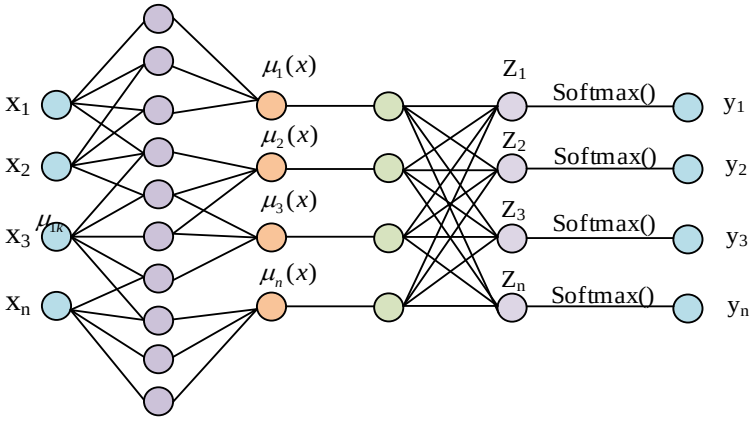


Fig. 5. CSA fuzzy neural network.

In Fig. 5, the first is the input layer, which completes the input of training data. The second is the rule mapping layer, which fuzzily divides each input component. The third is the rule fitness layer, where each node represents a fuzzy rule. The fourth is the normalization layer, which normalizes the fitness of each rule. The fifth is the anti-fuzzy layer, which completes the mapping from the fuzzy rule space to the output space through the Activation function. The last layer is the output layer, which outputs the network operational situation level. In the above CSA-FNN model, the network operational situation features include 10 indicators, that is, the input data is 10 dimensions. The number of nodes in each layer increases exponential type with the number of nodes in the input layer, resulting in too many nodes. Therefore, it is necessary to reduce the dimensions of the original data. This study uses PCA to reduce and reconstruct the dimensions of the original data. Fig. 6 is the process steps.

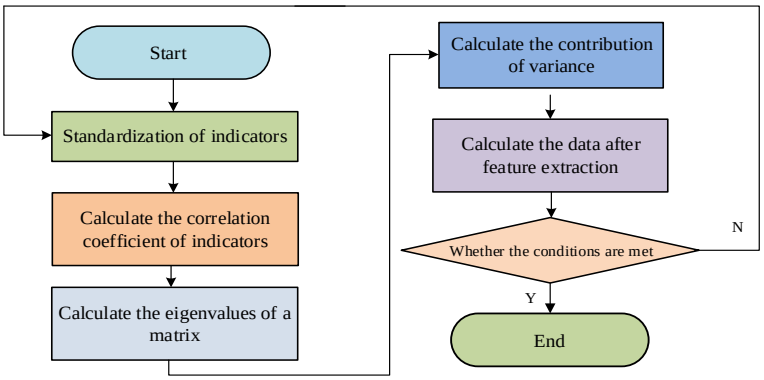


Fig. 6. CSA-FNN.

The first is to standardize the indicators. The dimensions of the original indicator data in computer networks are different, so it is necessary to normalize the indicator data and convert them into data under the same dimension. This article adopts the maximum-minimum normalization method, and its expression is shown in Eq. (10).

$$I=\begin{Bmatrix} I_{11} & I_{12} & \cdots & I_{1m} \\ I_{21} & I_{11} & \cdots & I_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ I_{n1} & I_{n2} & \cdots & I_{nm} \end{Bmatrix} = (I_1, I_2, ..., I_{\mathrm{m}}) \tag{10}$$

In Eq. (10), I represents the normalized network indicator data, and m represents that each data has m indicators. $\mathrm{I_m}$ represents the m-th dimensional vector of the data, normalized as Eq. (11).

$$i_{jp}{}' = \frac{i_{jp} - \overline{I_p}}{\sqrt{\sigma_p}} \tag{11}$$

In Eq. (11), $\overline{I_p}$ represents the average value of the p-th indicator. σ_p represents the mean square deviation of the p-th indicator. The correlation coefficient of each indicator in I is calculated to obtain the correlation coefficient matrix, as shown in Eq. (12).

$$A=\frac{1}{m}I^T I\begin{pmatrix} R_{11} & R_{12} & \cdots & R_{1m} \\ R_{21} & R_{11} & \cdots & R_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ R_{m1} & R_{m2} & \cdots & R_{mm} \end{pmatrix} \tag{12}$$

In Eq. (12), $R_{ij}(i, j=1,2,...,m)$ represents the correlation coefficient between the i-th and j-th indicators. The eigenvalue of matrix A and Orthogonalization is calculated, as shown in Eq. (13).

$$a=(a_1, a_2, ..., a_m) \tag{13}$$

In Eq. (13), a is the matrix obtained by Orthogonalization the characteristic matrix of matrix A. Then the contribution of each feature vector is calculated to the square difference, and the contribution calculation formula is Eq. (14).

$$a_i=\frac{\lambda_i}{\sum_{k=1}^{m}\lambda_k} \tag{14}$$

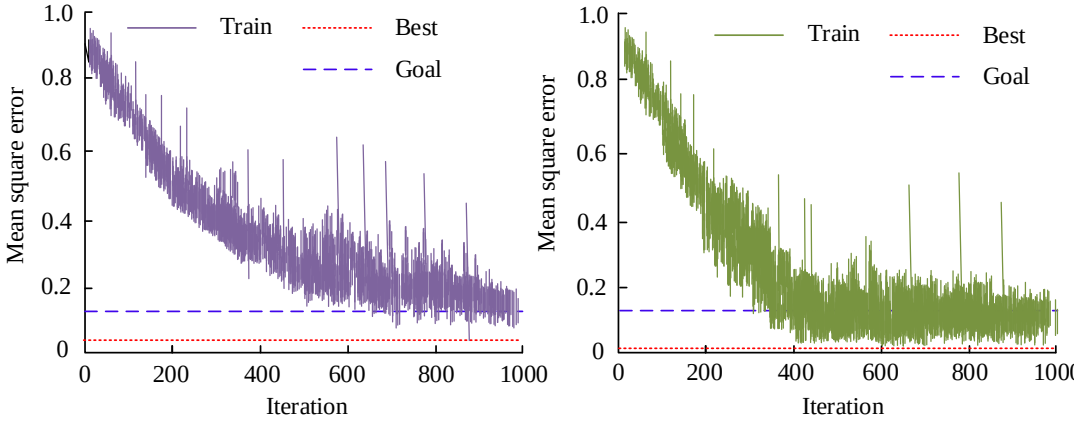
In Eq. (14), $a_i(i=1,2,...,m)$ is the contribution rate of the mean square deviation of the i-th eigenvector. Finally, the data after feature extraction based on the formula are calculated. After the above 5 steps of processing, the dimension of the data is reduced to 4 dimensions, and the number of nodes is significantly reduced, solving the problem of massive and multidimensional network data. Therefore, the improved CSA algorithm based on fuzzy logic has been successfully applied to computer network control systems.

IV. CSA-FNN Model Performance Testing

This chapter mainly verifies the optimization effect of the model. The first section mainly compares and analyzes CSA-FNN with other algorithms to verify the comparative advantages of the algorithm. Then, distinctive datasets are used to verify the generalization ability of the model. The second section mainly conducts simulation experiments to test the efficiency in practical environments.

A. Comparative Analysis of Algorithms and Validation of Generalization Ability

This study uses CSA combined with FNN to establish a computer network control system model, solving the problems caused by the uncertainty and fuzziness of computer network data. The original data dimension is reduced by PCA, which significantly reduces the number of nodes and complexity of the neural network [18]. To evaluate the optimization ability of fuzzy inference systems for computer network control systems, the experiment uses Python 3.8 on the Windows 10 platform and uses the Cooperative Association for Internet Data Analysis (CAIDA) dataset to perform 1000 iterations on the traditional CSA and CSA-FNN models, respectively. Fig. 7 shows the relationship between its training error and the iterations.



(a) CSA (b) CSA-FNN

Fig. 7. Comparison of CSA and CSA-FNN algorithm errors.

Fig. 7 shows the comparison of the effects of CSA before and after improvement on the training dataset. Compared to the standard CSA model, the rate of error reduction in the first 200 iterations of CSA-FNN is not significantly different. However, when the number of iterations reaches the range of [200, 400], the error of the FNN structure rapidly decreases and converges by the 400th iteration. The error of traditional CSA tends to converge after 700 iterations. Therefore, the CSA-FNN model proposed in the experiment has a faster convergence rate and a lower error in the final convergence. To eliminate the impact of dataset selection on experimental results and verify the generalization ability of the model, it is necessary to apply the Measurement and Analysis on the WIDE Internet (MAWI) dataset to train the above algorithms. Table I is the MAWI dataset parameter table.

Table I shows that the MAWI dataset is a very large dataset, including a large amount of network traffic data. Therefore, appropriate preprocessing and sampling are required when using this dataset for analysis and research [19]. The MAWI dataset usually contains more complex network traffic patterns, while the CAIDA dataset focuses more on reflecting the traffic characteristics of the actual Internet. For these two datasets, researchers preprocess the data, including data cleaning and normalization, to reduce noise and eliminate dimensional differences between features. Afterwards, a comparison is made between CSA-FNN and Whale Optimization Algorithm (WOA), Ant Colony Optimization (ACO), PSO, and Artificial Fish Swarm Algorithm (AFSA) to observe their training performance on different datasets. The results are shown in Fig. 8.

TABLE I. BASIC PARAMETERS OF THE ACTION DATASET

Parameter type	Parameter scale
Traffic data	735000
Time stamp	6900min
Source IP Address and Destination IP Address	900
Protocol	4
Packet size	512bit
Packet Marking	5

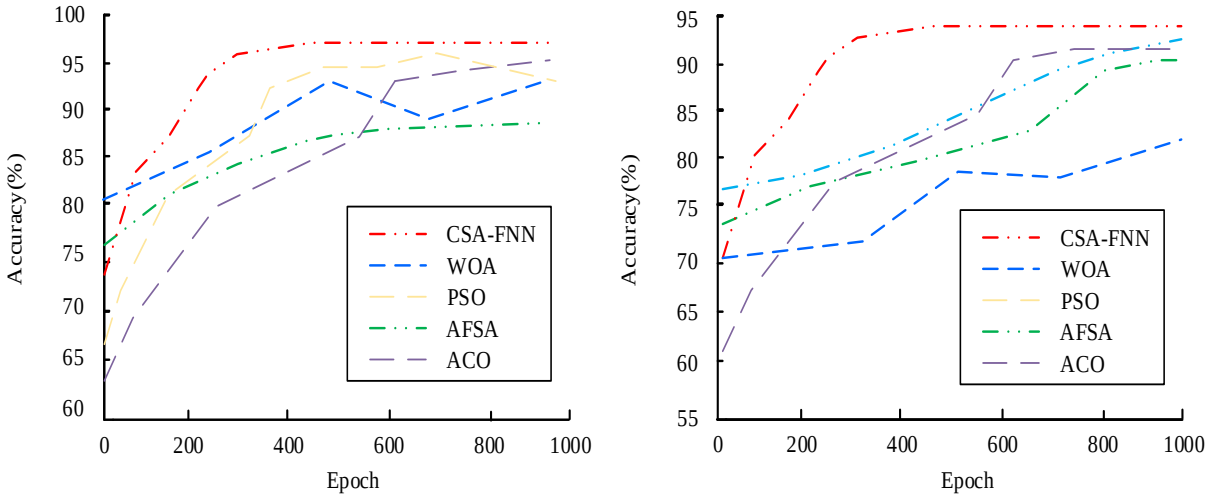


Fig. 8. Comparison of MAWI and CAIDA data set.

Fig. 8 shows the trend curve of the accuracy of various algorithms in two datasets as a function of the amount of training rounds. From Fig. 8 (a), the CSA-FNN model performs best, with accuracy tending to converge after 300 iterations, while the other four models only converge after training 600 times. In Fig. 8 (b), the accuracy of all algorithm models decreases when using the CAIDA dataset. However, the convergence of CSA-FNN accuracy does not change much, and it tends to converge after 400 iterations, ultimately converging to around 94%. The above results demonstrate that CSA-FNN has advantages over the other four algorithms, such as fast convergence, high accuracy, high stability, and strong generalization ability.

B. NOSA Simulation Experiment

The comparative analysis of algorithms has successfully verified the comparative advantages of CSA-FNN compared to other algorithm models, providing a solid theoretical foundation for its application in real computer network control systems. Although the superiority of the CSA has been verified, further simulation experiments are still necessary to evaluate the scalability, robustness, and stability of the algorithm. The experiment utilizes real traffic data from MAWILAB and obtains network traffic data through steps such as data cleaning and normalization [20]. Link utilization, IP packet rate, total number of IP packets, and TCP ratio are used as evaluation indicators. Link utilization refers to the degree to which network links are occupied by valid data. The IP packet rate and total number reflect the strength of network traffic. The TCP ratio represents the proportion of Transmission Control Protocol (TCP) packets to the total number of packets. Table II shows some data.

TABLE II. SOME NETWORK TRAFFIC DATA

Serial Number	Link Utilization	IP Packet Rate	Total Number of IP Packets	TCP Proportion
734	0.0113659	0.009562344	0.011231245	0.442123226
735	0.0132411	0.009878563	0.016324534	0.442661626
736	0.0114534	0.009758231	0.012313122	0.412336123
737	0.0164553	0.009431312	0.011229807	0.412286126
738	0.0174554	0.009341123	0.011123145	0.512326112
739	0.0142432	0.009234133	0.011212343	0.123146126
740	0.0111311	0.009124313	0.011224344	0.642666126
741	0.0141233	0.009413444	0.011224523	0.412312313
742	0.0143133	0.009512312	0.011253451	0.482626126
743	0.0143566	0.009413123	0.011231312	0.144567435
744	0.0163234	0.009434546	0.011212343	0.734341231
745	0.0178621	0.009223431	0.011213217	0.423123126
746	0.0115456	0.009254323	0.011221203	0.442231226
747	0.0112567	0.009256723	0.011231207	0.431234112

Normalizing the original data is helpful to eliminate the dimensional difference between features, improve the training effect and stability of the model, and improve the rate of convergence of the model. Then the normalized real network traffic data are used to verify the improved CSA-FNN structure. The simulation software used in the study is Pychar, and the simulation environment is Python 3.6. 1000 iterative training sessions are performed on CSA and CSA-FNN using the data shown in Table II. The link weights of the normalization layer and the anti-fuzzification layer are randomly generated by a Gaussian function to circumvent the issue of gradient disappearance or explosion due to an excessive data volume of the model. Furthermore, the introduction of randomness serves to enhance the stability of the model's learning and optimization process. The results are displayed in Fig. 9.

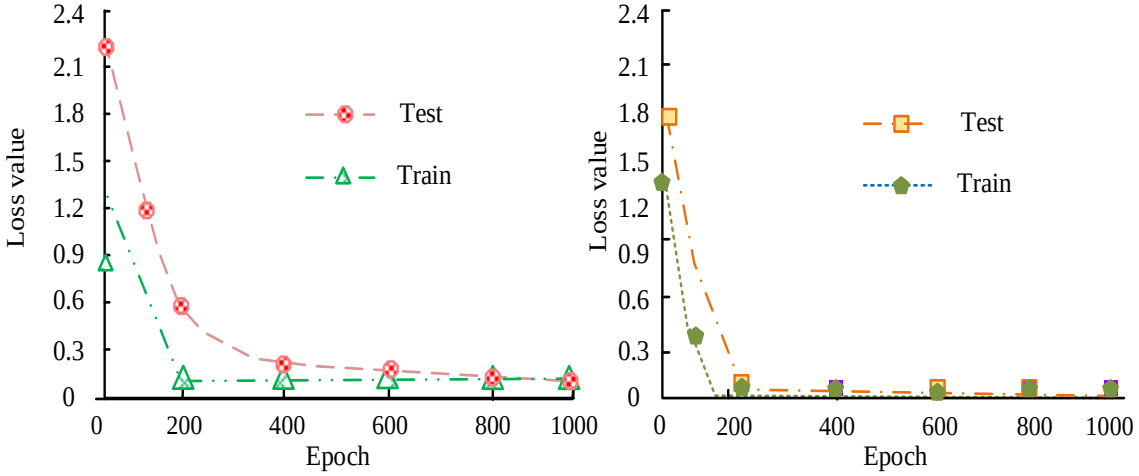


Fig. 9. Loss function curve of CSA and CSA-FNN.

In Fig. 9, the data gradient after the normalization operation is significantly reduced, and the stability of the model is significantly improved compared with the MAWI dataset, and the rate of convergence is accelerated. From Fig. 9(a), the CSA training set curve converges after about 200 iterations, while the test set converges after about 400 iterations. From Fig. 9(b) that during the first 120 iterations of the training set of the CSA-FNN model, the value of the Loss function rapidly decreases from 1.5 to 0.35, a decrease of about 76.7%. In the 220th iteration of the test set, the Loss function value is basically 0. Comparing the two figures, it can be found that CSA-FNN has increased by about 81.2% compared to CSA. Finally, Capsa software is used to capture real-time data traffic in the current local area network for analysis, intercepting network data traffic information for two consecutive hours. The experimental results are exhibited in Fig. 10.

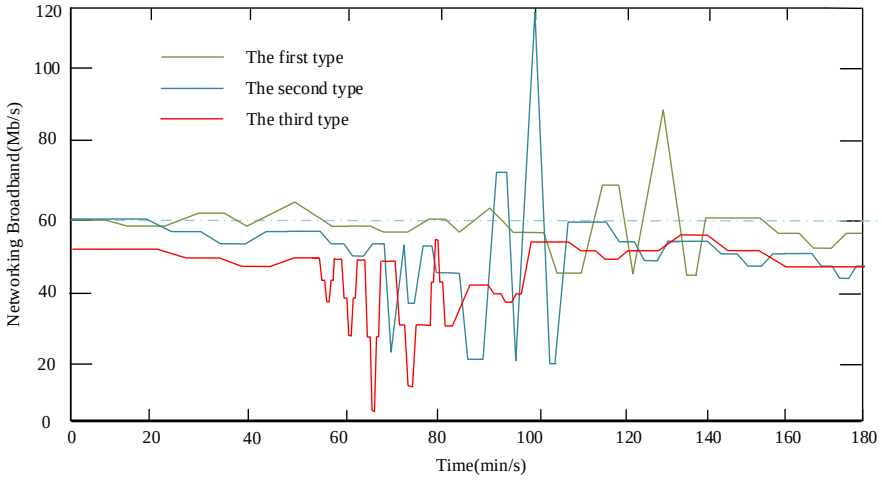


Fig. 10. Map of changes in network operational situational awareness.

Fig. 10 shows the changes in three types of NOSA. From the graph, the network bandwidth of the first type of network operation situation is mostly in a state of 40-60Mb/s, with occasional small fluctuations. This situation may be caused by accidental network changes. The second type of network operation situation is that the network is normal in the first 60 minutes, and severe fluctuations begin to occur in the first 60 minutes. The reason for this situation may be due to unstable factors in the local area network. The third type of network operation situation curve shows that the network is normal in the first 50 minutes, and then the network quality rapidly decreases. This situation may be due to the sudden addition of new tasks and drastic fluctuations in the network. Compared with the actual data, the network awareness model can accurately predict the network situation, with an accuracy rate of 98%. In addition, the study also records the changes in perception accuracy of the CSA-FNN model in different network environments, as shown in Fig. 11.

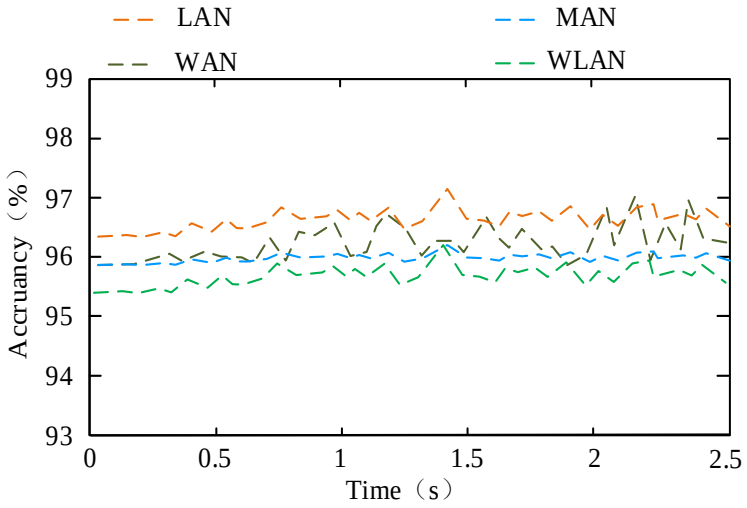


Fig. 11. Changes in perception accuracy under different network environments.

Fig. 11 shows the temporal variation of perception accuracy of the CSA-FNN model in different types of network environments (LAN, MAN, WAN, and WLAN). In Fig. 11, with the increase of time, the accuracy of various network environments shows an upward trend, among which LAN and WLAN have higher accuracy, maintaining above 98% and 99% respectively, while MAN and WAN have relatively lower accuracy, but also exceeding 97%. This phenomenon may be due to the relatively simple network environment of LAN and WLAN, which are easy to predict and classify, while MAN and WAN, due to their complex network structure, may have more uncertain factors leading to slightly lower accuracy. In addition, the fluctuation of each line type in the figure is relatively large, which may be caused by the dynamic changes in the network environment and the irregularity of the data flow. Overall, the CSA-FNN model can maintain high accuracy in various network environments, demonstrating its good adaptability and robustness. Finally, to verify the effectiveness of the algorithm in practical applications, the WSSD system proposed in study [21], the WEA-SA NSAM proposed in study [22], and the OWS-WOA model proposed in study [23] are introduced to compare with the CSA-FNN. The implementation phase comprises the deployment of infrastructure, the implementation of security controls, and the execution of security processes.

These include the installation of network and security hardware, the configuration of network devices and firewalls, the deployment of intrusion detection systems and SIEM systems, and the execution of key security processes such as security auditing, data backup and recovery, patch management, etc. The experimental results are shown in Table III.

TABLE III. COMPARISON BETWEEN CSA-FNN AND OTHER NETWORK SITUATIONAL AWARENESS SYSTEMS

Experimental Group	Detection Rate (%)	False Alarm Rate (%)	Response Time (ms)	Computing Resource Consumption (%)
CSA-FNN	94.2	2.5	35	12
WSSD	91.8	3.1	40	14
WEA-SA	93.0	2.8	38	13
OWS-WOA	92.5	3.0	42	15

According to the data in Table III, the CSA-FNN algorithm exhibits high performance in network situational awareness. Specifically, the detection rate of CSA-FNN reaches 94.2%, which is the highest among the four experimental groups, indicating that it can more accurately identify abnormal behaviors or threats in the network. Meanwhile, the false positive rate of CSA-FNN is 2.5%, which is also the lowest among the four experimental groups, indicating that it performs better in reducing unnecessary alarms and thus minimizing interference with normal network activity. In terms of response time, CSA-FNN is 35 milliseconds, faster than WSSD and WEA-SA, but slightly slower than OWS-WOA. However, a response time of 35 milliseconds is still very fast, which is already fast enough for real-time network situational awareness. In terms of computational resource consumption, CSA-FNN consumes 12% of resources, which is the lowest among all models, demonstrating its advantage in RUE. In contrast, OWS-WOA has the highest resource consumption, reaching 15%. Overall, CSA-FNN performs well in the three key indicators of detection rate, false positive rate, and computational resource consumption, especially in terms of detection rate and resource consumption, indicating that CSA-FNN is an efficient and accurate tool for network situational awareness.

V. CONCLUSION

To solve the problems of fuzziness, instability, large data volume, and difficulty in network situational awareness in network operation, this study designed a computer NSAM based on CSA and fuzzy logic fusion. The experiment used Python 3.8 on the Windows 10 platform and trained CSA and CSA-FNN using the CAIDA dataset. The results showed that the error of the proposed FNN structure rapidly decreased and converged at the 400th iteration. The error of the standard CSA only converged after 700 iterations, and the convergence rate increased by 75%. Then, the performance of CSA-FNN algorithm was evaluated through horizontal comparative experiments with algorithms such as WOA, ACO, PSO, AFSA, etc. The data showed that the accuracy of all algorithm models has decreased when using the CAIDA dataset. However, the accuracy convergence of the CSA-FNN model did not change much, and it tended to converge after 400 iterations, ultimately converging to around 94%. This proved that CSA-FNN had advantages over the other four algorithms such as fast convergence, high accuracy, high stability, and strong generalization ability. Finally, simulation experiments were conducted using MAWILAB: During the first 120 iterations of CSA-FNN, the testing accuracy improved with the improvement of training accuracy, and the speed was very significant. During this period, the value of Loss function decreased rapidly from 1.5 to 0.35, with a decrease of about 76.7%. Therefore, this model has good practical application capabilities. However, the model still requires a considerable amount of training and a longer training time, which is also an area that can be further improved in future research. The performance of the proposed model is an important consideration when the network size and complexity increase. The scalability of the model is key to ensuring its effective operation in larger or more complex network environments. As the scale of the network expands, models may need to handle larger amounts of data. Therefore, optimizing the data processing flow and algorithm efficiency is necessary to maintain or improve the response speed and accuracy of the model. In addition, the system for monitoring and measuring network status needs to effectively utilize computing resources. In the future, the concept of RUE can be studied and implemented to ensure that the system is efficient in resource utilization, especially in situations where multiple resources are limited.

REFERENCES

[1] Amin S N, Shivakumara P, Jun T X. An Augmented Reality-Based Approach for Designing Interactive Food Menu of Restaurant Using Android.Artificial Intelligence and Applications. 2023, 1(1): 26-34.

[2] Nsugbe E. Toward a Self-Supervised Architecture for Semen Quality Prediction Using Environmental and Lifestyle Factors.Artificial Intelligence and Applications. 2023, 1(1): 35-42.

[3] Oslund S, Washington C, So A, Multiview Robust Adversarial Stickers for Arbitrary Objects in the Physical World. Journal of Computational and Cognitive Engineering, 2022, 1(4): 152-158.

[4] Liu C, Wang J, Zhou L. Solving the multi-objective problem of IoT service placement in fog computing using cuckoo search algorithm. Neural Processing Letters, 2022, 54(3): 1823-1854.

[5] Abed-alguni B H, Alawad N A, Barhoush M, et al. Exploratory cuckoo search for solving single-objective optimization problems. Soft Computing, 2021, 25(15): 10167-10180.

[6] Cheng P, Wang J, Zeng X,Zeng, X. , Bruniaux, P. , & Tao, X..Motion comfort analysis of tight-fitting sportswear from multi-dimensions using intelligence systems.Textile Research Journal, 2022, 92(11-12):1843-1866.

[7] Eltamaly A M. Optimal control parameters for bat algorithm in maximum power point tracker of photovoltaic energy systems.International Transactions on Electrical Energy Systems, 2021, 31(4):1-22.

[8] Fan J, Xu W, Huang Y. Application of chaos cuckoo search algorithm in computer vision technology. Soft Computing, 2021, 25(18): 12373-12387.

[9] Li J, Yang Y H, Lei H, & Wang, G. G. Solving logistics distribution center location with improved cuckoo search algorithm. International Journal of Computational Intelligence Systems, 2021, 14(1): 676-692.

[10] Chen S Y, Yang M C. Nonlinear Contour Tracking of a Voice Coil Motors-Driven Dual-Axis Positioning Stage Using Fuzzy Fractional PID Control with Variable Orders.Mathematical Problems in Engineering, 2021, 2021(4):1-14.

[11] Muhammad K, Obaidat M S, Hussain T, Ser, J. D., & Doctor, F. Fuzzy Logic in Surveillance Big Video Data Analysis: Comprehensive Review, Challenges, and Research Directions.ACM Computing Surveys, 2021, 54(3):1-33.

[12] Ben Jabeur C, Seddik H. Design of a PID optimized neural networks and PD fuzzy logic controllers for a two-wheeled mobile robot. Asian Journal of Control, 2021, 23(1): 23-41.

[13] Costache R, Arabameri A, Moayedi H, Pham, Q. B., Santosh, M., Nguyen, H, Pham, B. T. Flash-flood potential index estimation using fuzzy logic combined with deep learning neural network, naïve Bayes, XGBoost and classification and regression tree. Geocarto International, 2022, 37(23): 6780-6807.

[14] Ulus Ş, Eski I. Neural network and fuzzy logic-based hybrid attitude controller designs of a fixed-wing UAV. Neural Computing and Applications, 2021, 33(14): 8821-8843.

[15] Katsikis V N, Stanimirović P S, Mourtas S D,Xiao, L., Karabašević, D., & Stanujkić, D. Zeroing neural network with fuzzy parameter for computing pseudoinverse of arbitrary matrix. IEEE Transactions on Fuzzy Systems, 2021, 30(9): 3426-3435.

[16] Dong S, Yu T, Farahmand H, et al. A hybrid deep learning model for predictive flood warning and situation awareness using channel network sensors data. Computer-Aided Civil and Infrastructure Engineering, 2021, 36(4): 402-420.

[17] Tan L, Yu K, Ming F, Cheng, X., & Srivastava. Secure and resilient artificial intelligence of things: a HoneyNet approach for threat detection and situational awarenes. IEEE Consumer Electronics Magazine, 2021, 11(3): 69-78.

[18] Huang D, Jiang F, Li K,Tong, G., & Zhou, G.. Scaled PCA: A new approach to dimension reduction. Management Science, 2022, 68(3): 1678-1695.

[19] Mawi H, Narine R, Schieda N. Adequacy of unenhanced MRI for surveillance of small (clinical T1a) solid renal masses. American Journal of Roentgenology, 2021, 216(4): 960-966.

[20] Aytekin A. Comparative analysis of the normalization techniques in the context of MCDM problems. Decision Making: Applications in Management and Engineering, 2021, 4(2): 1-25.

[21] Qidong Y, Jianbin G, Shengkui Z, et al. A dynamic Bayesian network-based reliability assessment method for short-term multi-round situation awareness considering round dependencies. Reliability engineering & system safety, 2024, 243(Mar.): 1109838.1-1109838.17.

[22] Zhitao C, Xiaodong Y, Yiyong Z. Research on hierarchical network security situation awareness data fusion method in big data environment. Journal of Cyber Security Technology, 2024, 8(1): 31-52.

[23] Guo, X, Jianing, Y, Zhanhui, G, et al. Research on Network Security Situation Awareness and Dynamic Game Based on Deep Q Learning Network. Journal of Internet Technology, 2023, 24(2): 549-563.

Application of Sanda-Assisted Teaching System Integrating VR Technology from a 5G Perspective

Zhaoquan Zhang*, Yong Ding
Police Command and Tactical College, Guangdong Police College, Guangzhou 510232, China

***Abstract*—Since the focus of Sanda teaching is to allow students to master martial arts techniques through confrontational exercises, Sanda teaching in colleges and universities commonly adopts contextualized teaching methods. However, this Sanda teaching method suffers from deficiencies such as poor teaching effectiveness and difficulty in reflecting the effectiveness of Sanda martial arts. In order to solve these problems and make up for the shortcomings of offline Sanda teaching, the study adopts the virtual reality technology and the fifth generation mobile communication technology to construct a Sanda-assisted teaching system for college students. In order to ascertain whether students complete the Sanda movement practice, the study has designed two models: a self-supervised model based on acceleration and angular velocity contrast learning and a multi-task semi-supervised model based on time-frequency contrast learning. These models aim to improve the analytical function of the Sanda-assisted teaching system and address the analytical deficiencies of the existing human movement identification algorithms. The results indicated that the maximum accuracy of the research-designed self-supervised model was 95.76% and 95.89% on the training and test sets, respectively. The multi-task semi-supervised model designed in the study plateaued after nearly 22 and 24 iterations on the training and test sets, respectively. The average response time of the research-designed system was 59ms, and the throughput could reach a maximum of 77651bit/s. The model and the research-designed system both worked well, and they can lower the risk of student injuries while offering technological support for Sanda-assisted teaching and learning in higher education institutions.**

***Keywords*—VR technology; Sanda; teaching system; motion recognition; feature extraction**

I. INTRODUCTION

Sanda, as an excellent traditional sport of the Chinese nation, contains valuable national spirit, reflects strong national traditional colors, and has strong vitality and epochal character. University Sanda teaching (UST) can promote the development of Sanda sport in the new era and make it glow with more beautiful colors [1-2]. Most colleges and institutions have embraced the contextual teaching style since it aligns with Sanda teaching's goal. However, this Sanda teaching method has shortcomings such as poor teaching effect and difficulty in reflecting the effectiveness of Sanda martial arts [3]. In order to solve this problem, constructing Sanda-assisted teaching system (SATS) becomes particularly important. Virtual reality (VR) has been progressively incorporated into the domains of education, healthcare, gaming, and entertainment with the advancement of computer technology [4]. Feng and other specialists created a model based on the combination of VR technology and Web application design to increase students' enthusiasm to learn. They then used the model's data to teach physical education (PET). The findings demonstrated how well VR and PET work together to increase students' excitement for athletics and interest in studying [5]. Gao and other researchers designed a medical system based on remote VR technology to address the problem of medical informatization. The system contained two major functional sections, namely, consultation management and system management. Among them, the consultation management section included all remote business consultation processes, and the system management version involved managing medical resources and user data. The findings demonstrated significant time and cost savings in the areas of remote VR technology adoption, safety, and rehabilitation-improvements of 85%, 92%, and 96%, respectively [6]. To help architecture students learn more effectively, researchers like Elgewely created a VR platform for architectural details based on VR technology and building information modeling. Additionally, the platform was validated in three main areas, namely pilot, system usability and immersion and learning gains. The results showed that students' learning progress increased by 30% after experiencing the VR environment [7]. Experts such as Almousa designed a VR-based virtual clinical simulation system with Oculus Quest headset in order to promote global academic collaboration. In addition, the system was able to connect and communicate in real time with an instructor control panel application. The results showed that this virtual clinical simulation system was able to provide realistic clinical training in a virtual space simulating a hospital environment and promote academic collaboration [8].

In SATS, the most important thing is to judge whether the student has completed the exercise or not, and this requires the use of human motion recognition (HMR) algorithm model. Common HMR methods include traditional machine learning, such as K-nearest neighbor classification algorithms, self supervised learning (SSL) methods, semi supervised learning, and wearable device based methods [9-10]. To solve the problem that the existing HMR-based algorithms cannot fully explore the spatio-temporal properties of motion, Yangzhi and other researchers designed a HMR algorithm based on a spatio-temporal attention graph convolutional network model that integrates spatial and temporal attention mechanisms. The comparison results show that the algorithm designed in the study has improved the accuracy of Top-1 and Top-5 by 5.0% and 4.5%, respectively [11]. To increase the precision of video HMR and the computational effectiveness of large-scale datasets, Gao and other specialists created a motion capture

multidimensional data model and deep learning framework based video image motion recognition. In addition, the study also used Gaussian mixture model and gradient histogram. The outcomes indicated that the average value (AV) of the classification accuracy of the method was 85.79% and the maximum running speed was 20 frames per second [12]. To improve the accuracy of the rehabilitation robot on HMR and to reduce the recognition time, Chen and other researchers designed an improved sparrow search algorithm based on multiple strategies. According to the results, this upgraded algorithm's recognition accuracy was 2.835% higher than that of the original classifier, which should make it easier for the rehabilitation robot to understand the intention behind a person's movements [13]. Ye and colleagues developed an enhanced time-slotted video down sampling technique based on Gaussian model and proposed a human interaction recognition algorithm based on parallel multi-feature fusion network to identify human actions. Convolutional kernels with various scales were employed in the study to extract features. The findings demonstrate that the approach can identify six interaction acts with an accuracy of 88.9% [14].

In summary, the current studies on VR-based systems and HMR are relatively rich and use a variety of methods. However, these studies also have certain shortcomings, such as not fully considering the complexity and specificity of human motion sensing signals, and still facing the problems of underutilization of human motion sensing data and difficulty in feature extraction. Therefore, in an attempt to compensate for the shortcomings of Offline Sanda teaching (OST), the study designed a SATS based on VR and 5th generation mobile communication technology (5G), constructed self-supervised model with contrastive learning of acceleration and angular velocity (SSMCLAA) and semi-supervised multi-task model with time-frequency (TF) contrastive learning (SSMTFCL). The study aims to improve the speed and accuracy of HMR, accelerate the judgment of whether students complete Sanda exercises, reduce the probability of student injuries, and improve the effectiveness of UST. The innovativeness of the study is that it combines 5G technology and VR technology to alleviate the problem of insufficient data labeling, and realize the consistency of TF characteristics of data and the improvement of HMR speed and accuracy.

There are five sections to the study overall. Section II is the design of the research methodology, which includes the construction of SATS based on 5G and VR technologies, the design of SSMCLAA model and SSMTFCL model. Section III is the performance validation of SATS, SSMCLAA model and SSMTFCL model. Discussion is given in Section IV. Section V is the conclusions, shortcomings and future prospects of the study.

II. Methods and Materials

To address the problems of OST, the study designed SATS based on 5G and VR technologies, and designed the structure and functions of teaching system (TS). In order to realize the judgment of whether students complete the exercises in TS, the research adopts the HMR algorithm, and designs the corresponding recognition model for the problems existing in the current HMR, respectively.

A. SATS Design Based on 5G and VR Technologies

UST suffers from deficiencies such as poor teaching effectiveness, difficulty in reflecting the effectiveness of Sanda martial arts, and the frequency of accidental injuries or even serious permanent injuries to students. To compensate for the shortcomings of OST, the study adopted 5G and VR technologies. Since VR technology can simulate real-world situations in three dimensions, it has been included into college and university curricula. The majority of these institutions have started to create VR teaching laboratories [15]. 5G technology has ultra-high speed rate, which can quickly enjoy 360° panoramic VR and display more high-definition VR images [16]. The structure of SATS is shown in Fig. 1.

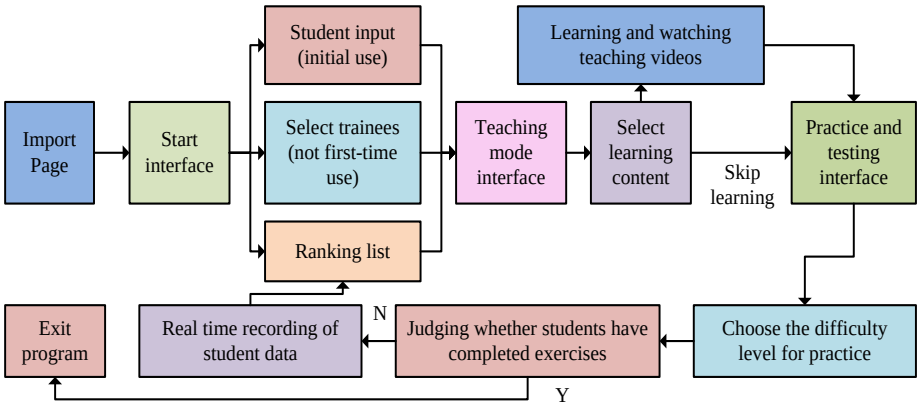


Fig. 1. The structure of sanda-assisted teaching system.

In Fig. 1, the structure of SATS involves import page, start interface, ranking list, student input (initial use), select trainees (not first-time use), teaching mode interface, select learning content, learning and watching teaching videos, skip learning, practice and testing interface, choose the difficulty level for practice, judging whether the difficulty level is too high or too low level for practice, judging whether students have completed exercises, real time recording of student data, and exit program. SATS involves four main functions, which are action demonstration function, practice and test function, interaction function, and data analysis function. Among these, the action demonstration function requires 5G technology to

guarantee that students can view the video with clarity and fluidity without experiencing any latency. Both the practice and test function and the interaction function need to allow students to follow the instructions on the VR interface to practice and get feedback from the system. The composition of the interaction function is shown in Fig. 2.

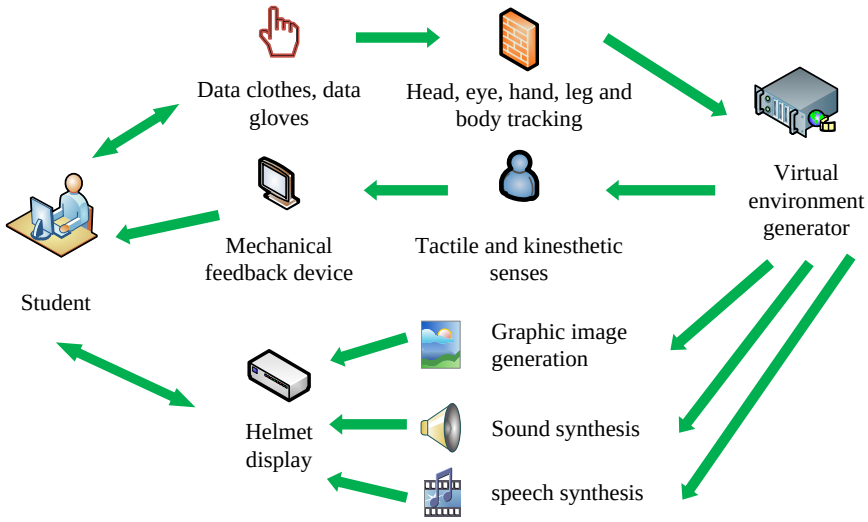


Fig. 2. Composition of interactive functions.

In Fig. 2, the interaction function requires the use of different wearable devices, a data suit, data gloves and a helmet display, as well as a mechanical feedback device. The camera in the helmet display is able to recognize the movement of the student's hand. The data suit and data gloves are able to track the student's head, eyes, hands, legs and body, and the mechanical feedback device mainly involves haptic and kinesthetic senses. The data analysis function is to collect the data changed in the VR scene through the data collection device in the background, and analyze and diagnose these data to determine whether the students have completed the exercise or not. In the parts that follow, the study's design will be thoroughly examined for the purpose of diagnosing these results.

B. Construction of the SSMCLAA Model

To make a judgment on whether a student has completed the exercise or not, the study uses the HMR algorithm. The human movement time series data collected using wearable sensing devices has a complex multidimensional spatial structure. To address the problems of complexity and variability of the collected data and the difficulty of feature extraction, the study designed the SSMCLAA model. SSL utilizes the input data itself as a supervisory signal and is beneficial for almost all types of downstream tasks [17]. Wearable sensing devices integrate various types of miniature sensing elements, such as accelerometers, gyroscopes, etc., which are capable of monitoring and tracking human activities in real time and continuously [18]. Therefore, the multidimensional time series data samples used in the study are composed of 3D acceleration data and 3D angular velocity data. The SSMCLAA model adopts the SSL framework, and its specific structure is shown in Fig. 3.

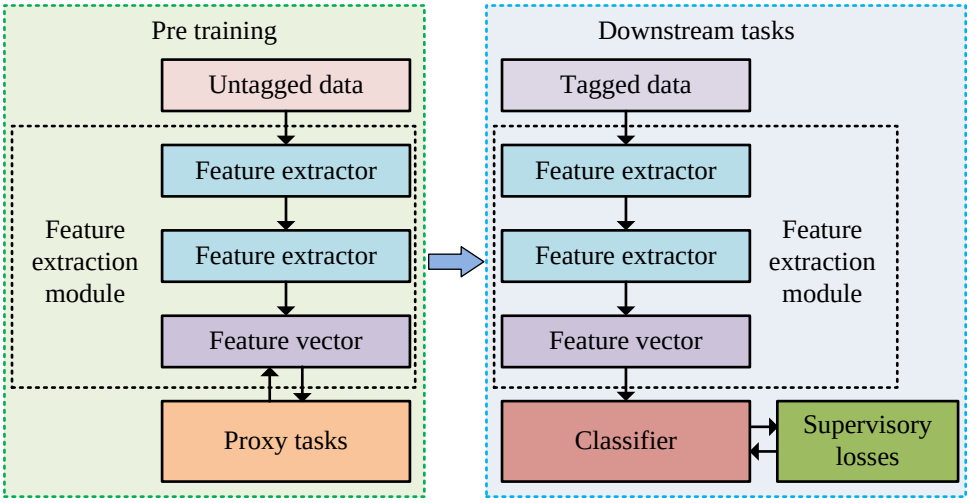


Fig. 3. The specific structure of the SSMCLAA model.

In Fig. 3, the SSL framework contains two main parts, which are pre-training and downstream tasks. Among them, the pre-training part contains unlabeled data (ULD), feature extraction module (FEM), and agent task, where FEM is divided into feature extractor (FE) + eigenvector. In order to carry out the agent task in the pre-training part, the study adopts the contrast learning technique. This technique is a common way to improve the model representation. The downstream task part contains labeled data (LD), FE trained in the pre-training part, eigenvector, classifier and supervised loss. Since the FE in the pre-training part only learns the feature representation of the data, the downstream task part needs to connect the FE trained in the pre-training part with the classifier, and then supervise the training with a small amount of LD to achieve action classification. In order to fully extract the acceleration and angular velocity features of the students when learning Sanda, the study constructs two independent FEs at the feature extraction module, and both of them use the deep residual network structure. In Fig. 4, the FEs' structure is displayed.

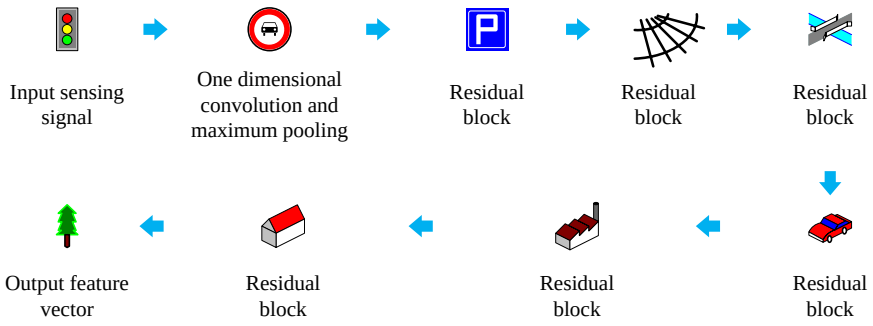


Fig. 4. The structure of a feature extractor.

In Fig. 4, the first step of the FE is the input sensing signal, i.e., the acceleration or angular velocity of the student while performing Sanda learning. The second step is a one-dimensional convolution and maximum pooling operation; after which it goes through six residual blocks. The combination of these six residual blocks together forms a deep residual network. The residual block is mainly composed of three consecutive one-dimensional convolutional layers and an activation function ReLU, and then the output and input of the convolution operation are directly added to solve the problem of gradient vanishing during the training process of deep neural networks, and to enable deep neural networks to learn deeper feature representations. The third step is the output eigenvector. The residual block mainly contains three 1D convolutional layers and ReLU activation function. The purpose of contrast learning pre-training is to enhance the FE representation. The commonly used formulaic definition of contrast learning is shown in Eq. (1) [19].

$$\max_{f_1,f_2}\sum_{i=1}^{N}\left\{s\left(z_1\left(w_{i1}\right)\right),\left(z_2\left(w_{i2}\right)\right)-\sum_{j=1,j\neq i}^{N}\left[s\left(z_1\left(w_{i1}\right)\right),\left(z_1\left(w_{j1}\right)\right)+s\left(z_1\left(w_{i1}\right)\right),\left(z_2\left(w_{j2}\right)\right)\right]\right\} \quad (1)$$

In Eq. (1), f_1 and f_2 represent different FEs, respectively. i and j are both ordinal numbers, and N is the total features. w_i and w_j represent different random samples, z_1 and z_2 are the projected heads of the upper and lower branches (ULB), respectively. $z_1\left(w_{i1}\right)$ and $z_2\left(w_{i2}\right)$ represent the features of the ULB of w_i, respectively, and both are set as positive samples. $z_1\left(w_{j1}\right)$ and $z_2\left(w_{j2}\right)$ are the features of the ULB of w_j, respectively, and are set as negative samples. s is the metric. The structure of the contrast learning pre-training is shown in Fig. 5.

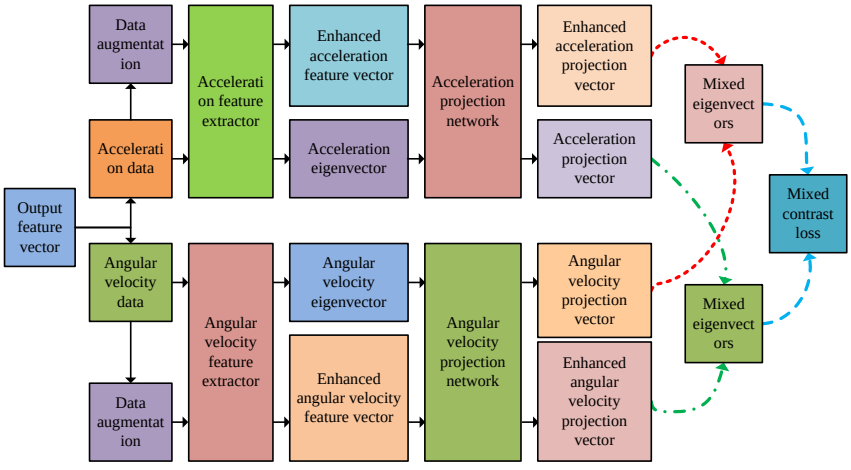


Fig. 5. Comparative learning pre training structure.

In Fig. 5, the contrast learning pre-training structure is mainly divided into the upper and lower halves. Among them, the upper half is the acceleration contrast learning process and the lower half is the angular velocity contrast learning process. In order to facilitate the subsequent unsupervised contrast learning, the study enhances both raw data. The study adds a projection network and obtains a new projection eigenvector in an effort to enhance the impact of FE even more. The acceleration contrast learning loss L_b is shown in Eq. (2).

$$L_b = -\sum_{i \in I} \log \frac{\exp\left(sim\left(z_i^b, z_i'^b\right)/\chi\right)}{\sum_{h \in H(i)} \exp\left(sim\left(z_i^b, z_h^b\right)/\chi\right)} \tag{2}$$

In Eq. (2), I represents the set of acceleration data after data augmentation, and $H(i)$ denotes the subscripts of the rest of the sample data except for the serial number i. χ is the parameter, z_i^b and $z_i'^b$ are the corresponding projected eigenvector before and after the acceleration data enhancement, respectively. z_h^b is the projected eigenvector corresponding to the acceleration data other than the serial number i. $sim\left(z_i^b, z_i'^b\right)$ is the cosine similarity, which is solved as shown in Eq. (3) [20].

$$sim\left(z_i^b, z_i'^b\right) = \frac{\left(z_i^b\right)^T \cdot z_i'^b}{\left\|z_i^b\right\| \left\|z_i'^b\right\|} \tag{3}$$

The expression for $H(i)$ is shown in Eq. (4).

$$H(i) = I \setminus \{i\} \tag{4}$$

The angular velocity comparison learning loss L_g is shown in Eq. (5).

$$L_g = -\sum_{i \in I} \log \frac{\exp\left(sim\left(z_i^g, z_i'^g\right)/\chi\right)}{\sum_{h \in H(i)} \exp\left(sim\left(z_i^g, z_h^g\right)/\chi\right)} \tag{5}$$

In Eq. (5), z_i^g and $z_i'^g$ are the projected eigenvectors corresponding to the angular velocity data before and after enhancement, respectively. z_h^g is the projected eigenvector corresponding to the angular velocity data other than the ordinal number i. The hybrid contrast loss function (LF) L_{bg} is shown in Eq. (6).

$$L_{bg} = -\sum_{i \in I} \log \frac{\exp\left(sim\left(z_i^{bg}, z_i'^{bg}\right)/\chi\right)}{\sum_{h \in H(i)} \exp\left(sim\left(z_i^{bg}, z_h^{bg}\right)/\chi\right)} \tag{6}$$

In Eq. (6), z_i^{bg} and $z_i'^{bg}$ represent the hybrid eigenvector corresponding to the data before and after data enhancement, respectively. z_h^{bg} is the hybrid eigenvector corresponding to the data other than the ordinal number i. The solution of z_i^{bg} and $z_i'^{bg}$ is shown in Eq. (7).

$$\begin{cases} z_i^{bg} = z_i'^b \oplus z_i^g \\ z_i'^{bg} = z_i^b \oplus z_i'^g \end{cases} \tag{7}$$

The total LF L_{sum} for the pre-training phase of contrast learning is shown in Eq. (8).

$$L_{sum} = L_b + L_g + \beta L_{bg} \tag{8}$$

In Eq. (8), β represents the parameter. The downstream task of the SSMCLAA model is to classify students' Sanda practice movements. In the downstream task, a simple two-layer linear mapping network and activation function are used for the classifier structure, and the cross-entropy loss function (CELF) L_{out} is used for the LF, as shown in Eq. (9) [21].

$$L_{out} = \frac{1}{m} \sum_{i, w_i \in G^d} \sum_{t=1}^{T} k_{it} \log\left(\hat{k}_{it}\right) \tag{9}$$

In Eq. (9), k_{it} and $\hat{k}_{it}$ represent the actual and predicted distribution probabilities, respectively. m is the size of the randomly selected data subset G^d. t and T are the action category ordinal number and total number.

C. Construction of the SSMTFCL Model

To make a judgment on whether a student has completed the Sanda exercise, research designed the HMR model of SSMCLAA to solve the problems of complex and variable collected data and the difficulty of feature extraction. However, the SSMCLAA model is unable to learn the common feature space structure of LD and ULD at the same time. In addition, human motion sensing signals are not only fluctuating and periodic, but also have multimodal characteristics such as time and frequency-domains (FDs) [22]. However, the current HMR only considers unilateral information in the time or frequency domain during feature extraction, ignoring the consistent relationship between time and frequency of human action sensing data, as well as the study of multi-task framework [23-24]. Based on these problems, in order to better judge whether students complete the Sanda exercise, the study also designed the SSMTFCL model and arranged it together with the SSMCLAA model in SATS to judge whether students complete the Sanda exercise. Fig. 6 depicts the SSMTFCL model's structure.

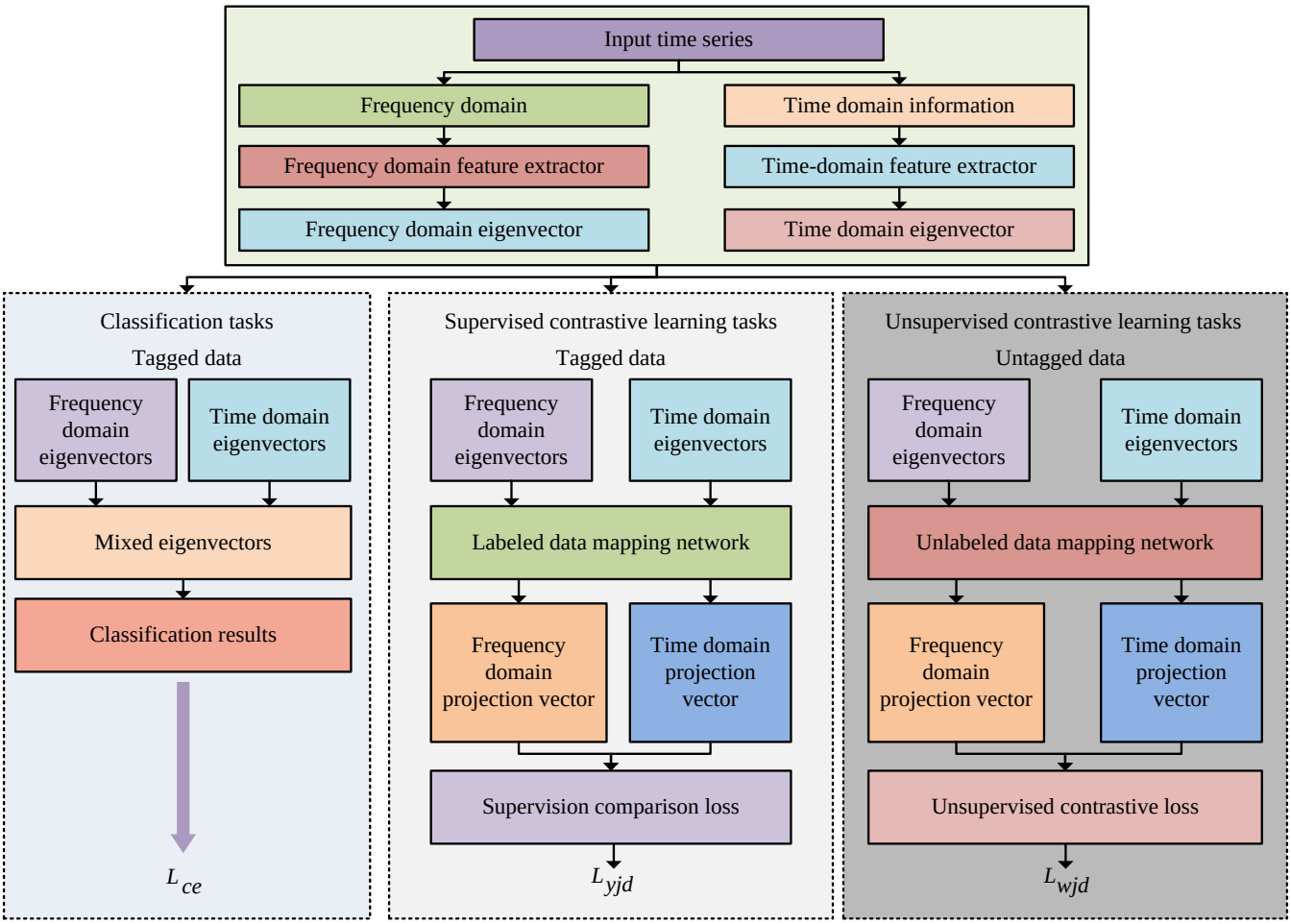


Fig. 6. The structure of SSMTFCL model.

In Fig. 6, the first step of the SSMTFCL model is to input the time series, and the second step is to obtain the FD information x_i' and the time-domain (TD) information x_i. The third step is to input x_i' and x_i into the FD FE V_F and the TD FE V_Q respectively. The fourth step is to output the corresponding FD eigenvector y_i' and the TD eigenvector y_i. The fifth step is to carry out the three tasks in the respective modules, i.e., classification task (CT), supervised contrast learning task and unsupervised contrast learning task, and output the CELF L_{ce}, supervised contrast learning LF L_{yjd} and unsupervised contrast learning LF L_{wjd}. In the CT, in order to maximize the consistency between the frequency domain and the TD of the samples, the study directly compares the TD projection vectors and the FD projection vectors of the data. To obtain the FD information, the study used Fourier transform. The Fourier transform has the advantage of good frequency localization and clearly shows the frequency components contained in the signal [25-26]. The process of Fourier transform is shown in Eq. (10) [27].

$$F(\omega)=\frac{1}{\sqrt{2\pi}}\int f(t)e^{i\omega t}dt \tag{10}$$

In Eq. (10), $f(t)$ represents a non-periodic function and $F(\omega)$ is the representation of the $f(t)$ function in the FD. $e^{i\omega t}$ is a complex exponential function and ω represents the angular frequency. The CELF L_{ce} for the CT is the same as the CELF L_{out} used in the downstream task of the SSMCLAA model, with only a slight difference in the values taken. The expression of L_{ce} is shown in Eq. (11).

$$L_{out}=\frac{1}{m}\sum_{i,x_i\in G^d}\sum_{t=1}^{T}k_{it}\log\left(\hat{k}_{it}\right) \tag{11}$$

To capture deeper TF information, the study constructed two independent FEs to extract the human action eigenvector from TD information and FD information, respectively. The TF feature extraction module is shown in Fig. 7.

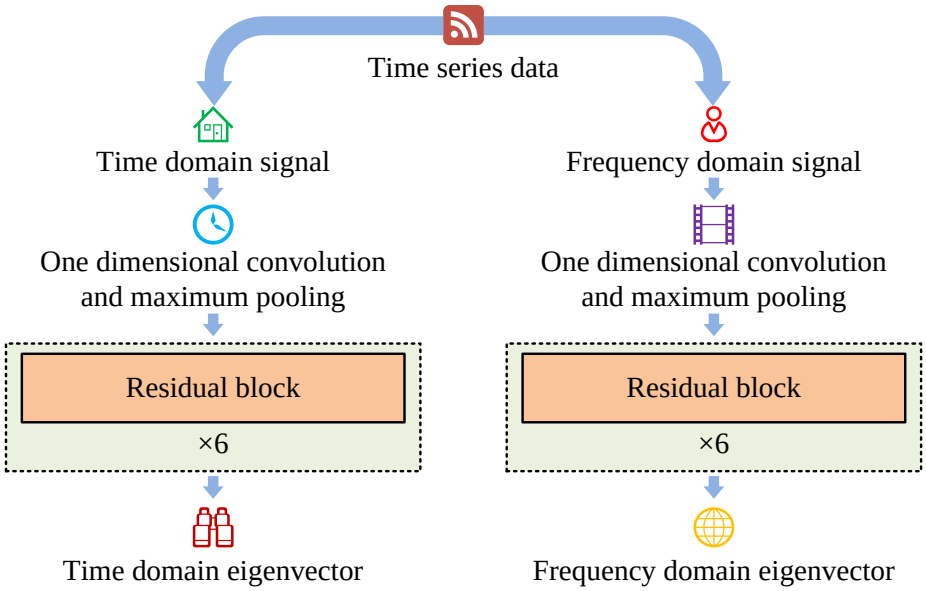


Fig. 7. Time-frequency feature extraction module.

In Fig. 7, the TF feature extraction module mainly consists of a TD signal FE and a FD signal FE. Both signal FEs involve one-dimensional convolution, maximum pooling, six residual block structure and eigenvector output. Both labeled and ULD are subjected to the same feature extraction method, and their respective obtained TF eigenvectors are used in subsequent multi-task learning modules. The supervised comparison learning task and the unsupervised comparison learning task in the SSMTFCL model are mainly used to mine the internal features of ULD and LD, and the supervised comparison learning task takes into account the data labeling information when calculating the loss. The unsupervised comparison LF L_{wjd} is shown in Eq. (12).

$$L_{wjd} = -\sum_{i \in R} \log \frac{\exp\left(sim\left(u_i^v, u''^v_i\right)/\chi\right)}{\sum_{h \in H(i)} \exp\left(sim\left(u_i^v, u_h^v\right)/\chi\right)} \tag{12}$$

In Eq. (12), R is the set of TD eigenvector and FD eigenvector, and u_i^v and u''^v_i denote the TD projection vectors and FD projection vectors of ULD, respectively. u_h^v is the TD projection vector corresponding to the ULD other than the ordinal number i. Minimization of unsupervised contrast loss can improve the model's representational ability. The study uses an unsupervised contrast learning task to learn the features of ULD, and also uses supervised contrast learning to perform deep mining of the deep structural features of LD. The study trains two contrast learning tasks in parallel to learn the feature space on the full data. The LF L_{yjd} for supervised contrast is shown in Eq. (13).

$$L_{yjd} = -\sum_{i \in R} \frac{1}{\left|K(i)\right|} \sum_{k \in K(i)} \log \frac{\exp\left(sim\left(u_i, u'_i\right)/\chi\right)}{\sum_{h \in H(i)} \exp\left(sim\left(u_i, u_h\right)/\chi\right)} \tag{13}$$

In Eq. (13), $K(i)$ represents the set of sample data of the same category as sample x_i. u_i and u'_i represent the TD projection vector and FD projection vector of the LD, respectively. u_h is the TD projection vector corresponding to the LD other than the serial number i. u_i and u'_i are solved as shown in Eq. (14).

$$\begin{cases} u_i = G_c\left(y_i\right) \\ u'_i = G_c\left(y_i'\right) \end{cases} \tag{14}$$

In Eq. (14), G_c represents the LD projection network. In order to allow the feature encoder to learn the intrinsic structure of the LD more deeply, the study minimizes the supervised comparison loss as well. In order to achieve overall consistency of data features, the study adopts a multi-task learning framework. The study co-trained the three tasks of the model so that they jointly participate in the optimization of the TF feature encoder. Eq. (15) displays the SSMTFCL model's total LF.

$$L_{\text{SSMTFCL}} = L_{ce} + \theta L_{yjd} + \varphi L_{wjd} \tag{15}$$

In Eq. (15), both θ and φ are scaling parameters.

III. RESULTS

To validate the performance of the research design SATS and the corresponding student action recognition classification model, the study sets up the experimental environment, experimental parameters and experimental dataset. In addition, the study also describes the comparison model and comparison system. The model comparison mainly involves accuracy, error and time consumption, while the system comparison mainly involves memory occupation, throughput and response time.

A. SSMCLAA Model Performance Validation

To validate the performance of the SSMCLAA model, the study uses the UCI-HAR dataset [28]. Since the length of the time series data is not uniform, the study uses a sliding window mechanism to segment the raw data, and the number of samples after processing is 6400. The study divides the dataset into a training set and a test set with a division ratio of 7: 3. The ratio of LD to ULD in the training set is 1: 9. The β parameter of the SSMCLAA model is set to 1, and the learning rates for the pre-training and downstream task phases are 1e-5 and 1e-3, respectively. Other SSL methods selected for comparison in the study are semi-supervised time model (SemiTime), self-supervised of human activity recognition (SelfHAR) and simple framework for contrastive learning of representations (SimCLR). The operating system used for the experiments is Windows 11 (64-bit), and the processor is Intel Core i5-12600 K with a maximum RWI of 4.9 GHz, a maximum accelerated power consumption of 130 W, and a maximum RAM of 128 GB. Fig. 8 compares the accuracy of the various models in the training and test sets.

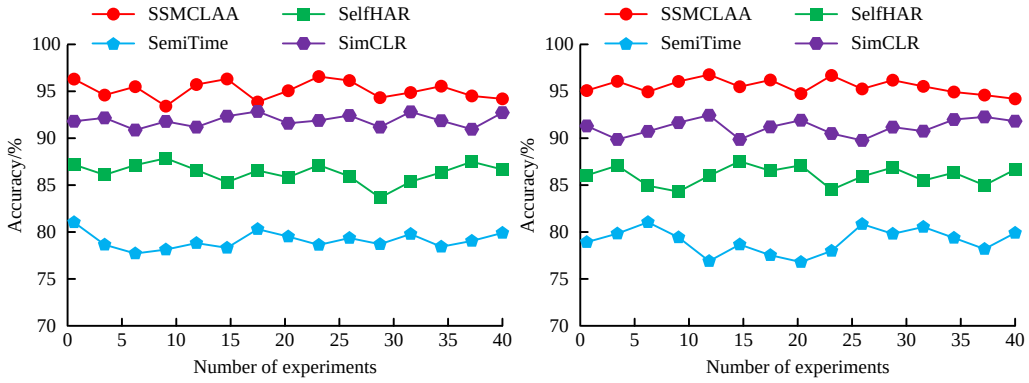


Fig. 8. Comparison of accuracy between different models on training and testing sets.

In Fig. 8(a), the maximum value (MaxV) of the accuracy of the research design SSMCLAA model is 95.76% and the minimum value (MinV) is 93.81% on the training set. The MaxVs of accuracy of SemiTime model, SelfHAR model and SimCLR model are 80.50%, 87.42% and 93.03%, respectively, and the MainVs are 77.64%, 83.89% and 90.71%, respectively. The accuracy of SSMCLAA model is significantly higher than the comparison models. In Fig. 8(b), on the test set, the MaxV of accuracy is appeared on SSMCLAA model with a value of 95.89%. The MainV of accuracy is appeared on SemiTime model with a value of 76.75%. The mean values of accuracy for SelfHAR model and SimCLR model are 86.13% and 91.24% respectively. The SelfHAR and SimCLR models outperformed the SemiTime model. In summary, the research design SSMCLAA model has a higher action recognition accuracy and is able to better recognize and classify students' Sanda actions in order to determine whether the students have completed Sanda training or not. Comparison of mean absolute error (MAE) and mean squared error (MSE) of different models are shown in Table I.

TABLE I. COMPARISON OF MAE AND MSE OF DIFFERENT MODELS

Model	MAE					MSE				
	Number of experiments					*Number of experiments*				
	1	*2*	*3*	*4*	*5*	*1*	*2*	*3*	*4*	*5*
SemiTime	1.21	1.37	1.19	1.32	1.28	1.31	1.29	1.38	1.42	1.46
SelfHAR	1.03	0.97	1.13	0.95	1.05	1.16	1.21	1.25	1.09	1.18
SimCLR	0.95	1.06	0.98	1.01	0.94	1.07	1.12	1.05	0.99	1.14
SSMCLAA	0.74	0.62	0.67	0.58	0.69	0.64	0.51	0.68	0.73	0.66

In Table I, the maximum and MainVs of MAE for the SSMCLAA model of the research design are 0.74 and 0.58, respectively. As a whole, the MAE values of the SSMCLAA model are significantly lower than those of the comparison models. The MaxVs of MAE for SemiTime model, SelfHAR model and SimCLR model are 1.37, 1.13 and 1.06, respectively, and the MainVs are 1.19, 0.97 and 0.94, respectively. In addition, the MainV of MSE occurs on the SSMCLAA model with a value of 0.51. The MaxV occurs on the SemiTime model with a value of 1.29. The mean values of MSE for the SemiTime model, SelfHAR model, SimCLR model and SSMCLAA model are 1.372, 1.178, 1.074 and 0.644, respectively. In summary, the values of MAE and MSE for the research-designed SSMCLAA model are significantly smaller than those of the comparison models, which suggests that the SSMCLAA model has a smaller classification error. This suggests that the SSMCLAA model can determine whether or not students have finished the Sanda training with a lower classification error. The time-consuming comparison of different models is shown in Fig. 9.

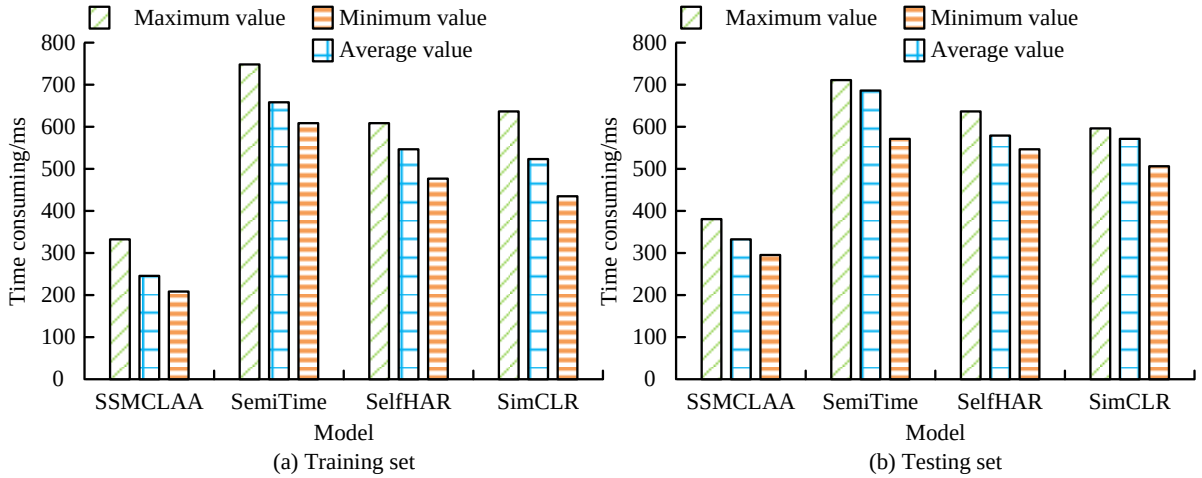


Fig. 9. Comparison of time consumption of different models.

In Fig. 9(a), on the training set, the time-consuming mean of the research design SSMCLAA model is 242 ms, and the time-consuming mean of the comparison models SemiTime, SelfHAR, and SimCLR are 670 ms, 543 ms, and 521 ms, respectively. The time-consuming mean of the SSMCLAA model is less than that of the comparison models, which are 328 ms, 201 ms, and 179 ms. The SSMCLAA model has a greater advantage in time. In Fig. 9(b), on the test set, the time-consuming AV of the SSMCLAA model is 337 ms, which is significantly lower than that of the comparison model. In conclusion, the research designed SSMCLAA model has less time consuming and can make a judgment on whether the students complete the Sanda training in a shorter period of time and reduce the waiting time of the students.

B. SSMTFCL Model Performance Validation

The dataset used in the study, the method the dataset is partitioned, and the experimental setup are all consistent with the SSMCLAA model performance validation, which is necessary to validate the SSMTFCL model's performance. The values of θ and φ in the LF of the SSMTFCL model are both 1.0, and the iterations of the model is 300. In addition, four classical semi-supervised algorithm models are selected for comparative validation of the research models, the double-Πmodel (Π-model), the MeanTeacher model, the multi task learning model (MTL), and the MixMatch model that combines Mixup and Fixmatch. Comparison of LF curves for different models is shown in Fig. 10.

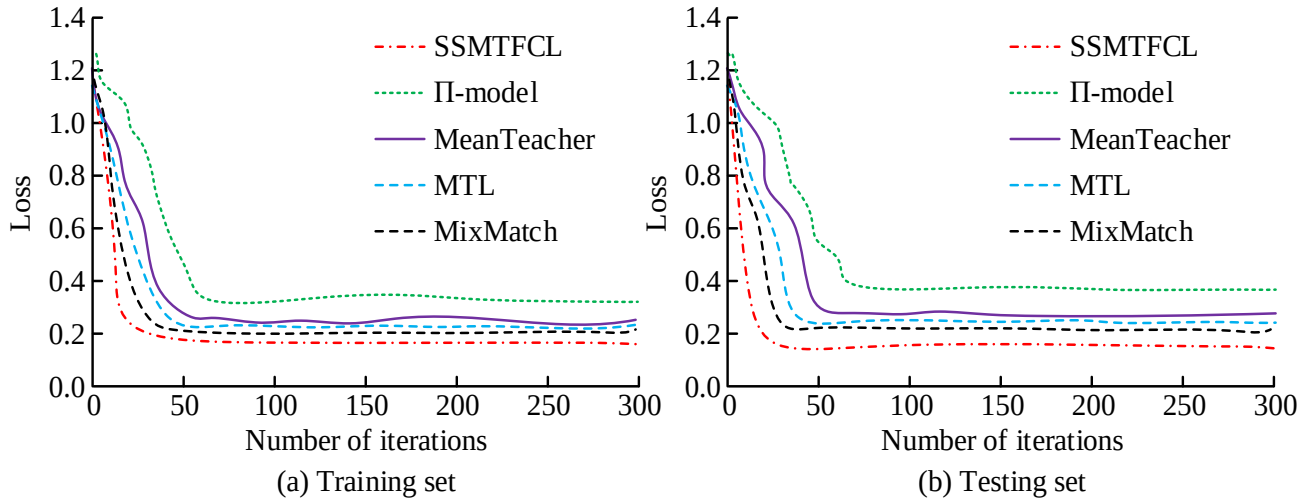
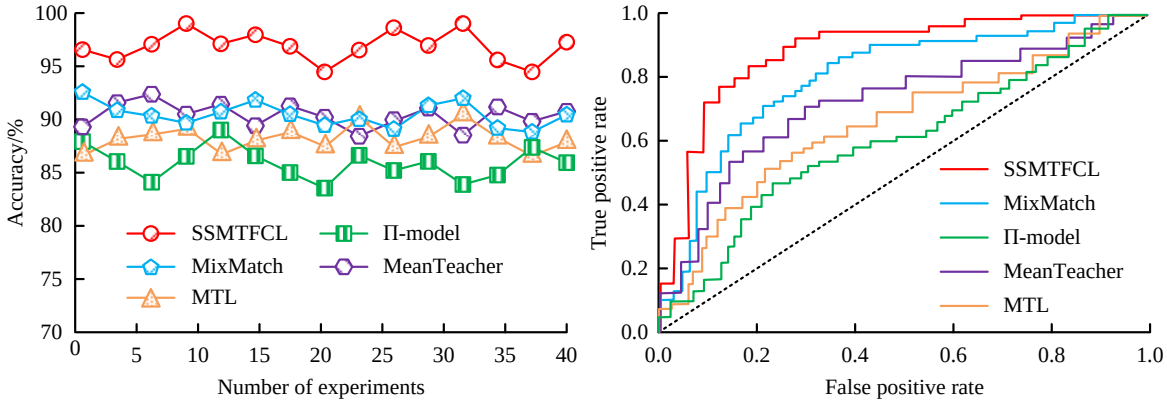


Fig. 10. Comparison of LF curves for different models.

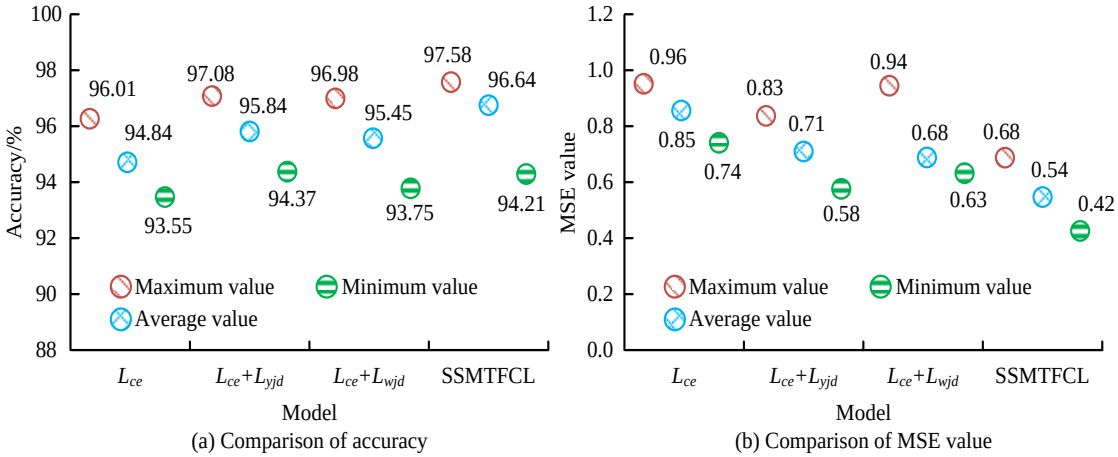
In Fig. 10(a), on the training set, the research design SSMTFCL model plateaus after almost 22 iterations, while the Π-model, MeanTeacher, MTL, and MixMatch models plateau after 57, 52, 48, and 43 iterations, respectively. The MainVs of loss values for the five models are 0.17, 0.33, 0.25, 0.23, and 0.21, respectively. In Fig. 10(b), on the test set, the MainV of loss values for the SSMTFCL model is 0.15, and the MainVs of loss values for the comparison models Π-model, MeanTeacher, MTL, and MixMatch are 0.36, 0.27, 0.24, and 0.22, respectively. The five models leveled off after nearly 24, 59, 54, 47, and 45 iterations, respectively. In summary, the research designed SSMTFCL model converges faster and has better performance with smaller loss values. A comparison of the accuracy and receiver operation characteristic, (ROC) curves for the different models is shown in Fig. 11.



(a) Comparison of accuracy of different models (b) Comparison of ROC curves of different models

Fig. 11. Comparison of accuracy and ROC curves of different models.

In Fig. 11(a), the MaxV of accuracy of the research design SSMTFCL model is 97.58%, the MainV is 94.21%, and the AV is 96.64%. The mean values of accuracy for the comparison models Π-model, MeanTeacher, MTL and MixMatch are 86.17%, 89.42%, 87.55% and 90.73% respectively. The accuracy of SSMTFCL model is significantly higher than the comparison model. In Fig. 11(b), the area under the ROC curve of the research design SSMTFCL model is the largest with a value of 0.945. It is followed by the MixMatch and MeanTeacher models with values of 0.899 and 0.857, respectively. The smallest are the MTL and Π-model models with values of 0.821 and 0.798, respectively. In summary, the research design SSMTFCL model has higher accuracy and better performance. To verify the effectiveness of the three tasks in the SSMTFCL model, the study conducted ablation experiments. L_{ce} is the LF for the CT, and L_{yjd} and L_{wjd} are the LFs for the supervised comparison learning task and the unsupervised comparison learning task, respectively. Fig. 12 displays the ablation experiment's outcomes.



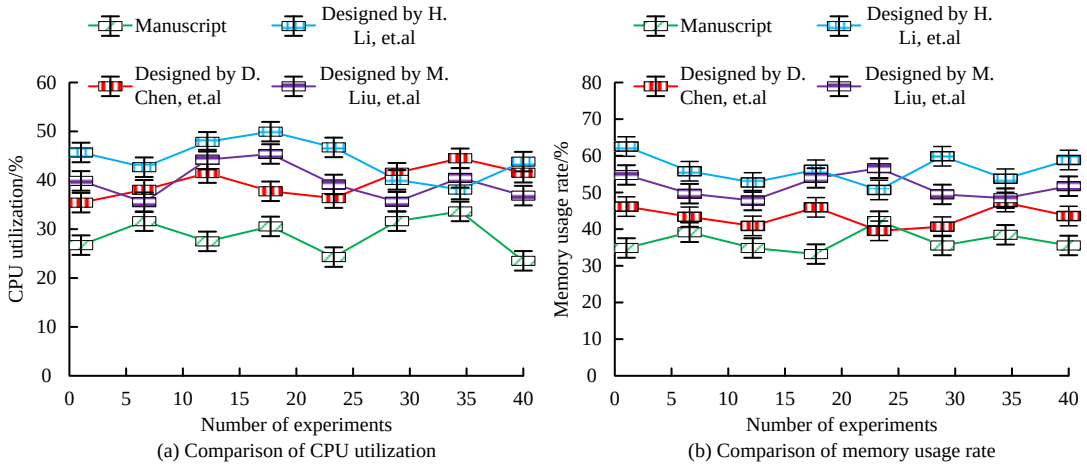
(a) Comparison of accuracy (b) Comparison of MSE value

Fig. 12. Results of SSMTFCL model ablation experiment.

In Fig. 12(a), the accuracy mean of the model using only the CT is 94.84%, which is 1.8% lower than the accuracy mean of the full SSMTFCL model of 96.64%. The average accuracy of the model using the CT and the supervised comparative learning task is 95.84%, while the average accuracy of the model using the CT and the unsupervised comparative learning task is 95.45%. In Fig. 12(b), the MSE mean values of the model using only the CT, the model using the CT and the supervised contrast learning task, and the model using the CT and the unsupervised contrast learning task are 0.85, 0.71, and 0.68, respectively. The MSE mean value of the full SSMTFCL model is 0.54. In conclusion, the multi-task learning framework used in the study is effective.

C. SATS Performance Validation

To validate the performance of the research design SATS based on 5G and VR technologies, the study selected other TSs for comparison. Among them, there are psychological virtual simulation experiment TS designed by experts such as D. Chen, PET aid network system designed by scholars such as Li, and piano playing TS designed by researchers such as Liu [29-31]. In addition, the experimental environment configurations used for system performance validation are consistent with those used for SSMCLAA model performance validation. A comparison of the central processing unit (CPU) utilization and memory occupancy of the different systems is shown in Fig. 13.



(a) Comparison of CPU utilization (b) Comparison of memory usage rate

Fig. 13. Comparison of CPU utilization and memory usage rate in different systems.

In Fig. 13(a), the average CPU utilization of the research design SATS is 27.8%, and the average CPU utilization of the psychological virtual simulation experiment TS, the PET aid network system, and the piano playing TS are 37.8%, 42.8%, and 39.7%, respectively. The CPU utilization of the research design SATS is significantly smaller than the comparison system. In Fig. 13(b), in terms of memory occupancy, the AV of memory occupancy of the research design system is 38%, and the AVs of memory occupancy of the other three comparison models are 43%, 57%, and 52%, which are 5%, 19%, and 14% higher than those of the research design system, respectively. In conclusion, the research design SATS has better performance. The response time and throughput comparisons of the different systems are shown in Table II.

TABLE II. COMPARISON OF RESPONSE TIME AND THROUGHPUT OF DIFFERENT SYSTEMS

System	Response Time/ms					Throughput/(bit/s)				
	Number of experiments					*Number of experiments*				
	1	*2*	*3*	*4*	*5*	*1*	*2*	*3*	*4*	*5*
Designed by D. Chen, et.al	136	143	132	158	132	63142	64498	65506	68504	63187
Designed by H. Li, et.al	145	162	157	140	142	60988	60543	61279	63885	64771
Designed by M. Liu, et.al	121	137	128	136	134	64312	67329	68634	66032	61062
Manuscript	63	57	61	55	59	74521	76282	75113	76895	77651

In Table II, in terms of response time, the AV of the research design system is 59 ms, and the AVs of the psychological virtual simulation experiment TS, the PET aid network system, and the piano playing TS are 140.2 ms, 149.2 ms, and 131.2 ms, respectively. The response time of the research design system is significantly lower than the comparison system. In addition, the throughput of the research design system can reach a maximum of 77651 bit/s and a minimum of 74521 bit/s. The MaxVs of the throughput of the psychological virtual simulation experiment TS, the PET aid network system, and the piano playing TS are 68504 bit/s, 64771 bit/s, and 68634 bit/s, respectively, and the MainVs are 63142 bit/s, 60543 bit/s and 61062 bit/s. In conclusion, the research design SATS has shorter response time, larger system throughput and better performance.

IV. DISCUSSION

In response to the shortcomings of offline Sanda teaching, this paper designs the SSMCLAA model and SSMTFCL model, and based on this, constructs a Sanda auxiliary teaching system that integrates 5G and VR technology. The results showed that the average memory usage of the Sanda assisted teaching system designed in the paper was 38%, which was 5%, 19%, and 14% lower than the average memory usage of the other three compared systems, respectively. The average response time is 59ms, which is 81.2ms, 90.2ms, and 72.2ms lower than the average response time of the other three compared systems, respectively. The paper design of a Sanda auxiliary teaching system can improve the effectiveness of Sanda teaching. Liu X et al. designed a decision support system to evaluate the functionality of 5G networks and artificial intelligence in higher education context teaching research in order to achieve teaching objectives. The results show that 5G networks and artificial intelligence algorithms can enhance the effectiveness of situational teaching in higher education [32]. This result is similar to the research findings.

V. CONCLUSION

To make up for the shortcomings of OST, the study constructs SATS based on 5G and VR technologies, and constructs SSMCLAA model and SSMTFCL model. The results revealed that the maximum accuracy of SSMCLAA model was 95.76% on the training set. The MaxVs of accuracy of the comparison models SemiTime, SelfHAR and SimCLR were 80.50%, 87.42% and 93.03%, which were 15.26%, 8.34% and 2.73% lower than the MaxVs of accuracy

of the SSMCLAA model, respectively. The MainVs of MAE and MSE for the SSMCLAA model were 0.58 and 0.51, respectively, and the mean values of time consumed for the training and test sets were 242ms and 337ms, respectively. The SSMCLAA model performed relatively well. On the training set, the SSMTFCL, Π-model, MeanTeacher, MTL, and MixMatch models plateaued after almost 22, 57, 52, 48, and 43 iterations, respectively, with minimum loss values of 0.17, 0.33, 0.25, 0.23, and 0.21, respectively. The AVs of accuracy for the five models were 96.64%, 86.17%, 89.42%, 87.55% and 90.73%. The SSMTFCL model performs better. The AV of CPU utilization for the research design SATS was 27.8%, the AV of memory occupancy was 38%, the AV of response time was 59 ms, and the maximum and MainVs of throughput were 77,651 bit/s and 74,521 bit/s, respectively.

The research design SATS has good performance. The study also has some shortcomings. One, the test of SATS on students in terms of Sanda is limited to the technical level, and future research can include a theoretical question-answering session to enhance students' theoretical knowledge base and further avoid students' injuries. Secondly, the helmet used in the study and other types of sensing devices can cause discomfort to some students after prolonged use, and future research can do in-depth exploration on the comfort of wearable devices. Thirdly, the Sanda auxiliary teaching system still lacks human-machine training, game mode, and competition mode. Future research can compensate for this module to further enrich the Sanda auxiliary teaching system. Fourthly, for the data collected by different data collection devices, future research can simply annotate their location information to improve the recognition efficiency of students' Sanda movements. Fifthly, some students are prone to multiple compound movements due to non-standard movements during Sanda practice, which increases the difficulty of judging Sanda movements. Future research can further improve and optimize the model for recognizing composite actions.

REFERENCES

[1] S. Feng, “Experimental study on attention blink of male sanda athletes of different sports levels under fatigue state,” J. Educ. Educ. Res., vol. 7, pp. 305-309, March 2024.

[2] H. Zhao, “Design of sanda action reconstruction model based on 3D images,” Wirel. Commun. Mob. Com., vol. 2021, pp. 1-6, January 2021.

[3] J. Xu and S. Ariyasajsiskul, “Sports injury of sanda wushu class of students in Xi'an city,” Int. J. Sociol. Anthropol. Sci. Rev., vol. 4, pp. 465-476, January 2024.

[4] Y. Asham, M. H. Bakr, and A. Emadi, “Applications of Augmented and Virtual Reality in Electrical Engineering Education: A Review,” IEEE. ACCESS., vol. 11, pp. 134717-134738, November 2023.

[5] Y. Feng, C. You, Y. Li, Y. Zhang, and Q. Wang, “Integration of computer virtual reality technology to college physical education,” J. Web Eng., vol. 21, pp. 2049-2071, December 2022.

[6] J. Gao, L. Chong, X. Qiao, and F. Tian, “Telemedicine virtual reality based skin image in children's dermatology medical system,” Comput. Intell., vol. 38, pp. 229-248, June 2021.

[7] M. H. Elgewely, W. Nadim, A. Elkassed, M. Yehiah, and S. Abdennadher, “Immersive construction detailing education: building information modeling (BIM)-based virtual reality (VR),” Open. House. Int., vol. 46, pp. 359-375, July 2021.

[8] O. Almousa, R. Zhang, M. Dimma, J. Yao, A. Allen, L. Chen, P. Heidari, and K. Qayumi, “Virtual reality technology and remote digital application for tele-simulation and global medical education: an innovative hybrid system for clinical training,” Simulat. Gaming., vol. 52, pp. 614-634, May 2021.

[9] P. P. Groumpos, “A critical historic overview of artificial intelligence: issues, challenges, opportunities, and threats,” AIA., vol. 1, pp. 197-213, July 2023.

[10] M. He, G. Song, and Z. Wei, “Human behavior feature representation and recognition based on depth video,” J. Web. Eng., vol. 19, pp. 883-902, December 2020.

[11] L. I. Yangzhi, J. Yuan, and H. Liu, “Human skeleton-based action recognition algorithm based on spatio-temporal attention graph convolutional network model,” J. Comput. Appl., vol. 41, pp. 1915-1921, July 2021.

[12] P. Gao, D. Zhao, and X. Chen, “Multi-dimensional data modelling of video image action recognition and motion capture in deep learning framework,” IET. Image. Process., vol. 14, pp. 1257-1264, April 2020.

[13] P. Chen, H. Wang, H. Yan, J. Du, Y. Ning, and J. Wei, “sEMG-based upper limb motion recognition using improved sparrow search algorithm,” Appl. Intell., vol. 53, pp. 7677-7696, July 2023.

[14] Q. Ye, H. Zhong, C. Qu, and Y. Zhang, “Human interaction recognition method based on parallel multi-feature fusion network,” Intell. Data. Anal., vol. 25, pp. 809-823, July 2021.

[15] M. Xu, “Application of human-computer interaction virtual reality technology to the design of ice and snow landscapes,” Int. J. Hum. Robot., vol. 19, pp. 74-88, April 2022.

[16] S. Makhsuci, S. M. Navidi, M. Sanduleanu, and M. Ismail. “A review of doherty power amplifier and load modulated balanced amplifier for 5G technology,” Int. J. Circ. Theor. App., vol. 51, pp. 2422-2445, July 2023.

[17] A. D. Desai, B. M. Ozturkler, C. M. Sandino, R. Boutin, M. Willis, S. Vasanawala, B. A. Hargreaves, C. Ré, J. M. Pauly, and A. S. Chaudhari, “Noise2Recon: enabling SNR-robust MRI reconstruction with semi-supervised and self-supervised learning,” Magn. Reson. Med., vol. 90, pp. 2052-2070, July 2023.

[18] I. Okpala, C. Nnaji, and I. Awolusi, “Wearable sensing devices acceptance behavior in construction safety and health: assessing existing models and developing a hybrid conceptual model,” Constr. Innov.: Inf. Process. Manage., vol. 22, pp. 57-75, January 2022.

[19] A. Tokhmetova and A. Y. Albagachiev, “Comparative analysis of a numerical method and machine learning methods of temperature determination of a doped lubricating layer with experimental data,” J. Mach. Manuf. Reliab., vol. 52, pp. 509-515, September 2023.

[20] P. Rathnasabapathy and D. Palanisami, “A theoretical development of improved cosine similarity measure for interval valued intuitionistic fuzzy sets and its applications,” J. Amb. Intel. Hum. Comp., vol. 14, pp. 16575-16587, June 2023.

[21] K. Li, B. Wang, Y. Tian, and Z. Qi, “Fast and accurate road crack detection based on adaptive cost-sensitive loss function,” IEEE. T. Cybernetics., vol. 53, pp. 1051-1062, February 2023.

[22] A. Natarajan, V. Krishnasamy, and M. Singh, “Device-free human motion detection using single link wiFi channel measurements for building energy management,” IEEE. Embed. Syst. Lett., vol. 15, pp. 153-156, September 2023.

[23] Q. Wu, Q. Huang, and X. Li, “Multimodal human action recognition based on spatio-temporal action representation recognition model,” Multimed. Tools. Appl., vol. 82, pp. 16409-16430, November 2023.

[24] D. Noh, H. Yoon, and D. Lee, “A Decade of Progress in Human Motion Recognition: A Comprehensive Survey From 2010 to 2020,” IEEE. ACCESS., vol. 12, pp. 5684-5707, January 2024.

[25] M. J. Grotevent, S. Yakunin, D. Bachmann, C. Romero, J. R. V. D. Aldana, and M. Madi, et al. “Integrated photodetectors for compact Fourier-transform waveguide spectrometers,” Nat. Photonics., vol. 17, pp. 59-64, October 2023.

[26] A. Khalili Golmankhaneh, K. K. Ali, R. Yilmazer, and M. K. A. Kaabar, “Local fractal fourier transform and applications,” Comput. Methods. Diffe., vol. 10, pp. 595-607, July 2022.

[27] K. Skrai, J. Petrovi, and P. Pale, “Classification of low-and high-entropy file fragments using randomness measures and discrete fourier transform coefficients,” Vietnam. J. Comput. Sci., vol. 10, pp. 433-462, July 2023.

[28] Y. Kaya and E. K. Topuz, "Human activity recognition from multiple sensors data using deep CNNs," MULTIMED. TOOLS. APPL., vol. 83, pp. 10815-10838, June 2024.

[29] D. Chen, X. Kong, and Q. Wei, "Design and development of psychological virtual simulation experiment teaching system," Comput. Appl. Eng. Educ., vol. 29, pp. 481-490, July 2021.

[30] H. Li, H. Zhang, and Y. Zhao, "Design of computer-aided teaching network management system for college physical education," Comput. Aided. Des. Appl., vol. 18, pp. 152-162, February 2021.

[31] M. Liu and J. Huang, "Piano playing teaching system based on artificial intelligence-design and research," J. Intell. Fuzzy Syst.: Appl. Eng. Technol., vol. 40, pp. 3525-3533, February 2021.

[32] X. Liu, M. Faisal, and A. Alharbi, "A decision support system for assessing the role of the 5G network and AI in situational teaching research in higher education," SOFT. COMPUT., vol. 26, pp. 10741-10752, October 2022.

Data Collection Method Based on Data Perception and Positioning Technology in the Context of Artificial Intelligence and the Internet of Things

Xinbo Zhao[1]*, Fei Fei[2]
School of Management, Liaoning University of International Business and Economics, Dalian, 116052, China[1]
School of Art & Design, Dalian Polytechnic University, Dalian, 116034, China[2]

***Abstract*—Wireless sensor networks are an important technical form of the underlying network of the Internet of Things. The energy of each node in the network is finite. When a node runs out of energy, it can cause network interruptions, which can affect the reliability of data collection. To reduce the consumption of communication resources and ensure the reliability of data collection, the study proposes data collection based on data compression perception positioning technology. This method first uses a Bayesian compression perception method to select nodes, and then adopts an adaptive sparse strategy to collect data. When selecting nodes using this proposed method, wireless sensor networks had the longest network lifespan. In the case of different degrees of redundancy and sparsity, the research method had the lowest reconstruction error, with reconstruction errors of 0.31 and 0.40, respectively. When the balance factor was set to 0.6, the reconstruction error of the research method was the lowest, with a minimum reconstruction error of 0.05. This proposed method has better reconstruction performance, effectively prolongs the lifespan of wireless sensor networks, and reduces the consumption of communication resources.**

Keywords*—Wireless sensor network; data collection; compression perception technology; Sparse Bayesian Learning; signal reconstruction

I. Introduction

The development and application of the Internet of Things (IoT) have brought great convenience to people's lives. IoT can connect any object in the physical world with the Internet to achieve real-time monitoring, intelligent control, remote operation and other functions, so as to achieve interconnection between things and people [1-2]. Wireless Sensor Network (WSN) is a key technological component of IoT. Since its inception, WSN has been a hot topic in information research and has been applied to various aspects of society, such as military, agriculture, industry, healthcare, intelligent transportation, and home furnishings [3-4]. However, the energy of nodes in WSN is limited. When energy is depleted, the network will be interrupted, affecting the lifespan of WSN [5]. In addition, there are still issues with anomalies and missing data collected by current wireless sensor network data collection methods, which can result in significant consumption of communication resources. Studying data collection methods can improve the performance of wireless sensor networks, making them more widely applicable in fields such as healthcare, military, and environmental monitoring. Then, compression perception perfectly reconstructs the signal through nonlinear reconstruction algorithms [6]. Therefore, to reduce the consumption of communication resources and ensure the reliability of data collection, the study proposes data collection based on data compression perception positioning technology.

When using data compression perception positioning technology for data collection, the innovative approach is to first use Bayesian compression perception for node selection. Furthermore, an adaptive sparse strategy was adopted for data collection in the experiment. The main contribution of this study is the proposal of data collection based on data compression perception positioning technology, which reduces the energy consumption of WSN and improves the reliability of data collection.

The study will investigate data collection methods based on data compression perception positioning technology from four aspects. Firstly, a review is conducted on the current research status of data collection methods based on data perception and localization technology in the context of artificial intelligence IoT. Next is the research on data collection methods based on data compression perception positioning technology. Then, experimental verification is conducted on the proposed method. Finally, a summary of the research content is provided.

II. Related Works

With the continuous development of IoT, the network scale is getting larger and the network environment is becoming more and more complex. This poses significant challenges to energy conservation, transmission efficiency, security, and other aspects of network communication. Compression perception helps to address these issues in intelligent network communication [7-8]. Wang Y et al. proposed a lightweight method based on deep learning to reduce the computational cost of traditional deep learning methods. The sparsity of the scaling factor was enforced through compression perception. This proposed method effectively reduced the model size and accelerated the calculation speed [9]. Cheng G et al. proposed a hyper chaotic image encryption scheme based on quantum genetic algorithm and compression perception. This eliminated the drawbacks of weak key flow, small key space, and small information entropy in chaotic image encryption schemes. This proposed method was more efficient in resisting statistical attacks and plaintext attacks [10]. Liang P et al. proposed a compression perception technique to address the overwhelming

*Corresponding Author

feedback overhead caused by a large number of antennas in base stations. Secondly, a deep learning-based signal recovery solver framework was used on the base station. This proposed method was superior to existing methods and reduced feedback overhead [11]. The remote sensing image size was relatively large in the new type of hyper chaotic system. Nan S et al. proposed a block compression encryption algorithm that combined a novel hyper chaotic system with block compression perception. This proposed algorithm had good encryption performance, reconstruction accuracy, and anti-attack ability [12]. Xu G et al. proposed a sparse synthetic aperture radar imaging technique based on compression perception and machine learning. This solved the limitation of data bandwidth on the resolution of synthetic aperture radar images. This proposed method further investigated sparse imaging in machine learning and improved image resolution [13].

Data collection is a fundamental and important operation in WSN applications [14]. Aziz A et al. proposed an efficient aggregation scheme for multi-hop clusters based on hybrid compression perception, which effectively combined compression perception and routing protocols to collect data. This reduced the energy consumption of sensor nodes and extended the WSN lifespan. This proposed method was more efficient in collecting data and prolonged the WSN lifespan [15]. Sekar K et al. proposed a data collection method based on compression perception. Random singular value decomposition was used in this experiment to achieve compressed matrix factorization. This reduced the amount of data collected and lowers communication costs. This proposed method had higher accuracy at lower sampling rates, reduced data transmission costs, and extended the lifespan of sensors [16]. Mei Y et al. proposed a sampling algorithm based on compression perception to sample conditioned signal data. This changed the design paradigm of low latency large-scale access requiring random access schemes. This proposed algorithm reduced data sample loss when data samples were transmitted wirelessly and could be used for wind turbine fault detection [17]. Lin C et al. proposed a hierarchical data collection scheme to improve the linear programming effect in agricultural detection and reduce energy consumption in data collection. Data sampling was carried out using precise and greedy methods using mixed compression. This proposed method effectively collected data and planed the path of drones with lower energy consumption [18]. Chang CY et al. proposed a mobile receiver for multi-rate data acquisition to improve receiver speed and obtain fresh data. This proposed method significantly shortened the path length of mobile receivers and collected data comprehensively and efficiently [19].

In summary, with the continuous development of the Internet of Things, wireless sensor networks are being applied in more fields. In the process of collecting data, many scholars and scientists have designed a large number of improved models based on compressed sensing to solve the problem of complex network environments and noise interference leading to shortened lifespan of wireless sensor networks. These models effectively reduce the amount of data collected, but ignore the impact of sensor nodes themselves on wireless sensor networks. Therefore, the study proposes data collection and sensor node selection based on compressive sensing positioning technology.

III. Data Collection Method based on Data Compression Perception Technology

A. Data Collection Method Based on Adaptive Sparse Strategy

Bayesian compression perception is a method that combines Bayesian theory and compression perception to solve signal reconstruction problems. Considering the prior distribution of the signal, Bayesian compression perception can more accurately estimate the original signal and adaptively match the signal sparsity. The main advantage is that it can better handle noise and uncertainty [20-21]. IoT and artificial intelligence technology are constantly evolving. Artificial intelligence IoT has been applied in smart homes, intelligent transportation, intelligent healthcare, and other fields, bringing great convenience to humanity. As the sensing terminal of IoT, WSN mainly consists of three parts, including nodes, sensor networks, and users. Each node in the network needs to reserve energy for long-term use. Energy depletion can cause network interruptions [22-23]. In addition, sensor networks suffer from noise interference, data anomalies, and missing issues, which affect the reliability of data collection. To reduce the consumption of communication resources and ensure the reliability of data collection, the study adopts an adaptive sparse strategy for data collection. The position of the model sensor nodes is fixed, the composition structure is the same, and the energy is sufficient. Fig. 1 shows the WSN structure.

Sparse Bayesian learning is a method for reconstructing sparse signals, which does not require setting regularization parameters. It first considers the observation of a vector and outputs the sample y affected by noise interference, expressed as Eq. (1).

$$y = \Phi\omega + \varepsilon \tag{1}$$

In Eq. (1), ε refers to the observation noise. $\Phi\omega$ refers to the mean. Φ refers to the observation matrix. ω refers to the weights used by model learning to construct each column in Φ. The purpose of Sparse Bayes is to improve the model's generalization ability, interpretability, and efficiency by introducing sparsity, that is, selecting only a few important features. The Gaussian likelihood function model is represented by Eq. (2).

$$p\left(y \mid \omega, \sigma^2\right) = (2\pi\sigma^2)^{-N/2} \exp\left\{-\frac{1}{2\sigma^2}\left\|y - \Phi\omega\right\|_2^2\right\} \tag{2}$$

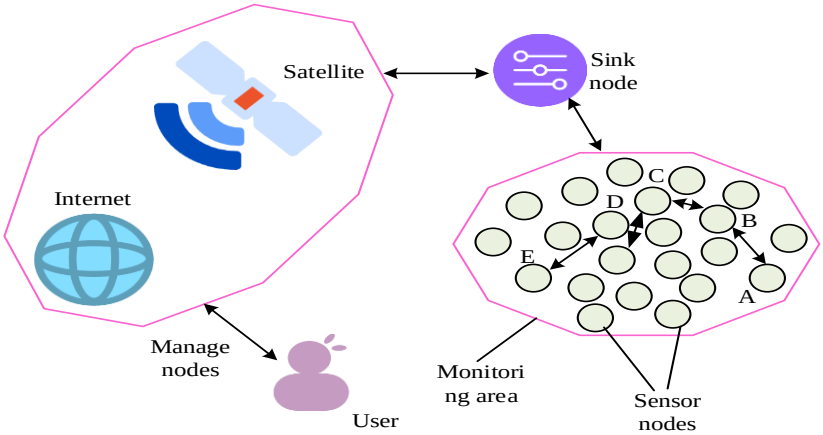


Fig. 1. Wireless sensor network architecture diagram.

In Eq. (2), y refers to the target variable, σ^2 refers to the variance, and N refers to the matrix of the sample. In this case, the task of obtaining maximum likelihood estimation is equivalent to finding the minimum norm solution while obtaining non-sparse solutions. Therefore, to find sparse solutions, the sparse Bayesian learning model estimates parameterized prior weights from the data, represented by Eq. (3).

$$p(\omega;\gamma)=\prod_{i=1}^{M}(2\pi\gamma_i)^{-\frac{1}{2}}\exp\left(-\frac{\omega i^2}{2\pi\gamma_i}\right) \quad (3)$$

In Eq. (3), γ represents M hyperparameter vectors. Secondly, Bayesian inference is performed. A posterior distribution is applied to all unknown variables to obtain the mean and covariance of the weight parameters. Finally, the hyperparameters are updated and estimated using Eq. (4).

$$\left(\sigma^2\right)^{new}=\frac{\left\|y-\Phi u\right\|_2^2}{N-\sum_i\gamma_i} \quad (4)$$

In Eq. (4), u represents the posterior mean. In order to solve the problem of excessive energy consumption caused by ineffective estimation in noisy environments, a compressed sensing adaptive sparse strategy was adopted to collect data. The steps are as follows: first, the sampling rate for data collection is determined to generate sampling random weights. Next, determine whether the data collection is initiated by a node. If the data collection is initiated by a node, then collect perception data and use the compressed perception adaptive sparse algorithm in the data sorting center to compress the original data. The data collection is complete. On the contrary, if there is no initiating node for data collection, it is determined whether the child node data has been received. If the child node data is received, the perception data is collected, and the compressed perception adaptive sparse algorithm is used in the data sorting center to compress the original data, and the data collection is completed. If no child node data is received, data collection ends. The data collection process is shown in Fig. 2.

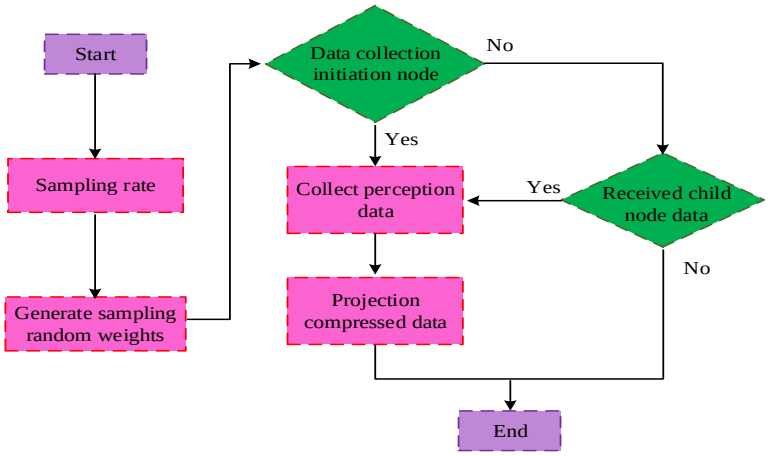


Fig. 2. Data collection flowchart.

The study adopts an adaptive sparse strategy based on compression perception for data collection. The data collection model is represented by Eq. (5).

$$f = BRx + v \tag{5}$$

In Eq. (5), f represents the observed signal. B refers to a random beta effort matrix. R is a sparse routing matrix. v represents the noise interference of the model. In data collection, there may be insufficient sparsity of the original signal. To obtain suitable sparse bases, the perception data are learned. The optimization process is represented by Eq. (6).

$$\min_{\Psi} \|X - \Psi S\|_F^2 \; s.t \forall i, \|S_i\| \le K \tag{6}$$

In Eq. (6), Ψ refers to the sparse basis in the model that needs to be optimized. K refers to the sparsity of optimizing sparse bases. S refers to a sparse vector matrix. $\|\bullet\|_F$ refers to the Frobenius norm of a matrix. If only one sparse basis is optimized, the constraint isometry condition needs to be satisfied. If not, multiple measurements need to be taken. Therefore, low coherence is used to compensate for the deficiency. Then, optimization and update are carried out through covariance, represented by Eq. (7).

$$\min_{\Psi} \left\| \left(\Psi^T \Phi^T V \Phi \Psi + A \right) \right\|_F^2 + \eta \left\| \Psi^T \Phi^T V \Phi \Psi - I \right\|_F^2 \tag{7}$$

In Eq. (7), η refers to the equilibrium factor in the model. A refers to the variance matrix of sparse signals. V refers to the inverse matrix of noise variance. $\left\| \Psi^T \Phi^T V \Phi \Psi - I \right\|_F^2$ refers to meeting low coherence. In a successfully deployed WSN, there is a significant amount of redundancy in the perception data collected from sensor nodes due to temporal and spatial correlations. Compression perception is a commonly used data collection technique. Therefore, the study adopts compression perception for data collection. The compression perception performance will improve with the increase of sample size, but the sparsity of sample data is uncertain. Fig. 3 shows a designed data collection framework to continuously update node sparsity.

B. Node Selection Based on Bayesian Compression Perception

The study adopts a method based on Bayesian compression perception adaptive sparsity to collect data. This can reduce the consumption of communication resources and ensure the reliability of data collection. However, this method ignores the performance differences of the nodes themselves. The WSN nodes produced by each company have different characteristics depending on the selected core processor, radio frequency communication chip, and expansion function. The sensor node mainly consists of four parts: sensing unit, processing unit, wireless transceiver unit, and power supply unit [24]. The differences in nodes can lead to a shortened network lifecycle. In this regard, the study adopts a method based on Bayesian compression perception to select nodes. Fig. 4 shows the composition of sensor nodes.

The research on data collection in WSN has been ongoing for many years and has achieved significant research results. However, sensor nodes still face issues such as imbalanced energy consumption, ineffective correlation measurement between nodes, and inability to perform effective spatial clustering. Therefore, when selecting sensor nodes, the study simplifies WSN. Because each sensor node is different, the sink node is affected by noise during the data collection process. Therefore, the model update is represented by Eq. (8).

$$g = \Phi x + a \tag{8}$$

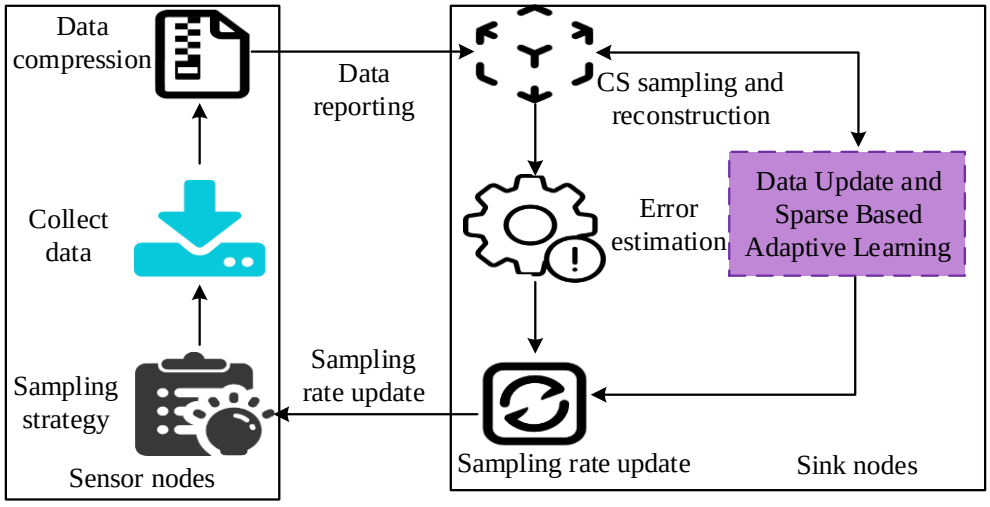


Fig. 3. Data collection framework diagram.

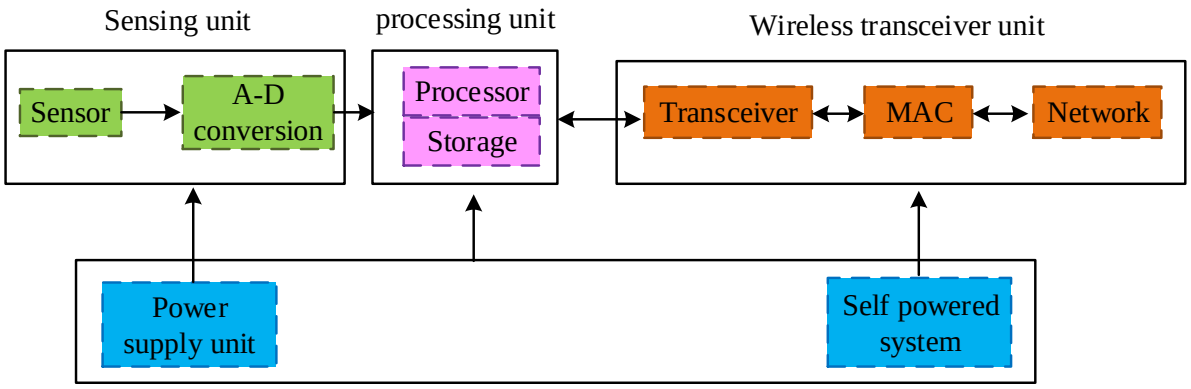


Fig. 4. Composition structure of sensor nodes.

In Eq. (8), a represents the noise vector measured by the model, x represents the data collected by each node, and g represents the measurement data. The model collects raw data without sparsity. Therefore, by utilizing the spatiotemporal correlation of data to transform the original data, sparse data can be obtained. The sparsity of the original data is represented by equation (9).

$$x = \Psi s = \sum_{i=1}^{N} \psi_i \cdot s_i \tag{9}$$

In Eq. (9), s refers to a sparse signal. x refers to raw data. Ψ refers to the sparse basis in the model that needs to be optimized. To better reconstruct the obtained data, assuming that the information of sparse basis and measurement matrix is known, Bayesian algorithm is used to reconstruct the measurement data. The posterior expectation of sparse signal is represented by equation (10).

$$\mu = \sum \Theta^T V^{-1} y \tag{10}$$

In Eq. (10), μ refers to the posterior expectation of the sparse signal. V refers to the inverse matrix of noise variance. Θ refers to the perception matrix. The posterior covariance of sparse signals is represented by Eq. (11).

$$\sum = \left(\Theta^T V^{-1} \Theta + A\right)^{-1} \tag{11}$$

In Eq. (11), $\sum$ refers to the posterior covariance of the sparse signal. A refers to the variance matrix of sparse signals. Θ refers to a perception matrix. In the framework of compression perception, selecting appropriate active sensor nodes is crucial for ensuring reconstruction performance when the data collection quality of each node is different. Bayesian compression perception utilizes prior knowledge to improve reconstruction performance, especially when there are differences in node quality. The node selection steps are as follows: first, all sensor nodes in the wireless sensor network obtain data sources and transmit them to the sink node through node selection. The sampled data is compressed, perceived, and reconstructed at the sink node. Then, the node selection is optimized based on the reconstruction information. Finally, the sink node distributes the optimized node selection information to various sensor nodes in the network through control signals, iteratively until the conditions are met. The architecture diagram for node selection is shown in Fig. 5.

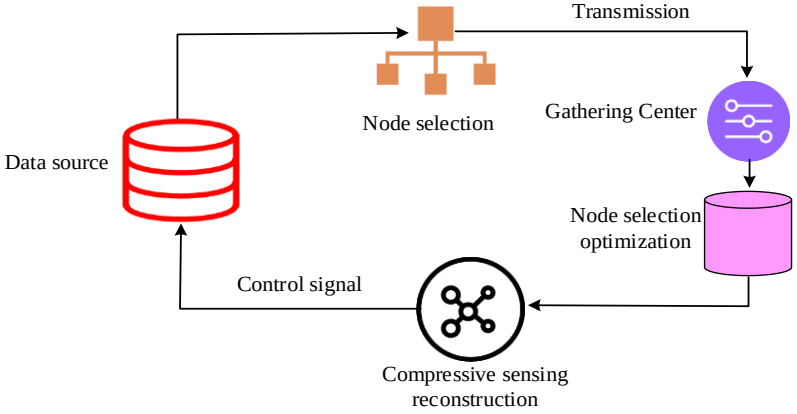


Fig. 5. Node selection framework diagram.

Assuming that the sink node provides complete information for hyperparameters and rewrites the mean and covariance, the observation matrix is an important parameter that affects the reconstruction effect. The reconstruction effect can be controlled by selecting different nodes. Mean square error is a commonly used indicator in machine learning to evaluate the model quality. It reflects the predicted and true values' error. This study uses mean square error to determine the reconstruction error, which is represented by Eq. (12).

$$MSE = \frac{\left\| \hat{x} - x \right\|_2^2}{\left\| x \right\|_2^2} \tag{12}$$

In Eq. (12), x represents the raw data. $\hat{x}$ represents the reconstructed original signal. The WSN lifecycle generally refers to the energy consumption of a node or a specific area of nodes. In order to extend the lifecycle of WSN, it is necessary to reduce the computational complexity. The average network lifetime is represented by Eq. (13).

$$E(L) = \frac{\xi_0 - E(E_u)}{P_c + \rho E(E_t)} \tag{13}$$

In Eq. (13), P_c refers to constant continuous power consumption. ρ refers to the average rate of reporting data. ξ_0 refers to the starting total energy. $E(E_t)$ refers to the average transmission energy consumption of sensors. $E(E_u)$ refers to the average idle energy during network downtime. The energy limitation of the node itself is represented by Eq. (14).

$$H_{i,i} = \frac{\varepsilon_i}{\vartheta_i} \tag{14}$$

In Eq. (14), ε_i represents the energy stored by a node. ϑ_i represents the energy consumed by a node when transmitting information. The energy consumption of sending data packets from each node to the sink node is represented by Eq. (15).

$$E_{\delta_1}^i = E^e \delta_1 + \bar{E}_{\delta_1} \delta_1 \hat{r}^\gamma \tag{15}$$

In Eq. (15), δ_1 refers to the length of the data packet. $\bar{E}_{\delta_1}$ refers to the energy consumption required to achieve the target signal strength. E^e refers to the energy related to the receiving radio wave device. γ refers to the loss in the process of information dissemination. $\hat{r}$ refers to the distance between nodes. To select active nodes, first, sink node sampling is performed. Then, Bayesian compression perception reconstruction is performed. Finally, it is determined whether the energy consumption of the node exceeds the limit. If exceeding this limit, the node selection is redone. If not exceeding this limit, the reconstructed original data are output. Fig. 6 shows the node selection process.

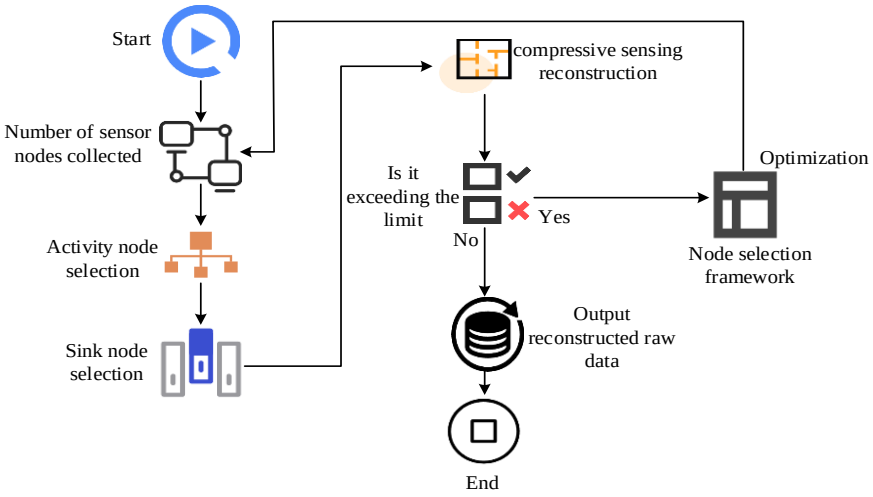


Fig. 6. Node selection flowchart.

IV. Effectiveness Analysis of Data Collection Methods Based on Data Perception and Positioning Technology

A. Experimental Parameter Setting and Efficiency Analysis

To verify the effectiveness of the adaptive sparse strategy and node selection strategy, analysis was conducted on a simulation platform. Set a square network monitoring area with a side length of L=10, and divide the monitoring area into 100 cells on average. Randomly deploy a sensor node in each cell, for a total of 100 sensor nodes. Deploy the sink node at the center position (5, 5) of the monitoring area, with an initial observation frequency of M=50. The length of the data packet is set to 20 bytes, the bandwidth is 2M/S, and the sparsity is 16. The simulation parameters and experimental environment settings are shown in Table I.

In the case of different sparsity, the reconstruction error of the adaptive sparsity strategy used in this study was compared with the reconstruction error of other methods. The balance factor was set to 0.6. Fig. 7 shows the statistical results. The reconstruction error decreased with increasing sparsity. The research method had the lowest reconstruction error. When the sparsity was 25, the reconstruction error curve tends to stabilize, and the minimum reconstruction error was 0.05. The reconstruction error of the overcomplete dictionary design method for sparse representation was slightly greater than that of the method used in the study. When the sparsity was 30, the reconstruction error curve tends to stabilize, and the minimum reconstruction error was 0.06. The wavelet transform method's reconstruction error was the largest. When the sparsity was 30, the reconstruction error curve tended to stabilize, and the minimum reconstruction error was 0.49. The reconstruction error of the discrete cosine transform method was slightly smaller than the wavelet transform method's. When the sparsity was 35, the reconstruction error curve tended to stabilize, and the minimum reconstruction error was 0.42.

TABLE I. SAMPLE PARAMETER SETTINGS

Adaptive sparse strategy		Node selection		Experimental environment	
Parameters	value	Parameters	value	Project	Environment
Number of sensor nodes	100	Sparsity	16	CPU	11th Gen Intel(R) Core(TM) i7-11800H @ 2.30GHz
Sink point location	(5,5)	Sink node sampling number	50	GPU	NVIDIA GeForce RTX 3070 Laptop GPU
Monitoring area	10*10	Sink point location	(5,5)	Memory	16GB
Bandwidth	2M/S	Number of sensor nodes	100	Video Memory	8GB
Packet length	20bytes	/	/	Operating system	Windows 10
Observation frequency	50	/	/	/	/

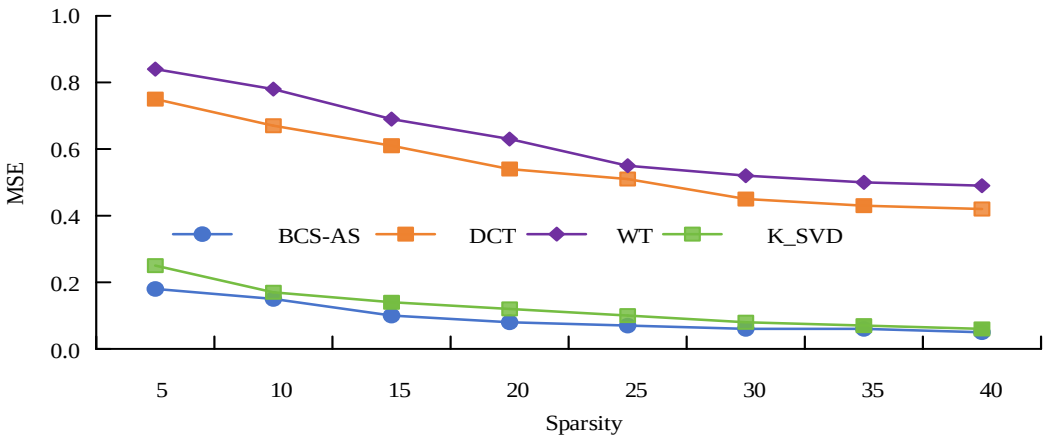
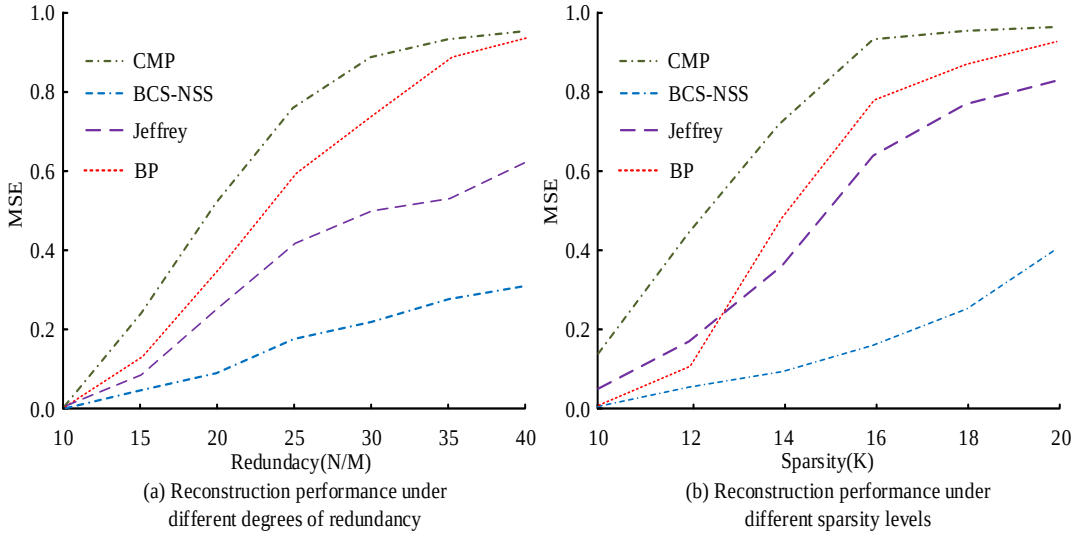


Fig. 7. Reconstruction performance of four sparse bases.

The reconstruction error of node selection based on Bayesian compression perception was compared with the reconstruction error of other methods under different degrees of redundancy and sparsity. Fig. 8 shows the statistical results. In Fig. 8(a), the reconstruction error increased with the increasing redundancy. The research method had the lowest reconstruction error and the maximum reconstruction error was 0.31. The orthogonal matching tracking algorithm had the highest reconstruction error, with a maximum reconstruction error of 0.96. The reconstruction error of the error backpropagation algorithm was slightly lower than the orthogonal matching tracking algorithm's, with a maximum reconstruction error of 0.93. The reconstruction error of Jeffrey's prior algorithm was higher than the research method's, with a maximum reconstruction error of 0.61. In Fig. 8(b), the reconstruction error increased with the increase of sparsity. The research method had the lowest reconstruction error and the maximum reconstruction error is 0.40. The orthogonal matching tracking algorithm had the highest reconstruction error, with a maximum reconstruction error of 0.98. The reconstruction error of the error backpropagation algorithm was slightly lower than that of the orthogonal matching tracking algorithm, with a maximum reconstruction error of 0.96. The Jeffrey's prior algorithm had higher reconstruction error than the research method, with a maximum reconstruction error of 0.82.

B. Quality Analysis of Data Collection Methods based on Data Perception and Positioning Technology

IoT signals are often affected by other noises during transmission, but this research does not consider noise interference in sensor networks. The reconstruction performance of the research adaptive sparse strategy was compared with other methods under different sampling quantities in Fig. 9. When the samples quantity reached 70, the reconstruction performance of different algorithms was basically the same in the two routing scenarios. The reconstruction performance decreased with increasing sampling. The adaptive strategy had the best reconstruction performance. When the sampling quantity was 70, the reconstruction error curve tended to stabilize, and the minimum reconstruction error was 0.04. The reconstruction performance of the discrete cosine transform method was the worst. When the sampling quantity was 60, the reconstruction error curve tended to stabilize, and the minimum reconstruction error was 0.42. The reconstruction performance of the overcomplete dictionary design method for sparse representation was slightly inferior to the research method. When the sampling quantity was 60, the reconstruction error curve tended to stabilize, and the minimum reconstruction error was 0.13.



(a) Reconstruction performance under different degrees of redundancy

(b) Reconstruction performance under different sparsity levels

Fig. 8. Comparison of reconstruction performance of four algorithms.

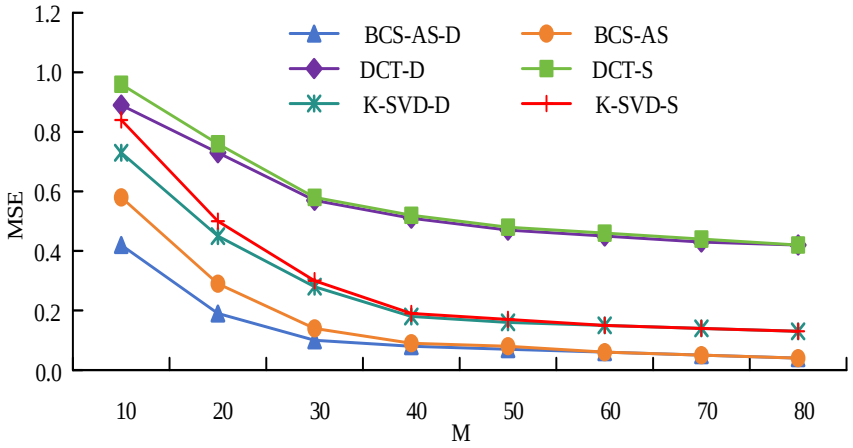
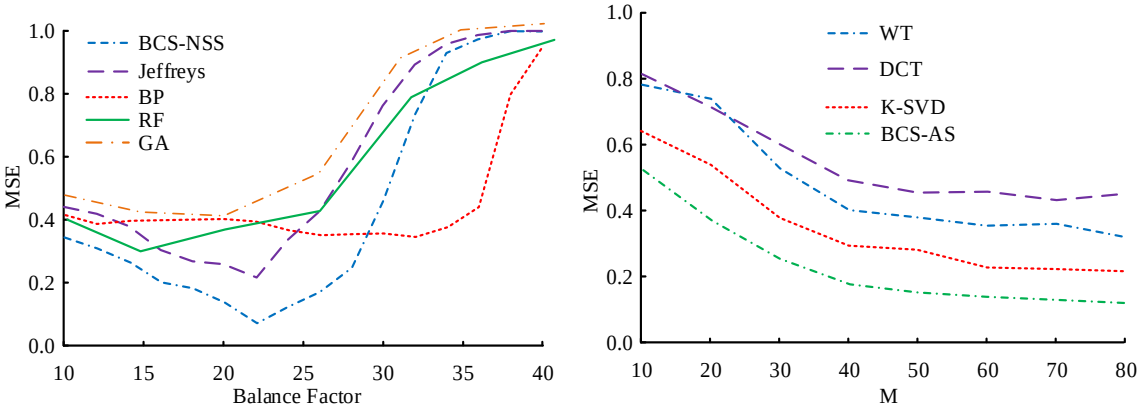


Fig. 9. Reconstruction performance of sparse routing and dense routing.

The reconstruction performance of node selection based on Bayesian compression perception was compared with other methods under different balance factors. In the case of unreliable links, the reconstruction performance of the adaptive sparse strategy was compared with that of other methods in Fig. 10. Observing Fig. 10(a), it can be seen that when the balance factor is small, the reconstruction performance of node selection based on Bayesian compressive sensing is the best, with a minimum reconstruction error of 0.09. When the balance factor is small, the reconstruction performance of genetic methods is the worst. The reason is that for some complex, high-dimensional, and nonlinear problems, genetic algorithms may have difficulty effectively searching for the optimal solution. At this time, the minimum reconstruction error is 0.47. When the balance factor is large, the reconstruction performance stability of the five algorithms is insufficient. In

Fig. 10 (b), the adaptive sparse strategy had better and more stable reconstruction performance, with a reconstruction error of 0.16. The reconstruction performance of the overcomplete dictionary design method for sparse representation was slightly inferior to the research method, with a reconstruction error of 0.22.

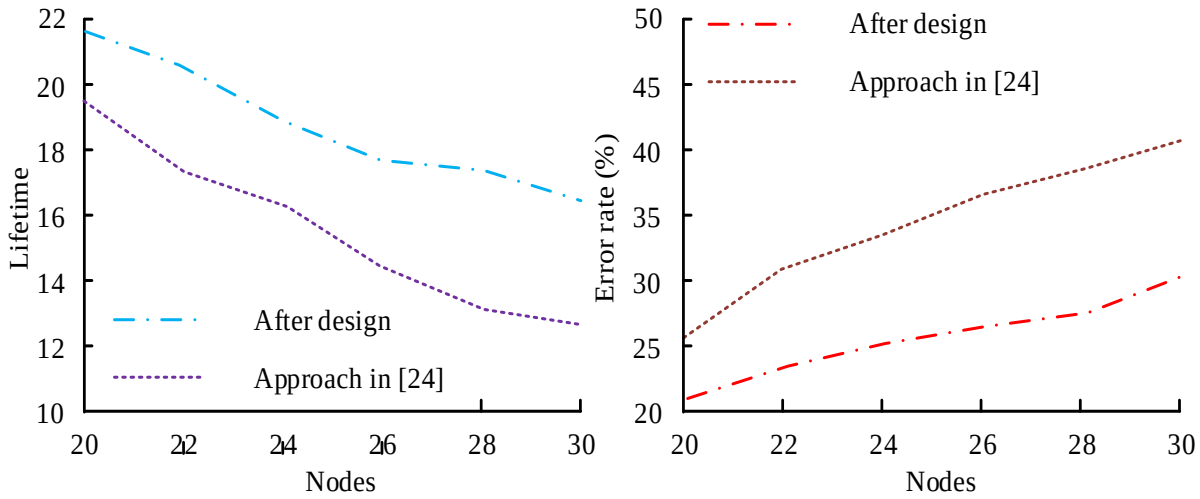
In the case of different numbers of nodes, the study compared the network lifetime and error rate of data collection using methods with the method proposed in study [24]. The statistical results are shown in Fig. 11. Observing Fig. 11(a), it can be seen that the network lifetime of the wireless sensor network data collected by the method used in the study has been extended. When the number of nodes is 28, the network lifetime is 17, and the network lifetime has been extended by 30.8%. Observing Fig. 11 (b), it can be seen that the error rate of wireless sensor network data collection has been reduced in the study. When the number of nodes is 30, the error rate of data collection has been reduced the most, with an error rate of 29%. The error rate of data collection has been reduced by 12%.



(a) Comparison of Reconstruction Performance under Different Balance Factors

(b) Reconstruction performance under unreliable links

Fig. 10. Comparison of reconstruction performance of different methods.



（a）Comparison of network lifespan for data collection using different methods

（b）Comparison of error rates in data collection using different methods

Fig. 11. Comparison of error rates and network lifespan of data collected by different methods.

The adaptive sparse strategy was compared with the discrete cosine transform method in terms of lifespan in sparse and dense routing. The reconstruction error and network lifetime of node selection based on Bayesian compression perception were compared with those of other methods under different numbers of nodes. Table II shows the statistical results. The lifespan of WSN with adaptive sparse strategy was longer than that of discrete cosine transform method. WSN had the longest lifespan under sparse routing. When the nodes were the same, the reconstruction error of node selection based on Bayesian compression perception was minimized. When nodes were 28, the minimum reconstruction error of the strategy used in this study was 0.04. Under the same reconstruction error, the network lifespan of node selection based on Bayesian compression perception was longer. When the reconstruction error was 0.6, the longest network lifetime of the strategy used in this study was 28.

TABLE II. COMPARISON OF RECONSTRUCTION ERROR AND NETWORK LIFESPAN

Same number of nodes			Same reconstruction error		
method	Number of active nodes	Reconstruction error	method	Reconstruction error	Network lifespan
BCS-NSS	22	0.35	BCS-NSS	0.3	25
BCS-DSSR	22	0.55	BCS-DSSR	0.3	6
BCS-MA-DR	22	0.62	BCS-MA-DR	0.3	22
BCS-NSS	24	0.15	BCS-NSS	0.4	26
BCS-DSSR	24	0.46	BCS-DSSR	0.4	20
BCS-MA-DR	24	0.32	BCS-MA-DR	0.4	23
BCS-NSS	26	0.08	BCS-NSS	0.5	27
BCS-DSSR	26	0.24	BCS-DSSR	0.5	22
BCS-MA-DR	26	0.37	BCS-MA-DR	0.5	24
BCS-NSS	28	0.04	BCS-NSS	0.6	28
BCS-DSSR	28	0.36	BCS-DSSR	0.6	26
BCS-MA-DR	28	0.22	BCS-MA-DR	0.6	27

V. DISCUSSIONS

The proposed digital sensing technology has unique advantages in data collection in wireless sensor networks. The study in [15] uses an efficient aggregation method for multi hop clusters based on hybrid compressive sensing to collect data. By combining compressive sensing and routing protocols, although the effect is significant, it requires multiple compression and decompression operations on the signal, thus requiring a large amount of computing resources and time, and the computational cost is high. The method in study [16] achieves compressive matrix decomposition through random singular value decomposition, thereby achieving higher accuracy at lower sampling rates and reducing data transmission costs. However, stochastic singular value decomposition is based on random sampling, and its results are sensitive to the randomness of the initial sampling matrix, which may lead to instability and uncertainty in the results. The method proposed in study [17] can reduce the loss of data samples when transmitted wirelessly, but its applicability is limited. In contrast, when studying the collection of wireless sensor network data based on digital sensing technology, the selection of sensor nodes reduces the amount of data collected and extends the lifespan of the wireless sensor network. In the experiment, as the amount of data continued to increase, the minimum reconstruction error of the research method was 0.04, which was 0.42 lower than DCT and 0.14 lower than SVD. The proposed method achieves significant performance improvement while maintaining low error, making it suitable for large-scale wireless sensor network data collection.

In summary, the research on data collection based on data aware positioning technology reduces the amount of data collected and extends the lifespan of wireless sensor networks by selecting sensor nodes. The research has provided theoretical support and practical guidance for wireless sensor network data collection, further improving the efficiency and accuracy of wireless sensor network data collection. And this method requires high-performance computer resources, so future research directions will focus on how to reduce the demand for computer resources in Bayesian compressive sensing methods.

VI. CONCLUSION

The development of IoT enables wireless communication between objects and automates data exchange. The foundation of IoT is wireless sensor technology. There is noise interference and abnormal or missing data in WSN, which leads to excessive consumption of communication resources and insufficient data reliability. The study proposes a data collection method based on data compression perception positioning technology. Firstly, a Bayesian compression perception-based method is adopted to select nodes to address the shortened network lifecycle caused by differences in node performance. Secondly, an adaptive sparse strategy based on Bayesian compression perception is adopted to collect data and reduce communication resource consumption. When nodes were 28, the minimum reconstruction error based on Bayesian compression perception node selection strategy was 0.04. When the reconstruction error was 0.6, the longest lifespan of the network based on Bayesian compression perception node selection strategy was 28. Under different sampling quantities, the adaptive strategy had the best reconstruction performance with a minimum reconstruction error of 0.04. When the balance factor was small, the reconstruction performance of node selection based on Bayesian compression perception was the best, with a minimum reconstruction error of 0.09. Compared with existing algorithms, the proposed method has been effectively applied in data acquisition, reducing energy consumption of wireless

sensor networks, extending their lifecycle, and improving the reconstruction performance of compressed sensing. However, Bayesian compressive sensing requires a prior distribution of known signals and has a high computational complexity, requiring a large amount of computing resources. Subsequent research will adopt other methods to optimize the Bayesian compressive sensing method, in order to avoid problems such as long computation time and large storage space.

ACKNOWLEDGMENT

The research is supported by Scientific Research Project of Liaoning University of International Business and Economics (No.2023XJLXZD01); Scientific research project of Liaoning Provincial Department of Education (No.JYTMS20230406).

REFERENCES

[1] Zhou I, Makhdoom I, Shariati N, Raza M A, Keshavarz R, Lipman J. Internet of things 2.0: Concepts, applications, and future directions. IEEE Access, 2021, 9: 70961-71012.

[2] Ghaffari K, Lagzian M, Kazemi M, Malekzadeh G. A comprehensive framework for Internet of Things development: A grounded theory study of requirements. Journal of Enterprise Information Management, 2020, 33(1): 23-50.

[3] Liu J, Zhao Z, Ji J, Hu, M. Research and application of wireless sensor network technology in power transmission and distribution system. Intelligent and Converged Networks, 2020, 1(2): 199-220.

[4] Luo J, Chen Y, Wu M, Yang Y. A survey of routing protocols for underwater wireless sensor networks. IEEE Communications Surveys & Tutorials, 2021, 23(1): 137-160.

[5] Antony S M, Indu S, Pandey R. An efficient solar energy harvesting system for wireless sensor network nodes. Journal of Information and Optimization Sciences, 2020, 41(1): 39-50.

[6] Zhang J, Zhao C, Gao W. Optimization-inspired compact deep compressive sensing. IEEE Journal of Selected Topics in Signal Processing, 2020, 14(4): 765-774.

[7] Ye G, Liu M, Wu M. Double image encryption algorithm based on compressive sensing and elliptic curve. Alexandria engineering journal, 2022, 61(9): 6785-6795.

[8] Bhuiyan M N, Rahman M M, Billah M M, Saha D. Internet of things (IoT): A review of its enabling technologies in healthcare applications, standards protocols, security, and market opportunities. IEEE Internet of Things Journal, 2021, 8(13): 10474-10498.

[9] Wang Y, Yang J, Liu M, Gui G. LightAMC: Lightweight automatic modulation classification via deep learning and compressive sensing. IEEE Transactions on Vehicular Technology, 2020, 69(3): 3491-3495.

[10] Cheng G, Wang C, Xu C. A novel hyper-chaotic image encryption scheme based on quantum genetic algorithm and compressive sensing. Multimedia Tools and Applications, 2020, 79(39): 29243-29263.

[11] Liang P, Fan J, Shen W, Qin Z, Li G Y. Deep learning and compressive sensing-based CSI feedback in FDD massive MIMO systems. IEEE Transactions on Vehicular Technology, 2020, 69(8): 9217-9222.

[12] Nan S, Feng X, Wu Y, Zhang H. Remote sensing image compression and encryption based on block compressive sensing and 2D-LCCCM. Nonlinear dynamics, 2022, 108(3): 2705-2729.

[13] Xu G, Zhang B, Yu H, Chen J, Xing M, Hong W. Sparse synthetic aperture radar imaging from compressed sensing and machine learning: Theories, applications, and trends. IEEE Geoscience and Remote Sensing Magazine, 2022, 10(4): 32-69.

[14] Chinthamu N, Karukuri M. Data Science and Applications. Journal of Data Science and Intelligent Systems, 2023, 1(1): 83-91.

[15] Aziz A, Osamy W, Khedr A M, El-Sawy A A, Singh K. Grey Wolf based compressive sensing scheme for data gathering in IoT based heterogeneous WSNs. Wireless Networks, 2020, 26(5): 3395-3418.

[16] Sekar K, Devi K S, Srinivasan P. Compressed tensor completion: A robust technique for fast and efficient data reconstruction in wireless sensor networks. IEEE Sensors Journal, 2022, 22(11): 10794-10807.

[17] Mei Y, Gao Z, Wu Y, Chen W, Zhang J, Ng D W K, Di Renzo M. Compressive sensing-based joint activity and data detection for grant-free massive IoT access. IEEE Transactions on Wireless Communications, 2021, 21(3): 1851-1869.

[18] Lin C, Han G, Qi X, Du J, Xu T, Martínez-García M. Energy-optimal data collection for unmanned aerial vehicle-aided industrial wireless sensor network-based agricultural monitoring system: A clustering compressed sampling approach. IEEE Transactions on Industrial Informatics, 2020, 17(6): 4411-4420.

[19] Chang C Y, Chen S Y, Chang I H, Yu G J, Roy D S l. Multirate data collection using mobile sink in wireless sensor networks. IEEE Sensors Journal, 2020, 20(14): 8173-8185.

[20] Lin Z, Chen Y, Liu X, Jiang R, Shen B. A Bayesian compressive sensing-based planar array diagnosis approach from near-field measurements. IEEE Antennas and Wireless Propagation Letters, 2020, 20(2): 249-253.

[21] Jiang X, Li N, Guo Y, Yu D, Yang S. Localization of multiple RF sources based on Bayesian compressive sensing using a limited number of UAVs with airborne RSS sensor. IEEE Sensors Journal, 2020, 21(5): 7067-7079.

[22] Singh J, Kaur R, Singh D. Energy harvesting in wireless sensor networks: A taxonomic survey. International Journal of Energy Research, 2021, 45(1): 118-140.

[23] Wei X, Guo H, Wang X, Wang X, Qiu M. Reliable data collection techniques in underwater wireless sensor networks: A survey. IEEE Communications Surveys & Tutorials, 2021, 24(1): 404-431.

[24] Xue Z. Routing optimization of sensor nodes in the Internet of Things based on genetic algorithm. IEEE sensors journal, 2021, 21(22): 25142-25150.

Hyperparameter Optimization in Transfer Learning for Improved Pathogen and Abiotic Plant Disease Classification

Asha Rani K P*, Gowrishankar S
Department of Computer Science and Engineering, Dr. Ambedkar Institute of Technology,
Bengaluru – 560056, Karnataka, India, Affiliated to VTU, Belagavi – 590018, Karnataka, India

***Abstract*—The application of machine learning, particularly through image-based analysis using computer vision techniques, has greatly improved the management of crop diseases in agriculture. This study explores the use of transfer learning to classify both spreadable and non-spreadable diseases affecting soybean, lettuce, and banana plants, with a special focus on various parts of the banana plant. In this research, 11 different transfer learning models were evaluated in Keras, with hyperparameters such as optimizers fine-tuned and models retrained to boost disease classification accuracy. Results showed enhanced detection capabilities, especially in models like VGG_19 and Xception, when optimized. The study also proposes a new approach by integrating an EfficientNetV2-style architecture with a custom-designed activation function and optimizer to improve model efficiency and accuracy. The custom activation function combines the advantages of ReLU and Tanh to optimize learning, while the hybrid optimizer merges feature of Adam and Stochastic Gradient Descent (SGD) to balance adaptive learning rates and generalization. This innovative approach achieved outstanding results, with an accuracy of 99.96% and an F1 score of 0.99 in distinguishing spreadable and non-spreadable plant diseases. The combination of these advanced methods marks a significant step forward in the use of machine learning for agricultural challenges, demonstrating the potential of customized neural network architectures and optimization strategies for accurate plant disease classification.**

Keywords*—Spreadable diseases; non-spreadable diseases; transfer learning; Keras; optimizers; CNN; underfitting and overfitting; retraining the models; base models; finetuning; abiotic; biotic; infectious and non-infectious diseases; custom optimization techniques; hyperparameter tuning in neural networks; hybrid activation functions

I. Introduction

Over 80,000 plant diseases are known to exist in the world. Crop diseases often harm crop plants, which can result in major economic and agricultural losses [1]. If a plant disease is induced by an environmental element and is not spread from one plant to another, it is referred to be abiotic, or non-infectious. Diseases classified as biotic or infectious are those brought on by pathogens like viruses, fungi and nematodes.

Pathogens that affect animals as well as humans mimic plant illnesses. Fungus, organisms that mimic fungal, bacterial, phytoplasmas, viral, viral vectors, nematodes and parasitic higher plants are examples of plant diseases.

In order to establish and maintain food security and revenue sources for a developing world, it is more crucial to safeguard plants against diseases. Severity of plant diseases can be reduced with the aid of early identification.

The first lettuce farms were established in ancient Egypt as it was the most widely consumed salad produce and has substantial economic worth. Lettuce is more susceptible to biotic than abiotic illnesses. The study highlights crucial and substantial diseases, such as downy mildew, which may spread swiftly to impact most plants in a crop. Lettuce diseases can cause significant damage and occasionally full crop loss [2]. Some diseases, such as downy and powdery mildews, can spread swiftly and harm the majority of the plants in a crop.

Diseases and insect pests are the main issues in soybean production. To get a broader perspective on spreadable and non-spreadable diseases, soybean plant is chosen. This calls for careful diagnosis and prompt handling to prevent the soybean crops from suffering significant losses. The world's soybean production is projected to be 333.67 million tonnes in 2019–2020 from a total area of 120.50 million hectares [3].

It is crucial to learn more about the spreadable and non-spreadable diseases that affect various plant parts, including the leaves, fruits, stems, nodes and roots. The banana crop fits best into this category because each part of the plant has a variety of uses, including medicinal properties for the stem and roots and maintaining the health of the soil. Additionally, with an output of 97.5 million tonnes, bananas are a significant fruit crop on a worldwide scale. It has a total yearly output of 490,710,000 hectares producing 16.91 million tonnes [4].

Bananas are an important fruit crop in India.They are a staple food for many people in the country and are also widely used in various dishes. Panama disease, also known as Fusarium wilt, is a serious threat to banana production worldwide. It is caused by a fungus that infects the root system of the banana plant, ultimately causing the plant to wilt and die. Aphids are a common pest that can also infect banana plants. They feed on the sap of the plant, which can weaken it and make it more susceptible to other pests and diseases. It is important for farmers to monitor their banana crops closely and take steps to prevent and control these pests and diseases to protect their yield.

Food security is at risk from plant diseases because they can harm crops, lowering food production and driving up food

prices. Leaf blight, septoria blight, powdery mildew and downy mildew, which can be fungal, are the main diseases impacting the lettuce crop. Bacterial rust and downy mildew are the diseases that affect soybeans, whereas weevil, soft rot, aphids, and few may affect bananas [5]. Fig. 1 and 2 shows the illnesses of lettuce and soybean.

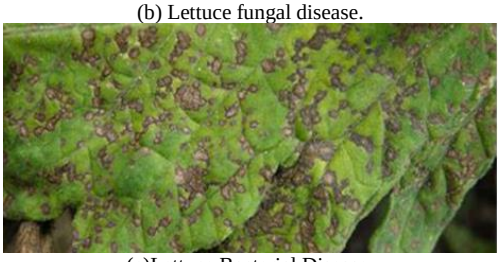
Deep learning is a cutting-edge technique for object recognition and image processing that improves categorisation of numerous crop diseases [6]. One well-liked method in deep learning where pre-trained models are modified to perform a new job is transfer learning. Deep Transfer Learning (DTL) creates a applied novel framework for predictive analytics and digital image processing that is more accurate and has enormous potential for crop disease identification. A potential method for recognising diseases onsite is the DTL technique, which also offers a quick way to adapt created models to the constraints imposed by mobile applications [7]. This would be very useful in a real-world field scenario.



(a) Lettuce viral disease.



(b) Lettuce fungal disease.



(c)Lettuce Bacterial Disease.



(d) Lettuce Non-Infectious Disease (Salt burn).

Fig. 1. Some major plant diseases found in lettuce plant dataset.

A variety of factors, including that of the high-definition camera, high efficient processing and many built-in accessories, enable automatic disease identification. The accuracy of the outcomes has increased because of the use of cutting-edge techniques like deep learning and machine learning. Our experimental results represent significant advances in the understanding of the severity of plant diseases. The paper is organised as follows for the following sections: Section I Introduction, Section II Literature Review, Section III highlights Methodology that includes expanded dataset description, Augmentation and Activation functions, Section IV describes Performance evaluation. Performance reviews go into great detail, Section V is concerned with implementation, results are providedd in tabular format and Section VI acts as a conclusion.



(a) Soybean Iron deficiency.



(b) Soybean fungal disease.



(c) Soybean Diabrotica speciosa disease.

Fig. 2. Some major plant diseases found in soybean plant dataset.

II. Literature Survey

N. Saranya et al. [8] have categorized many ailments that affect the leaves and fruits of the banana plant. Fuzzy c-means, histogram-based equalization and artificial neural networks all have important roles in the proposed approach. The image is divided using fuzzy c-means and the histogram is then utilized to transform it without losing any of the details of the banana plant. In this study, a better categorization strategy is recommended in order to deliver the best return.

Michael Gomez Selvaraj et al. [9] applied pixel-based banana classification using the Random Forest (RF) model utilizing integrated features of Vegetative Indices (VI) and Principal Component Analysis (PCA) to map banana under mixed-complex African settings. Gomez Selvaraj et al. provided higher & low-resolution aerial (UAV and satellite) photos with cutting-edge computer vision algorithms to achieve more than 90% accuracy under actual settings (Smart phone-based AI applications).

A high-definition camera is used to take photos of the early and intermediate phases of soybean disease and the photos are then expertly batched into uniform sizes. The picture

segmentation methods used by E. Miao et al. [10] include lab grayscale map, ultragreen feature approach, genetic algorithm and threshold segmentation. Next, the results are filtered using the median and corroded expansion. A Convolutional Neural Network (CNN) that uses a MultiLayer Perceptron (MLP) framework to execute supervised learning of the network and achieves an average recognition rate of 94.87% is seen in the soybean illness picture identification experiment.

Three disease groups of soybean leaves were examined by Sachin B. Jadhav et al. [11] bacterial blight, frogeye leaf spot, and septoria brown spot. The diseased leaf area is segmented using incremental K-means clustering. Color and texture data are recovered using the R, G, B color space and the Gray Level Co-occurrence Matrix (GLCM), respectively. SVM and K-Nearest Neighbors Algorithm (KNN) are used in a classification technique to identify the exact kind of leaf disease. The results demonstrate that the SVM classifier approach outperforms the KNN methodology with efficiencies of 87.3% and 83.4%, respectively.

Elham Khalili et al. [12] examined and compared six ML methods for identifying the ailment known as charcoal rot in their study that was published in science. Healthy plants were gathered from the stem and root of soybean plants during the ripening stage based on the maturity's symptomless qualities. R7(Yellowing of the leaves and yellow pods at 50% growing stage) was chosen for sick plants based on physical criteria indicates the presence of bright grey and mycelium on the root and stem. Gradient Tree Boosting and Support Vector Machines performed better than Regularized Logistic Regression, MultiLayer Perceptron and Random Forest techniques.

Miao Yu et al. [13] used the OTSU technique, which decreases the effect of the background on the disease images. Using ResNet18 and RANet, the model's effectiveness in the test set was confirmed and assessed. The response time was 0.0514 seconds, the F1-value was 98.52, and the RANet average recognition rate was 98.49%. Compared to ResNet18, the identification rate increased by 1.15 percent, the F1-value increased by 1.17 and 0.0133 seconds were saved while identifying illnesses from images.

Lack of calcium makes tip-burn, which is common in lettuce plants cultivated indoors, worse. Photos of tip-burn lettuce were illuminated using white, red and blue LEDs, and these images served as the training, validation and testing datasets for a deep-learning detection method. The detection approach developed by Munirah Hayati Hamidon et al. [14] was based on three detectors: CenterNet, YOLOv4 and YOLOv5. YOLOv5 beat the other two models tested, with an accuracy of 84.1% mAP.

Positive and negative samples from each kind of weed and crop were chosen by Kavir Osorio et al. [15] and taken. There are just a few of weed characteristics that remain consistent, making identification difficult. The identification of the vegetation was done alternatively using multispectral bands. The R-CNN model distinguished itself for its accuracy in detecting the crop and showing the edges, making it a tactic that may be recommended for addressing problems like fruit detection. The RCNN and HOG-SVM-based algorithms were shown to be the most trustworthy using the Bland-Altman approach. The YOLO strategy exaggerates the high levels of cannabis coverage in contrast to the other two.

According to J. Amara et al. [16], who used the LeNet architecture as a Convolutional Neural Network to classify the data, banana leaf disease may now be classified using deep learning. This strategy stabilized after 25 iterations. This research demonstrated its effectiveness in a variety of picture situations, including ones with a complex background and various sizes and orientations.

W. Liao et al. [17] proposed using the SVM classifier in a machine learning-based strategy for early identification of banana disease. Hyperspectral images taken at close range are utilized in this instance. When using spectral and morphological data, the classifiers' outputs have an overall accuracy of 96% for early detection, 90% for mid-detection, and 92% for late detection.

More research is being done to detect and classify the disease, not just for banana leaves but also for the majority of food crops including rice, maize, apple, cheerio and other well-known plants. Here are a few of these judgements.

A superior convolutional neural network should be used to classify apple plant and cherry plant diseases, according to [18].

In the extremely packed growing conditions of indoor settings, early diagnosis of tip-burns in lettuce is vital in order to reduce the cost of human identification and boost lettuce quality and production. Shimamura et al. [19] created a system for tip-burn identification in plant factories by using GoogLeNet to classify two different types of tip-burn from a single picture of lettuce.

The most recent Neuron Compute Stick pretrained Movidius of deep CNN model from Intel provided an accuracy rate of 88.46% for Mishra et al. [20]'s system for classifying and diagnosing maize leaf diseases. The system was implemented on a Raspberry Pi 3.

K-means segmentation and multiclass support vector machines were used by Kumar et al. [21] to identify and classify different plant leaf diseases (SVM-based classification). Compared to other approaches, the detection accuracy is much higher.

To eliminate manual feature stage modelling, Mazzia et al. [22] created an LC&CC deep learning model that blends Recurrent Neural Networks (RNN) with Convolutional Neural Networks (CNN).

Researchers Cetin et al. [23] used six different machine learning algorithms to analyze and classify six sunflower varieties (105 single seeds) based on their fatty acid and mineral composition, biochemical traits and physical characteristics. These algorithms included Decision Tree (DT), Random Forest (RF), Support Vector Machine (SVM), Multiple Linear Regression (MLR), Naive Bayes (NB) and MultiLayer Perceptron (MLP).

Sharif et al. [24] proposed a hybrid feature selection method, which included the principal components analysis score, entropy, skewness-based covariance vector and Multiclass-SVM (MSVM), produced true positive rates of 96.9% and

97.1% for the detection of anthracnose disease and melanose disease on citrous leaves.

Machine learning models including K-Nearest Neighbor (KNN), Support Vector Machine (SVM), and other machine learning models have been widely used as classifiers to find anomalous areas on crop leaves. With a detection rate of 90.5%, Lu et al. [25] used Fisher discriminant analysis to identify anthracnose crown rot in the early stages of affected strawberry leaves inside.

To detect early blight on potato leaves, Vijver et al. [26] used partial least squares discriminant and discovered a positive predictive value of 0.92. This study demonstrated that artificial intelligence can accurately identify aberrant leaves in a range of crops. Therefore, using machine learning algorithms to detect yellow and wilted lettuce leaves in hydroponic systems is encouraging.

III. METHODOLOGY

A. Dataset Description

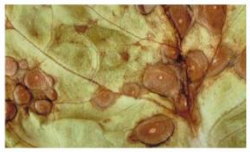
Performance evaluation of existing transfer learning pretrained models for plant disease classification is done using the Lettuce, Soybean and Banana dataset. Non-Spreadable diseases caused by abiotic factors include herbicide injury which turn leaves or leaf veins yellow or red. Calcium strengthens plant cell walls and salt burn is a result of the plant's inability to supply enough calcium for developing leaves during periods of rapid growth. The dataset for the proposed research consists of images of Lettuce plant and Soybean Plant infected by non-spreadable diseases and images affected by spreadable diseases. These images have been obtained from CrowdAI [27] and PlantVillage dataset [28]. The images have been augmented and brought up to 628 images for the former and 1845 images for the latter, as represented in Fig. 3 and 4. Each image maintains a fixed width and height of 256x256 pixels.

Two mobile phones and a UAV were used to take pictures of soybeans. Three groups make up the dataset: (I) photos of healthy plants, (II) pictures of plants harmed by caterpillars, and (III) pictures of plants harmed by Diabrotica speciosa. To meet our demands, the photos have undergone processing and augmentation [29].

The banana plant is vulnerable to bacterial, fungal, and viral diseases that can affect different parts of the plant. For research purposes, the PSFD-Musa dataset [30] was utilized, which contains pre-existing enhanced photos. Fig. 5 to 11 represent the dataset, showcasing the spreadable disease affected regions such as Node, Leaf, Banana, and Fruit.

Lettuce Dataset			
Non-Spreadable Disease	Herbicide Injury	Salt Burn	Healthy
No. of Original images	26	12	26
No. of Augmented Images	208	212	208

Fig. 3. Lettuce (Non-Spreadable) Dataset Description.

Lettuce Dataset				
Spreadable Disease	Bacterial	Fungal Downy Mildew	Funga Powdery Mildew	Fungal Septoria Blight
No. of Original images	35	32	26	19
No. of Augmented Images	220	261	464	238

(a)

Lettuce Dataset			
Spreadable Disease	Fungal wilt and leaf blight	Viral	Healthy
No. of Original images	10	17	26
No. of Augmented Images	212	232	208

(b)

Fig. 4. Lettuce (Spreadable) Dataset Description.

Banana Dataset			
Stem Diseases (Spreadable)	Bacterial Soft Rot	Healthy	Pseudostem Weevil
No. of Images	543	95	561

Fig. 5. Dataset classification of Banana Stem.

Banana Dataset				
Leaf Diseases (Spreadable)	Black Sigatoka	Healthy Leaf	Panama Disease	Yellow Sigatoka
No. of Images	474	129	102	528

Fig. 6. Dataset classification of Banana Leaf (Spreadable).

Banana Dataset		
Node Diseases (Spreadable)	Aphids	Healthy
No. of Images	366	408

Fig. 7. Dataset classification of Banana Node

Banana Dataset		
Leaf Diseases (Non-Spreadable)	Healthy Leaf	Potassium Deficiency
No. of Images	129	555

Fig. 8. Dataset classification (non-Spreadable) disease in Banana crop.

Banana Dataset		
Fruit Diseases (Spreadable)	Healthy	Scarring Beetle
No. of Images	144	150

Fig. 9. Dataset classification spreadable disease in Banana Fruit.

Soyabean Dataset					
Non spreadable Disease	Caterpillar	Diabrotica Speciosa	Iron Deficiency	Sunscald	Healthy
No. of Original images	273	136	14	9	252
No. of Augmented Images	273	136	214	209	252

Fig. 10. Dataset classification of Soybean Non-Spreadable Diseases.

Soyabean Dataset			
Spreadable Disease	Rust	Downy Mildew	Healthy
No. of Original images	61	27	252
No. of Augmented Images	261	227	252

Fig. 11. Dataset classification of Soybean Spreadable Diseases.

The dataset includes images of leaves affected by diseases like Black Sigatoka, Panama, and Yellow Sigatoka (Fig. 5-7), with a total count of 1104. Stem diseases, namely Bacterial Soft Rot and Pseudostem Weevil, also have 1104 corresponding images. The Banana node is affected by Aphids (Fig. 8), and the Banana Fruit is affected by Scarring Beetle (Fig. 9), with 366 and 150 images obtained, respectively.

To enhance the model's efficiency and performance, self-captured healthy banana images were incorporated. Fig. 7 illustrates images of healthy nodes. Additionally, potassium deficiency, caused by abiotic factors, has an adverse impact on banana leaves (Fig. 8). The dataset also includes 150 images of spreadable diseases in banana fruit (Fig. 9).

In addition to the banana dataset, 740 photos of soybean plants infected with spreadable diseases (Fig. 10) and 1084 non-spreadable images (Fig. 11) were obtained for comparative analysis.

B. Augmentation

Convolutional Neural Networks (CNN) are widely used for image classification. However, the quality and quantity of training data significantly impact model performance. Insufficient or imbalanced data can lead to poor generalization. Techniques like oversampling or undersampling can address class imbalance [31]. Data augmentation, including affine transformations and color manipulation, is a popular method to increase dataset size. Classical approaches may not always improve accuracy or address overfitting effectively. Affine transformations include rotation, reflection, scaling, and shearing. Additional techniques like permutation rotate, random zoom, variation in shear, random crop, and flip can be used. After data augmentation and balancing, the dataset consisted of 200 augmented images per class [32].

C. Activation Function

An activation function is a mathematical function that is applied to the input of a neural network node or a layer of nodes. The activation function is used to introduce non-linearity into the network, which is necessary for the network to learn complex patterns in the input data. Without activation functions, a neural network would essentially be a linear model, which is limited in its ability to learn complex relationships.

Activation methods that are frequently employed based on a few desirable characteristics include:

1) Nonlinear: Whenever the activation function is nonlinear, it has been shown that a two-layer neural network is an excellent approximator of any function. The identical activation function does not satisfy this condition. When many layers employ the same activation function, the network as a whole is equivalent to a single-layer model.

2) Range: Gradient-based training techniques have a tendency to be more stable when the activation function's range is finite, since only a small number of weights are significantly affected by pattern presentations. Since most of the weights are strongly affected by pattern presentations when the range is unlimited, training is often more effective. Short learning rates are often required in the latter scenario.

3) Continuously differentiable: For the purpose of allowing gradient-based optimization approaches, this property is desirable (ReLU is not continuous differentiable and has some challenges with it, but it is still achievable). Because the binary step activation function is not differentiable at zero and differentiates to zero for all future values, gradient-based techniques cannot advance with it.

4) Monotonic: A single-layer model's related error surface is always guaranteed to be convex when the activation function is monotonic.

5) Approximates near the origin: When activation functions have this property, the neural network can learn efficiently when its weights are initialized with low-level random values. If the activation function differs from identity near to the origin while initializing the weights, more care must be taken.

Each activation function has advantages and disadvantages, so we must be cautious when choosing one. Following are some frequent considerations to make while selecting an activation function:

1) When it comes to classification issues, sigmoid [33] functions (including softmax) and their combinations often perform better.

2) Due to the vanishing gradient issue, sigmoid and tanh functions continue to be avoided in hidden layers.

3) Tanh is typically avoided because of the dead neuron issue [34].

4) Because it produces superior results, ReLU activation function is frequently employed and is the default option (than sigmoid and tanh) [35].

5) However, the ReLU function should only be utilized in the buried layers (and not in the output layer).

6) In cases of regression issues, an output layer's activation function can be linear, however nonlinear activation functions are required for classification tasks.

7) The leaky ReLU function is the ideal option if we come into an instance of dead neurons in our networks.

8) For any kind of neural network, the ReLU activation function is presently the one that is most frequently employed for the hidden layers (but never for the output layer).

9) Swish activation should only be utilized for bigger neural networks with depths of more than 50 layers, even though it does not consistently beat ReLU in complicated applications [36].

10) The output (top-most) layer should be triggered by the sigmoid function for 2-class applications, as well as for multi-label classification.

11) The output layer must be triggered using the softmax activation function for multi-class applications.

12) A basic regression neural network should just employ the linear activation function in the output layer.

13) The tanh activation function is recommended for the hidden layer in Recurrent Neural Networks (RNN). By default, it is configured by TensorFlow.

14) In some circumstances, switching to a leaky ReLU might produce better outcomes and overall performance if ReLU is unable to deliver the desired results.

Some of the known Activation functions are:

1) The sigmoid function: Logistic regression and simple neural network implementation both use sigmoid functions. The fundamental activation units in machine learning are sigmoid functions. However, because of a number of limitations, it is simply not advisable to use complicated neural network sigmoid functions (vanishing gradient problem). Given that it is among the most frequently used activation functions, it serves as an excellent introduction for those who are naïve to data science and machine learning. Whilst the sigmoid function and its derivative are simple to use and help reduce the time required to develop models, there is a considerable downside of data lost since the derivative has a constrained range.

2) Tanh function: The tanh function partially addresses the drawback of the sigmoid function. Its key feature is that its curve is symmetric across the origin and has coefficients that range from -1 to 1 [34]. This does not, however, mean that the fading or bursting gradient problem does not occur. It does exist for tanh, however unlike Sigmoid, it is centered at zero, making it more ideal than Sigmoid Function.

3) ReLU (Rectified Linear Units) and Leaky ReLU: ReLU functions, as opposed to Logistic Activation functions, are currently used in the majority of Deep Learning applications, such as computer vision, natural language processing, speech recognition, deep neural networks, etc [35]. ReLU outperforms tanh or sigmoid functions in terms of application-level manifold convergence speed. Among the ReLU variations are Leaky ReLU, Parametric ReLU, Parametric Softplus (SmoothReLU), Noisy ReLU, and ExponentialReLU (ELU) [36].

4) Softmax function: The Softmax activation function which not only turns our output into a [0, 1] range but also changes each outcome so that the sum of each is 1 [37], is extremely fascinating. Softmax produces probability distribution as a result. In logistic regression model (multivariate), Softmax is used for multi-classification while Sigmoid is employed for binary classification.

D. Mathematical Approach for the Considered Procedure

The field of machine learning relies heavily on mathematical principles and techniques to design, train, and optimize models that can make predictions or learn patterns from data. This mathematical approach enables us to create powerful algorithms capable of solving a wide range of tasks, from image recognition and natural language processing to financial predictions and recommendation systems.

At the core of the mathematical approach in machine learning is the idea of formulating the learning problem as an optimization task. The goal is to find the model's parameters that minimize a certain objective function, such as the mean squared error in regression tasks or the cross-entropy loss in classification tasks. This process involves using various mathematical tools to represent the model, compute gradients, and iteratively update the parameters to approach the optimal solution. let's go through each of these layers, providing a brief introduction and their mathematical formulas:

1) Convolutional Layer (Conv layer): Convolutional layers are the fundamental building blocks of Convolutional Neural Networks (CNNs). They are designed to automatically and adaptively learn spatial hierarchies of features from input data such as images. A convolutional layer applies convolutional operations to input data using learnable filters (kernels) to detect local patterns and features.

The output of a convolutional layer can be represented as follows:

Given an input feature map X with dimensions (height, width, channels), and a set of learnable filters W of size (filter_height, filter_width, input_channels, output_channels), the convolution operation can be represented as:

$$Y[i, j, k] = \Sigma \Sigma \Sigma X[p + i, q + j, r] * W[p, q, r, k]$$

Here,

- Y[i, j, k] is the value of the output feature map at position (i, j) in the k-th channel.
- X[p+i, q+j, r] is the value of the input feature map at position (p+i, q+j) in the r-th channel.
- W[p, q, r, k] is the value of the learnable filter at position (p, q) in the r-th input channel and k-th output channel.
- The summation is performed over all spatial positions (p, q) of the filter and all input channels (r).

2) MaxPooling Layer (MaxPooling): MaxPooling is a downsampling technique commonly used in CNNs to reduce the spatial dimensions of the feature maps while retaining the most important information. It works by dividing the input

feature map into non-overlapping regions and taking the maximum value within each region.

The output of a MaxPooling layer can be represented as follows:

Given an input feature map X with dimensions (height, width, channels), and a pooling window of size (pool_height, pool_width), the MaxPooling operation can be represented as:

$$Y[i,j,k] = max(X[i * pool_{height} : (i+1) * pool_{height}, j * pool_{width} : (j+1) * pool_{width}, k])$$

Here,

- Y[i, j, k] is the value of the output feature map at position (i, j) in the k-th channel.
- The max function takes the maximum value within the pooling window.

3) SeparableConv Layer (Depthwise Separable Convolution): The SeparableConv layer is an alternative to standard convolutions designed to reduce computation and model size while maintaining representational capacity. It splits the convolution operation into two steps: depthwise convolution and pointwise convolution.

The output of a SeparableConv layer can be represented as follows:

Given an input feature map X with dimensions (height, width, channels), a depthwise kernel DW of size (filter_height, filter_width, channels), and a pointwise kernel PW of size (1, 1, channels, output_channels), the SeparableConv operation can be represented as:

$$Y[i,j,k] = \Sigma\,\Sigma\, X[i+p, j+q, r] * DW[p,q,r] * PW[1,1,r,k]$$

Here,

- Y[i, j, k] is the value of the output feature map at position (i, j) in the k-th channel.
- X[i+p, j+q, r] is the value of the input feature map at position (i+p, j+q) in the r-th channel.
- DW[p, q, r] is the value of the depthwise kernel at position (p, q) in the r-th channel.
- PW[1, 1, r, k] is the value of the pointwise kernel at position (1, 1) in the r-th input channel and k-th output channel.
- The summation is performed over all spatial positions (p, q) of the depthwise kernel and all input channels (r).

4) GlobalAveragePooling2D Layer: Global Average Pooling 2D is another downsampling technique used in CNNs, often as an alternative to fully connected layers at the end of the network. It computes the average value of each channel of the feature map, reducing the spatial dimensions to a single value per channel.

The output of a GlobalAveragePooling2D layer can be represented as follows:

Given an input feature map X with dimensions (height, width, channels), the Global Average Pooling operation can be represented as:

$$Y[k] = \left(\frac{1}{(height * width)}\right) * \Sigma\,\Sigma\, X[i,j,k]$$

Here,

- Y[k] is the value of the output for the k-th channel.
- The summation is performed over all spatial positions (i, j) of the feature map.

5) Dense Layer (Fully Connected Layer): The Dense layer is the standard fully connected layer in neural networks. It connects every neuron from the previous layer to every neuron in the current layer. The Dense layer performs a linear transformation followed by an activation function.

The output of a Dense layer can be represented as follows:

Given an input vector X of size (input_units) and the weight matrix W of size (input_units, output_units), the Dense layer operation can be represented as:

$$Y = activation_function(X * W + b)$$

Here,

- Y is the output vector.
- activation_function is the non-linear activation function applied element-wise to the linear transformation.
- b is the bias vector of size (output_units).

6) BatchNormalization layer: BatchNormalization is a normalization technique applied to intermediate layers in neural networks to stabilize and accelerate training. It normalizes the activations of each layer's mini-batch, making the network more robust and less sensitive to the scale of the input.

The output of a BatchNormalization layer can be represented as follows:

Given an input feature map X with dimensions (batch_size, features), and learnable scaling and shifting parameters γ and β, the BatchNormalization operation can be represented as:

$$\mu = \frac{1}{batch_{size}} * \Sigma\, X$$

$$\sigma^2 = \frac{1}{batch_{size}} * \Sigma\,(X - \mu)^2$$

$$X_{normalized} = \frac{(X - \mu)}{\sqrt{\sigma^2 + \varepsilon}}$$

$$Y = \gamma * X_normalized + \beta$$

Here,

- μ and σ^2 are the mean and variance of the mini-batch.

- X_normalized is the normalized input.
- γ and β are learnable scaling and shifting parameters, respectively.
- ε is a small constant (usually a small value like 1e-5) added for numerical stability.

7) Flatten layer: The Flatten layer is used to reshape the high-dimensional feature maps into a 1D vector, which is then fed into a Dense (fully connected) layer for further processing.

Let's consider an input tensor X with dimensions (batch_size, height, width, channels), where:

batch_size: The number of samples in the batch.

height: The height dimension of the feature maps.

width: The width dimension of the feature maps.

channels: The number of channels (depth) of the feature maps.

The Flatten layer reshapes the input tensor X into a 1D vector with size (batch_size, height * width * channels). This is achieved by simply concatenating all the elements of each feature map in X into a single long vector.

$$Flatten_output = Reshape(X, (batch_size, height * width * channels))$$

Here, Reshape(X, (batch_size, height * width * channels)) represents the operation of reshaping the input tensor X into the specified dimensions.

IV. Performance Evaluation and Results

A. Model and Dataset Selection

On datasets for lettuce, soybeans and bananas, 11 Transfer Learning models have been used to identify and categorize disease occurrence. The dataset that was collected from web sources are treated as raw data and organized as indicated in Fig. 12 and 13.

The original dataset was reshuffled, and the resulting dataset is used to train the transfer learning models in Keras module. Divided 38 TL Keras models into 11 different groups, grouped them as families, and primarily selected a member from each group for further study.

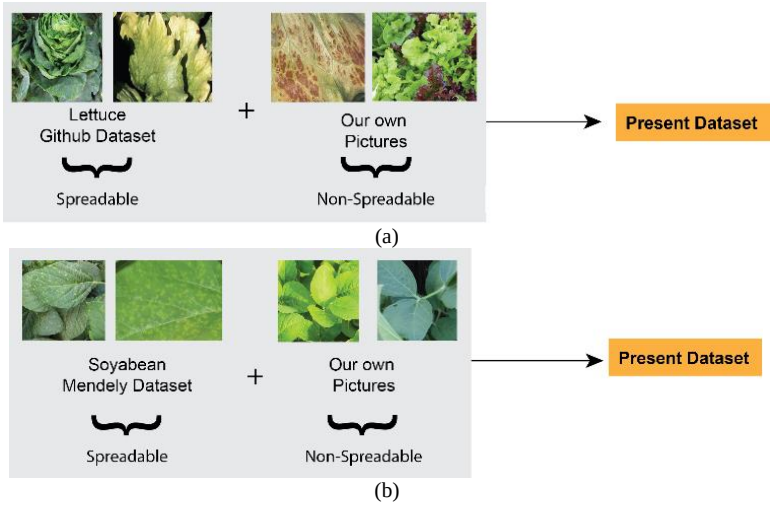


Fig. 12. (a) Modification of Lettuce Dataset, (b) Modification of Soybean Dataset.

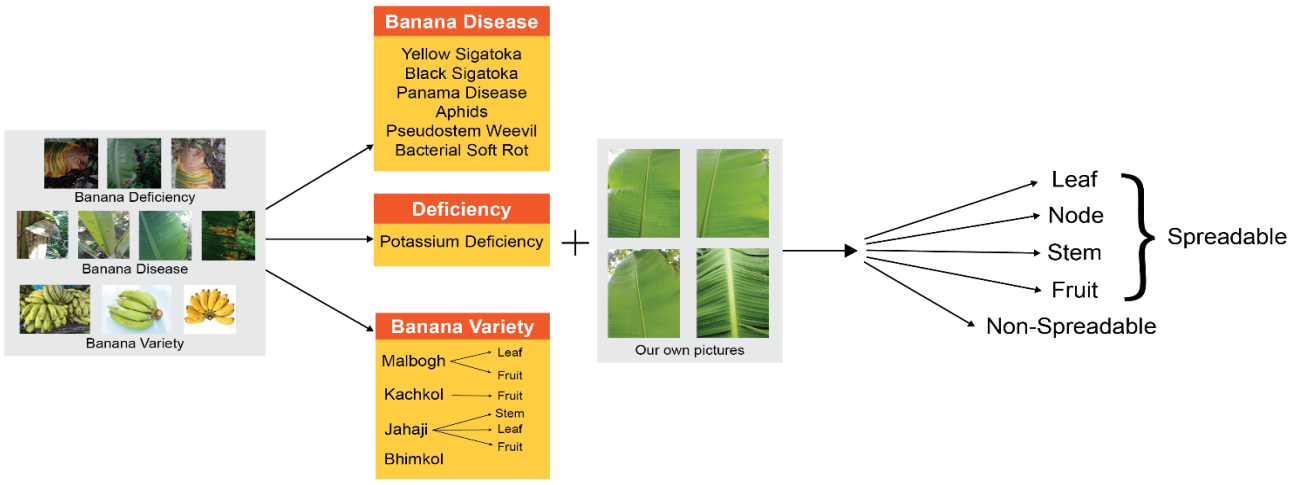


Fig. 13. Variety of banana dataset considered.

The bold models in Table I are chosen for research.

TABLE I. MODEL CLASSIFICATION ACCORDING TO MODELS FAMILY CONCEPT

Family	Model
Xception family	**Xception**
VGG Family	VGG16
	VGG19
ResNet Family	ResNet50
	ResNet50V2
	ResNet101
	ResNet101V2
	ResNet152
	ResNet152V2
Inception Family	**InceptionV3**
	InceptionResNetV2
MobileNet Family	MobileNet
	MobileNetV2
DenseNet Family	DenseNet121
	DenseNet169
	DenseNet201
NASNet Family	**NASNetMobile**
	NASNetLarge
EfficientNet Family	**EfficientNetB0**
	EfficientNetB1
	EfficientNetB2
	EfficientNetB3
	EfficientNetB4
	EfficientNetB5
	EfficientNetB6
	EfficientNetB7
EfficientNetV2 Family	EfficientNetV2B0
	EfficientNetV2B1
	EfficientNetV2B2
	EfficientNetV2B3
	EfficientNetV2S
	EfficientNetV2M
	EfficientNetV2L
ConvNext Family	ConvNeXtTiny
	ConvNeXtSmall
	ConvNeXtBase
	ConvNeXtLarge
	ConvNeXtXLarge

To retrain a transfer learning model, you will need to follow these steps:

1) Choose a pre-trained model: Start by choosing a pre-trained model that you want to use as the base for your model. There are many pre-trained models available in various libraries and frameworks, such as TensorFlow, PyTorch and Keras.

2) Freeze the base model: The pre-trained model will likely have many layers, and you will want to "freeze" the weights of these layers so that they are not updated during training. This will allow you to take advantage of the knowledge learned by the pre-trained model on a large dataset, while still training a new model that is customized for your specific task.

3) Add new layers: Next, you will want to add one or more layers to the model that you can train specifically for your task. These layers should be added on top of the frozen base model.

4) Train the model: Once you have added your new layers, you can compile and train your model using your own dataset. This will allow the model to learn task-specific features that are relevant to your problem.

5) Fine-tune the model: After training, you may want to fine-tune your model by unfreezing some of the layers in the base model and training them along with the new layers. This can help to further improve the performance of your model.

6) Evaluate the model: Once the model has been trained and fine-tuned, it is important to evaluate its performance on a validation set to ensure that it is not overfitting to the training data. You can use metrics such as accuracy, precision, recall and F1 score to evaluate the performance of your model.

7) Tune hyperparameters: You may need to tune hyperparameters such as learning rate, batch size, and number of epochs to optimize the performance of your model. This can be done using techniques such as grid search or random search.

8) Deploy the model: Finally, once the model has been trained and evaluated, it can be deployed in a production environment to make predictions on new, unseen data. This can be done using various deployment strategies such as containerization or serverless functions.

Maintaining a modest learning rate during fine-tuning is an important strategy to avoid over and under-distorting the CNN weights. The learning rate determines the step size at each iteration during the optimization process and a high learning rate can result in large weight updates that may cause the weights to diverge or oscillate. On the other hand, a low learning rate may result in slow convergence or getting stuck in local minima. A modest learning rate strikes a balance between these two extremes, allowing the model to converge towards an optimal solution without over-distorting the weights. The identification of both communicable and non-communicable diseases is done using Transfer learning techniques as shown in Fig. 14.

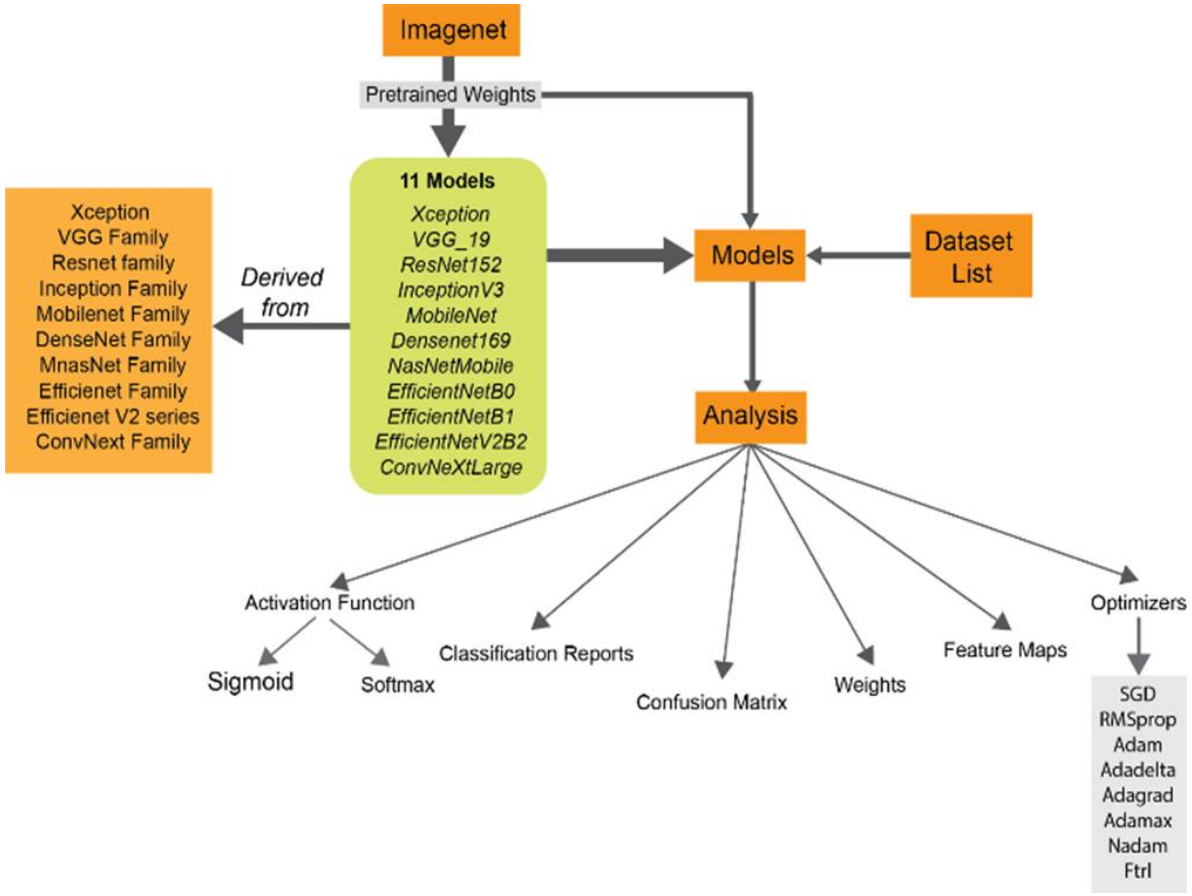


Fig. 14. Approach and analysis of the Transfer Learning Models for plant disease detection in Lettuce, Soybean and Banana.

V. Implementation and Results

The accuracy of 11 TL models selected from Table II, as well as average F1 score for each class belonging to both spreadable and non-spreadable kinds, are shown in Table III. Table IV considers the average F1 score and accuracies for the Lettuce dataset, which contains 7 classes for infectious diseases.

Accuracy is employed when True Positives as well as True Negatives are more necessary, but F1-score is employed when False Negatives but also False Positives are essential.

While F1-score is a superior measure when there are unbalanced classes, as in the example above, accuracy may be utilized when the class distribution is similar. Due to the uneven class distribution that characterizes the majority of real-world classification tasks, F1-score is a superior statistic to use when assessing the model.

A. What is ConvNext?

The science of computer vision has long employed residual networks like ResNets. Because of its smaller Residual Block design, it is considerably simpler to train deep neural networks employing skip connections. ResNet will serve as the beginning point because of their incredible accomplishment. The network will be gradually improved from this starting point, and after each enhancement, its performance will be assessed using the dataset and compared to vision transformers.

1) ConvNext: A ConvNet that outperforms Vision Transformers in terms of accuracy, performance and scalability while having the structural simplicity of Convolutional Neural Networks.

B. Assessment of ConvNext

In comparison to its vision transformer contemporaries, the new ConvNet, termed ConvNeXt, is not only more accurate but also more scalable. The graph of Fig. 15 compares ConvNext models to their equivalent vision transformers in ImageNet-1K [38].

The Table V displays the accuracy of the 11 models that were trained on the Banana dataset for all 5 classes. The ConvNeXtXLarge model performs best by yielding the most accurate findings, but the baseline model, NASNetMobile, is inappropriate for datasets based on plants, as can be seen in Table V.

The outcomes of Transfer Learning models developed for three separate datasets—lettuce, soybean and banana under various categorizations, infectious and non-infectious illnesses, were covered in the section above. The behavior of the models trained on the same three datasets but combined is covered.

TABLE II. MODEL DESCRIPTION

Model	Description
Xception	It makes use of the Xception 71-layer deep convolutional neural network. More than one million pictures from the Imagenet dataset may be used to preload a network that has previously been pretrained. The pretrained network will categorize photos into far more than a thousand more object categories in addition to keyboards, mice, pencils, and other animals.
VGG_19	The total number of layers in the convolutional neural network VGG-19 is 19. More than one million pictures from ImageNet database may be used to preload a network which has previously been pretrained. The pretrained network will categorize photos into far more than a thousand more object categories in addition to keyboards, mice, pencils and other animals.
ResNet152	Detailed Retention Learning, recognizing images with ResNet-152. The bottleneck in TorchVision occurs at the second 3x3 convolution, as opposed to initial 1x1 convolution in the original work. ResNet V1 is the modification that improves accuracy.
InceptionV3	On ImageNet dataset, it has been demonstrated that InceptionV3 image recognition model achieves greater than 78.1% accuracy.
MobileNet	The MobileNet model is a network model that uses depthwise separable convolution as its fundamental unit. It has two layers in its depthwise separable convolution: depthwise convolution and point convolution.
Densenet169	The suggested model includes 4 convolutional layers, 2 maxpool layers, 1 fully connected layer, and three dense layers.
NasNetMobile	More than a million photos from the ImageNet collection were used to train the NASNet-Mobile convolutional neural network. There are more than 1000 different object categories which the network can identify in images, including keyboards, mouse, pens and other animals.
EfficientNetB0	The architecture EfficientNetB0 is launched. The output of this function is a Keras image classification model that can be trained using weights from ImageNet.
EfficientNetB1	The CNN construction and scaling approach EfficientNetB1 equally scales all depths, width and resolution parameters using a compound coefficient.
EfficientNetV2B2	It is a brand-new class of convolutional networks that train faster and more efficiently than older models. We develop this family of models by combining training-aware neural architecture search with scaling to jointly improve training speed and parameter efficiency. A search region that had been widened to include fresh processes like Fused-MBConv was utilized to hunt up the models. Our testing show that EfficientNetV2 models train up to 6.8 times faster than state-of-the-art models despite being much smaller.
ConvNeXtXLarge	ConvNeXT, is said to exceed Vision Transformers in terms of performance (ConvNet).

TABLE III. F1 SCORE AND ACCURACIES OF SOYBEAN DATASETS

Models	F1 Score (Non-Spreadable)	Accuracy (Non-Spreadable)	F1 Score (Spreadable)	Accuracy (Spreadable)
Xception	0.666	69.7248	0.916	91.4634
VGG_19	0.608	62.8440	0.86	85.9756
ResNet152	**0.826**	**85.3211**	**1**	**100.0000**
InceptionV3	0.578	62.8440	0.89	89.0244
MobileNet	0.704	73.3945	0.976	97.5610
Densenet169	0.816	83.0275	0.983	98.1707
NasNetMobile	0.32	44.4954	0746	75.6098
EfficientNetB0	0.894	90.3670	0.993	99.3902
EfficientNetB1	0.896	90.8257	0.993	99.3902
EfficientNetV2B2	**0.86**	**87.6147**	**1**	**100.0000**
ConvNeXtXlarge	**0.904**	**92.2018**	**1**	**100.0000**

TABLE IV. F1 SCORE AND ACCURACIES OF LETTUCE DATASETS

Models	F1 Score(Spreadable)	Accuracy (Spreadable)	F1 Score(Non Spreadable)	Accuracy(Non-Spreable)
Xception	0.751	77.0889	1.00	100.0000
VGG_19	0.744	76.8194	0.993	99.2366
ResNet152	0.945	95.1482	1.00	100.0000
InceptionV3	0.764	77.6280	0.926	92.3664
MobileNet	0.677	72.2371	0.993	99.2366
Densenet169	0.955	96.2264	1.00	100.0000
NasNetMobile	0.710	71.6981	0.96	96.1832
EfficientNetB0	0.820	94.0700	0.97	96.9465
EfficientNetB1	0.935	94.0700	0.993	99.2366
EfficientNetV2B2	0.942	94.6091	1.00	100.0000
ConvNeXtXLarge	**0.961**	**96.4959**	**1.00**	**100.0000**

TABLE V. ACCURACY OF BANANA DATASET OF ALL KINDS OF PARTS OF THE PLANT

Models	Accuracy (Stem)	Accuracy (node)	Accuracy (Leaf)	Accuracy (Fruit)	Accuracy (Non-Infectious)
Xception	87.4477	100.000	85.6557	60.344	93.3824
VGG_19	80.3347	100.000	82.7869	93.103	81.6176
ResNet152	94.1423	100.000	98.3607	100.00	99.2647
InceptionV3	82.0084	100.000	82.3770	98.275	94.1176
MobileNet	88.2845	100.000	94.6721	51.724	99.2647
Densenet169	93.3054	100.000	98.3607	100.0000	99.2647
NasNetMobile	**44.3515**	**92.2078**	**74.1803**	**94.827**	**76.4706**
EfficientNetB0	94.1423	100.000	95.4918	100.0000	98.5294
EfficientNetB1	94.9791	100.000	97.1311	100.0000	98.5294
EfficientNetV2B2	95.3975	100.000	98.3607	100.0000	99.2647
ConvNeXtXLarge	**95.3975**	**99.3506**	**99.5902**	**98.275**	**100.0000**

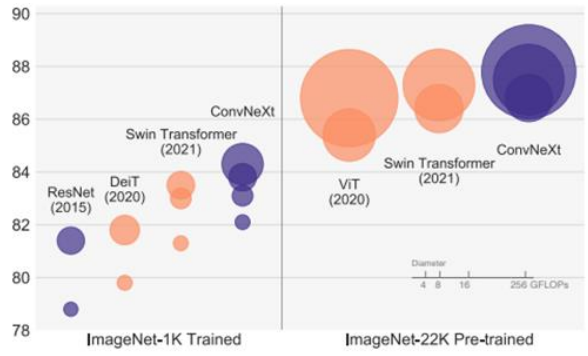


Fig. 15. Classification results for ConvNets and vision Transformers [38].

The accuracy achieved after combining all disease categories from the datasets of lettuce, soybean, and banana, which include both infectious and non-infectious conditions, is displayed in Tables VI through Table VIII.

EfficientNetV2B2 has the maximum accuracy of 95.81% from the Table VI, compared to a baseline accuracy of 56.38% from the NASNetMobile model.

C. What is EfficientNetV2?

The successor to EfficientNets is EfficientNetV2. EfficientNet is a family of models that was unveiled in 2019 and is optimised for FLOPs and parameter efficiency [39]. It makes use of neural architecture search to find the EfficientNet-B0 baseline model with the best possible accuracy and FLOPs trade-off. EfficientNets aggressively increase picture size, which results in high memory use and slow training. To overcome this problem, slightly altered the scaling rule and limited the maximum picture size to a lower amount. EfficientNetV2's technology, Deep learning models and training set both are becoming bigger and bigger. Efficiency in training is crucial in this situation. For instance, few-shot learning is demonstrated by the GPT-3 model, which has unheard-of model and training data volumes. However, retraining or enhancing the model is challenging because it takes weeks to train with thousands of GPUs. To construct this model, it combines scaling and training-aware Neural Architecture Search (NAS)to maximize training time and parameter effectiveness.

The findings of the models are remarkably comparable to those of the lettuce dataset, as can be seen from the Table VII, where EfficientNetV2B2 achieves a 93.97% accuracy while NASNetMobile achieves a baseline accuracy of 40.66%.

TABLE VI. ACCURACY OF THE COMBINED DATASET(SPREADABLE + NON SPREADABLE) OF LETTUCE

Models	F1 Score	Accuracy(Lettuce)
Xception	**0.7566**	**77.3128**
VGG_19	**0.7366**	**73.3480**
ResNet152	0.9411	94.7137
InceptionV3	0.7794	79.7357
MobileNet	0.8477	85.4626
Densenet169	0.9522	95.3744
NasNetMobile	**0.5533**	**56.3877**
EfficientNetB0	0.9511	95.3744
EfficientNetB1	0.9544	95.5947
EfficientNetV2B2	**0.9555**	**95.8150**
ConvNeXtXLarge	**0.9511**	**95.1542**

TABLE VII. ACCURACY OF THE COMBINED DATASET(SPREADABLE + NON SPREADABLE) OF SOYBEAN

Models	F1 Score	Accuracy(Soybean)
Xception	**0.7028**	**72.5904**
VGG_19	**0.6585**	**69.5783**
ResNet152	0.9142	93.3735
InceptionV3	0.6914	70.4819
MobileNet	0.6971	72.5904
Densenet169	0.8600	88.2530
NasNetMobile	**0.3414**	**40.6627**
EfficientNetB0	0.9057	92.1687
EfficientNetB1	0.8828	90.3614
EfficientNetV2B2	**0.9242**	**93.9759**
ConvNeXtXLarge	**0.9185**	**93.3735**

TABLE VIII. ACCURACY OF THE COMBINED DATASET (SPREADABLE + NON SPREADABLE) OF BANANA

Models	F1 Score	Accuracy(Banana)
Xception	**0.9362**	**95.1865**
VGG_19	**0.8877**	**89.7714**
ResNet152	0.9762	98.3153
InceptionV3	0.9408	95.9085
MobileNet	0.9677	97.1119
Densenet169	0.9746	98.0746
NasNetMobile	**0.6562**	**65.3430**
EfficientNetB0	0.9608	97.2323
EfficientNetB1	0.9685	97.7136
EfficientNetV2B2	**0.8992**	**98.0746**
ConvNeXtXLarge	**0.9100**	**95.3069**

The Table VIII makes it very evident that ResNet model has the maximum accuracy of 98.31% whereas NASNetMobile Baselines are at 65.34%.

D. What is ResNet?

ResNet was created with the goal of resolving computer vision issues. Deep residual networks that take advantage of remaining blocks to increase model precision. The concept of "skip connections," which is the foundation of the residual blocks, is the strength of this type of neural network.

1) 'Skip Connections' in ResNet: There are two ways that these skip connections work. The gradient is given a new shortcut to employ in order to address the issue of the fading gradient. Additionally, they give the model the capacity to pick up an identity function. This ensures that the performance of the model's top tiers is equal to or better than that of its lower layers. In conclusion, the residual blocks lets the layers acquire identity functions considerably more quickly. ResNet therefore decreases errors while boosting the efficiency for deep neural networks with far more neural layers. In other words, the skip connections integrate the outputs of older layers with outputs from stacked layers, enabling the training of networks that are far deeper than was previously possible.

Final point: ResNet, sometimes referred to as residual network, was a crucial development that changed how deep convolutional neural networks are trained for computer vision tasks. The venerable Resnet featured 34 layers with 2-layer bottleneck blocks, while more advanced models, like the Resnet50, used 3-layer bottleneck blocks that guarantee greater accuracy and faster training.

Tables VI to VIII shows the models in bold that are being examined for improvement of outcomes by modifying a hyperparameter, particularly the optimizer.

The optimizers that are considered for research work are:

- Adadelta
- Adagrad
- Adam
- Adamax
- Ftrl
- Nadam
- RMSprop
- SGD

2) Stochastic Gradient Descent (SGD): This is a typical 'base' optimizer, and many others are variations on it [40],[41]. It is adjustable by varying the learning rate, momentum and decay.

a) Learning rate: The learning rate controls the magnitude of parameter updates at each iteration of the optimization algorithm. A higher learning rate allows for larger updates, potentially leading to faster convergence but also increasing the risk of overshooting the optimal solution. On the other hand, a lower learning rate results in smaller updates, which may slow down convergence but can help the model settle into a more accurate and stable solution.

b) Momentum: propels SGD in the desired direction while dampening oscillations. Essentially, it allows SGD to push past local optima, resulting in quicker convergence and reduced oscillation. A normal momentum value is between 0.5 and 0.9.

c) Decay: For the learning rate, you can provided a decay function. As training advances, this will alter the learning rate. Decay functions are- Time delay, Step delay and Exponential delay.

d) Nesterov: Nesterov momentum is a variant of the momentum method that provided better theoretical converge guarantees for convex functions. In practice, it is somewhat more effective than conventional momentum.

3) Adaptive learning rate optimizers

a) Adagrad: Adagrad is an optimizer with variable parameter-specific learning rates based on how frequently a parameter is altered during training [42].

b) Adadelta: Adadelta is an optimizer that dynamically adapts the learning rate during training without the need for a predefined initial learning rate. It uses a combination of the gradient information and a moving average of the past gradients to adjust the learning rate at each iteration, allowing for efficient convergence [43]. The learning rate in Adadelta is not explicitly set by the user but is internally calculated based on the algorithm's parameters and the gradient history.

c) RMSprop: RMSprop, like Adadelta, modifies the Adagrad technique in a very easy way to lessen its aggressive, monotonically declining learning rate [44].

d) Adam: Adam is an RMSProp optimizer update. It's essentially RMSprop with momentum [45].

e) Adamax: It is a first-order gradient-based optimization approach and a version of Adam based on the infinite norm. It is well suited to learning time-variant processes, such as voice data with dynamically changing noise circumstances, because to its capacity to alter the learning rate based on data features.

f) Nadam: Similarly, to how Adam is RMSprop with momentum, Nadam is Adam with Nesterov momentum.

4) Ftrl: "Follow The Regularized Leader" (FTRL) is an optimization technique created by Google in the early 2010s for click-through rate prediction [46]. It works well with shallow models with vast and sparse feature areas, this was discussed by McMahan et al. (2013). Both online L2 regularization (the L2 regularization described in the study above) and shrinkage-type L2 regularization are supported in the Keras version (which is the addition of an L2 penalty to the loss function).

The results of the cluster of models are not as linear as shown in Tables IX to XI, and the behavior of each model with its corresponding optimizer differs for different datasets.

The adamax optimizer dominates in every other model, whereas SGD optimizers have low efficiency rate due to its poor processing performance. SGD is a very fundamental technique that is seldom employed in applications nowadays. Another issue with the method is, its constant learning rate for each epoch. Furthermore, it is not particularly good at dealing with saddle points. Because of the frequent modifications in the learning rate, Adagrad performs better than stochastic gradient descent in general. It works well when dealing with sparse data. RMSProp produces comparable results to the gradient descent technique using momentum; the only difference is how the gradients are calculated. Finally, the Adam optimizer inherits the best aspects of RMSProp and other algorithms.

The Adamax optimizer provideds a faster computation time, provided better results than other optimization techniques, it requires fewer tuning parameters. Adam is recommended as default optimizer for majority of applications as a consequence of all of this. Any application may have the highest chance of producing the finest outcomes if Adamax optimizer is used.

Finally, we discovered that even Adamax optimizer had certain drawbacks. In some circumstances, algorithms like as SGD may be more useful and perform better than the Adam optimizer. To pick the finest optimization method and obtain great results, it is critical to understand the needs and the type of data dealt with.

Since the performance of NASNet model with Adamax as optimizer , VGG19 model with Adagrad as optimizers and Xception model with Adamax as optimizer for Lettuce dataset NASNet model with Ftrl as optimizer, VGG19 model with Adagrad as optimizer and Xception model with Adamax as optimizer for soybean dataset and NASNet model with Adam as optimizer, Xception model with Adamax as optimizer for Banana dataset, although is highest with respect to other optimizer, yet its accuracies are incompatible for practical consideration.

Hence it is required to improve its performance.

One way to enhance performance is to train the model so that it is exclusively prepared for this dataset by initializing the weights to 0.

TABLE IX. ACCURACY OF MODELS FOR DIFFERENT OPTIMIZERS (LETTUCE DATASET)

Lettuce Dataset			
Sl No	**Model**	**Optimizer**	**Accuracy**
1	ConvNeXtXtLarge	Adadelta	80.17621
2	ConvNeXtXtLarge	Adagrad	92.73128
3	ConvNeXtXtLarge	Adam	95.15419
4	**ConvNeXtXtLarge**	**Adamax**	**96.9163**
5	ConvNeXtXtLarge	Ftrl	20.26432
6	ConvNeXtXtLarge	Nadam	92.95154
7	ConvNeXtXtLarge	RMSprop	95.15419
8	ConvNeXtXtLarge	SGD	92.73128
9	EfficientNetV2B2	Adadelta	32.59912
10	EfficientNetV2B2	Adagrad	89.86784
11	EfficientNetV2B2	Adam	95.81498
12	EfficientNetV2B2	Adamax	95.81498
13	EfficientNetV2B2	Ftrl	95.81498
14	EfficientNetV2B2	Nadam	95.81498
15	**EfficientNetV2B2**	**RMSprop**	**96.25551**
16	EfficientNetV2B2	SGD	89.86784
17	NASNetMobile	Adadelta	10.79295
18	NASNetMobile	Adagrad	36.78414
19	NASNetMobile	Adam	56.38767
20	**NASNetMobile**	**Adamax**	**58.81057**
21	NASNetMobile	Ftrl	56.38767
22	NASNetMobile	Nadam	56.38767
23	NASNetMobile	RMSprop	41.18943
24	NASNetMobile	SGD	46.69604
25	VGG_19	Adadelta	64.97797
26	**VGG_19**	**Adagrad**	**84.14097**
27	VGG_19	Adam	74.6696
28	VGG_19	Adamax	83.70044
29	VGG_19	Ftrl	20.26432
30	VGG_19	Nadam	78.19383
31	VGG_19	RMSprop	22.68722
32	VGG_19	SGD	9.69163
33	Xception	Adadelta	30.17621
34	Xception	Adagrad	68.06167
35	Xception	Adam	75.11013
36	**Xception**	**Adamax**	**83.48018**
37	Xception	Ftrl	20.26432
38	Xception	Nadam	73.78855
39	Xception	RMSprop	78.85463
40	Xception	SGD	35.46256

TABLE X. ACCURACY OF MODELS FOR DIFFERENT OPTIMIZERS (SOYBEAN DATASET)

Soybean Dataset			
Sl No	**Model**	**Optimizer**	**Accuracy**
1	ConvNeXtXtLarge	Adadelta	82.22892
2	ConvNeXtXtLarge	Adagrad	89.15663
3	ConvNeXtXtLarge	Adam	93.37349
4	**ConvNeXtXtLarge**	**Adamax**	**94.57831**
5	ConvNeXtXtLarge	Ftrl	14.71698
6	ConvNeXtXtLarge	Nadam	93.1677
7	ConvNeXtXtLarge	RMSprop	93.6747
8	ConvNeXtXtLarge	SGD	90.66265
9	EfficientNetV2B2	Adadelta	93.9759
10	EfficientNetV2B2	Adagrad	88.55422
11	EfficientNetV2B2	Adam	93.9759
12	EfficientNetV2B2	Adamax	92.46988
13	EfficientNetV2B2	Ftrl	86.74699
14	**EfficientNetV2B2**	**Nadam**	**96.08434**
15	EfficientNetV2B2	RMSprop	94.57831
16	EfficientNetV2B2	SGD	88.55422
17	NASNetMobile	Adadelta	12.3494
18	NASNetMobile	Adagrad	31.3253
19	NASNetMobile	Adam	38.55422
20	NASNetMobile	Adamax	43.07229
21	**NASNetMobile**	**Ftrl**	**43.9759**
22	NASNetMobile	Nadam	43.6747
23	NASNetMobile	RMSprop	40.66265
24	NASNetMobile	SGD	40.66265
25	VGG_19	Adadelta	59.33735
26	**VGG_19**	**Adagrad**	**80.72289**
27	VGG_19	Adam	58.43373
28	VGG_19	Adamax	70.18072
29	VGG_19	Ftrl	23.79518
30	VGG_19	Nadam	68.9759
31	VGG_19	RMSprop	24.6988
32	VGG_19	SGD	16.26506
33	Xception	Adadelta	40.66265
34	Xception	Adagrad	64.75904
35	Xception	Adam	66.86747
36	**Xception**	**Adamax**	**74.6988**
37	Xception	Ftrl	19.27711
38	Xception	Nadam	73.19277
39	Xception	RMSprop	70.48193
40	Xception	SGD	48.79518

TABLE XI. ACCURACY OF MODELS FOR DIFFERENT OPTIMIZERS (BANANA DATASET)

Banana Dataset			
Sl No	**Model**	**Optimizer**	**Accuracy**
1	ConvNeXtXtLarge	Adadelta	95.29653828
2	**ConvNeXtXtLarge**	**Adagrad**	**98.31528279**
3	ConvNeXtXtLarge	Adam	95.4356249
4	ConvNeXtXtLarge	Adamax	97.35258724
5	ConvNeXtXtLarge	Ftrl	95.30685921
6	ConvNeXtXtLarge	Nadam	95.16727657
7	ConvNeXtXtLarge	RMSprop	95.3078993
8	ConvNeXtXtLarge	SGD	95.30685921
9	EfficientNetV2B2	Adadelta	43.92298436
10	EfficientNetV2B2	Adagrad	97.83393502
11	EfficientNetV2B2	Adam	97.71359807
12	**EfficientNetV2B2**	**Adamax**	**98.0746089**
13	EfficientNetV2B2	Ftrl	85.19855596
14	EfficientNetV2B2	Nadam	97.95427196
15	EfficientNetV2B2	RMSprop	97.87426738
16	EfficientNetV2B2	SGD	97.95427196
17	NASNetMobile	Adadelta	8.664259928
18	NASNetMobile	Adagrad	37.18411552
19	**NASNetMobile**	**Adam**	**88.56799037**
20	NASNetMobile	Adamax	15.04211793
21	NASNetMobile	Ftrl	13.35740072
22	NASNetMobile	Nadam	86.64259928
23	NASNetMobile	RMSprop	69.7954272
24	NASNetMobile	SGD	36.70276775
25	**VGG_19**	**Adadelta**	**98.31528279**
26	VGG_19	Adagrad	98.0746089
27	VGG_19	Adam	77.61732852
28	VGG_19	Adamax	67.99037304
29	VGG_19	Ftrl	14.92178099
30	VGG_19	Nadam	90.49338147
31	VGG_19	RMSprop	76.89530686
32	VGG_19	SGD	8.784596871
33	Xception	Adadelta	31.28760529
34	Xception	Adagrad	47.17208183
35	Xception	Adam	90.97472924
36	**Xception**	**Adamax**	**93.50180505**
37	Xception	Ftrl	13.47773767
38	Xception	Nadam	87.36462094
39	Xception	RMSprop	89.89169675
40	Xception	SGD	39.95186522

Why didn't the process initially explore retraining or fine-tuning the model from scratch?

Due to the small amount of data, creating new models from start would be a resource and time-intensive operation with no assurance of performance. These models were previously trained quite effectively. In order to improve performance efficacy in our case, it is thus preferable practice to load those pertained models and use the information that the two models have previously acquired in the course of their original work. Transfer learning and fine-tuning are commonly confused with one another since they are both parts of the same process. Many refer to the entire process as fine-tuning since we often do so after transfer learning.

However, fine-tuning involves more than just applied the weights from the pre-trained models. In order to adjust the model to the present job, it is also using prior information but freezing some layers while training the last layers at a slow learning rate. Convolution deep learning model results are shown to provide a better understanding of the entire process.

TABLE XII. ACCURACIES OF MODELS BEFORE AND AFTER RETRAINING WITH THE DATASETS

Sl No	Model	Optimizer	Accuracy before retraining	Accuracy after retraining
Lettuce Dataset				
1	NASNetMobile	Adamax	58.81057269	71.36564
2	VGG_19	Adagrad	84.14096916	90.30837
3	**Xception**	**Adamax**	**83.48017621**	**96.69604**
Soybean Dataset				
4	NASNetMobile	Ftrl	43.97590361	19.27711
5	VGG_19	Adagrad	80.72289157	82.22892
6	**Xception**	**Adamax**	**74.69879518**	**92.46988**
Banana Dataset				
7	NASNetMobile	Adam	88.56799037	33.81468
8	**Xception**	**Adamax**	**93.50180505**	**98.19495**

E. Concept of Underfitting and Overfitting

Why poor accuracy is viewed for few models over other models with various optimizers?

If a model adequately generalizes all new input data from the problem domain, it is considered to be a good machine learning model. Additionally, underfitting and overfitting are the main reasons why machine learning algorithms perform poorly [47].

F. Concept of Bias, Variance, Underfitting and Overfitting

1) Bias: Essentially, it is the error rate of the training data. Whenever the error margin is high, we say the bias is strong, while when it is low, the bias is low.

2) Variance: The variance is the difference in the error margin between the training and test sets of data. The variance is described as being high when it is large and low whenever the difference between both the errors is small. Usually, we want to expand our model with the least amount of variance.

3) Underfitting: Underfitting is the term used whenever a statistical model as well as machine learning algorithm fails to capture the overall pattern of the data, i.e., when it performs well on training data but poorly on testing data. Its recurrence simply shows that model or method doesn't really adequately fit the data. It frequently happens when there are not enough data to build a solid model or when we try to build a linear model with too little non-linear data. Because its rules are too basic and flexible to be applied to such scant data, a machine learning model will probably make a number of inaccurate predictions under these circumstances. Underfitting may be avoided by utilizing more data and restricting the features through feature selection [48].

Underfitting, is when a model is unable to perform satisfactorily on the training dataset or generalize to new data.

G. Justifications for underfitting

- Low variance and high covariance.
- The used training dataset's size is insufficient.
- The model is rather basic.
- Training data has noise in it and is not being eliminated.

Methods to reduce underfitting

- Amplify model complexity.
- Boost feature count by doing feature engineering.
- Data noise should be removed.
- To achieve better outcomes, increase the period of training or the number of epochs.

1) Overfitting: An overfitted statistical model is one that cannot accurately predict events, based on test data [49]. When a model has been trained with a massive quantity of data, it begins to gain knowledge from the disturbance and incorrect data entries in the given dataset and when test data is used for testing, there is a lot of diversity. The model is unable to correctly recognize the data because of the overabundance of characteristics and distortion. Since these give machine learning algorithms greater freedom to build the model depending on the dataset, non-parametric and non-linear techniques are the primary sources of overfitting and can result in extremely illogical models [50]. Using a linear approach to analyze linear data is one strategy to avoid overfitting.

Overfitting, is a problem when the evaluation of machine learning algorithms on unknown data varies from the analysis on training data.

Overfitting has the following causes:

- Both variance and bias are high.
- The model is very sophisticated.
- The volume of training data.

Methods to reduce overfitting

- Expand the training data.
- Simplify the model.
- During the training phase usage of early stopping stay updated on the loss during the training period, and cease training as soon as it starts to rise.
- Regularization of the Ridge and the Lasso [51].
- To mitigate overfitting in neural networks, using a dropout technique.

H. How to Fit a Statistical Model Well?

When a statistical model makes predictions that are error-free, that is the ideal situation, it is said to be a good match with the data. This situation may exist anywhere between overfitting and underfitting. To understand the model, we need to look at how it performs over time as it learns from the training dataset.

As time goes on, the model would continue to learn, as a result, the model's accuracy just on test & training data will decrease over time. If the model is given an abnormally lengthy time to learn, the accumulation of junk and less important characteristics can make it more susceptible to overfitting. The model's performance will therefore suffer. In order to obtain a good match, you must stop just as the error starts to grow worse. In both our concealed testing dataset & training datasets, the model is judged as being competent at this point.

I. How to find whether the model chosen is overfit or underfit?

When the validation accuracy increases after retraining and subsequently sharply decreases, the model is overfit. While in the event of underfit, there is just a slow, lower value rise in validation accuracy.

Models with excellent accuracy prior to retraining but significantly reduced accuracy after retraining are shown in Table XII. The validation accuracy and test accuracy in Table XIII are used to determine whether the model is overfit or underfit.

The model training details show that none of the models under consideration are overfit, but NASNetMobile model with Ftrl as optimizer for soybean dataset and NASNetMobile model with Adam as optimizer for banana dataset is underfit. This is because the mentioned model has good training accuracy but low validation and test accuracy making it underfit.

Also, the presence of overfit model does not affect the accuracies in present research case because of consideration of best fit model and early stopping criteria.

TABLE XIII. ACCURACIES OF MODEL WITH VALIDATION ACCURACIES

Sl.No	Model	Optimizer	Accuracy (when retrained)	Validation accuracy (when retraing)	Accuracy (after retraining)
Lettuce Dataset					
1	NASNet Mobile	Adamax	99.49	73.951	71.365
2	VGG_19	Adagrad	99.93	87.23	90.308
3	Xception	Adamax	100	96.689	96.696
Soybean Dataset					
4	**NASNet Mobile**	**Ftrl**	**75.89**	**19.219**	**19.277**
5	VGG_19	Adagrad	100	83.784	82.228
6	Xception	Adamax	99.61	96.096	92.469
Banana Dataset					
7	**NASNet Mobile**	**Adam**	**92.9**	**33.454**	**33.814**
8	Xception	Adamax	96.75	91.095	98.194

VI. PROPOSED METHODOLOGY

In the context of proposing a methodology for developing a high-accuracy convolutional neural network (CNN) model, particularly for tasks like image classification, the method integrates cutting-edge architectural principles from EfficientNetV2B2 with custom-designed elements, specifically a novel activation function and optimizer. Here's a detailed breakdown of the proposed methodology.

To further elucidate the rationale behind choosing the specific combination of ReLU and Tanh for the custom activation function and the integration of Adam and SGD characteristics in the custom optimizer, we can refer to the previously discussed implementation and general principles in neural network optimization and activation functions as shown in Fig. 16.

A. Components of the Methodology

1) Base Architecture (EfficientNetV2):

- EfficientNetV2B2 [52] is chosen as the foundational architecture due to its state-of-the-art performance in image classification tasks.
- It utilizes a compound scaling method that uniformly scales the depth, width, and resolution of the network, making it highly efficient and effective.

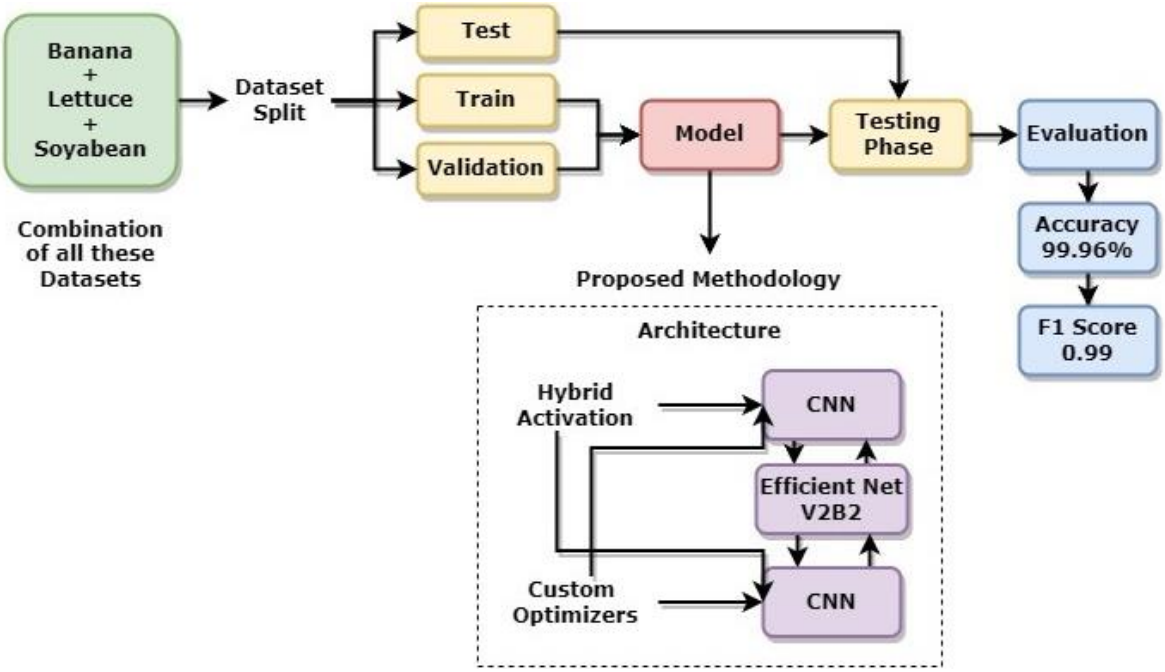


Fig. 16. Proposed methodology.

2) Custom Activation Function:

- A novel activation function is introduced to potentially improve the learning process.
- The function, custom activation, is a combination of ReLU and Tanh functions. ReLU ensures non-saturation of gradients for positive values, enhancing learning speed, while Tanh provided output normalization, potentially aiding in stabilizing the learning process.

3) Custom Optimizer:

- Developing a custom optimizer aims to enhance the training efficiency and convergence rate.
- The optimizer combines elements of Adam (adaptive learning rates) and SGD (stochastic gradient descent), attempting to utilize the benefits of both.

4) Integration and Training:

- The EfficientNetV2B2 base model is loaded with pre-trained ImageNet weights to leverage transfer learning, accelerating the training process and improving initial accuracy.
- The top layers of the model are replaced with custom layers, including Global Average Pooling and Dense layers, utilizing the custom activation function.
- The model is compiled with the custom optimizer, and categorical cross-entropy is used as the loss function, suitable for multi-class classification tasks.

5) Training Strategy:

- Initially, the EfficientNetV2B2 base layers are frozen to preserve the pre-trained features, and only the custom top layers are trained.
- Subsequently, fine-tuning can be performed by unfreezing some of the top layers of the base model and continuing the training, allowing for refined feature extraction tailored to the specific dataset.

6) Evaluation:

- The model's performance is evaluated using accuracy metrics on a validation dataset.
- Regular checkpoints and monitoring are employed to track the training progress and prevent overfitting.

B. Integration of Concepts Based on Previous Results

The decision to combine these specific elements from ReLU, Tanh, Adam, and SGD is not only based on their individual strengths but also on empirical observations from previous implementations:

1) ReLU and Tanh: The combined use of ReLU and Tanh in various architectures has shown promising results in terms of faster convergence and improved accuracy, as these functions complement each other's properties.

2) Adam and SGD: Similarly, the integration of Adam's adaptive learning rate and SGD's generalization capabilities aims to create a more robust and efficient optimizer. This is based on observations where models trained with Adam initially show rapid improvement but sometimes fail to achieve the level of generalization that SGD can provided.

3) Outcomes:

- Enhanced Model Performance: By integrating the architectural efficiency of EfficientNetV2 with the novel elements of the custom activation function and optimizer, the model has demonstrated remarkable performance in image classification tasks. Notably, this approach has achieved a remarkable accuracy of 99.96%, positioning

it at the forefront of current image classification models. This high level of accuracy indicates an exceptional ability of the model to correctly classify images, minimizing both false positives and false negatives.

- Superior F1 Score: Alongside accuracy, the model has achieved an F1 score of 0.99. The F1 score is a more complex metric that considers both precision and recall, providing a more holistic view of the model's performance. An F1 score of 0.99 implies that the model not only accurately classifies the positive cases but also maintains a high rate of successfully identifying true negatives. This balance is crucial in scenarios where both types of classification errors carry significant consequences.
- Improved Learning Dynamics: The custom activation function and optimizer have played a pivotal role in enhancing the training process. These custom elements have contributed to improved training stability and accelerated convergence speed, enabling the model to quickly adapt and optimize its performance. The combination has been instrumental in achieving the high accuracy and F1 score, underlining the effectiveness of these custom components in handling complex learning tasks.

VII. Conclusion

This paper focuses on the use of transfer learning for the identification of spreadable and non-spreadable plant diseases. The study considers three different plant types, namely lettuce, soybean, and banana, and addresses the classification of the most prevalent diseases in these plants. The diseases are categorized into spreadable and non-spreadable diseases, treated as distinct classes in the analysis.

To evaluate the classification accuracy of different models, a comparative research approach is employed. The performance of 11 transfer learning models available in Keras are assessed on separate datasets for spreadable and non-spreadable diseases. Additionally, the models are evaluated on a combined dataset that includes five different portions of the plants, comprising both healthy and diseased parts. Early stopping criteria are set at a minimum of 20 to 30 epochs with a patience of 6 for comparison. It is observed that the metrics and accuracy of the models vary depending on the dataset being used. However, some of the selected models did not exhibit the anticipated high accuracy after training on the datasets.

To improve the model performance, various techniques such as optimizing the models and retraining them from scratch can be employed. These strategies aim to enhance the accuracy and effectiveness of the classification models for identifying spreadable and non-spreadable plant diseases. The available optimizers in Keras are taken into consideration in order to increase accuracy, that includes SGD, RMSprop, Adam, Adadelta, Adagrad, Adamax, Nadam, and Ftrl. However, this strategy only worked for higher models (like EfficientNet models, ConvNeXt); smaller models (like VGG-19, Xception) showed less improvement. The paper's major goal was to select a lower model since it is less complicated to train for and has a smaller width. Improving them suggests an upgrade to the foundational CNN model, making the study more flexible. With certain models, there is a rapid fall in accuracy, leading it to be considered as either underfit or overfit. In the current instance, the NASNetMobile model is underfit, and situations of overfit are not evident because of the early stopping approach. VGG_19 model with Adadelta as optimizer without retraining and Xception model with Adamax as optimizer when retrained from scratch, outperform in terms of classification metrics for the datasets under consideration.

In addition to these strategies, the paper proposed a novel methodology focusing on the integration of an EfficientNetV2-style architecture with a custom-designed activation function and optimizer. The custom activation function, a hybrid of ReLU and Tanh, aims to enhance learning dynamics by combining the benefits of non-saturation (from ReLU) and output normalization (from Tanh). The custom optimizer, blending elements of Adam and SGD, is designed to achieve a balance between adaptive learning rates and effective generalization. This proposed methodology, especially with the EfficientNetV2's efficient scaling and advanced architecture, is expected to yield even higher accuracy and robustness in classifying plant diseases. Notably, this approach has achieved remarkable performance with an accuracy of 99.96% and an F1 score of 0.99 in the classification tasks, setting a new standard in the field and underscoring the effectiveness of combining advanced neural network architectures with innovative custom components for complex classification challenges.

VIII. Future Work

Moving forward, several key areas offer promising opportunities to extend and enhance the research presented in this study. One significant direction is the expansion and diversification of the dataset. By including a broader range of plant species, diseases, and environmental conditions, the models could be made more robust and generalizable across different agricultural contexts. Additionally, incorporating data from various geographical regions and employing data augmentation techniques could help address issues of overfitting and improve the model's performance on smaller or imbalanced datasets.

Integrating real-time data from environmental sensors is another avenue that could significantly enhance the predictive accuracy of the models, especially in relation to both biotic and abiotic plant stressors. By developing models capable of adapting to dynamic environmental conditions, the relevance and effectiveness of AI-driven solutions in agriculture could be substantially improved. Moreover, refining the custom activation function and optimizer introduced in this study remains an important task. Testing these components across different deep learning architectures and applications, such as pest detection or crop yield prediction, could assess their versatility and broader applicability.

Ethical considerations and societal impacts also warrant close attention. As AI-driven plant disease identification systems are deployed, it is crucial to address potential ethical issues, such as data privacy, fairness, and the implications for small-scale farmers. Moreover, the societal impacts, including potential job displacement and the need for upskilling agricultural workers,

should be carefully considered to ensure responsible and equitable deployment of these technologies.

Finally, exploring the scalability of the proposed models for large-scale farming operations is essential, particularly in terms of computational efficiency and resource constraints. Investigating cloud-based or edge-computing solutions could facilitate real-time disease detection in remote or resource-limited settings. Collaborative, multi-disciplinary research involving AI experts, agronomists, plant pathologists, and agricultural economists will be critical in developing holistic solutions that effectively address the complexities of plant disease management.

CONFLICTS OF INTEREST

The authors declare no conflict of interest.

FUNDING STATEMENT

This research did not receive any specific funding or financial support.

ETHICAL STATEMENT

This research was conducted in accordance with ethical principles, including informed consent, confidentiality, and adherence to relevant guidelines.

AUTHORS' CONTRIBUTION

Conceptualization: Asha Rani K P and Gowrishankar S jointly conceived and designed the research project.

Methodology: Asha Rani K P developed the methodology for the study.

Validation: Asha Rani K P and Gowrishankar S collectively validated the results and ensured their accuracy.

Formal Analysis: Asha Rani K P conducted the formal analysis of the data.

Investigation: Asha Rani K P and Gowrishankar S carried out the investigation and collected the necessary data.

Resources: Asha Rani K P and Gowrishankar S provided the required resources for the project.

Data Curation: Asha Rani K P curated and prepared the dataset for analysis.

Writing—Original Draft Preparation: Asha Rani K P wrote the initial draft of the manuscript.

Writing—Review and Editing: Gowrishankar S critically reviewed and Asha Rani K P edited the manuscript.

Visualization: Asha Rani K P created the visualizations used in the paper.

Both the authors have substantially contributed to the work reported and have approved the final version of the manuscript.

DATA AVAILABILITY STATEMENT

Images used for the study are obtained from CrowdAI [27] and PlantVillage dataset [28].

REFERENCES

[1] Raid, Richard N. "Lettuce Diseases and Their Management." Diseases of Fruits and Vegetables: Volume II, n.d., 121–47.

[2] Carrasco, Gilda A., and S. W. Burrage. "Diurnal Fluctuations in Nitrate Accumulation and Reductase Activity In Lettuce (LACTUCA SATIVA L.) Grown using Nutrient Film Technique" Acta Horticulturae, no. 323, International Society for Horticultural Science (ISHS), Feb. 1993, pp. 51–60. Crossref, https://doi.org/10.17660/actahortic.1993.323.3.

[3] Singh, Gaurav, Garima Dukariya, and Anil Kumar. "Distribution, Importance and Diseases of Soybean and Common Bean: A Review." Biotechnology Journal International, 2020, 86–98.

[4] Wahome, C. N., Maingi, J. M., Ombori, O., Kimiti, J. M., & Njeru, E. M. (2021). Banana production trends, cultivar diversity, and tissue culture technologies uptake in Kenya. International Journal of Agronomy, 2021, 1–11. https://doi.org/10.1155/2021/6634046

[5] K . Lakshmi Narayanan ,1 R. Santhana Krishnan ,2 Y. Harold Robinson , E. Golden Julie , S. Vimal , V. Saravanan , and M. Kaliappan5. "Banana Plant Disease Classification Using Hybrid Convolutional Neural Network"

[6] Andreas Kamilaris, Francesc X. Prenafeta-Boldú, "Deep learning in agriculture: A survey", Computers and Electronics in Agriculture, Volume 147, Pages 70-90, ISSN 0168-1699, 2018.

[7] Ramcharan, Amanda, Kelsee Baranowski, Peter McCloskey, Babuali Ahmed, James Legg, and David P. Hughes. "Deep Learning for Image-Based Cassava Disease Detection." Frontiers in Plant Science 8 (2017)

[8] N.Saranya, L. Pavithra, N. Kanthimathi, B. Ragavi, P. Sandhiyadevi . "Detection of Banana Leaf and Fruit Diseases Using Neural Networks"

[9] Michael Gomez Selvaraj, Alejandro Vergara, Frank Montenegro , Henry Alonso Ruiz, Nancy Safari , Dries Raymaekers , Walter Ocimati , Jules Ntamwira , Laurent Tits , Aman Bonaventure Omondi , Guy Blomme "Detection of banana plants and their major diseases through aerial images and machine learning methods: A case study in DR Congo and Republic of Benin".

[10] E. Miao, Guixia Zhou, and Shengxue Zhao "Research on Soybean Disease Identification Method Based on Deep Learning"

[11] Sachin B. Jadhav, Vishwanath R. Udupi, Sanjay B. Patil, "Soybean leaf disease detection and severity measurement using multiclass SVM and KNN classifier" 26 April 2019

[12] Elham Khalili, Samaneh Kouchaki, Shahin Ramazi, Faezeh Ghanati "Machine Learning Techniques for Soybean Charcoal Rot Disease Prediction" 14 December 2020

[13] Miao Yu, Xiaodan Ma, Haiou Guan, Meng Liu, Tao Zhang "A Recognition Method of Soybean Leaf Diseases Based on an Improved Deep Learning Model" 31 May 2022

[14] Munirah Hayati Hamidon, Tofael Ahamed "Detection of Tip-Burn Stress on Lettuce Grown in an Indoor Environment Using Deep Learning Algorithms" 24 September 2022

[15] Kavir Osorio, Andrés Puerto, Cesar Pedraza, David Jamaica, Leonardo Rodríguez "A Deep Learning Approach for Weed Detection in Lettuce Crops Using Multispectral Images" 28 August 2020

[16] J. Amara, B. Bouaziz, and A. Algergawy, "A Deep Learning-based Approach for Banana Leaf Diseases Classification" in B. Bernhard Mitschang, Norbert Ritter, Holger Schwarz, Meike Klettke, Andreas Thor, Oliver Kopp, Matthias Wieland (Hrsg.): BTW 2017 – Workshopband, Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2017 79.

[17] W. Liao, D. Ochoa, L. Gao, B. Zhang, and W. Philips, "Morphological Analysis for banana disease detection in close range hyperspectral remote sensing images," in Proceedings of the IGARSS 2019 - 2019 IEEE International Geoscience and Remote Sensing Symposium, pp. 3697–3700, Yokohama, Japan, 28 July-2 August 2019.

[18] S. kaur, G. Babbar, and Gagandeep, "Image processing and classification, A method for plant disease detecion," International Journal of Innovative Technology and Exploring Engineering, vol. 8, no. 9S, 2019.

[19] Shimamura, S.; Uehara, K.; Koakutsu, S. Automatic Identification of Plant Physiological Disorders in Plant Factory Crops. IEEJ Trans. Electron. Inf. Syst. 2019, 139, 818–819.

[20] S. Mishra, R. Sachan, and D. Rajpal, "Deep convolutional neural network based detection system for real-time corn plant disease recognition," Procedia Computer Science, vol. 167, pp. 2003–2010, 2020.

[21] D. A. Kumar, P. S. Chakravarthi, and K. S. Babu, "Multiclass Support Vector Machine Based Plant Leaf Diseases Identification from Color, Texture and Shape Features," in Proceedings of the 2020 ;ird International Conference on Smart Systems and Inventive Technology (ICSSIT), pp. 1220–1226, IEEE, Tirunelveli, India, August 2020.

[22] V. Mazzia, A. Khaliq, and M. Chiaberge "Improvement in land cover and crop classification based on temporal features learning from sentinel-2 data using recurrent-convolutional neural network (R-CNN)"

[23] N. Çetin, K. Karaman, E. Beyzi, C. Sa˘glam, and B. Demirel, "Comparative evaluation of some quality characteristics of sunflower oilseeds (helianthus annuus L.) through machine learning classifiers," Food Analytical Methods, vol. 14, no. 8, pp. 1666–1681, 2021.

[24] Sharif M, Khan MA, Iqbal Z, Azam MF, Lali MIU, Javed MY. Detection and classification of citrus diseases in agriculture based on optimized weighted segmentation and feature selection. Comput Electron Agric 2018;150:220–34. https://doi.org/10.1016/j.compag.2018.04.023.

[25] Lu J, Ehsani R, Shi Y, Abdulridha J, de Castro AI, Xu Y. Field detection of anthracnose crown rot in strawberry using spectroscopy technology. Comput Electron Agric 2017.

[26] Ruben Van De Vijver, Koen Mertens, Kurt Heungens, Ben Somers, David Nuyttens, Irene Borra-Serrano, Peter Lootens, Isabel Roldán-Ruiz, Jürgen Vangeyte, Wouter Saeys, In-field detection of Alternaria solani in potato crops using hyperspectral imaging, Computers and Electronics in Agriculture, Volume 168, 2020, 105106, ISSN 0168-1699, https://doi.org/10.1016/j.compag.2019.105106.

[27] D. Zhang, Y. Zhang, Q. Li, T. Plummer and D. Wang, "CrowdLearn: A Crowd-AI Hybrid System for Deep Learning-based Damage Assessment Applications," 2019 IEEE 39th International Conference on Distributed Computing Systems (ICDCS), Dallas, TX, USA, 2019, pp. 1221-1232, doi: 10.1109/ICDCS.2019.00123.

[28] Noyan, Mehmet Alican. "Uncovering bias in the PlantVillage dataset." arXiv preprint arXiv:2206.04374 (2022).

[29] A. Mikołajczyk and M. Grochowski, "Data augmentation for improving deep learning in image classification problem," 2018 International Interdisciplinary PhD Workshop (IIPhDW), 2018, pp. 117-122, doi: 10.1109/IIPHDW.2018.8388338.

[30] Medhi, Epsita, and Nabamita Deb. "PSFD-Musa: A Dataset of Banana Plant, Stem, Fruit, Leaf, and Disease." Data in Brief 43 (2022): 108427.

[31] A. Mikołajczyk and M. Grochowski, "Data augmentation for improving deep learning in image classification problem," 2018 International Interdisciplinary PhD Workshop (IIPhDW), 2018, pp. 117-122, doi: 10.1109/IIPHDW.2018.8388338.

[32] Maraghehmoghaddam, Armin. "Synthetic Data Generation for Deep Learning Model Training to Understand Livestock Behavior," n.d. https://doi.org/10.31274/etd-20200902-98

[33] Pratiwi, Heny, Agus Perdana Windarto, S. Susliansyah, Ririn Restu Aria, Susi Susilowati, Luci Kanti Rahayu, Yuni Fitriani, Agustiena Merdekawati, and Indra Riyana Rahadjeng. "Sigmoid Activation Function in Selecting the Best Model of Artificial Neural Networks." Journal of Physics: Conference Series 1471, no. 1 (2020): 012010.

[34] Namin, Ashkan & Leboeuf, Karl & Muscedere, Roberto & Wu, Huapeng & Ahmadi, Majid. (2009). Efficient hardware implementation of the hyperbolic tangent sigmoid function. Proceedings - IEEE International Symposium on Circuits and Systems. 2117 - 2120. 10.1109/ISCAS.2009.5118213.

[35] Bodyanskiy, Yevgeniy, Anastasiia Deineko, Viktoria Skorik, and Filip Brodetskyi. "Deep Neural Network with Adaptive Parametric Rectified Linear Units and Its Fast Learning." International Journal of Computing, 2022, 11–18.

[36] Abien Fred Agarap. (2018). Deep Learning using Rectified Linear Units (ReLU).https://doi.org/10.48550/arXiv.1803.08375

[37] I. Kouretas and V. Paliouras, "Simplified Hardware Implementation of the Softmax Activation Function," 2019 8th International Conference on Modern Circuits and Systems Technologies (MOCAST), Thessaloniki, Greece, 2019.

[38] Zhuang Li, Hanzi Mao, Chao-Yaun, Christoph Feichtenhofer, Trevor Darrell and Saining Xie, "A ConNet for the 2020s" , 2020.

[39] Menghani, Gaurav. "Efficient Deep Learning: A Survey on Making Deep Learning Models Smaller, Faster, and Better." ACM Computing Surveys 55, no. 12 (2023): 1–37.

[40] Ruder, S. (2016). An overview of gradient descent optimization algorithms. arXiv preprint arXiv:1609.04747..

[41] Bottou L. (2012) Stochastic Gradient Descent Tricks. In: Montavon G., Orr G.B., Müller KR. (eds) Neural Networks: Tricks of the Trade. Lecture Notes in Computer Science, vol 7700. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-35289-8_25

[42] Duchi, J., Hazan, E., & Singer, Y. (2011). Adaptive Subgradient Methods for Online Learning and Stochastic Optimization. Journal of Machine Learning Research, 12, 2121–2159.

[43] Zhang, Rui, Weiguo Gong, Victor Grzeda, Andrew Yaworski, and Michael Greenspan. "An Adaptive Learning Rate Method for Improving Adaptability of Background Models." IEEE Signal Processing Letters 20, no. 12 (2013): 1266–69. https://doi.org/10.1109/lsp.2013.2288579.

[44] Peto, Levente, and Janos Botzheim. "Parameter Optimization of Deep Learning Models by Evolutionary Algorithms." 2019 IEEE International Work Conference on Bioinspired Intelligence (IWOBI), 2019.

[45] Diederik P. Kingma and Jimmy Lei Ba. Adam: A method for stochastic optimization. 2014. arXiv:1412.6980v9 (2014)

[46] "Follow-the-regularised-leader and Mirror Descent" (2020) Bandit Algorithms, pp. 286–305. Available at: https://doi.org/10.1017/9781108571401.035.

[47] ALHAWAS, Nagham, and Zekeriya TÜFEKCİ. "The Effectiveness of Transfer Learning and Fine-Tuning Approach for Automated Mango Variety Classification." European Journal of Science and Technology, European Journal of Science and Technology, Mar. 2022. Crossref, https://doi.org/10.31590/ejosat.1082217.

[48] Kundjanasith Thonglek, Keichi Takahashi, Kohei Ichikawa, Chawanat Nakasan, Hidemoto Nakada, Ryousei Takano, Hajimu Iida, "Retraining Quantized Neural Network Models with Unlabeled Data," 2020 International Joint Conference on Neural Networks (IJCNN), 2020, pp. 1-8, doi: 10.1109/IJCNN48605.2020.9207190.

[49] A. Ghasemian, H. Hosseinmardi and A. Clauset, "Evaluating Overfit and Underfit in Models of Network Community Structure," in IEEE Transactions on Knowledge and Data Engineering, vol. 32, no. 9, pp. 1722-1735, 1 Sept. 2020, doi: 10.1109/TKDE.2019.2911585.

[50] S. K. Noon, M. Amjad, M. A. Qureshi and A. Mannan, "Overfitting Mitigation Analysis in Deep Learning Models for Plant Leaf Disease Recognition," 2020 IEEE 23rd International Multitopic Conference (INMIC), 2020, pp. 1-5.

[51] Keith, Michael. "Ridge and Lasso." Machine Learning with Regression in Python, 2020. https://doi.org/10.1007/978-1-842-6583-3_4.

[52] K. P. Asha Rani and S. Gowrishankar, "Pathogen-Based Classification of Plant Diseases: A Deep Transfer Learning Approach for Intelligent Support Systems," in IEEE Access, vol. 11, pp. 64476-64493, 2023, doi: 10.1109/ACCESS.2023.3284680.

Design and Application of Intelligent Visual Communication System for User Experience

Chao Peng
School of Art and Design, Henan Industry and Trade Vocational College, Zhengzhou, Henan 450000, China

***Abstract*—The design and application of visual communication system should be human-oriented, but currently this is often ignored by designers, resulting in poor user experience of visual communication. In order to improve the experience effect of visual communication system, combined with the existing computer technology, this paper proposes an intelligent visual communication system for user experience. First, for the problem of extracting multimodal features of users, considering the characteristics of different modal data, long and short-term memory networks are used to extract features with contextual information, and multi-scale convolutional neural networks are used for visual modality to extract low-level features from video frames. In the cross-modal stage, the low-level features in the source modality are used to enhance the target modality features. Then, for the personalized recommendation problem of users, a graph information extractor is constructed based on the graph convolutional neural network to fuse the recommended user-item bipartite graph node neighborhood information and generate a dense vector representation of nodes, which can enhance the recommendation effect in the form of incorporating the graph information representation in the deep recommendation model with Transformer as the sequence feature extractor. The proposed method is experimentally validated to shorten the response time and improve the performance of the system, which can increase the user experience of the visual communication system. The system designed in this article is user experience oriented, combined with Multimodal Features and intelligent recommendation algorithms, effectively meeting the personalized needs of users and has certain practical significance.**

Keywords*—*Visual communication system; user experience-oriented; multimodal features; recommendation algorithm

I. INTRODUCTION

Visual communication was first used in the 1920s and refers to a design activity that conveys visual information by means of all visual media. Visual communication design is defined as a design that uses visual symbols to communicate information. Visual communication is included in design, which is also a purposeful and creative human practice. However, visual communication is different from other designs in that its primary function is to transmit information, which is different from product design and environmental design, where the use of perception is the main focus [1]. It is different from the transmission of abstract concepts that is done by language, and it involves a wide range of fields.

Visual communication enables the exchange of ideas and information between individuals and individuals. Vision is a physiological term. Vision is created by the action of light, the cells in the visual organ become active and excited, and the external visual content is processed by the visual nervous system to form vision. Through vision, human beings can perceive the size of external objects, the color of light and dark, color and the movement and stillness of objects, and in the process of perception, they can obtain various types of information that are important for the existence of the body. Vision is the most important of the five senses of human perception, and at least 80% of natural information in the natural environment is obtained through vision.

From the external form of visual communication, any act of communication must be carried out by means of physical symbolic media loaded with information, and these media must be visible. In the past, the most important medium of visual communication design was print, and the form of bearing was mainly two-dimensional plane [2].

Visual communication design is composed of three basic elements: text, logo and illustration. In order to communicate their thoughts and feelings and necessary information, humans have gradually developed language. The medium of spoken language is the beginning of the transformation and development of our communication behavior and the most fundamental medium in the progress of our social communication activities [3]. However, oral language has certain limitations. The first is the limitation of distance, because dialogue can only exist in close communication; the second is the constraint of time, because the oral language is not easy to record and store since there is no more export.

The visual communication system is a way for people to communicate through seemingly, using visual language to spread information. And the application of computer technology is to retouch and process the pictures. Nowadays, although people's living standards are improving, many people are busy with their lives and rarely have time to enjoy various scenery. In this case, the advantages of the visual tradition system emerge [4]. People do not have to go out to enjoy the scenery of various regions, which not only saves people's time but also satisfies their diverse visual experience. The relationship between visual communication among people, nature and society is shown in Fig. 1.

The visual communication system is an art form created by treating text, pictures, and colors as basic elements, and using advanced computer graphics software to process these elements so as to achieve a good artistic presentation effect, transforming a single performance into a more acceptable and favorite art form [5]. Nowadays, professional graphic image

processing software such as Photoshop software, CAD series. This software will be based on the needs of visual communication design, text and other elements of the processing, to achieve a good visual communication effect.

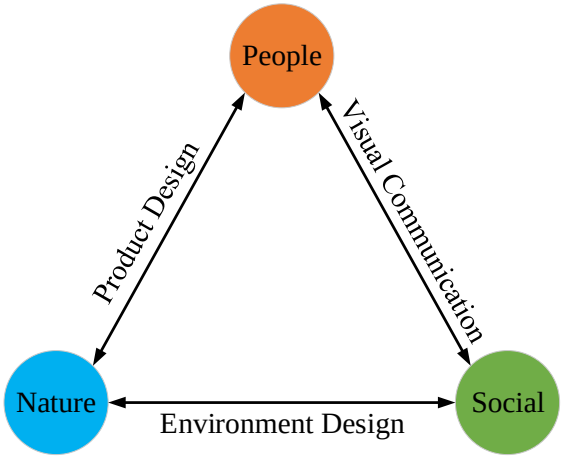


Fig. 1. Visual communication in the relationship between human, nature and society.

The application of these calculations in the visual communication system is discussed below: 1) Packaging design; 2) Advertising interface design; 3) Image processing technology. Nowadays, people like to use cell phones to take pictures, and in the process of taking pictures, the role of graphic image processing technology should not be underestimated, such as PS technology [6].

Computer technology is a way to use computer technology to modify, adjust and present. To achieve its function, it is necessary to build a specific scene of graphic description and create virtual optical effects with the help of light simulation technology, which is very closely related to computer geometry design. For computer graphics image processing technology, its data information is mainly from the subjective and objective world, the focus of image processing is to enhance the presentation of graphic images, which includes geometric image transformation, texture enhancement, color conversion and so on. At the same time, computer processing is based on digital signals, and it is widely used in the fields of medicine and aviation.

Based on the above background, this paper proposes an intelligent visual communication system for user experience in order to better improve the experience of visual communication system and combine with computer technology. Firstly, for the problem of multimodal feature extraction of users, an emotion recognition algorithm model with multimodal feature fusion is proposed, and each module and component of the proposed network model is introduced in detail. Then, for the user personalized recommendation problem, a new model architecture is designed by combining graph neural network with Transformer model to improve the recall of recommendation matching model. It is experimentally verified that the method in this paper has advantages in speed and performance.

The main research content of this article is as follows: First, for the problem of extracting multimodal features of users, considering the characteristics of different modal data, long and short-term memory networks are used to extract features with contextual information, and multi-scale convolutional neural networks are used for visual modality to extract low-level features from video frames. In the cross-modal stage, the low-level features in the source modality are used to enhance the target modality features. Then, for the personalized recommendation problem of users, a graph information extractor is constructed based on the graph convolutional neural network to fuse the recommended user-item bipartite graph node neighborhood information and generate a dense vector representation of nodes, which can enhance the recommendation effect in the form of incorporating the graph information representation in the deep recommendation model with Transformer as the sequence feature extractor.

II. Related Work

A. The Development Status of Visual Communication System

Visual communication system usually refers to the sender of information using visual symbolic elements to convey various information to the audience of information design, referred to as visual design visual communication includes two basic concepts, visual symbolic elements refers to what our eyes can see and can express the certain nature of things symbolic elements [7]. Visual communication design is also known as graphic design. The expressive design that we convey to the viewer's eyes and then shape the relevant content is collectively called visual communication design.

At this stage of analysis, the main content of visual communication design is still graphic design, professionals used to call it graphic design, visual communication design [8]. There is no great difference in the scope of design between graphic design and graphic design at this stage, and the relationship between them is a progressive one in terms of conceptual scope, not a contradictory and opposing relationship.

The study of non-planar elements of visual communication design is to break through the purely one-dimensional space to think and express around graphics, text and layout, and to advocate that all factors related to our design works should be included in the designer's scope of thinking and expression [9]. Visual communication design in the new era is not only about human vision, but also involves the sense of hearing and even touch.

In the context of the rapid development of computer media, visual communication technology, with its unique and innovative design images and language, has been gradually applied to the field of computer media, contributing to the development of the media. Visual communication technology is a kind of design activity that transmits and presents information to people through visual symbols, which are mostly expressed in the form of images [10]. Therefore, image design and processing is of great value in visual communication analysis, especially image fusion technology, which is very important for clear and complete presentation of visual images.

Computer graphic image processing techniques and visual communication systems are difficult to divide using standard

lines in many aspects. The commonality between the two is reflected in the basic professional curriculum, color and pattern design, which can be said to converge between them. The aesthetics, design methods, design concepts and development backgrounds are the same, and there are also similar concepts in the arrangement of lines and surfaces [11]. In addition, the design techniques produced by both are based on the same design techniques, use and depth of exploration. In this condition, the interpolation of the two can be better realized.

Many visual image fusion related research results have appeared. Some researchers use visual weight map for image fusion, decompose multi-scale images using cross bilateral filters, calculate visual weight values at different decomposition layers, and complete image fusion by integrating the results of weight value calculation, but the quality of fusion still needs further improvement. Recently, some researchers use adaptive PCNN to extract image information and fuse images by inverse NSCT, and use different fusion strategies to fuse images several times to increase the fusion accuracy, which has better fusion quality but longer computation process [12]. There are also researchers who enhance the infrared image based on the guiding filter, inject the infrared image information into the visible image effectively, complete the image fusion, and post-process the image to enhance the fusion effect, the process is more complex and time-consuming.

At this stage, the society is developing in the direction of diversification, and in the future, the use of computer processing technology and visual communication design will gradually expand, it is necessary to discuss and study the problems of technology application in depth, and under this condition, propose effective solutions to promote the application of technology to obtain a better visual communication effect.

B. Application of Multimodal Features in Recommendation Algorithm

The essence of recommendation algorithm is to connect users and items in a certain way, and search and recommendation should be divided into at least two stages: recall and sorting. In the recall phase, because of the large amount of data processed, fast speed, simple models and few features are required.

Due to the need of business scenarios, online evaluation is likely to be the evaluation of multi-objective fusion results. The evaluation of offline experiments on search ranking focuses on two aspects: efficiency evaluation and effectiveness evaluation. Efficiency evaluates response time and space consumption. Effectiveness is comparing the ranking results with the standard results. The evaluation metrics include: accuracy, recall, F-value, average accuracy [13]. The accuracy of the system can be evaluated according to three different paradigms. The first is that the recommendation algorithm can be used as a rating prediction model, predicting the user's rating for all the subject matter that did not produce an action. The second is to view the recommendation algorithm as a classification problem.

All of these algorithms currently have the following drawbacks: the recommended items are very similar to the items previously purchased by the user, the recommended results can only cover a small portion of the items, and they cannot explore the user's changing interests [14]. With the increasing popularity of posting contents on social platforms, people will post images, texts and videos on microblogs and other platforms to exchange their thoughts and express their emotions.

Based on this scenario, we propose to extract information from multiple modalities to give personalized recommendations through sentiment computing. A novel DNN approach is proposed to exploit the features and class relationships in video classification by fusing speech, video, and text modalities using a joint architecture for video classification. The classification performance is improved by imposing regularization based on the trajectory paradigm on the specially tailored fusion and output layers and exploiting the commonality shared among semantic classes, however, the feature representation and classification models are not learned jointly [15].

Multimodal learning refers to learning information from each of the multiple modalities and enabling the exchange and transformation of information from each modality. Multimodal deep learning refers to building neural network models that can perform multimodal learning tasks. The prevalence of multimodal learning and the heat of deep learning have given multimodal deep learning a vivid vitality and development potential. The fusion architectures are divided into joint architectures, coordination architectures and codec architectures according to the different ways of feature fusion. The fusion methods include three model-independent methods, early, late and hybrid, and two model-independent methods, multi-core learning and image model. Modal alignment has been the difficulty of multimodal fusion technology, and the two commonly used methods are display alignment and implicit alignment.

As a technique that enables machines to possess more human intelligence, multimodal deep learning is expected to gain significant momentum in the future. The next step can be to further investigate the insufficiently researched issues such as semantic conflict of modalities, multimodal combination evaluation criteria, and modal generalization ability, to deeply explore the difficult issues such as cross-modal transfer learning and non-convex optimization, and to promote the application of this technology in some new areas of deep learning [16]. In order to better capture the inter-modal relationships, this paper proposes to adopt a model-independent approach. A weighted sentiment feature fusion model based on is proposed, thus providing support for the later calculation of sentiment similarity.

Based on the above analysis, it can be concluded that traditional methods suffer from poor clarity after image reconstruction or excessive redundancy in visual information feature extraction, resulting in unsatisfactory extraction results, low extraction efficiency, and long training time. Therefore, considering the characteristics of different modal data, this paper uses long short-term memory networks to extract

features with contextual information, and uses multi-scale convolutional neural networks to extract low-level features from video frames. In the cross modal phase, low-level features in the source modality are used to enhance the target modality features, Integrating graph information representation into deep recommendation models to enhance recommendation performance. The experimental verification shows that the proposed method can shorten response time, improve system performance, and enhance the user experience of visual communication systems.

III. Algorithm Design

A. User Multimodal Feature Extraction

The model proposed in this section implements feature fusion of three modalities: image, acoustic and text. For the features of different modalities, a suitable deep neural network is selected for low-level feature extraction, and then the cross-modal attention mechanism proposed in this paper is used for feature fusion.

With the progress of machine learning and deep learning, machines that can achieve learning cognition and express emotions in complex real-world environments are often more versatile and effective in human-computer interaction-related research in artificial intelligence applications, which makes the role of affective computing more and more important. The purpose of this paper is to learn effective modal representations for feature fusion of textual, acoustic and visual multimodal data by using a cross-modal attention mechanism approach. The model structure of this paper is shown in Fig. 2.

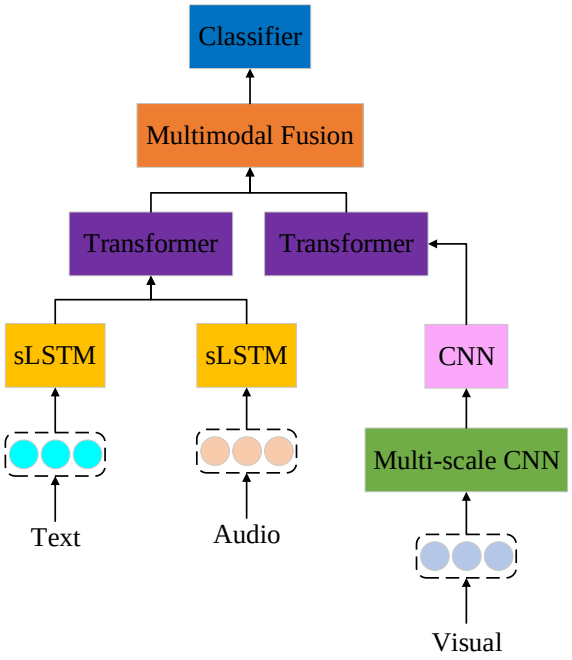


Fig. 2. Multi-modal feature fusion model.

The algorithm process is divided into four steps:

Step 1. In this paper, we use multimodal signals to detect emotions in videos, and use three modal data as the input of the model. The inputs include three low-level feature sequences from transcribed text, acoustic and visual modalities.

Step 2. Given the different characteristics of each modal feature, different neural networks are used to extract the features.

Step 3. After getting the different modal feature vectors, they are divided into two groups, text-acoustic and acoustic-visual.

Step 4. The output results of the two Transformer networks are spliced to obtain the fused features, and finally the prediction results are obtained through the fully connected network.

The advantages of LSTM networks in processing temporal data are: global processing and memory units. By processing data with time- or space-dependent characteristics, global information containing context-dependent information is obtained, and memory units are used to control the retention and removal of key historical information, enabling long- and short-term memory functions [17]. Therefore, it is more reasonable to use LSTM networks to capture the feature information of contextual cues. The textual feature vector obtained by using a one-way long and short-term memory network to capture temporal features, is calculated as follows:

$$t_i = sLSTM\left(T_i; \theta_i^{lstm}\right) \tag{1}$$

Since speech is highly random and correlated between adjacent frames, it is mainly reflected in the phenomenon of co-articulation when speaking, where words with connections before and after have an impact on the currently spoken word. Therefore, in this paper, for acoustic modality, the same LSTM network is used for feature extraction, and the acoustic feature vector is obtained and calculated as follows:

$$a_i = sLSTM\left(A_i; \theta_i^{lstm}\right) \tag{2}$$

For visual features, it is necessary to segment the video to obtain each video frame containing key feature information, sample it, and then use the face images in all video frames as the output information of acoustic modality. Therefore, for the visual modality considering the size of the input image data and the performance of the experimental equipment in this paper, MSCNN is applied and a set of feature maps are obtained, calculated as:

$$G_a^{MSCNN}\left(V, \theta\right) = \left\{v_a \middle| a \in \Omega\right\} \tag{3}$$

Where V denotes the input features.

For the different characteristics of the dataset, the first two are regression tasks, i.e., dichotomous tasks, and the commonly used loss functions are L1 Loss (Mean absolute loss, MAE) and L2 Loss (Mean Square Error, MSE). The MAE is calculated with absolute error as the distance, and is computed as shown below:

$$MAE = \frac{1}{n}\sum_{1}^{n}\left|y_i - \hat{y}_i\right| \tag{4}$$

MSE is also often used as a regular term, but because of the presence of the squared term, the gradient tends to explode when the predicted value differs significantly from the target value. Therefore, this paper uses MAE as the loss function and also introduces a varying learning rate.

B. User Recommendation Algorithm Design

Recommendation system is a research direction closely integrated with industrial applications, and the research of recommendation system from the academic point of view has certain limitations, therefore, the research on it needs to focus on the sub-problems of the recommendation task, or simplify the complex scenario, and optimize the sub-structure of the whole system from a certain point of view.

With the development of new neural network structures and efficient information extractors in recent years, the problem is solved by introducing graphical neural networks and Attention mechanism. Therefore, this chapter will introduce recommendation models based on graphical neural networks and Transformer structures, and show the effect of their application on classical datasets of recommendation systems and comparative experiments [18]. In addition, although the timestamps of interactions are introduced in the training process to constitute sequential information, it is necessary to distinguish this approach from intra-session recommendation, where individual sessions occur for a short period of time, while this model faces a long time span of data, which is still essentially using relatively static data for matching user preferences and item traits.

The core of the recommendation model proposed in this section is shown in Fig. 3, which incorporates information from the graph perspective of user-item interaction and applies an efficient feature extractor Transformer structure in order to obtain an effective method for item characterization with certain industrial practical capabilities.

Compared with graph embedding methods, the biggest advantage of graph neural networks in recommender system construction is their ability to be trained as part of an end-to-end model for collaborative filtering based on the introduction of graph structure. It is possible to use both node information and graph structure information.

The way node information is updated in a graph neural network is divided into two parts, i.e., message construction and message delivery. In each layer of GNN, messages can be passed along the edges of the graph. Each node receives messages from its neighbors and uses appropriate aggregation and mapping functions to construct new messages, which are used as node information in the next layer and also represent the new node features aggregated to the neighboring nodes' information. The message construction part mainly consists of the F mapping function, and the Sigma aggregation function. The Sigma function can be replaced by any aggregation function, and the F mapping function can be set as a single layer neural network:

$$M_i^r = F\left(\sum M_u^{r-1}, M_i^{r-1}\right) \tag{5}$$

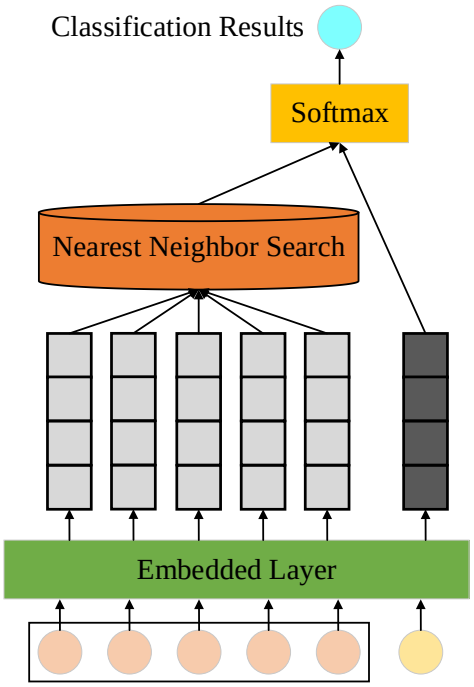


Fig. 3. Recommended model.

The difference between item-based and user-based recall algorithms lies in the pre-calculation of which of the two representations is used for nearest neighbor search. For most of the recommendation scenarios including movies and e-commerce, the recommended items are often relatively fixed, while the users are in constant high-frequency dynamic changes, and for large-scale e-commerce sites, the size of the users may be more than 10 times the size of the items.

The graph neural network structure used in the model is GCN, which uses the adjacency matrix, the degree matrix and the hidden feature matrix of the current layer to obtain the connection relationship between the nodes and the node features, and iteratively updates the node features:

$$H^{(l+1)} = Leaky\,\mathrm{Re}\,lu\left(D^{-\frac{1}{2}}AD^{-\frac{1}{2}}H^{(l)}W^{(l)}\right) \tag{6}$$

where the node hidden feature matrix H is the X-feature matrix in layer 0, the input layer.

In the training phase, the integrated features will be interacted with the candidate items, also embedded via items, in order to train the model parameters through a binary classification task. In the online service phase of the matching model, this feature tensor will be used as the retrieval target to obtain the matching result set by performing a K-NN nearest neighbor search from the full set of candidate items.

IV. Experiments

A. Experiment of Feature Extraction

Simulating pulse neural networks using computers requires powerful computing power. This article relies on the hardware devices listed in Table I for experiments:

TABLE I. HARDWARE TABLE OF DEVELOPMENT ENVIRONMENT

Development Environment	Parameter
GPU	V100S 32GB x 4
CPU	Intel(R) Xeon(R) Gold 6240R
Memory	38GB
Disk	7TB

The experiments first require setting hyper-parameters based on previous experience, and then tuning the parameters. The trainable parameters are obtained by back propagation, and the hyper-parameters must be set before building the entire network architecture. Each hyper-parameter related to the network structure needs to be specified first. For hyper-parameters such as discard rate and number of heads in the cross-modal attention module, a basic grid search method is performed in this paper. When the verification performance reaches a steady state, the learning rate is decayed by a factor of 10, making it more likely that the model will converge to a locally optimal solution, and the experimental results are cross-validated by a factor of 10. Two metrics, accuracy and F1 score values, are used to evaluate the performance of the algorithm model.

In order to fully verify the performance of the model proposed in this paper, the baseline models for experimental comparison are representative and recent algorithmic models with excellent performance. The baseline models are basically as follows: 1) EF-LSTM: a traditional network model using early fusion network and late fusion. 2) RMFN: a recurrent multi-stage fusion network that decomposes the fusion problem into multiple stages, each stage focusing on a subset of multimodal signal on a subset of the signal to achieve specialized and efficient fusion. 3) MFM: A multimodal decomposition model optimized for the joint generation of discriminative objectives for multimodal data and labels.

Table II gives the results obtained from the baseline model and the model proposed in this paper in the dataset, where the higher values of evaluation metrics Acc and F1 scores are better, and the comparison with the evaluation metrics Acc and F1 scores of different baseline models is used to demonstrate the optimal performance of the multimodal feature fusion sentiment recognition model proposed in this paper. In order to display the comparison results more intuitively, the best data shown in each column are bolded. It can be obtained that the sample recognition accuracy reaches 91.38%, the accuracy reaches 88.64%, and the average accuracy reaches 87.18%. The recognition accuracy of happy, angry and neutral emotions is the highest.

TABLE II. COMPARISON OF FEATURE EXTRACTION EFFECT OF EACH MODEL

Model	Acc	F1	Average
EF-LSTM	85.26%	83.72%	82.57%
RMFN	87.49%	86.43%	83.49%
MFM	90.54%	86.81%	85.36%
Ours	91.38%	88.64%	87.18%

Taking the Last.FM and Movielens datasets as examples, the distribution of the number of model training interactions is shown in Fig. 4. Although the overall interaction data scale is above 10M, the interaction volume has an obvious long-tail effect, which leads to the fact that the interaction volume of most items cannot meet the training scale of the model, which is reflected in the training results, only a very small number of items can be fully trained in a short time, and the overall training speed is slow. Therefore, negative sampling was introduced to expand the data in the training. When the negative sampling ratio is set to 5, that is, whenever a positive interaction is put into the training set, 5 samples of items that have not been interacted with the user will be collected at the same time, and during training, these 5 negative samples will be processed negatively, which greatly increases the training efficiency.

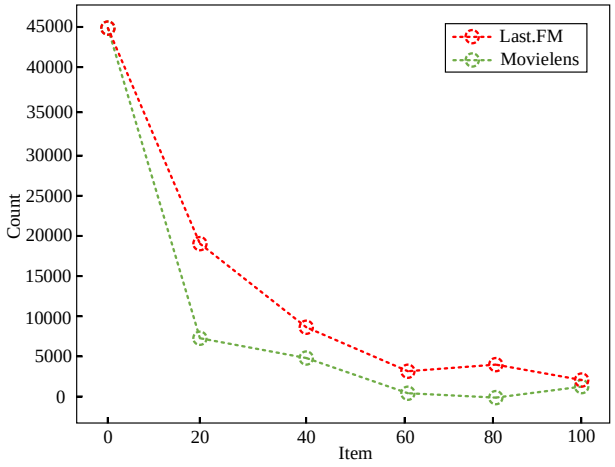


Fig. 4. Experimental results of model training interaction.

B. Personalized Recommendation Experiments

In the personalized recommendation experiments for users, the datasets used are: 1) Amazon-Music: containing 125,998 shopping reviews generated by 5,453 users and 65,833 items, with an interaction density of 0.35%. 2) Amazon-App: containing 341,784 shopping reviews generated by 18,328 users and 32,684 items, with an interaction density of 0.57%. The interaction density was 0.57%. 3) Movielens movie rating dataset, which contains user ratings of movies from the Movielens movie rating website collected by GroupLens.

The comparison methods are: 1) Popular+ClaSS(PopClass): a dynamic recommendation strategy based on popularity, which introduces the consideration of user preference categories. 2) DeepWalk: - a classical graph embedding technique, which constructs a sequence of nodes by random wandering on an undirected graph and applies (3) YoutubeDNN: In the matching phase, YoutubeDNN tries to use the information of items that users have interacted with and additional auxiliary information such as time information to input into the deep neural network as the user's representation, and then train the network through the multi-category classification task of videos. The steps can be broken down into two phases: network pre-training and model service.

The hit rate performance of our model and the comparison algorithm are shown in Table III. Since the Amazon-Music dataset is the sparsest, it indicates that there is incomplete training and therefore the overall hit rate is lower than the other datasets. In DeepWalk, the overfitting may be caused by the sparsity of the training data, but the graph-based embedding method is able to achieve the best hit rate performance based on a simple structure. However, the graph-based embedding method is able to achieve excellent performance based on a simple structure, which fully illustrates the importance of using the graph structure information in the field of recommendation recall.

In contrast to Transformer-Encoder, which also extracts sequence information, it should be noted that although this model achieves a good hit rate level, in industrial applications, the recall of target items by a single method only reflects a limited perspective of the recall strategy, and multiple fusion recall is needed in conjunction with business scenarios.

TABLE III. EXPERIMENTAL RESULTS OF HIT RATE OF EACH MODEL

Model	Amazon-Music	Amazon-App	Movielens
PopClass	13.25%	8.56%	4.52%
DeepWalk	24.63%	6.49%	12.61%
YoutubeDNN	22.84%	7.54%	10.84%
Ours	27.38%	11.27%	15.06%

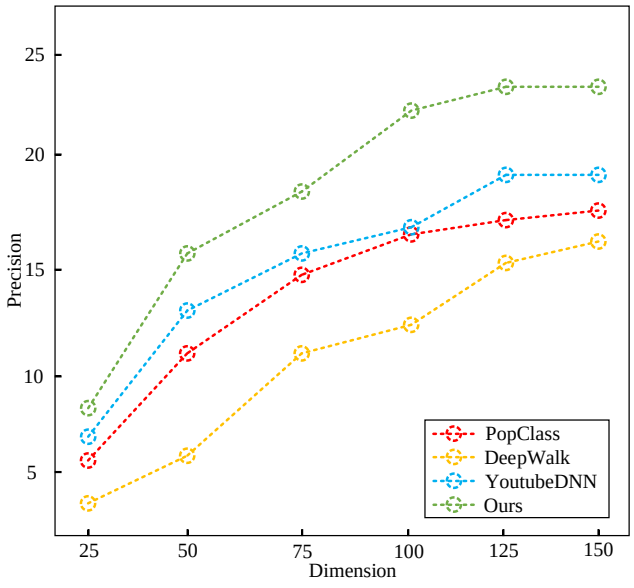


Fig. 5. Experimental results of each model under different implicit vector dimensions.

The experimental results of each model with different implied vector dimensions are shown in Fig. 5. It can be seen from the figure that the larger the implied vector dimension of the model, the better the performance of the model, because the larger the implied vector dimension, the more information the nodes contain, the finer the granularity of the representation of the user's interest, and thus the deeper and finer the granularity of the user's interest preference can be captured, and the recommendation effect will be better. However, due to the nature of the neural network structure, the implicit vector dimension increases the time and space complexity of the model, so the implicit vector dimension should be set to balance performance and resource consumption.

V. CONCLUSION

In order to improve user experience, this paper proposes an intelligent visual communication system oriented to user experience, and applies user experience to the algorithm design of the visual communication system to realize the design and application of personalized visual communication. First, for the user multimodal feature extraction problem, based on the cross-modal attention mechanism in its module, the target modal features are enhanced by including low-level features from other modalities. The output of multiple attention heads in it enhances the representational capability of the network and further better improves the effect of modal fusion. Then, for the personalized recommendation problem, the Transformer is used as the main structure of the sequence feature for extractor to replace the DNN in the base model, and the graph convolutional neural network module with end-to-end graph learning capability is combined with the dense vector representation of items in the recommendation matching stage to improve the hit rate of the matching algorithm. It is experimentally verified that the method in this paper achieves a large performance improvement on three different datasets compared with other advanced network models.

Although convolutional neural networks have developed relatively maturely in the field of visual recognition, capturing subtle changes in visual images that are difficult to detect, efficiently extracting facial expression features, and effectively utilizing these features are key issues; Meanwhile, how to fully consider the changes in key facial features, as well as the fusion of global and local features, is also a major key factor in the process of facial expression recognition. Although the paper proposes a visual feature recognition method based on convolutional neural networks to address the above issues, which has improved the recognition accuracy to some extent, our work still has some shortcomings and needs to be improved and perfected in the following aspects in the future.

Firstly, for static image recognition, due to the limitations of the dataset itself, the accuracy of recognizing certain expression categories is still lacking. Therefore, future work will focus on training and validating the proposed method on datasets with more evenly distributed expression labels.

Secondly, the method proposed in this article has not yet been applied to datasets collected in real environments. Future work will train and validate the method using datasets collected in real environments.

Finally, for the local collection end, when performing facial expression recognition, only one person's facial image can be recognized, and the problem of storing expressions when multiple people appear simultaneously has not been solved; For online management, we hope to achieve real-time facial expression collection on web pages in the future to reduce costs.

REFERENCES

[1] Alhayani, B. S., & Llhan, H. (2021). RETRACTED ARTICLE: Visual sensor intelligent module based image transmission in industrial manufacturing for monitoring and manipulation problems. Journal of Intelligent Manufacturing, 32(2), 597-610.

[2] Tan, J. K., & Sato, A. (2020). Human-robot cooperation based on visual communication. International Journal of Innovative Computing, Information and Control, 16(2), 543-554.

[3] Hassan, M. K., Hassan, M. R., Ahmed, M. T., Sabbir, M. S. A., Ahmed, M. S., & Biswas, M. (2021). A survey on an intelligent system for persons with visual disabilities. Aust. J. Eng. Innov. Technol, 3(6), 97-118.

[4] Ma, C., Li, X., Li, Y., Tian, X., Wang, Y., Kim, H., & Serikawa, S. (2021). Visual information processing for deep-sea visual monitoring system. Cognitive Robotics, 1(2), 3-11.

[5] Kountouris, M., & Pappas, N. (2021). Semantics-empowered communication for networked intelligent systems. IEEE Communications Magazine, 59(6), 96-102.

[6] Wang, B., Xu, K., Zheng, S., Zhou, H., & Liu, Y. (2022). A deep learning-based intelligent receiver for improving the reliability of the MIMO wireless communication system. IEEE Transactions on Reliability, 71(2), 1104-1115.

[7] Yang, J., Wang, C., Jiang, B., Song, H., & Meng, Q. (2020). Visual perception enabled industry intelligence: state of the art, challenges and prospects. IEEE Transactions on Industrial Informatics, 17(3), 2204-2219.

[8] Liu, R. W., Guo, Y., Lu, Y., Chui, K. T., & Gupta, B. B. (2022). Deep network-enabled haze visibility enhancement for visual IoT-driven intelligent transportation systems. IEEE Transactions on Industrial Informatics, 19(2), 1581-1591.

[9] Lan, Q., Wen, D., Zhang, Z., Zeng, Q., Chen, X., Popovski, P., & Huang, K. (2021). What is semantic communication? A view on conveying meaning in the era of machine intelligence. Journal of Communications and Information Networks, 6(4), 336-371.

[10] Dodda, S., Chintala, S., Kanungo, S., Adedoja, T., & Sharma, S. (2024). Exploring AI-driven Innovations in Image Communication Systems for Enhanced Medical Imaging Applications. Journal of Electrical Systems, 20(3s), 949-959.

[11] Imoize, A. L., Adedeji, O., Tandiya, N., & Shetty, S. (2021). 6G enabled smart infrastructure for sustainable society: Opportunities, challenges, and research roadmap. Sensors, 21(5), 1709-1719.

[12] Fu, Y., Li, C., Yu, F. R., Luan, T. H., & Zhang, Y. (2021). A survey of driving safety with sensing, vehicular communications, and artificial intelligence-based collision avoidance. IEEE transactions on intelligent transportation systems, 23(7), 6142-6163.

[13] Alhayani, B., Abbas, S. T., Mohammed, H. J., & Mahajan, H. B. (2021). Intelligent secured two-way image transmission using corvus corone module over WSN. Wireless Personal Communications, 120(1), 665-700.

[14] Alhayani, B., Abbas, S. T., Mohammed, H. J., & Mahajan, H. B. (2021). Intelligent secured two-way image transmission using corvus corone module over WSN. Wireless Personal Communications, 120(1), 665-700.

[15] Liu, R. W., Nie, J., Garg, S., **ong, Z., Zhang, Y., & Hossain, M. S. (2020). Data-driven trajectory quality improvement for promoting intelligent vessel traffic services in 6G-enabled maritime IoT systems. IEEE Internet of Things Journal, 8(7), 5374-5385.

[16] Njoku, J. N., Nwakanma, C. I., Amaizu, G. C., & Kim, D. S. (2023). Prospects and challenges of Metaverse application in data-driven intelligent transportation systems. IET Intelligent Transport Systems, 17(1), 1-21.

[17] Gao, X., Wang, Y., Chen, X., & Gao, S. (2021). Interface, interaction, and intelligence in generalized brain–computer interfaces. Trends in cognitive sciences, 25(8), 671-684.

[18] Lazaroiu, G., Androniceanu, A., Grecu, I., Grecu, G., & Neguriță, O. (2022). Artificial intelligence-based decision-making algorithms, Internet of Things sensing networks, and sustainable cyber-physical management systems in big data-driven cognitive manufacturing. Oeconomia Copernicana, 13(4), 1047-1080.

Synchronous Update and Optimization Method for Large-Scale Image 3D Reconstruction Technology Under Cloud-Edge Fusion Architecture

Jian Zhang[1]*, Jingbin Luo[2], Yilong Chen[3]
Guangdong Power Grid Co., Ltd, Guangzhou 510030, China[1, 2]
Guangdong Power Grid Co., Ltd. Guangzhou Power Supply Bureau, Guangzhou 510000, China[3]

***Abstract*—Aiming at the problems of limited bandwidth and network delay in the traditional centralized cloud computing mode during large-scale image processing of transmission and distribution digital corridors, a synchronous updating and optimization method for large-scale image 3D reconstruction technology under cloud-edge fusion architecture is proposed. Based on the cloud-side fusion architecture, the image data of the transmission and distribution corridor is preprocessed, feature extraction is performed by deep learning, synchronous updating is performed by using the cloud-side cooperative network, and matching and 3D reconstruction are performed according to the order of the point cloud data; given the dynamically changing characteristics of the image data in the cloud-side fusion environment, the incremental learning is combined with the continuous learning and synchronous updating of the model parameters, to realize the adaptive updating mechanism. The research method utilizes the advantage of cloud-edge fusion architecture to distribute the computational tasks to the cloud and edge, realizing parallel processing and load balancing, and improving the accuracy and efficiency of 3D reconstruction. The experimental results show that the research method in this paper has an image feature point matching rate as high as 96.72%, a lower network latency rate, and a higher real-time performance, which provides strong technical support for the optimization of the transmission and distribution digital corridor 3D reconstruction technology.**

***Keywords*—Cloud-edge fusion; cloud-edge collaboration; 3D reconstruction; synchronized update**

I. INTRODUCTION

With the in-depth promotion of intelligent transformation in the power industry, 3D reconstruction technology is widely used in the construction of digital corridor models, and the study of 3D reconstruction technology with efficient synchronization performance is of great significance to enhance the construction and management efficiency of transmission and distribution corridors. The study in [1] designed a point cloud data extraction process based on tilt photography technology, combined with the optical distortion problem for 3D modeling of transmission lines, which can quickly obtain the target shape, but the throughput of remote image communication is not high and the transmission rate is low. The study in [2] fuses transmission line channel information based on visible image sequences, sets the model geometry through standardized mapping space, establishes mapping relationship with power line point cloud data and images, and constructs 3D reconstruction model under the constraint relationship function, but the technique lacks double-ended synchronous communication system and centralized network information processing efficiency is not high. The research in [3] introduces the cloud-side cooperative function to realize the global resource sensing and cooperative scheduling of star-earth fusion network, and combines the arithmetic network technology to provide a new technical path for the synchronous transmission of large-scale image 3D reconstruction. The study in [4], on the other hand, provides a comprehensive overview of the development history of image 3D reconstruction technology, evaluation methods and datasets. Image 3D reconstruction techniques have made remarkable progress, from early geometric model-based methods to today's advanced algorithms based on deep learning, the accuracy and efficiency of reconstruction have been continuously improved. However, current 3D reconstruction techniques still face many challenges. On the one hand, the synchronized processing of large-scale image data puts forward extremely high requirements on computational resources and storage capacity, which are often difficult to be met by traditional centralized architectures. On the other hand, images from different data sources differ in quality, angle, illumination, etc., how to effectively fuse these images for accurate 3D reconstruction is still a difficult problem. Therefore, this paper provides new ideas and methods to solve the difficulties in large-scale image 3D reconstruction. By fully utilizing the advantages of cloud-edge fusion architecture, it can effectively integrate computational resources, improve reconstruction efficiency and accuracy, and meet the urgent needs of image 3D reconstruction in different fields. Innovative use of cloud computing and edge computing technology for organic combination, cloud-edge cooperative network can effectively improve the double-end communication throughput, providing powerful technical support for large-scale image 3D reconstruction. At the same time, the 3D reconstruction algorithm is optimized and improved by combining deep learning to improve the reconstruction accuracy and efficiency, providing a more efficient technical path and solution for the 3D reconstruction of transmission and distribution corridors, which improves the efficiency accuracy and enhances the robustness of large-scale image reconstruction from the technical level.

*Corresponding Author.

II. Transmission and Distribution Digital Corridor IoT Cloud Edge Convergence Architecture

A. Cloud-edge Fusion Network Architecture

The transmission and distribution of digital corridor image 3D reconstruction system mainly adopts the cloud-edge fusion system architecture, including three key components: cloud center, edge domain, and edge nodes (see Fig. 1).

Cloud Center. The cloud center is the core of the whole architecture, responsible for processing and fusion of multi-dimensional data of the system, which can support massive 3D reconstruction image and data storage as well as synchronous data-parallel update. The cloud center aggregates the edge nodes and resource pools through the high-speed network and provides big data applications for the system through deep learning and association algorithms [5].

Edge domain. The edge domain is the "bridge" connecting the cloud center and edge nodes, with the function of data processing and forwarding. It is deployed in each key position of the transmission and distribution corridor according to the actual needs, and acquires real-time corridor information by collecting sensor data, video streams, etc., extracts key features and information through preliminary pre-processing, and then transmits them to the cloud center to realize "data in the cloud" [6].

Edge node. Edge nodes are multi-dimensional sensing front-end devices deployed at the site of transmission and distribution corridor, with the ability to data acquisition, processing, and execution of control commands. The front-end node acquires the image, temperature, humidity, and other environmental information of the corridor in real time, adjusts the parameters of the equipment to carry out special operations, and sends the front-end information to the cloud center through the convergence of the edge domain.

B. Cloud Data Center Structure Optimization

Large-scale image three-dimensional reconstruction technology requires parallel processing of massive data, the cloud data center should take large-capacity, high-speed, and stable equipment to optimize the system hardware structure. The cloud data server cluster selects Dell PowerEdge R430 dual-channel rackmount servers with powerful computing capability and scalability, which can meet the needs of large-scale data processing and computing. AS13000 storage system is selected to provide high-capacity, highly reliable storage with data redundancy and backup functions to prevent data loss and ensure data security and fast reading [7]. The switch model is S5130S-28S-EI, which ensures a high-speed and stable network connection and supports the transmission and exchange of large amounts of data. The NGFW4000-UF firewall with a wide coverage domain is selected to guarantee the security of the Cloud Edge Convergence network, preventing external attacks and intrusions, reducing the risk of data loss, and ensuring the stability and efficiency of data transmission. Fig. 2 shows processor architecture.

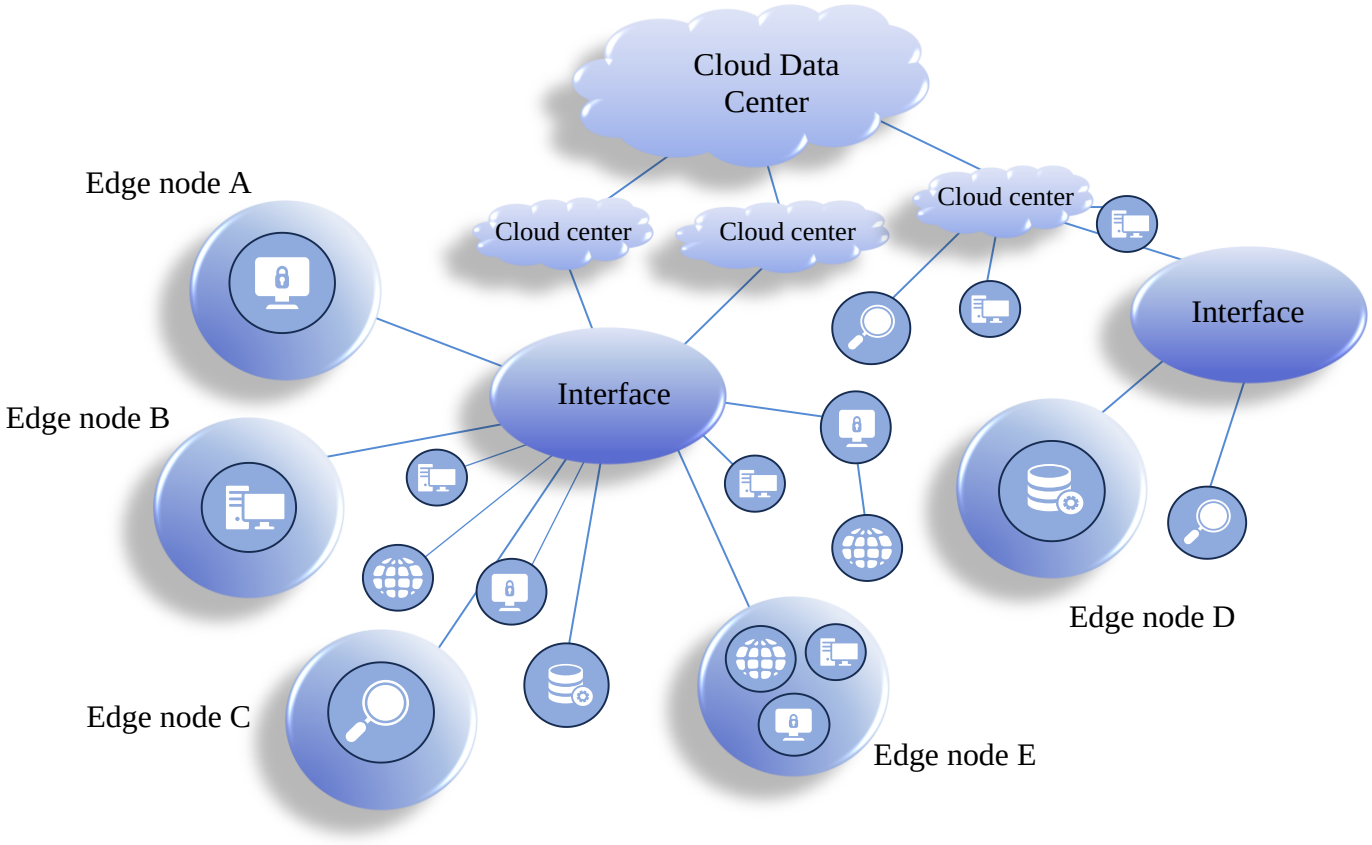


Fig. 1. Cloud edge converged network architecture.

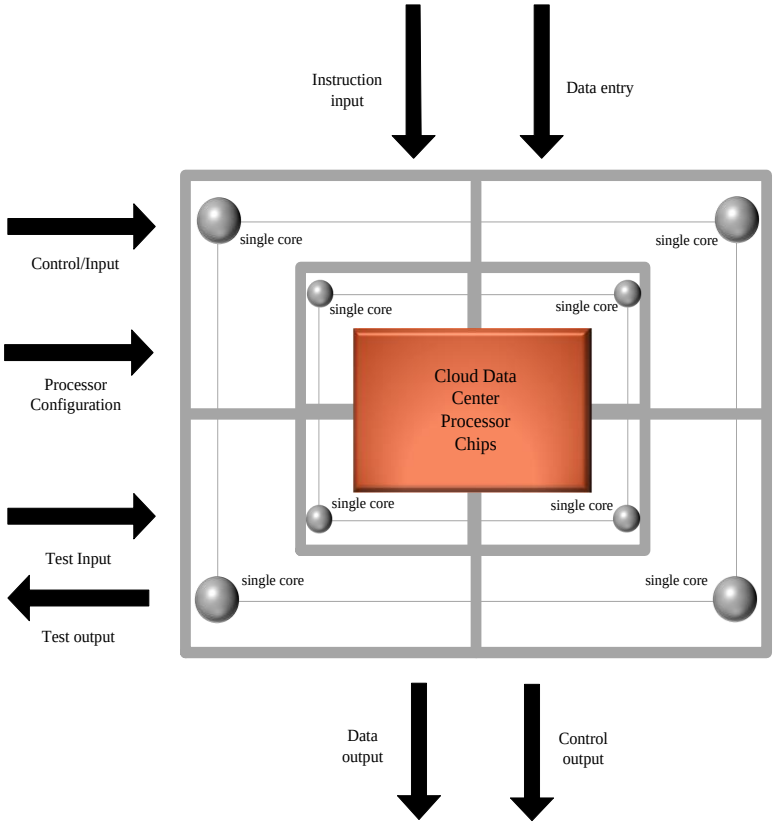


Fig. 2. Processor architecture.

C. Edge Node Structure Optimization

The front-end edge devices mainly include the SICK LMS111 high-precision laser distance sensor, ON Semiconductor MT9V034 high-definition image sensor, DHT11 temperature and humidity sensor, ADXL345 acceleration sensor, which can detect the distance of the transmission and distribution corridor, image, temperature and humidity and motion status [8]. The core terminal equipment of the edge domain is the Lenovo tower server, which is rugged and suitable for complex environments, and can aggregate massive data in the domain for transmission and pre-processing. Moxa EDS-408A-MM-SC switches provide real-time synchronized edge network connection for the system, which is characterized by high bandwidth and low latency and supports a variety of communication protocols with strong adaptability to guarantee the high efficiency and stability of the edge domain network. Fig. 3 shows edge-core server CPU architecture.

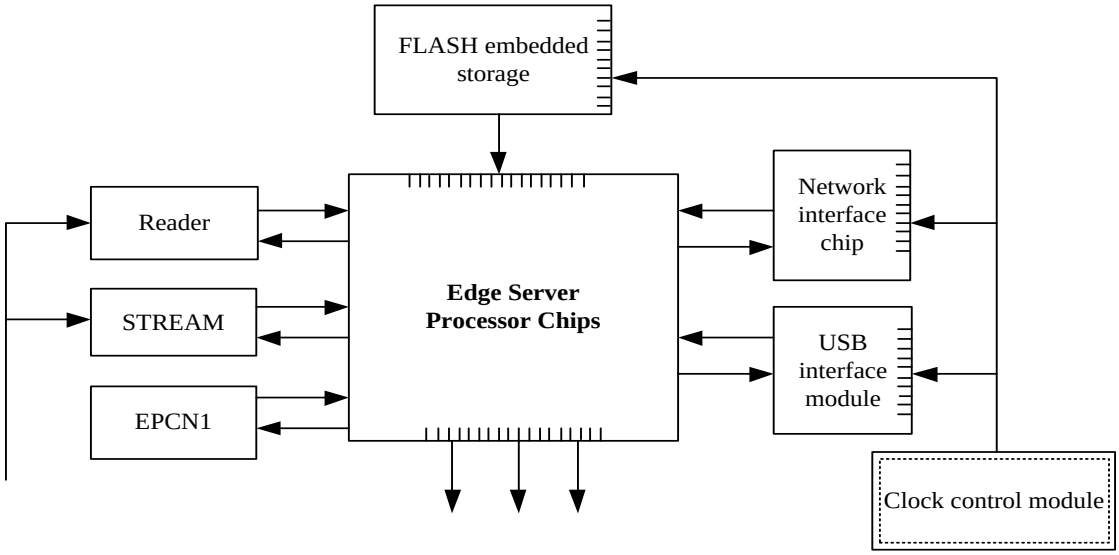


Fig. 3. Edge-Core server CPU architecture.

III. Optimization of Image 3D Reconstruction Technique Based on Cloud Edge Fusion

A. Non-Sampling Transform Fusion

Aiming at the complex reality of transmission and distribution corridors, non-sampling directional filters are installed at the front end, which can perform filtering operations on signals or images in a specific direction to extract and enhance the signal features in a specific direction while suppressing the interference and noise in other directions, which has important applications in the field of three-dimensional image reconstruction and other fields [9]. Firstly, the image is decomposed at the second level according to the directional component to obtain the quadrant sampling formula:

$$\begin{cases} S_l(i) = 2*[i/2] - 2^{l-3} + 1 \\ S^p(i) = 2*I_0^{p-3} - I_3^{S_l(i)} \end{cases} \tag{1}$$

In the formula, $S_l(i)$ and $S^p(i)$ are the quadrant filter function formulas for random pixel points i in the horizontal and vertical directions, respectively, l is the horizontal quadrant parameter, p is the vertical quadrant parameter, and I is the matrix threshold of the sampling points. The decomposition is performed iteratively for three or more times according to the above steps, and the source image is decomposed into low-frequency sub bands and high-frequency sub bands, which are normalized by wavelet transform. The low-frequency sub band signals are average weighted according to the low-frequency criterion:

$$\begin{cases} P_l = \sqrt{\frac{1}{LP}\sum_{L}^{i=1}\sum_{P}^{j=2}[f(i,j) - f(i,j-1)]^2} \\ P_p = \sqrt{\frac{1}{LP}\sum_{L}^{i=2}\sum_{P}^{j=1}[f(i,j) - f(i-1,j)]^2} \end{cases} \tag{2}$$

$$P_s = \sqrt{P_l^2 + P_p^2} \tag{3}$$

In the formula, P_l and P_p are the spatial row and column frequencies of the low-frequency sub band averaged and weighted, respectively, and P_s is the low-frequency spatial frequency criterion. The high-frequency sub band signals are regionally energy maximized according to the high-frequency criterion:

$$\begin{cases} E_l(i,j) = \sum_{1}^{i=-1}\sum_{1}^{j=-1} W(i,j)\left[l(i+L, j+P)\right]^2 \\ E_p(i,j) = \sum_{1}^{i=-1}\sum_{1}^{j=-1} W(i,i)\left[p(i+L, j+P)\right]^2 \end{cases} \tag{4}$$

In the formula, $E_l(i,j)$ and $E_p(i,j)$ are the energy maxima of the high-frequency sub band pixel point (i,j) in the horizontal and vertical regions of space, and W is the regional energy weighting coefficient. The transform fusion is performed by the filtered dual channel, so that the source image is finely preprocessed at the edge end. Fig. 4 shows convolutional neural network workflow visualization.

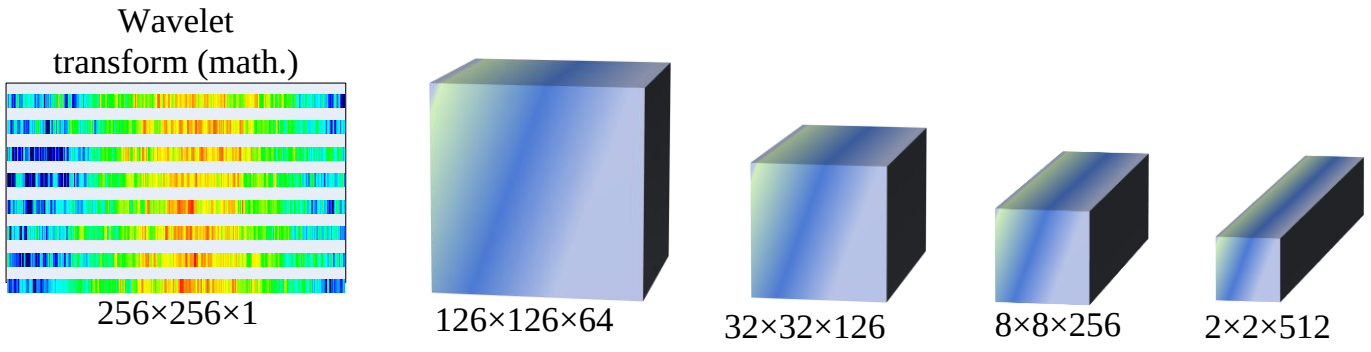


Fig. 4. Convolutional neural network workflow visualization.

B. Edge Detection Constraints

For the source image collected by the infrared laser equipment, there is an edge variability problem, to further guarantee the authenticity of the fused image, the edge image information needs to be evaluated constraints. The richness of the image elements is evaluated through the entropy formula:

$$R = \sum_{H-1}^{i=0} P_i \ln P_i \tag{5}$$

In the formula, R represents the entropy value of image pixel content richness, H is the image resolution grayscale, and P_i is the distribution frequency of pixel point i. The image normalized mean weight δ is introduced for standard deviation calculation to evaluate the image grayscale clarity level:

$$\Delta C = \sqrt{\sum_{L-1}^{i=0}\sum_{P-1}^{j=0}[f(i,j) - \delta]^2 / lp} \tag{6}$$

The larger the standard deviation calculation results,

indicating that the fusion image grayscale dispersion value is higher, the worse the fusion effect, the authenticity and relevance do not meet the standard, need to repeat the above steps of convolution iterative optimization [10]. Fig. 5 shows grayscale test of edge image.

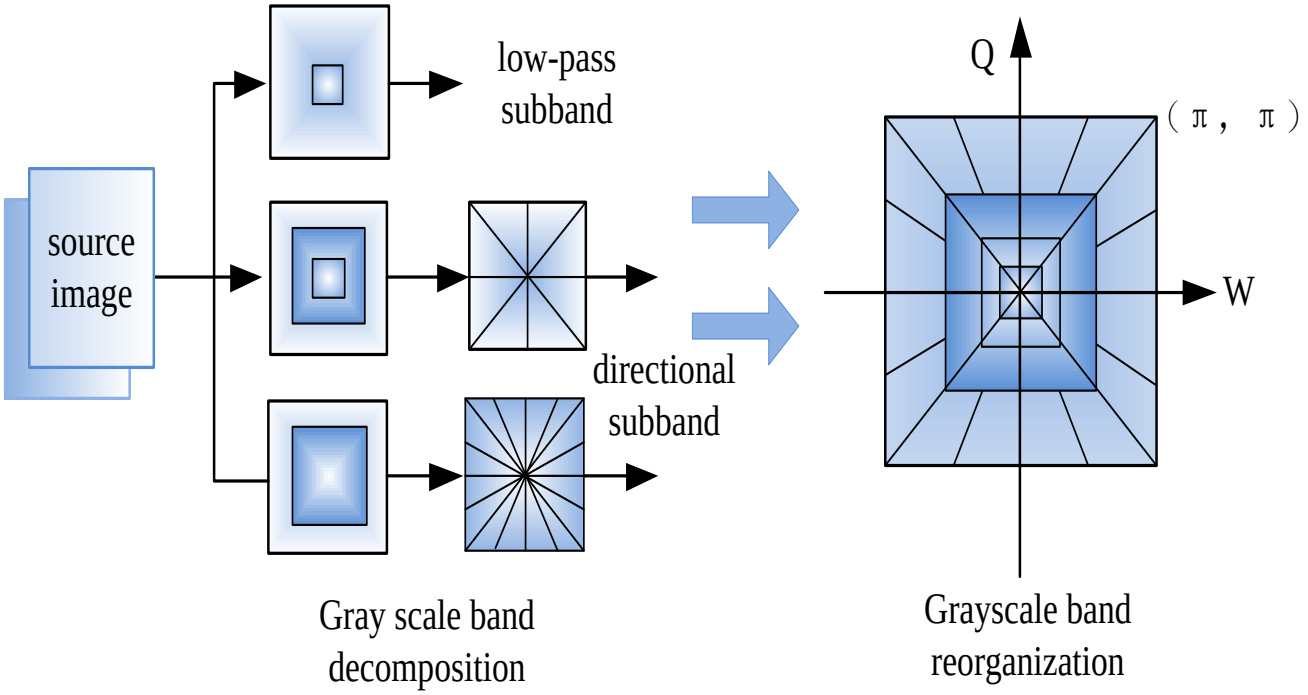


Fig. 5. Grayscale test of edge image.

C. Feature Point Extraction Matching

The preprocessed image information at the edge end is transmitted to the cloud data center for source data archiving in real time via the edge domain network. For the image of the transmission and distribution corridor edge points, line points, corner points, bright spots, dark spots, and other key feature points for localization and extraction, through the Gaussian function convolution to generate three-dimensional scale space:

$$K(x,y,z)=G(x,y,z)\times(x,y) \tag{7}$$

In the formula, $K(x,y,z)$ is the 3D scale space mapping coordinates of the image 2D coordinates (x,y) and $G(x,y,z)$ is the Gaussian function convolution coordinate parameter. The Gaussian difference is utilized to detect the three-dimensional spatial coordinate point poles:

$$J(x,y,z)=(G(x,y,z\beta)-G(x,y,z))*(x,y) \tag{8}$$

According to the calculation results, the three-dimensional spatial discrete extreme points are extracted, and then the feature points are localized by three-dimensional quadratic function fitting to eliminate the edge points with poor stability [11]. Due to the existence of dynamics in the real-time monitoring of cloud edge fusion, the gradient method is used to carry out a balanced and stable simulation of the moving structure of the feature points in a specific direction, and the description of the baseline feature direction is obtained as:

$$D(x,y)=\sqrt{(K(x+1,y)-K(x-1,y))^2+(K(x,y+1)-K(x,y-1))^2} \tag{9}$$

$$Dir(x,y)=\tan^{-1}\left(\frac{K(x,y+1)-K(x,y-1)}{K(x+1,y)-K(x-1,y)}\right) \tag{10}$$

In the formula, $D(x,y)$ denotes the gradient value of the feature point and $Dir(x,y)$ denotes the spatial reference direction of the feature point. The feature points are extracted from the 3D reconstructed spatial point cloud to generate a descriptor. The descriptor contains the local geometric information of the feature point, and the SIFT (Scale Invariant Feature Transform) algorithm is used to generate the feature point description vector:

$$M(x_i,y_i)=Dir(x,y)*\Delta T(x_i+y_i) \tag{11}$$

In the formula, $M(x_i,y_i)$ is the dynamic scale description vector in the reference direction and ΔT is the 3D coordinate mapping dynamic time difference. Based on the description vector, dynamic image 3D point cloud reconstruction can be performed, which utilizes the camera position and point cloud information to project each pixel point of the image into the 3D space [12].

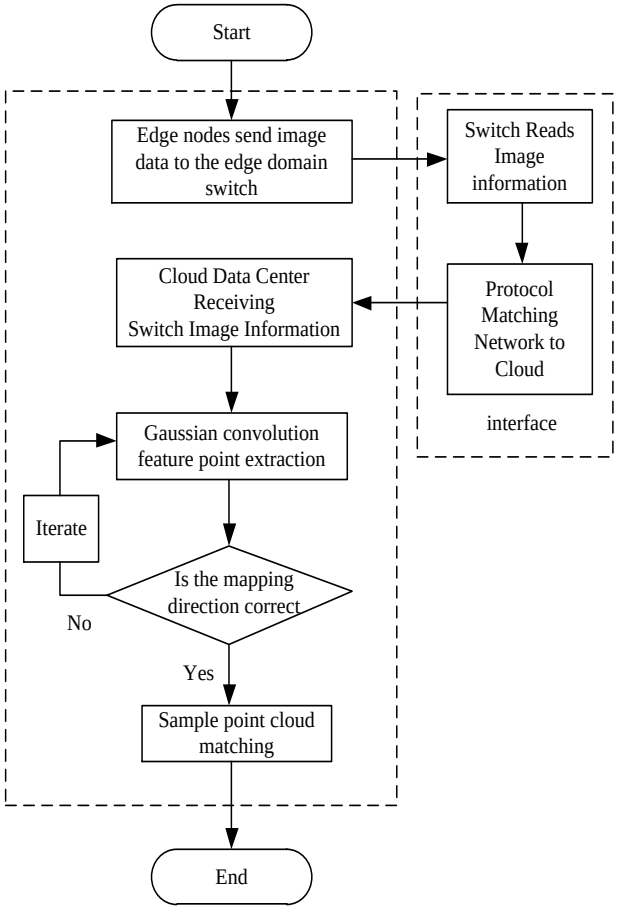


Fig. 6. Cloud edge collaborative feature point extraction process.

D. Point Cloud Synchronized 3D Reconstruction

Based on the basic geometric structure and moving position described by the feature point vectors, the key feature points with high edge dispersion are extracted to construct a sparse point cloud for localization and reconstruction in 3D space [13]. However, the sparse only contains a small number of salient feature points, which cannot accurately describe the 3D details of the transmission and distribution corridor, and further dense point cloud reconstruction is required [14]. First, the data collected by the edge devices are synchronously transmitted to the cloud data center, and the sparse point cloud is arranged from near to far according to the camera positional distance to obtain the feature-matching initialization sequence:

$$A(w)=\left\{ w_i \middle|\ w_i \in w_m, n(w)\cdot\frac{O(w)O(w_i)}{\left|O(w)O(w_i)\right|}>\cos\alpha \right\} \tag{12}$$

In the formula, $A(w)$ is the set of sequence images, w_i denotes the image i, $O(w)$ is the center point of the surface of the fused image, $O(w_i)$ is the center of the focused light of the camera of the fused image, and α is the maximum value of the angle between the largest edge point within the image and the center of the camera. Face sheet expansion is performed on the neighboring images in the sequence:

$$Q(w)=\left\{ Q_i(x',y') \middle| w\in Q_i(x,y), |x-x'|+|y-y'|=1 \right\} \tag{13}$$

In the formula, (x,y) and (x',y') are the corresponding point clouds of neighboring image facets. To ensure the authenticity of the extended image, the facets with anomalies in the extended fusion are filtered, and the neighboring image facets are screened by the consistency test:

$$\left|Q^*(w)\right|\left(1-hid^*(w)\right)<\sum_{w_i\in W(w)}1-hid^*(w) \tag{14}$$

In the formula, $hid^*(w)$ is the grayscale consistency test function. If the tested image facets satisfy the above conditions, it means that the two image facets have large differences and are not suitable for extended reconstruction. Based on the consistency fusion test, the dense point cloud can accurately capture the key geometric structures in the corridor, and the processed dense point cloud data is converted into a 3D mesh model using the Power Crust algorithm [15]. Fig. 6 shows cloud edge collaborative feature point extraction process.

A set of discrete extreme values highlighted in the image is extracted as the simulation center axis, the point cloud is linearly estimated and regionally divided, adjacent points are connected to generate a Voronoi diagram, the extreme points

of each cell Voronoi diagram are extracted, i.e., the feature points of each region that are farthest away from the sampling points, and the weighted computation is performed to get the description of the reconstruction of the three-dimensional Power diagram [16]:

$$U_{pow(x,\varphi_{o,u})} = U^2(o,x) - \eta^2 \tag{15}$$

In the formula, $U_{pow(x,\varphi_{o,u})}$ is the weighted stereo Power map distance description of the extreme point, U is the distance of the extreme point from the sampling center, $\varphi_{o,u}$ represents a sphere with o as the center and a radius of u, and η^2 is the reconstruction weighting value. Each region of the image features a point cloud in turn for three-dimensional reconstruction, to get the three-dimensional framework, comprehensively improve the image's three-dimensional reconstruction accuracy (see Fig. 7). The texture information in the original image is extracted and mapped onto the surface of the 3D model, which makes the transmission and distribution corridor model more realistic and vivid and can effectively enhance the realism of the digital model [17].

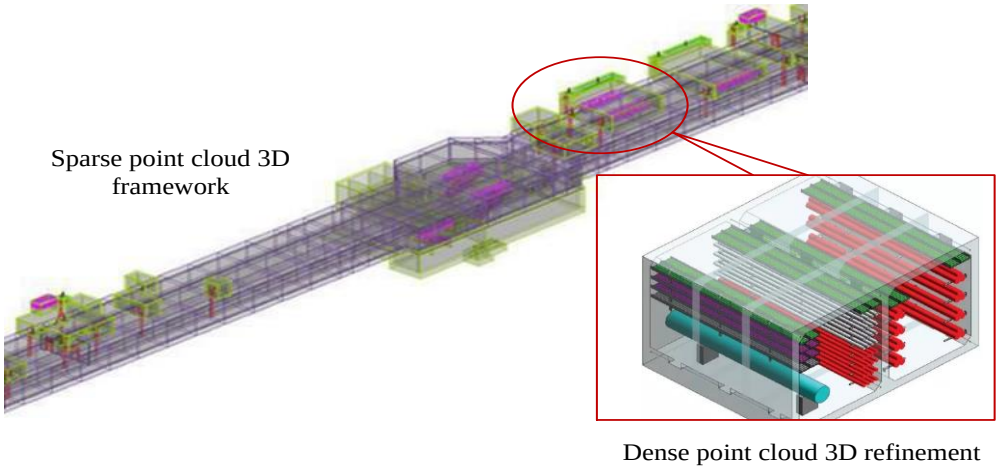


Fig. 7. 3D reconstructed view of point cloud.

IV. Synchronized Update of 3D Reconstruction Data for Cloud Edge Collaborative Images

A. Dynamic Estimation of Feature Points

The transmission and distribution corridor monitoring equipment is subjected to remote manipulation for displacement, and the images need to be evaluated for feature point motion. Assuming that the equipment moves at the same speed as the pre-set parameters, the dynamic estimation model is expressed as:

$$\begin{bmatrix} v_{\gamma 1} \\ v_{\gamma 2} \\ v_{\gamma 3} \\ v_{\gamma 4} \end{bmatrix} = \frac{1}{r} \begin{bmatrix} 1 & 1 & long_n \\ -1 & 1 & -long_n \\ -1 & 1 & long_n \\ 1 & 1 & -long_n \end{bmatrix} \begin{bmatrix} v_x \\ v_y \\ v_\gamma \end{bmatrix} \tag{16}$$

In the formula, v_γ represents the angular velocity parameter of the device moving different gears, $long_n$ is the distance of the device running for one week, v_x and v_y are the moving speeds of the device in the x and y directions, and r is the radius of the McNamee wheel [18]. The time difference variable Δt of the dynamic image facets is calculated to recognize the dynamic position of the sequence image feature points, and the displacement coordinate variable is extracted to be fused with the 3D reconstruction data:

$$dt_{ni} = LinarInterp\left((x_i, y_i), \frac{long_{ni+1} - long_{ni}}{\Delta t}\right) \tag{17}$$

Through dynamic estimation, the synchronized transmitted images can be analyzed and compared in a more detailed way, and the optimal data can be selected for fusion and three-dimensional reconstruction to reduce the distortion of information due to light and shadow and noise in the moving process.

B. Incremental Learning Synchronization Update

After the initial reconstruction of the 3D image, continuous optimization of the 3D reconstructed data is achieved by incremental learning allowing the 3D model to continuously learn from and update new data while retaining knowledge of the existing data [19]. The edge network transmits new image or scan data to the cloud data center and continuously adds new image frames using the Incremental SFM (Structure for Recovery of Motion) algorithm to 3D map the real-time data onto the existing 3D knowledge [20]:

$$\tilde{R}(x,y,z \mid \Delta t) = \sum_{N^t}^{i=1} y_n \log \frac{\exp(o_{i^*})}{\sum_{N^t}^{i=1} \exp(o_i)} \tag{18}$$

In the formula, $\tilde{R}(x,y,z \mid \Delta t)$ represents the cross entropy of 3D reconstructed feature points under uniform

synchronization conditions, $\exp(o_i)$ and $\exp(o_{i'})$ represent the original feature point center position and the incremented feature point center position, respectively. The dynamic synchronization data is introduced to iterate the network weights λ to calculate the extra data loss existing in the synchronization process:

$$loss\left(\tilde{R}^{t}\right)=\frac{1}{2}\sum_{\left|\tilde{R}^{t-1}\right|}^{i=1}\lambda\left(\tilde{R}_{i}^{t-1}-\tilde{R}_{i}^{t}\right)^{2} \quad (19)$$

It solves the problem of data synchronization inconsistency in 3D reconstruction through incremental learning, and supports long-time continuous learning and optimization to eliminate image data discrepancies and improve the consistency and accuracy of large-scale image 3D reconstruction (see Fig. 8).

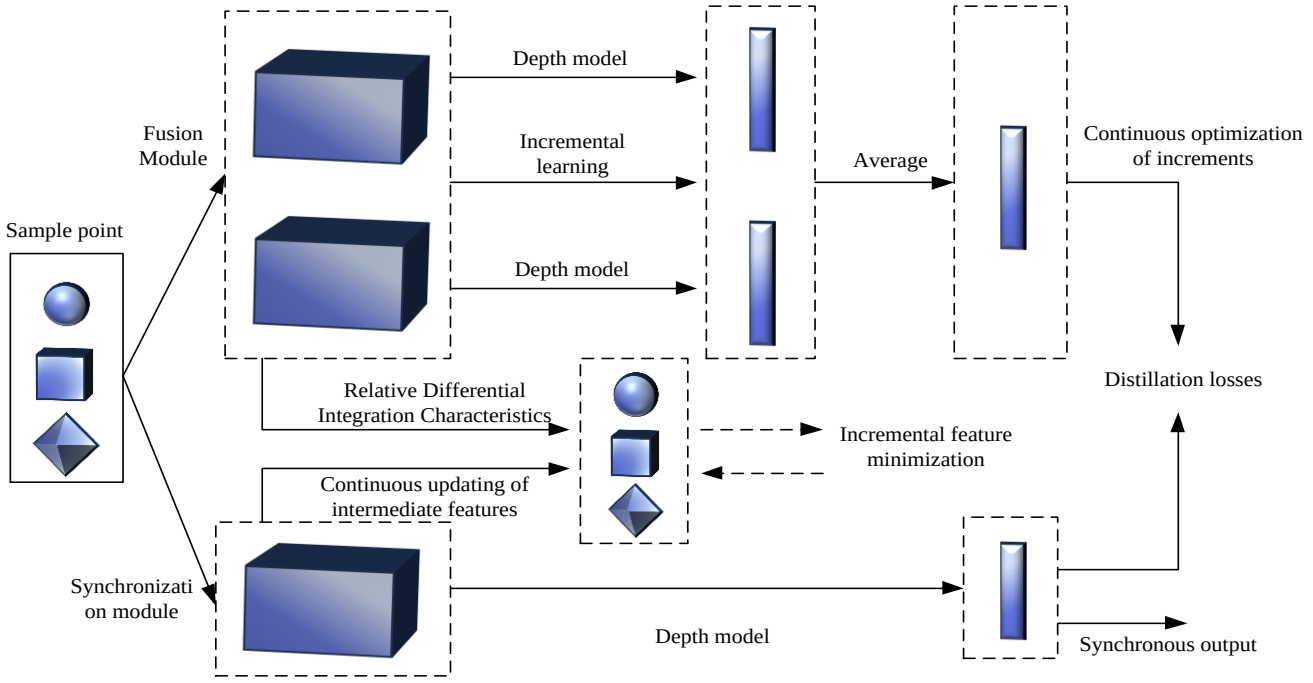


Fig. 8. Incremental learning synchronization mechanism.

V. Experimental Research

To verify the application effectiveness of large-scale image 3D reconstruction technology based on cloud edge fusion architecture and the optimized performance of cloud edge data synchronous update technology, a comparison experiment is designed for image acquisition and 3D reconstruction of transmission and distribution corridors. The experimental cloud data center equipment is selected from Dell Power Edge R430 dual-channel rack servers, and the edge core equipment is selected from Lenovo tower servers, with a unified setup of 8Mbps upstream broadband, and the collected images of transmission and distribution corridors are visualized using MATLAB software, to carry out comprehensive technical evaluation of the network latency state, matching performance, and reconstruction time of the large-scale image three-dimensional reconstruction method. A comprehensive technical evaluation is carried out.

A. Synchronization Update Network Latency Evaluation

The changing dynamics of the synchronization update network state of the image transmission from the front-end device to the system host computer is shown in Fig. 9.

According to the above figure, it can be seen that the data throughput of the cloud edge network based on cloud edge fusion technology is larger, the average throughput is always above 20000 Kbps, and the highest throughput can reach 42000 Kbps. The centralized cloud communication technology used for tilt photography has an average network throughput of about 10000 Kbps to 20000 Kbps due to the lack of a two-end synergistic mechanism. The larger the network throughput, the faster the synchronization and updating speed for large-scale images, so the image synchronization transmission delay time of the cloud-side cooperative network is the shortest, with a maximum of no more than 85 ms and a minimum of only 19 ms, which is significantly better than the other two methods in terms of synchronization transmission speed.

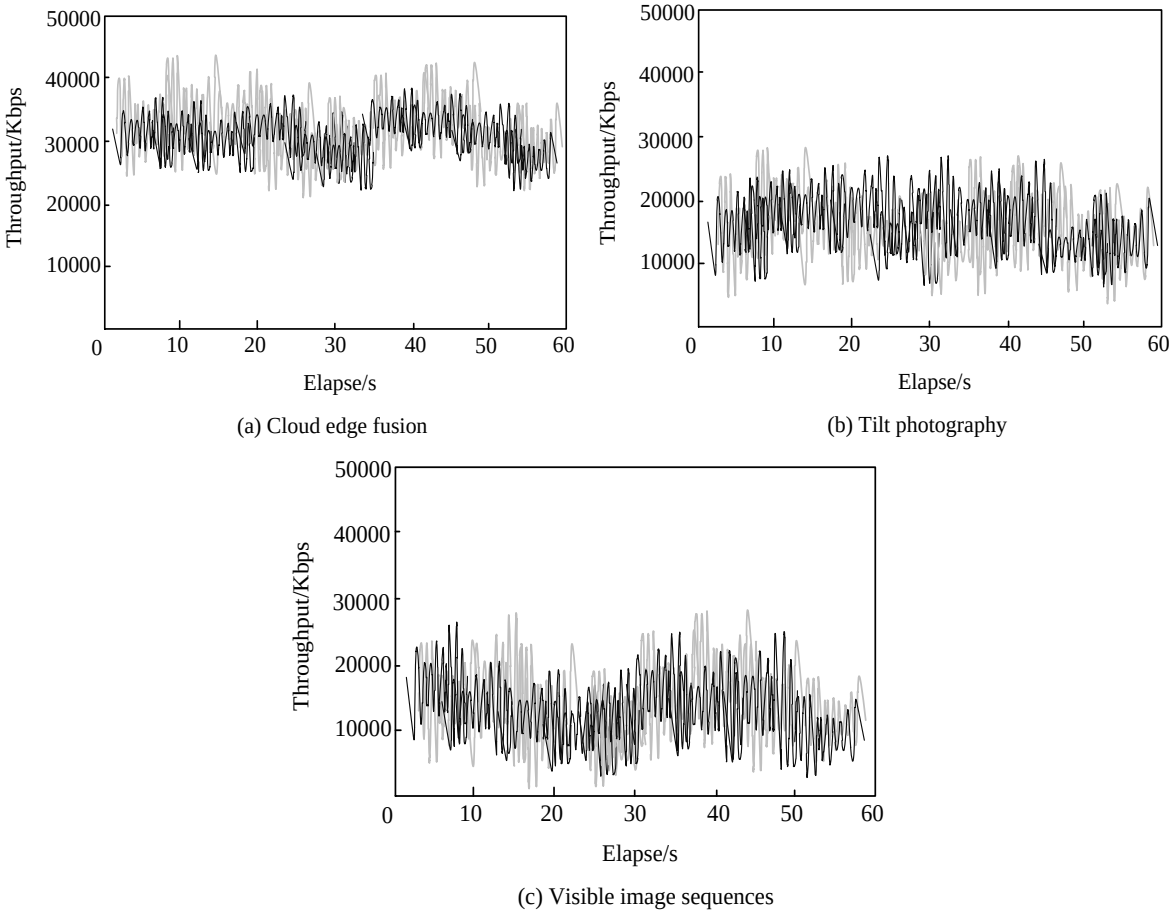


Fig. 9. Network communication throughput.

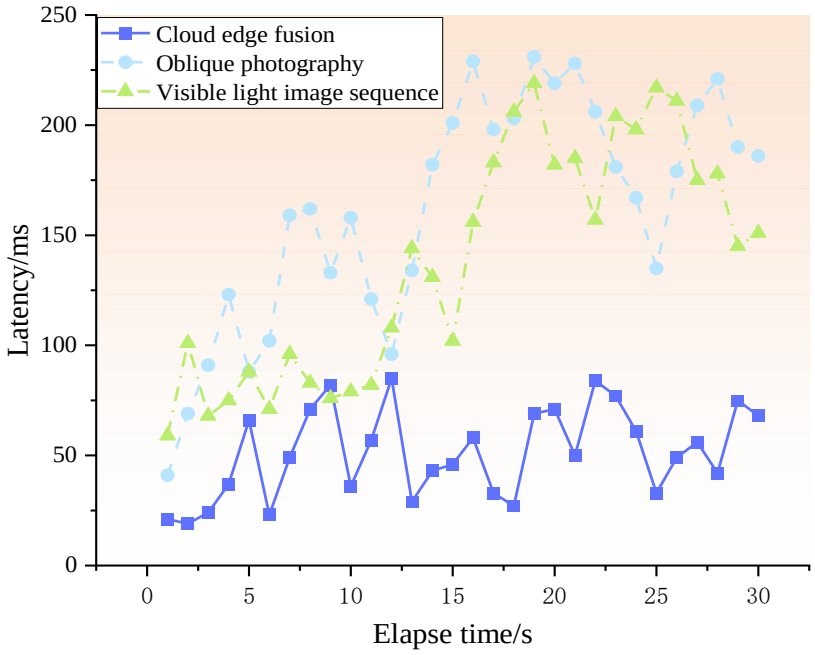


Fig. 10. Network synchronization delay.

B. Image 3D Reconstruction Time Evaluation

To check the processing time of large-scale image 3D reconstruction stages, 50, 100, 200, and 300 images are fused and reconstructed respectively, and the time required for each stage is statistically shown below.

TABLE I. STATISTICS OF 3D RECONSTRUCTION DURATION FOR CLOUD-EDGE FUSION

Number of images	Duration of the first phase(min:s)	Point cloud upload time(min:s)	Duration of the second phase (min:s)	Total processing time (min:s)
50	01:48	02:15	03:23	07:26
100	03:21	03:46	06:58	14:05
200	10:17	07:31	09:18	27:06
300	21:39	17:28	19:42	58:49

TABLE II. STATISTICS ON THE DURATION OF 3D RECONSTRUCTION FOR TILT PHOTOGRAPHY

Number of images	Duration of the first phase (min:s)	Point cloud upload time (h:min:s)	Duration of the second phase (min:s)	Total processing time (h:min:s)
50	06:28	04:10	04:33	15:11
100	11:45	25:26	07:12	44:23
200	25:59	1:09:41	21:37	1:57:17
300	40:27	1:25:58	31:02	2:37:27

TABLE III. STATISTICS OF 3D RECONSTRUCTION DURATION OF THE VISIBLE IMAGE SEQUENCE

Number of images	Duration of the first phase (min:s)	Point cloud upload time (h:min:s)	Duration of the second phase (min:s)	Total processing time (h:min:s)
50	07:18	12:15	06:23	25:55
100	15:21	35:33	13:08	64:02
200	31:37	1:17:21	24:14	2:13:11
300	56:32	1:37:05	39:42	3:33:19

The data in Tables I to III show that cloud-edge fusion uses a cloud-edge cooperative network with a higher synchronization rate, and the time taken in the image point cloud upload phase is much lower than that of the tilt photography and visible sequence techniques. When the number of fused reconstructed images is 300, the first stage of image feature extraction and preprocessing of cloud edge fusion takes 21:39 min, point cloud uploading takes 17:28 min, and the second stage of point cloud reconstruction takes 19:42 min. The total processing time of the whole stage of 3D reconstruction of cloud edge fusion is only 58:49 min, whereas tilt photography requires a total time of 2:37:49 min. The total processing time for the whole phase of 3D reconstruction is only 58:49 min, while the total time required for tilt photography is 2:37:27 h, and the total time used for the visible light sequence is 3:33:19 h. Thus, it can be seen that the cloud-edge fusion technology not only has good network synchronization performance but also has obvious advantages in image feature extraction and point cloud reconstruction functions.

C. Reconstruction Matching Performance Evaluation

It is verified that the cloud-edge fusion 3D reconstruction technique has good network synchronization performance, and further evaluation of its matching performance is still needed.

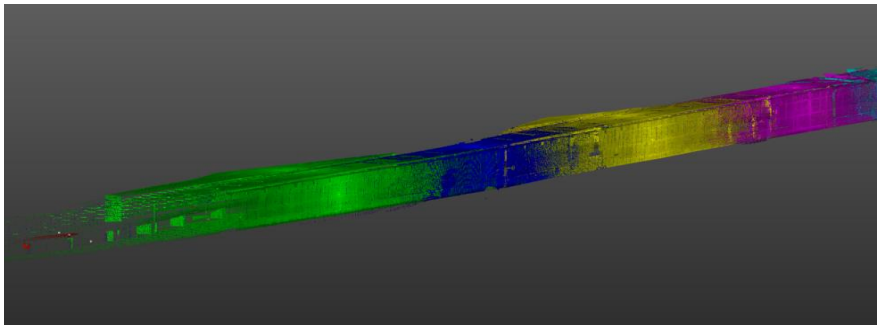


Fig. 11. 3D reconstruction of the visualized image.

Completing the 3D reconstruction of the transmission and distribution digital corridor, the statistics of the number of matches, the number of matching points, the success rate of matching, and the length of time taken for the three methods of identification are shown in the following Table IV.

TABLE IV. COMPARISON OF THREE-DIMENSIONAL RECONSTRUCTION MATCHING PERFORMANCE

Methodologies	Number of matches/times	Match Points/Each	Match rate/%	Timing
Cloud Edge Fusion	5981	3925	96.72	1 h 11 min
Oblique photography	25883	1098	80.35	3 h 05 min
Visible Image Sequence	30768	875	72.64	3 h 58 min

According to the data in Table IV, it can be seen that for 3D reconstruction of the same transmission and distribution corridor image, the cloud-edge fusion technique takes the shortest time and carries out the least number of matches, but successfully recognizes and matches the greatest number of point clouds, with a total of 3925 matching points identified, and the matching success rate is as high as 96.72%. In summary, the cloud-edge fusion 3D reconstruction technique has efficient synchronization efficiency of the cooperative network and also has a good application effect in the point cloud recognition and matching function, which has significant advantages over other methods.

VI. DISCUSSION

The research results show that the synchronous update and optimization method for large-scale image 3D reconstruction technology based on cloud-edge fusion architecture has significant application advantages. On the one hand, the architecture can efficiently integrate cloud and edge computing resources, dynamically allocate them according to the task demand, hand over the computationally intensive work to the cloud, and the edge end is responsible for data acquisition and pre-processing, which greatly improves the resource utilization rate. On the other hand, the edge end can perform local preliminary processing of image data, reducing the time delay of transmission to the cloud. The synchronous update mechanism gives the technology adaptability and flexibility to adapt to changing image data and application scenarios, and the cloud-edge convergence architecture can be flexibly expanded and deployed according to actual needs to meet the needs of different scales and types of applications.

The author believes that the technology has great potential for development and has a bright future. In the future, it can further explore smarter resource allocation strategies and synchronized update mechanisms, improve the degree of system automation and efficiency, and achieve more innovative breakthroughs by combining advanced technologies such as artificial intelligence and 5G communication. In addition, research on data security and privacy protection should be strengthened to ensure the safe storage and transmission of large-scale image data under the cloud-side convergence architecture, so as to bring more efficient, accurate, and reliable image 3D reconstruction solutions to many fields.

VII. CONCLUSION

Aiming at the problem of three-dimensional reconstruction of power transmission and distribution corridors, this paper optimizes the design of large-scale image three-dimensional reconstruction technology and its synchronous update performance under the cloud-edge fusion architecture, and mainly completes the following research: Establish the transmission and distribution digital corridor IoT cloud-edge fusion network architecture, optimize the cloud data center and edge node structure; Optimize the accuracy of cloud-edge fusion 3D reconstruction point cloud matching by using non-sampling filtering, edge checking, and Gaussian convolution; Enhance the synchronous updating performance of cloud-edge cooperative networks through dynamic evaluation of feature points and incremental learning algorithm.

It can be seen through experiments that the cloud-edge fusion 3D reconstruction technology can realize accurate image 3D reconstruction in a short time, the information synchronization update delay rate is low, and the processing time is shorter. Although the results of this paper have the above advantages, there are still some shortcomings: Cloud edge fusion technology requires higher compatibility equipment configuration, higher application costs, and relatively greater technical difficulties; 3D reconstruction software for image fusion detail processing and model training balance control is insufficient, there are still errors and distortion. Future research needs to be deepened from the above perspectives to comprehensively improve the efficiency and accuracy of the application of cloud edge fusion 3D reconstruction systems.

REFERENCES

[1] Feng Xiao; Li Bingran; Bai Chenxu. Three-dimensional reconstruction of transmission line based on oblique photogrammetry. Electrotechnical Application, 2020, 39(3):51-54.

[2] Zhang Lu, Yuan Wei, Liu Xiaolin.Research on Three-dimensional Reconstruction Model of Transmission Line Corridors based on Visible Light Image Sequences. Electric Power Equipment Management, 2023(11):96-98.

[3] Song Yaqin; Xu Hui; Liu Xianfeng; Wang Yapeng; Cheng Zhimi; Wang Hucheng; Chen Shanzhi. Cloud-Edge Collaboration Architecture and Key Technologies for 6G Integrated Satellite and Terrestrial Network. Space-Integrated-Ground Information Networks, 2023, 4(3):3-11.

[4] Yang Hang; Chen Rui; An Shipeng; Wei Hao; Zhang Heng. The growth of image-related three dimensional reconstruction techniques in deep learning-driven era:a critical summary. Journal of Image and Graphics, 2023, 28(8):2396-2409.

[5] Zhang Lei; Shi Yan; Lu Wenyong; Xu Rui; Jin Zhan; Luo Weijie; Cehn Yi; Zhao Chunliu; Zhan Chunlian.3D reconstruction technique based on SURF-OKG feature matching. Optics and Precision Engineering, 2024, 32(6):915-929.

[6] Su Yu; Zhang Zexu; Yuan Mengmeng; Xu Tianlai; Deng Hanzhi; Wang Jing. A Point Cloud Fusion Method for Space Target 3D Laser Point Cloud and Visible Light Image Reconstruction Method. Journal of Deep Space Exploration, 2021, 8(5):534-540.

[7] Ren Mengxin, Yang Jianfeng, Deng Zhougray, Zou Qiong, Tong Tianle. 3D Reconstruction Method Based on PP-Matting and Incremental Structure-from-Motion. Modeling and Simulation, 2023, 12(4):4116-4126.

[8] Zhou X, Xu X, Liang W, et al. Deep-Learning-Enhanced Multitarget Detection for End–Edge–Cloud Surveillance in Smart IoT. IEEE internet of things journal, 2021(8-16):12588-12596.

[9] Zhang E, Hu K, Xia M, et al. Multilevel feature context semantic fusion network for cloud and cloud shadow segmentation. Journal of Applied Remote Sensing, 2022, 16(4):1-26.

[10] Shang X. Enabling Data-intensive Workflows in Heterogeneous Edge-cloud Networks. Performance Evaluation Review, 2022,50(03):36-38.

[11] Jiang S, Gao H, Wang X, et al. Deep reinforcement learning based multi-level dynamic reconfiguration for urban distribution network: a cloud-edge collaboration architecture. Global Energy Interconnection, 2023, 6(1):1-14.

[12] Wang Z, Ko I Y. Edge-Cloud Collaboration Architecture for Efficient Web-Based Cognitive Services. 2023 IEEE International Conference on Big Data and Smart Computing (BigComp), 2023:124-131.

[13] Wang M, Liu R, Yang J, et al. Traffic Sign Three-Dimensional Reconstruction Based on Point Clouds and Panoramic Images. Photogrammetric record, 2022(Mar. TN.177):37.

[14] Gaige D, Rongwu W, Chengzu Li, et al. Three-Dimensional Model Reconstruction of Nonwovens from Multi-Focus Images. Journal of Donghua University (English version), 2022(003):185-192.

[15] Iacoviello P, Bacigaluppi S, Gramegna M, et al. Microsurgical Three-Dimensional Reconstruction of Complex Nasal and Midfacial Defect: Multistep Procedure Respecting Aesthetic Unit Criteria. Journal of Craniofacial Surgery, 2021,(32):1517-1520.

[16] Shan C, Yao Q, Cao S, et al. Measurement of fracture development evolution of coal samples under acid-alkaline by three-dimensional reconstruction and AE time-frequency characteristic analysis. Measurement, 2023(217):1-23.

[17] Zhang C, Guo Y, Meng D, et al. Hybrid iteration and optimization-based three-dimensional reconstruction for space non-cooperative targets with monocular vision and sparse lidar fusion. Aerospace Science and Technology, 2023(140):1-16.

[18] Zhao L, Pan J, Xu L. Cone-Beam Computed Tomography Image Features under Intelligent Three-Dimensional Reconstruction Algorithm in the Evaluation of Intraoperative and Postoperative Curative Effect of Dental Pulp Disease Using Root Canal Therapy. Scientific programming, 2022(3):1-8.

[19] Zhao M, Lin M, Xu P. A Study on the Whole-skin Peeling System of a Snakehead Based on Three-dimensional Reconstruction. Applied Engineering in Agriculture, 2022(38):741-751.

[20] Zhang F, Pan H, Zhang X, et al. Three-dimensional reconstruction for flame chemiluminescence field using a calibration enhanced non-negative algebraic reconstruction technique. Optics Communications: A Journal Devoted to the Rapid Publication of Short Contributions in the Field of Optics and Interaction of Light with Matter, 2022(520):1-10.

The Application of Anti-Collision Algorithms in University Records Management

Ying Wang[1]*, Ying Mi[2]
Archives of the Party and Administration Office, North China Electric Power University, Baoding, 071000, China[1]
Department of Economic Management, Baoding Vocational and Technical College, Baoding, 071000, China[2]

***Abstract*—University records management has grown in importance as a result of the quick growth of big data, artificial intelligence, and other technologies. However, university archives management is prone to data loss, redundancy, and errors. Moreover, the use of scientific management systems and algorithms can effectively improve such problems. To create an effective and secure archive management system and run simulation tests, the study suggests an RFID-based archive management system and uses nested random time slot ALOHA (RS0) and binary tree (BT) anti-collision algorithms to solve the collision problem between tags in the created system. The test results showed that the average query coefficient, recognition efficiency, and communication volume of the proposed algorithm were 1 and 1.2 times, 95% and 90%, 50 Bit and 180 Bit in two scenarios, continuous and uniform, respectively. 0.91% and 3.92%, 24.21% and 31.14% of the system CUP and memory occupation were achieved when the number of clients was 10 and 100, respectively. The average response time of the system was 0.112s and 1.244s when 100 and 1000 users were accessed, respectively. The information extraction accuracy of the system was 94% at 1000 accessed users. This suggests that the approach used in the study can significantly improve the operational effectiveness of the records management system and the accuracy of information extraction, as well as provide technical support for improving the university records management system.**

Keywords*—Anti-collision algorithms; archive management systems; information networks; RFID technology

I. INTRODUCTION

The number of archives has greatly increased as the twenty-first century has progressed due to the growth in both students and faculty in higher education. As a result, many schools have established electronic information-based records management systems [1-2]. Additionally, radio frequency identification technology can be extensively employed in the gathering and extraction of archival information, among other things, due to its benefits of high recognition efficiency and long usage time [3]. However, one of the key issues with radio frequency recognition technology is the collision problem. In this technology, the collision problem is further broken down into collisions between readers and collisions between tags. In most cases, it is the collision between tags that is encountered. While common RFID tag anti-collision algorithms are effective in improving the system's automatic identification performance, they also have several shortcomings. In high collision situations, each time slot or query will be subject to a significant amount of duplication, which can lead to a reduction in system efficiency or lag. Additionally, these algorithms are not well-suited for big data applications. When the data volume is very high and the number of tags exceeds the system's processing capacity, the system may experience performance issues and substantial delays [4-6]. The standard anti-collision technique is therefore no longer able to satisfy the requirements of archive management in the big data era. Therefore, to maintain the security, integrity, efficiency, and accuracy of archive management, it is important to upgrade the conventional anti-collision algorithm. Based on this background, it is proposed that when a tag collision occurs, a little time slot is first released to handle the tags that have collided. Depending on the number of tags that have collided, the nested random time slot ALOHA (SR) algorithm and the nested choice binary tree (BT) anti-collision algorithm are used to process them in anticipation of further improving the recognition efficiency and accuracy of the system.

II. OVERVIEW

Data management of archives is one of the key issues in the construction of universities today. The huge number and variety of types of archives in universities require a standardized management system to ensure the security, integrity and accuracy of archives management as well as the efficiency and accuracy of archives access. Numerous specialists have studied data management extensively in this area. Uka K. K. et al. created an online system for document sharing and data management at higher education institutions. The programme was a decentralized cloud-based file management and sharing programme. The results showed that the proposed system was able to enable file sharing, replication and data management permission transactions between accessing users in higher education institutions [7]. To make viewing of platform picture data more safe, Gamido et al. created an image file management system. The system was encrypted and kept on the server, and image files were encrypted using the AES technique to give the file owner more security. The outcomes demonstrated that the encryption of image files on the server was successful, and the technique for exchanging images was now reliable and efficient [8]. Veena et al. constructed a hybrid image and document archiving system in order to increase the volume of document storage and make document retrieval more efficient. The system used a client-server architecture, which utilized Opencv to identify objects in images and Tesseract to identify text in images and generate labels. The results showed that the system could be used to store and retrieve large numbers of documents [9]. Zhou et al. proposed a metadata service approach to the problem of dealing with large amounts of data requiring long, continuous and uninterrupted data access. The strategy made

use of a novel shared storage pool and main-standby fault-tolerant design. The outcomes demonstrated that the method significantly decreased the average recovery time of the data service and enhanced the availability of the file system [10]. In the context of currently utilized distributed file systems that differentiate between the file system and network layers, Zhu et al. proposed a distributed persistent memory file system with RDMA support. The system introduced self-identifying remote procedure calls and offered direct access to a shared pool of persistent memory. The results demonstrated that the file system outperformed other distributed file systems by several orders of magnitude [11].

The anti-collision algorithm, which is extensively employed in many sectors, can significantly enhance the system's automatic recognition performance. Choi et al. proposed an anti-collision algorithm for mobile robots to address the problem that robots were prone to collision during the distribution process. The algorithm will reset a new path and motion when the robot collides. The outcomes demonstrated that the system could successfully prevent a robot-robot collision [12]. The multi-tag collision problem in RFID systems was addressed by the hybrid ALOHA and tree algorithm (HAMT) proposed by Zhou et al. For tag recognition, the method employed the DFSA algorithm. The outcomes demonstrated that the algorithm's recognition efficiency could increase to about 0.72 [13]. For the collision problem of RFID tag detection in the Internet of Things, Qiu et al. suggested an improved group adaptive query tree technique (IGAQT). The approach fixed the issues with the original algorithm's temporal complexity and communication overload. The outcomes demonstrated that the proposed strategy had a significantly higher recognition efficiency [14]. For the RFI multi-tag collision avoidance problem, Qu J et al. offered an adaptive frame-time slot ALOHA collision avoidance algorithm based on IGA. The algorithm counted the amount of tags before grouping them to identify tags. According to the findings, the algorithm was approximately 71% efficient, which is 90% more effective than the conventional ALOHA algorithm [15]. To address the collision issue, Jing C et al. devised an anti-collision method based on blind source separation (BSS). FastICA, PowerICA, ICA_p, and SNR_MAX were among the BSS algorithms that were used to separate and test the mixed signals in RFID systems. The ICA_p method had the best overall performance among the aforementioned algorithms, according to the results [16].

From the research of numerous scholars mentioned above, anti-collision algorithms have been widely applied in multiple fields, most of which are applied in the Internet of Things, robot path research, and other fields. Few scholars have combined anti-collision algorithms with data archive management systems. Therefore, this research innovatively introduces anti-collision algorithms, a widely used technology in fields such as the Internet of Things, into university archive management systems, thus filling the technical gap in RFID tag recognition in this field. At the same time, the existing anti-collision algorithms are improved and optimized, and a combination scheme of nested SR algorithm and BT anti-collision algorithm is proposed. This scheme has shown significant advantages in reducing tag collisions, improving recognition efficiency and accuracy, and providing new ideas and methods for the application of RFID technology in the field of records management.

III. Design of an RFID-Anti-Collision Algorithm Based Archive Management System For Universities

This chapter first elaborates on the functions and design scheme of a university archives management system based on RFID technology. In response to the issues of data loss, redundancy, and errors that are inherent to university archives management, a nested SR algorithm and BT anti-collision algorithm have been developed to enhance the efficiency and accuracy of tag collision recognition, mitigate the risk of collisions between tags, and ultimately improve the operational efficacy of the university archives management system.

A. Model Design of RFID Technology-based University Archive Management System

In colleges and universities, there are many distinct kinds of records management, and each college has various records management requirements. The basic functions required for a general university records management system are shown in Fig. 1.

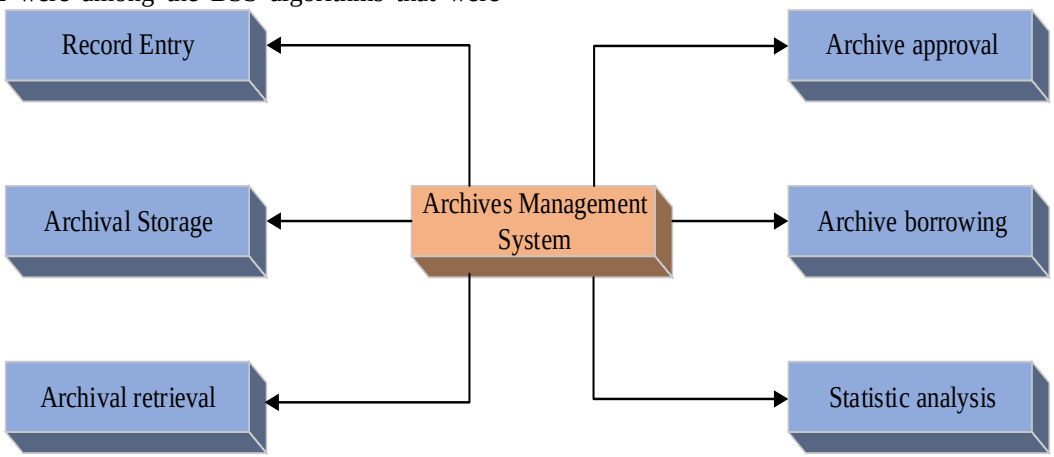


Fig. 1. Basic functions of university archives management system.

The basic operations of the university's records management system, including record entry, record storage, record retrieval, record approval, record borrowing, and statistical analysis, are shown in Fig. 1. According to the different needs of major universities for records management, the corresponding functions can be added or deleted. According to the needs of general universities for archive management systems, a network connection structure of university archive management systems based on RFID technology is studied and designed, as shown in Fig. 2.

As shown in Fig. 2, the network structure of the system uses a reader to first read the file information collected by the antenna, a WEB server to send it to the database server for backup, and a terminal server to implement file information management for students and teachers [17]. Four key components make up the function of the file management system developed for the project, which are shown in Fig. 3.

As demonstrated in Fig. 3, the archive management system designed in the study mainly includes four modules: user management, archive management, information collection and arrangement, and database management. The user management module mainly focuses on the access rights of different users to the archives. The database module mainly backs up and destroys the archives. The information collection and arrangement module mainly collects and processes information from different sources of archives. The archive management mainly manages the information of archives such as access, borrowing and returning, and whether they are in place. The database module is used for backing up and destroying the archives. The information collection and arrangement module is used for collecting and processing information from different sources. The archives management module is used for managing the information on access, borrowing and returning, and whether the archives are in place [18]. Six major divisions and fourteen subsections have been created based on the needs of the system's users to better serve their needs. The specific structure of the system functions for the management of archives by users is shown in Fig. 3.

As shown in Fig. 4, the basic functions of archive management in the university archive management system include user rights management, organization management, classification management, archive management, and borrowing management. According to the different user rights, the system divides users into three categories: system administrator, file administrator, and ordinary user. The system administrator has the highest authority and can perform all the functions of the system and modify the user rights to modify the file information. The file administrator is responsible for the daily maintenance of the system. Ordinary users can only query their own file information. The study employs a three-tier architecture for the software system structure in order to meet the many functional needs of the users and to make the management system convenient and effective with lower running costs. Fig. 5 shows the exact technical design of this system.

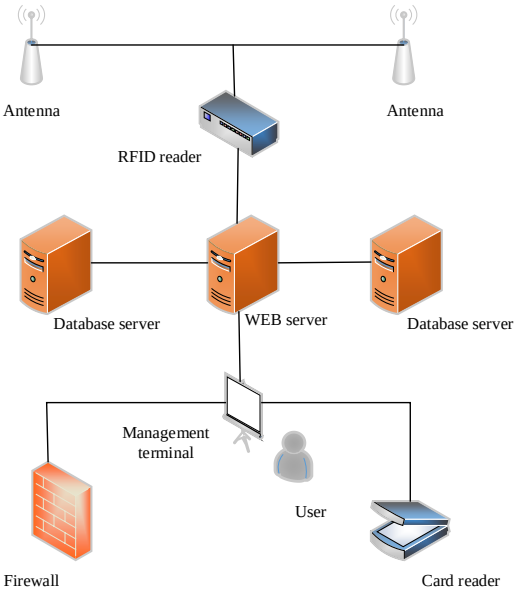


Fig. 2. System for managing university archives connections to the internet.

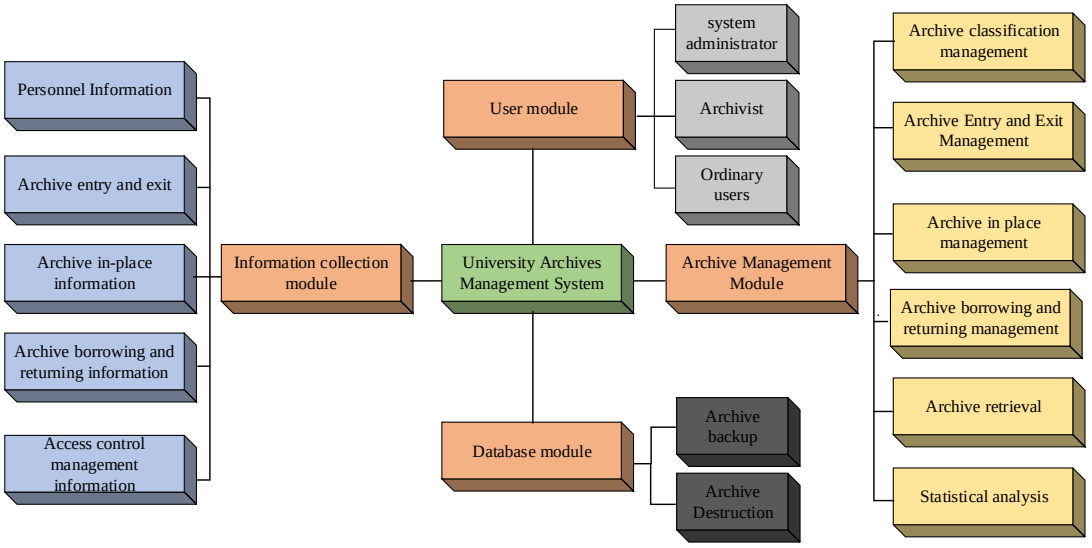


Fig. 3. System for managing university archives, showing its functional structure.

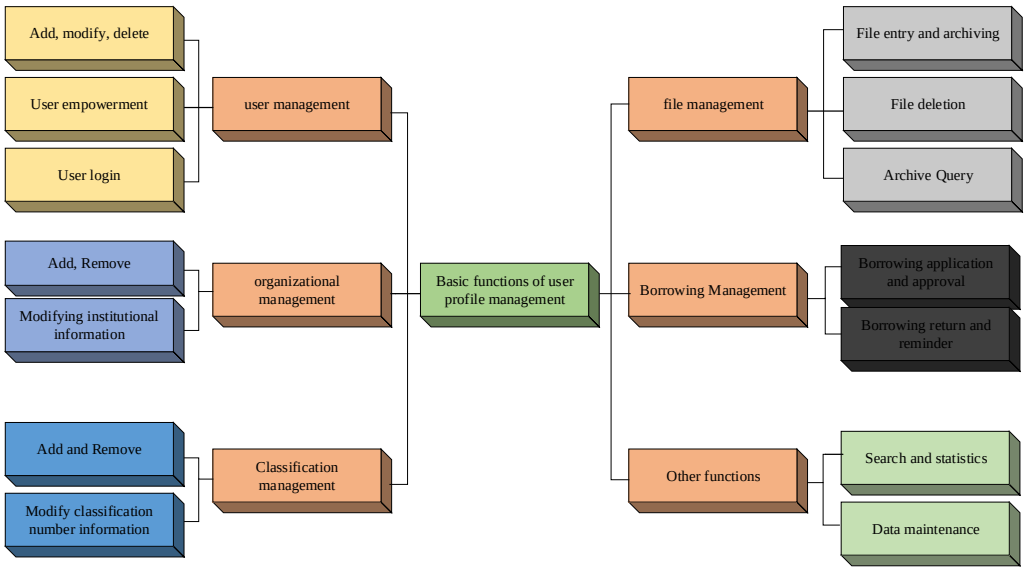


Fig. 4. Schematic representation of the fundamental file management operations carried out by users.

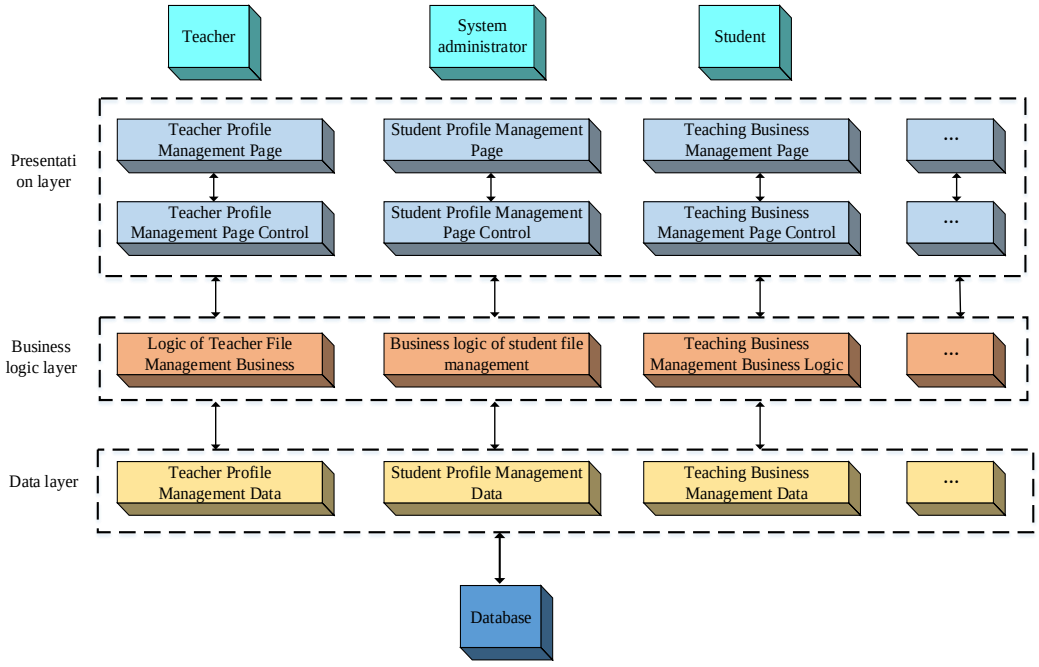


Fig. 5. Technical architecture of university archives management system.

Fig. 5 illustrates the three-layer B/S architecture used by the university file management system, with the presentation layer serving as the user operation interface via which users can manage, query, and alter files as needed. The business logic layer is the system operation logic processing, the main function of which is to parse the information sent by the user operation interface and execute the parsed instructions, and then send the instructions to the data layer. The data layer serves as the system's data processing layer. Its primary function is to filter out the necessary data in accordance with the instructions from the business logic layer, after which it sends the data back to the business logic layer for processing before sending it to the presentation layer for the users to see. Then, according to the needs of this file management system, the corresponding running tools and development environment are selected, as shown in Table I.

TABLE I. OPERATING CONDITIONS AND TOOLS OF ARCHIVE MANAGEMENT SYSTEM

Operating system	WIN 8
Database	MYSQL
B/S architecture development tool	Eclipse

B. *Design of Improved Anti-Collision Algorithms Based on RS and BT*

The university's RFID-based file management system has a sizable number of tags. When data is transmitted between

these tags and the reader, collision of data information is likely to occur, resulting in failure of information transmission [19]. There are two situations in which data information transmission collision occurs, as shown in Fig. 6.

Fig. 6(a) shows a tag-to-reader collision, which is caused by a tag being unable to accept command requests from multiple readers at the same time. A tag-to-tag collision is depicted in Fig. 6(b), where a single reader is unable to extract the data from several tags at once. While the former collision scenario can be avoided by adjusting transmit power, the latter scenario requires strengthening the anti-collision algorithm. The ALOHA-based algorithm and the tree-based algorithm are the two anti-collision algorithms between tags that are most frequently utilized. ALOHA algorithms include the pure ALOHA algorithm (PA algorithm), the time slot ALOHA algorithm (SA algorithm), and the frame time slot ALOHA algorithm (FSA algorithm) [20]. The FSA algorithm decomposes the recognition process into separate recognition frames. Since the FSA algorithm can break down the recognition process into individual frames and utilize the time slot ALOHA algorithm one at a time, this method has greater recognition accuracy. The SR algorithm is devised since the amount of time slots per frame cannot be altered. This algorithm has the ability to change the total number of time slots in a recognition cycle and allow unrecognized tags to select a different time slot. The algorithm flow is shown in Fig. 7.

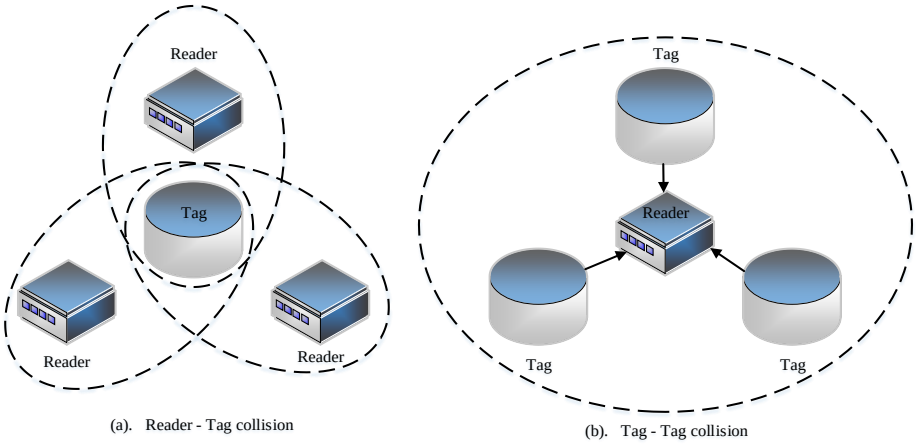


Fig. 6. Collision situation.

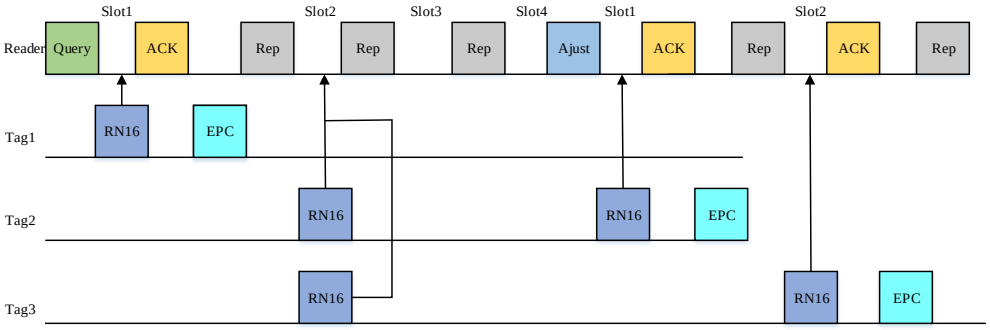


Fig. 7. SR algorithm flowchart.

The SR algorithm's main flow is depicted in Fig. 7. The reader transmits a command to the tag with the parameter Q, which the tag receives and uses to create a random number as the response command when the time period is up. Finally, the reader sends a separate instruction and moves on to the following time slot based on the corresponding amount of tags. The success rate of a specific tag identification in a particular time slot of the cycle is displayed in Eq. (1).

$$Pe = C_n^1 \frac{1}{2^Q}\left(1-\frac{1}{2^Q}\right)^{(n-1)} \tag{1}$$

In Eq. (1), Q is the parameter in the command. n is the total number of tags. When the reader accepts signals from several tags at the same time, the probability of collision between tags is shown in Eq. (2).

$$P_i = C_n^i \left(\frac{1}{2^Q}\right)^i \left(1-\frac{1}{2^Q}\right)^{(n-i)} \left(i \geq 2\right) \tag{2}$$

Let the total number of time slots be 2^Q , then the total expected value of tag collisions over the entire recognition cycle is calculated as shown in Eq. (3).

$$E_0 = n\left(1-\frac{1}{2^Q}\right)^{(n-1)} \tag{3}$$

In Eq. (3), E_0 is the total expected value of collisions for a given cycle of tags, based on which the algorithm efficiency

of SR can be calculated in Eq. (4)

$$\eta_{SR} = \frac{n\left(1-\frac{1}{2^Q}\right)^{(n-1)}}{2^Q} \tag{4}$$

In Eq. (4), η_{SR} is the efficiency of SR algorithm. The tree structure-based collision prevention algorithms include the BT algorithm, the query algorithm (QT algorithm), and the spanning tree algorithm. Fig. 8 displays the schematic diagram of the BT anti-collision algorithm among them.

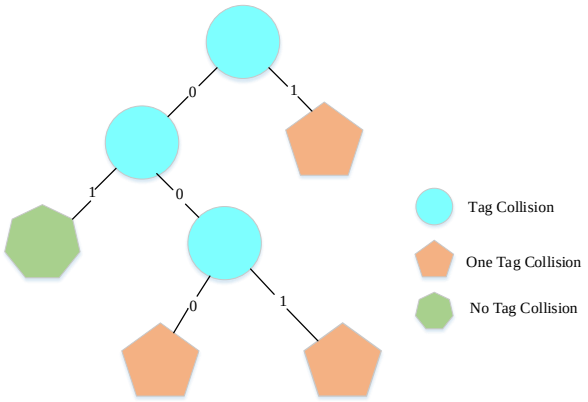


Fig. 8. Schematic diagram of binary tree anti-collision.

As shown in Fig. 8, the algorithm starts with a counter for each reader and tag, each with a value of 0. The tag then sends a message to the reader, and the reader adjusts the value of the counter according to the message to drive the algorithm. Finally, the algorithm terminates when the reader's counter value reaches 0. The formula for the h level of the binary tree, which has 2^k cycles at the h level, is given in Eq. (5).

$$\begin{cases} S_{BT}(n,h) = 2^h\left(1-\frac{1}{2^h}\right)^n \\ A_{BT}(n,h) = n\left(1-\frac{1}{2^h}\right)^{(n-1)} \\ H_{BT}(n,h) = \sum_{h=0}^{\infty} 2^h\left[1-\left(1-\frac{1}{2^h}\right)^{(n-1)} - n\frac{1}{2^h}\left(1-\frac{1}{2^h}\right)^{(n-1)}\right] \end{cases} \tag{5}$$

In Eq. (5), $S_{BT}(n,h)$ is the number of idle cycles. $A_{BT}(n,h)$ is the number of readable cycles. $H_{BT}(n,h)$ is the number of collision cycles for the h layer. In this algorithm, the number of queries $L(n)$ and the amount of communication $S(n)$ for the reader to recognize the tag are shown in Eq. (6).

$$\begin{cases} L(n) = n(\log_2 n+1) \\ S(n) = Y\square n(\log_2 n+1) \end{cases} \tag{6}$$

Based on the basic properties of bifurcation numbers, the total period identified by the algorithm can be obtained as $1+H_{BT}$. Both of these algorithms are common anti-collision algorithms for handling tag collisions, but both have shortcomings, so the study proposes a sub-nesting algorithm to improve them. In the event of a collision between tags when a command is sent in the SR algorithm, the reader processes the tag directly into the subsequent time slot, thereby bypassing the tag that collided. To address this problem, the study proposes that instead of going to the next gap immediately after a collision occurs, the colliding tag is processed first and the SR algorithm is nested with SR. Let the number of time slots to be released by this algorithm be m, and the SR nested SR algorithm is analyzed by the formula, the probability of success of identifying a tag per time slot in this algorithm is shown in Eq. (7).

$$Pe_1 = \left(1-\frac{1}{m}\right)^{(m-1)} \tag{7}$$

Since there is a m time slot, the expectation of successful tag recognition at m is expressed in Eq. (8).

$$E_1 = m\left(1-\frac{1}{m}\right)^{(m-1)} \tag{8}$$

In contrast, the traditional SR algorithm's expectation of successful tag recognition in m time slots is shown in Eq. (9).

$$\frac{n}{2}\left(1-\frac{1}{2^Q}\right)^{(n-1)} \tag{9}$$

Therefore, the nested SR algorithm label recognition success expectation optimization rate can be obtained as shown in Eq. (10).

$$\eta = \frac{\frac{n}{2}\left(1-\frac{1}{2^Q}\right)^{(n-1)}}{m\left(1-\frac{1}{m}\right)^{(m-1)}} \times 100\% \tag{10}$$

Moreover, for the improvement of the BT algorithm, the study uses the SR nested BT algorithm to improve it. Let the number of tags to be recognized be L ,the depth of the binary tree is $\left\lceil \log_2^L \right\rceil$, combined with the traditional BT algorithm, Eq. (11) can be used to determine the time required to recognize the tags and the formula for calculating expectations.

$$\begin{cases} E_A = \sum_{a=1}^{\left\lceil \log_2^L \right\rceil}\left(1-\frac{1}{2^a}\right)^{L-1} \\ T_A = \sum_{a=1}^{\left\lceil \log_2^L \right\rceil} 2^a \end{cases} \tag{11}$$

In Eq. (11), E_A is the expectation of successful tag recognition. T_A is the time taken to recognize the tag. The efficiency of the algorithm can be obtained from the

expectation and the time taken. Next, the SR nested SR algorithm and the SR nested BT algorithm are compared. Let the total number of tags identified by the SR nested BT algorithm be E_A and the time taken be T_A. For the same usage time, the total number of tags identified by the SR nested SR algorithm is E_{SR}, and the tag identification expectation is given in Eq. (12).

$$E_{SR} = T_A\left(1-\frac{1}{T_A}\right)^{(T_A-1)} \tag{12}$$

The label recognition expectations according to the two algorithms are compared in size using the do-quotient method, as shown in Eq. (13).

$$\lambda = \frac{T_A\left(1-\frac{1}{T_A}\right)^{(T_A-1)}}{\sum_{a=1}^{\lceil \log_2^L \rceil}\left(1-\frac{1}{2^a}\right)^{L-1}} \tag{13}$$

The time used in Eq. (13) can be brought in with the time used in Eq. (10) and calculated to give Eq. (14).

$$\lambda = \frac{\sum_{a=1}^{\lceil \log_2^L \rceil} 2^a\left(1-\frac{1}{\sum_{a=1}^{\lceil \log_2^L \rceil} 2^a}\right)^{\left(\sum_{a=1}^{\lceil \log_2^L \rceil} 2^a\right)-1}}{\sum_{a=1}^{\lceil \log_2^L \rceil}\left(1-\frac{1}{2^a}\right)^{L-1}} \tag{14}$$

According to Eq. (14), it is concluded that the SR nested SR algorithm has more tag recognition when $L \geq 5$ and the SR nested BT algorithm has more tag recognition when $L < 5$ [21]. Based on this conclusion, the nested RS and BT anti-collision algorithm flow is shown in Fig. 9.

The upgraded algorithm's reader delivers a command with a Q value, as seen in Fig. 9, which kick-starts the cycle of tag recognition. Each tag receiving the command randomly generates its own tag number and selects a corresponding time slot. When there is no tag response, the reader issues a command to move on to the next time slot. When there is one tag response, the reader issues a command to read and write the tag before moving on to the next time slot. Moreover, when there are multiple tag responses, the reader issues a nested BT algorithm command if there are $L < 5$ or $L \geq 5$ tag collisions, respectively.

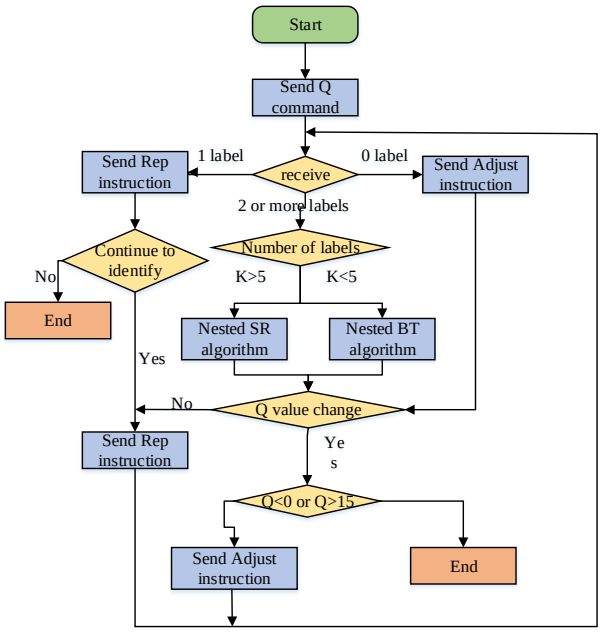


Fig. 9. Flow diagram of nested RS and BT anti-collision algorithm.

IV. Performance Analysis and Simulation Application of RFID-Based Anti-Collision Algorithm for Archive Management System

This chapter mainly elaborates on the performance of nested RS and BT anti-collision algorithms and the design of file management system detection experiments.

A. Performance Analysis of Nested RS and BT Anti-Collision Algorithms

The study divides the tag identity codes into two distribution scenarios—continuous and uniform—to test the effectiveness of the recommendation algorithms developed in this study. The performance of the five QT algorithms, DBS algorithm, RS algorithm, BT algorithm, nested RS and BT algorithms are simulated and tested in these two scenarios respectively. Table II displays the experimental parameter settings.

TABLE II. Experimental Parameter Settings

Number of readers	1
Label length	100
Transfer rate	100kbps
Number of labels	4~2048
Software	Matlab
Number of experiments	100

The average number of queries for each algorithm is first tested in two distribution scenarios for the tag identity code. The results of which are shown in Fig. 10.

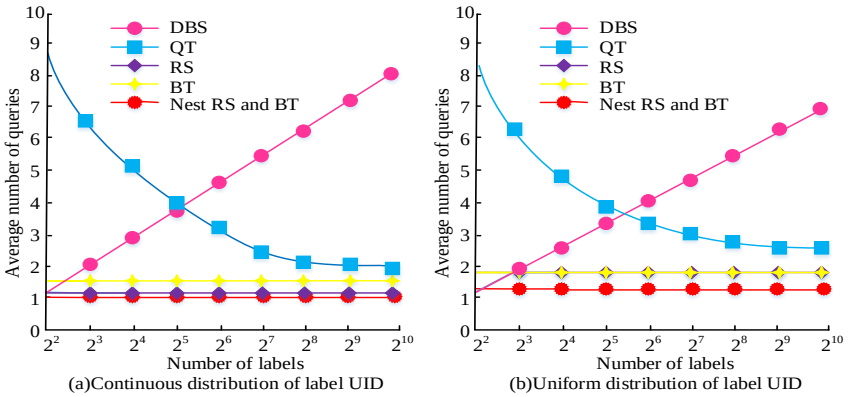


Fig. 10. Average query times of different algorithms in two scenarios.

As shown in Fig. 10(a), the average number of queries for the five algorithms in the continuous scenario is shown. It can be concluded that the average query counts of the three algorithms, nested RS and BT algorithms, RS algorithm and BT algorithm, remain stable. Therefore, the average number of times to identify a tag is 1, 1.2 and 1.5 times, respectively. The average query count of DBS algorithm increases with the number of tags. The average query count of QT algorithm decreases with the number of tags. The average number of inquiries for the five methods in the uniform scenario are shown in Fig. 10(b). It can be concluded that the DBS algorithm and QT algorithm query counts vary similarly to (a). The RS algorithm and BT algorithm both have the same average query count of 1.7. The nested RS and BT algorithms have the lowest query count of 1.2. The average recognition efficiency of the various algorithms is then tested in two distribution scenarios of tag identity codes, and the specific results are shown in Fig. 11.

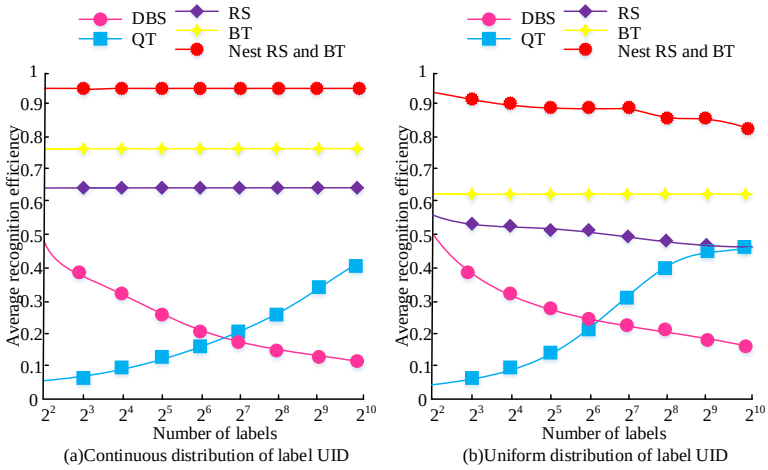


Fig. 11. Average recognition efficiency of different algorithms in two scenarios.

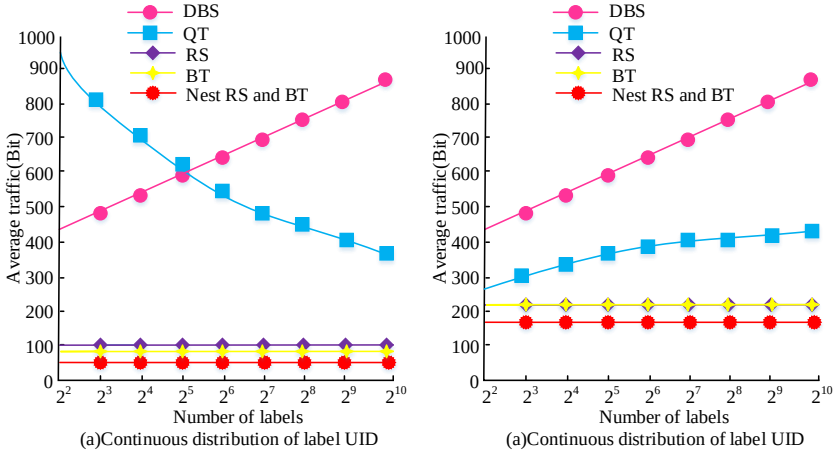


Fig. 12. Average traffic of different algorithms in two scenarios.

As shown in Fig. 11(a), the average recognition efficiency of the five algorithms in the continuous scenario is shown. As can be shown, the average recognition efficiency of the three algorithms—nested RS and BT algorithms, RS algorithm, and BT algorithm—remains stable as the number of tags increases, at 95%, 64%, and 78%, respectively. The average recognition efficiency of the DBS algorithm and QT algorithm is constantly changing, with the highest being around 50%. Fig. 11(a) shows the average recognition efficiency of the five algorithms in the uniform scenario. As can be observed, the average recognition efficiency of the BT algorithm stays constant at 62%, whereas that of the nested RS and BT + RS algorithms varies between 90% and 55%, respectively. The average recognition efficiency of the DBS and QT algorithms is similar to the case in (a), with a maximum of 50%. Finally, the average communication of the various algorithms is tested in two distribution scenarios of tag identity codes, and the specific results are shown in Fig. 12.

As shown in Fig. 12(a), the average communication volume of the five algorithms in the continuous scenario is shown. It can be concluded that the average communication volume of three algorithms, namely nested RS and BT algorithm, RS algorithm, and BT algorithm, remains stable, with 50, 100, and 80 Bit, respectively. While the average communication volume of the QT algorithm is inversely proportional to the number of tags and has a minimum communication volume of 400 Bit that of the DBS method is proportionate to the number of tags and has a minimum communication volume of 430 Bit. Fig. 12(b) shows the average communication volume of the five algorithms in the continuous scenario. The average communication volume of the five algorithms is shown in Fig. 12(b). The average communication volume of the nested RS and BT algorithms, RS algorithm and BT algorithm remains stable but increases slightly to 180, 220 and 220 Bit respectively. The average communication volume of the DBS algorithm is the same as in scenario (a). The average communication volume of the QT algorithm increases slowly to a minimum of 280 Bit and uniform distribution scenarios.

B. *Simulation Application Analysis of a File Management System Based on Nested RS and BT Anti-Collision Algorithms*

To further verify the system designed by the research in practical applications, a university is selected to conduct simulation experiments with the research system. To provide a point of comparison, a traditional campus network file management system and a card-based file management system are also set up. The number of clients is set to 100, and the CPU occupancy and memory occupancy of the three systems are first tested. The results are shown in Fig. 13.

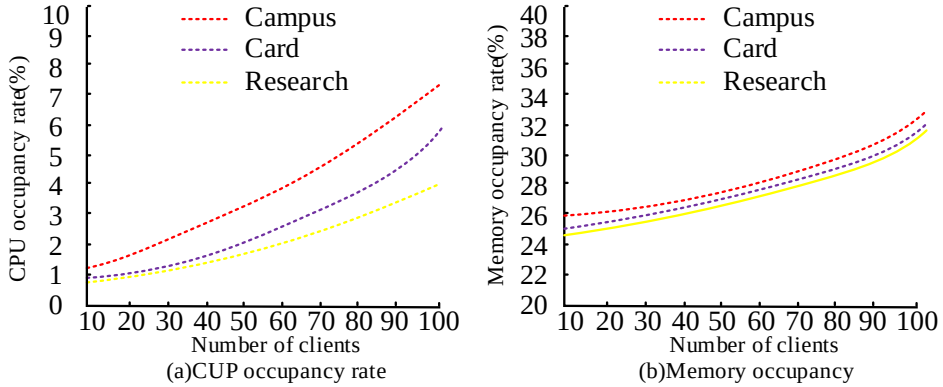


Fig. 13. CPU usage and memory usage of three systems.

As shown, Fig. 13(a) shows the change of CP U occupancy with client growth for the three systems. Fig. 13(b) shows the change of memory occupancy with client growth for the three systems. It can be concluded that the CPU occupancy and memory occupancy of the study system are smaller than those of the other two systems throughout the client growth. When the number of clients is 10, the CPU occupancy of the research system is 0.91% and the memory occupancy is 24.21%, while the CPU occupancy and memory occupancy of the other systems are as low as 0.97% and 25.13%. When the number of clients is 100, the CPU occupancy of the research system is 3.92% and the memory occupancy is 31.14%, while the CPU occupancy and memory occupancy of the other systems are as low as 5.98% and 32.13%. The average response time of the three systems is then evaluated once more, with the lowest values being 5.98% and 32.12%. Fig. 14 displays the average findings from 100 tests on the three systems' average reaction times with 100, 500, 1000, and 2000 users.

Fig. 14 shows that when the number of users increases, the average response time of the three algorithms increases as well. The average response time of the three systems is roughly the same at 0.112 seconds when there are 100 users being accessed? The average response time of the studied system is 1.244 seconds, which is 0.723 seconds faster than the card-based archive management system campus network archive management system and 1.036 seconds faster than the campus network archive management system. However, as the total amount of users increased, the difference in average response time between the three systems grow increasingly larger. When the number of access users reaches 2000, the studied system's average response time is 1.244 seconds. Finally, the accuracy of archive information extraction is tested for the three systems. A total of 1,000 users are permitted to access the system at any given time, with the system operating continuously for a period of either 12 or 24 hours per day. The results are shown in Fig. 15.

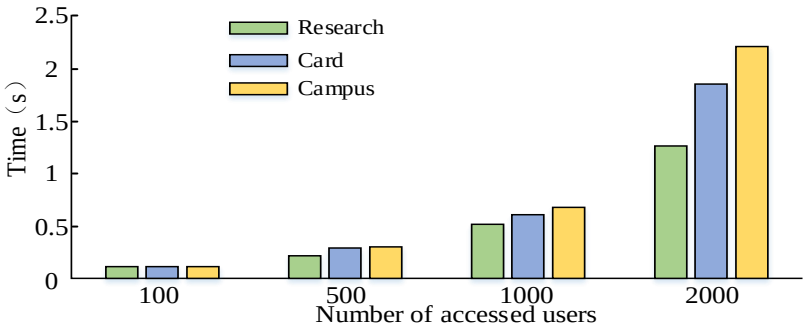


Fig. 14. Average response time of three systems.

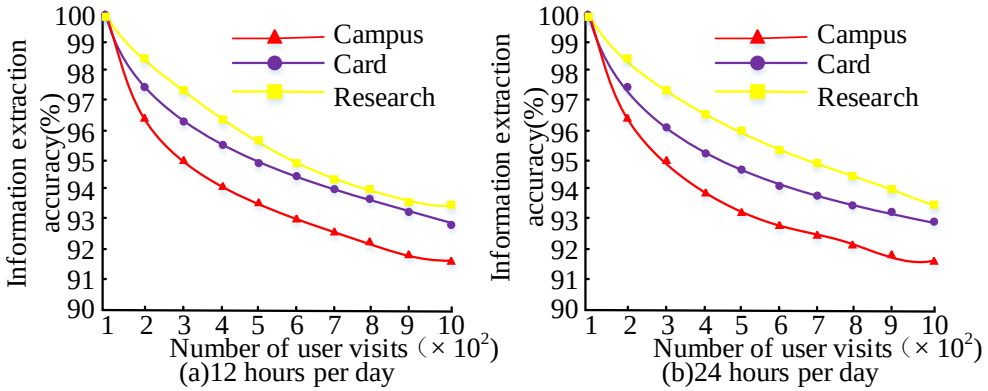


Fig. 15. Accuracy of extracting archive information for three systems.

Fig. 15(a) depicts the correlation between the number of users visiting the archives and the accuracy of information extraction from the archives for the three systems operating for 12 hours each day. It is clear that for all three systems, information extraction accuracy is high when the number of access users is low. The accuracy rate starts to decline as the user base grows, but the research system ages the least quickly. When the number of users is 1000, the accuracy of the research system algorithm is 94%, while the other two algorithms are accurate up to 93%. Fig. 15(b) shows the accuracy of archive information extraction versus the number of users accessed for the three systems running for 24h a day. The comparison shows that there is only a slight difference between the two scenarios. The duration of operation of all systems does not affect this result. In conclusion, it is clear that the research system offers quick performance, excellent information correctness, and minimal memory utilization for managing university records.

V. Discussion

Research on the effective integration of RFID technology, Internet big data and artificial intelligence technology is an important trend in the field of Internet of Things. Internet big data provides a rich database for RFID system, while AI technology further improves the intelligence level of RFID system through data analysis, prediction and optimization. This technological integration can not only improve the efficiency of university archives management, but also improve the system's adaptability and decision-making ability. In the process of RFID tag recognition, nested random slot ALOHA and binary tree anti-collision algorithms were introduced and optimized, significantly improving the efficiency and accuracy of tag recognition. The combination of these two algorithms not only leveraged the advantages of the random slot ALOHA algorithm in reducing collision probability, but also utilized the characteristics of the binary tree algorithm in quickly locating and solving collision problems. This dual optimization of recognition efficiency and communication volume is achieved through the effective integration of these two algorithms. From the test results, this study outperformed other algorithms in terms of average query coefficient, recognition efficiency, and communication volume, which fully demonstrated the effectiveness and superiority of the proposed algorithm. Especially in continuous and uniform scenarios, its performance advantages were more pronounced, which was crucial for stability and reliability in practical applications. This study is of great significance for fields such as university archive management that require high-precision and high-efficiency record management.

VI. Conclusion

An archive management system based on RFID technology was researched, designed and implemented, effectively integrating big data and artificial intelligence technologies to improve the intelligence level of archive management. The study introduced nested SR and binary tree anti-collision algorithm into the archive management system to solve the collision problem in the RFID tag recognition process. The test results showed that in both continuous and uniform scenarios, the proposed algorithm outperformed other algorithms in terms of average query coefficient, recognition efficiency, and communication volume, with values of 1.2

times and 1.2 times, 95% and 90%, and 50Bit and 180Bit, respectively. In practical application testing, when the number of clients was 10 and 100, the CPU and memory occupancy of the research system were 0.91% and 3.92%, 24.21% and 31.14%, respectively. At different numbers of clients, the CPU and memory usage of the system remained at a low level, indicating high resource utilization and the ability to meet the needs of large-scale file management. When accessing 100 and 1000 users, the average response time of the research system was 0.112 seconds and 1.244 seconds, respectively. The average response time of the system was short. Especially in high concurrency access scenarios, it still maintained good response speed and improved user experience. When accessing 1000 users, the information extraction accuracy of the research system was 94%. This indicated that the method could reduce recognition practice, improve recognition efficiency, and have high recognition accuracy. However, this method has not fundamentally solved the label collision problem, and the optimization level of information extraction accuracy is not obvious, which needs further improvement. Future research should concentrate on the fundamental resolution of the RFID tag collision issue. This should be achieved through the implementation of innovative algorithms and technological solutions, which will significantly enhance the accuracy of information extraction. Furthermore, continuous optimization of system performance is essential to guarantee the efficiency, accuracy, and stability of the archive management system.

REFERENCES

[1] Watanabe O, Delmo C, Ridho T. Analysis Analysis of Mobile-Based Archive Management Information System Design. Journal of Information Systems and Technology Research, 2023, 2(2): 91-101.

[2] Reynaldo G. Alvez. “The Development of A Cloud-Based University Research Repository Software Using A Configurable Subscription Model.” Acta Informatica Malaysia. 2022; 6(1): 07-12.

[3] Alfath A N, Taningngsih I, Suharto E. “Designed an Inactive Archives Management Information System Using Visual Basic 2010.” The Department of Journal of Applied Engineering and Technological Science (JAETS), 2022, 3(2): 105-115.

[4] Ai Y, Bai T, Xu Y, Zhang, W. “Anti-collision algorithm based on slotted random regressive-style binary search tree in RFID technology.” IET Communications, 2022, 16(10): 1200-1208.

[5] Zan J. “Research on robot path perception and optimization technology based on whale optimization algorithm.” Journal of Computational and Cognitive Engineering, 2022, 1(4): 201-208.

[6] Fan Zhang. “Research On Keyword Mining of Academic Library Subject Service. Acta Informatica Malaysia.” 2023; 7(2): 97-100.

[7] Uka K K, Oguoma S I, Chuma-Uba U P. “Analysis of Blockchain Architecture in File Sharing Management for Tertiary Institution.” Intelligent Information Management, 2020, 12(3):88-104.

[8] Gamido H V, Gamido M V, Sison A M. “Developing a secured image file management system using modified AES.” Bulletin of Electrical Engineering and Informatics, 2019, 8(4):1461-1467.

[9] Veena B, Dhiraj S. “A Personalized and Scalable Machine Learning-Based File Management System.” Tehnički glasnik, 2022, 16(2): 288-292.

[10] Zhou J, Chen Y, Wang W, He S, Meng, D. “A highly reliable metadata service for large-scale distributed file systems.” IEEE Transactions on Parallel and Distributed Systems, 2019, 31(2): 374-392.

[11] Zhu B, Chen Y, Wang Q, Lu, Y., & Shu, Jl. “Octopus+: An rdma-enabled distributed persistent memory file system.” ACM Transactions on Storage (TOS), 2021, 17(3): 1-25.

[12] Choi Y I, Cho J H, Kim Y T. “Collision Avoidance Algorithm of Mobile Robots at Grid Map Intersection Point.” International Journal of Fuzzy Logic and Intelligent Systems, 2020, 20(2):96-104.

[13] Zhou W, Jiang N. “Research on hybrid of ALOHA and multi-fork tree Anti-collision algorithm for RFID.” Procedia Computer Science, 2021, 183(5):389-394.

[14] Qiu Y, Lu J, Yang L, Xu, Y. “Adapted RFID anti-collision algorithm and its application in sharing economy.” International Journal of Internet and Enterprise Management, 2020, 9(3): 248-260. Enterprise Management, 2020, 9(3): 248-260.

[15] Qu J, Wang T. “An adaptive frame slotted ALOHA anti-collision algorithm based on tag grouping.” Cognitive Computation and Systems, 2021, 3(1): 17-27.

[16] Jing C, Luo Z, Chen Y, et al. “Blind anti-collision methods for RFID system: a comparative analysis.” Infocommunications Journal, 2020, 12(3):8-16.

[17] Fahmi H, Fadli S, Ashari M, Ramadhon, M. S. Development of Mail Archive Management Information System at Lombok Tengah District Education Office. JISA (Jurnal Informatika dan Sains), 2022, 5(2): 165-172.

[18] Watanabe O, Delmo C, Ridho T. Analysis Analysis of Mobile-Based Archive Management Information System Design. Journal of Information Systems and Technology Research, 2023, 2(2): 91-101.

[19] Choudhuri S, Adeniye S, Sen A. “Distribution Alignment Using Complement Entropy Objective and Adaptive Consensus-Based Label Refinement for Partial Domain Adaptation.” Artificial Intelligence and Applications, 2023, 1(1): 43-51.

[20] Bertram J, Wei P, Zambreno J. A fast Markov decision process-based algorithm for collision avoidance in urban air mobility[J]. IEEE transactions on intelligent transportation systems, 2022, 23(9): 15420-15433.

[21] Shen K, Shivgan R, Medina J, Dong Z, Rojas-Cessa R. Multidepot drone path planning with collision avoidance. IEEE Internet of Things Journal, 2022, 9(17): 16297-16307.

Advancements in Deep Learning Architectures for Image Recognition and Semantic Segmentation

Dr. Divya Nimma[1], Arjun Uddagiri[2]
PhD in Computational Science, University of Southern Mississippi, USA[1]
Gloom Dev Pvt Ltd, Penamaluru, Vijayawada 521139 Andhra Pradesh, India[2]

***Abstract*—This paper focuses on using Convolutional Neural Networks (CNNs) for tasks such as image classification. It covers both pre-trained models and those that are built from scratch. The paper begins by demonstrating how to utilize the well-known AlexNet model, which is highly effective for image recognition due to transfer learning. It then explains how to load and prepare the MNIST dataset, a common choice for testing image classification methods. Additionally, it introduces a custom CNN designed specifically for recognizing MNIST digits, outlining its architecture, which includes convolutional layers, activation functions, and fully connected layers for capturing handwritten numbers' details. The paper also guides starting the model, running it on sample data, reviewing outputs, and assessing the accuracy of predictions. Furthermore, it delves into training the custom CNN and evaluating its performance by comparing it with established benchmarks, utilizing loss functions and optimization techniques to fine-tune the model and assess its classification accuracy. This work integrates theory with practical application, serving as a comprehensive guide for creating and evaluating CNNs in image classification, with implications for both research and real-world applications in computer vision.**

Keywords*—*Convolutional Neural Networks (CNNs); AlexNet; image classification; transfer learning; MNIST Dataset; Custom CNN Architecture; deep learning; model training and evaluation; neural network optimization; activation functions; feature extraction; machine learning; pattern recognition; data preprocessing; loss functions; model accuracy

I. Introduction

Convolutional Neural Networks (CNNs) have revolutionized the field of deep learning, proving particularly effective in tasks such as image classification. Their ability to automatically learn hierarchical feature representations from raw input data makes them highly suitable for processing images and videos across various applications. This paper focuses on leveraging CNNs for image classification tasks, examining both pre-trained models and those constructed from scratch.

A landmark advancement in CNN architecture is AlexNet, which gained prominence during the 2012 ImageNet Large Scale Visual Recognition Challenge (ILSVRC). Developed by Alex Krizhevsky, Ilya Sutskever, and Geoffrey Hinton, AlexNet utilizes a deep architecture consisting of multiple convolutional layers followed by fully connected layers. As depicted in Fig. 1, the model's layered structure enhances its capacity to learn complex patterns in images while employing ReLU activation and dropout techniques to prevent overfitting. This efficiency in feature extraction and classification establishes AlexNet as a powerful tool for image recognition tasks, especially through the application of transfer learning.

Fig. 1 illustrates the architecture of AlexNet, highlighting the convolutional and fully connected layers that work in tandem to enhance learning and mitigate overfitting. Following the introduction of AlexNet, numerous custom CNN architectures have been developed to address specific challenges in image classification. One such architecture, designed for the MNIST dataset, focuses on recognizing handwritten digits. This dataset is a standard benchmark in image classification and contains a diverse set of examples for evaluating model performance. The architecture of the custom CNN includes essential components, such as convolutional layers for feature extraction, activation functions to introduce non-linearity, and fully connected layers to classify the extracted features. Fig. 2 illustrates sample images from the MNIST dataset, demonstrating the variety of handwritten digits that the model aims to classify.

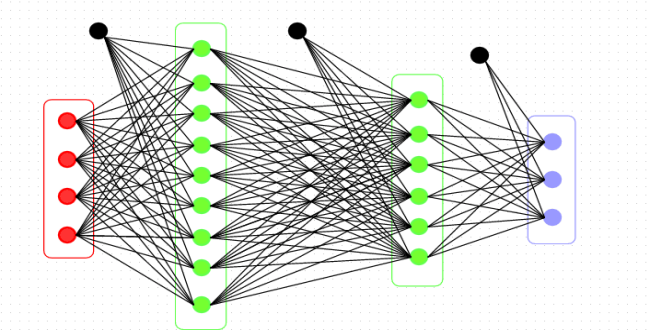


Fig. 1. A pioneering architecture in convolutional networks for AlexNet.

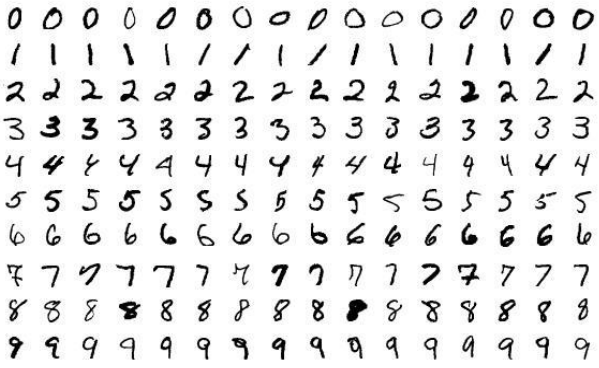


Fig. 2. MNIST dataset samples.

Fig. 2 showcases samples from the MNIST dataset, illustrating the diversity of handwritten digits that the custom CNN model is designed to recognize. To facilitate the practical implementation of these models, this paper will outline the steps required to load and prepare the MNIST dataset for training. It will provide a detailed explanation of the custom CNN architecture, covering its layers and functionality, as well as guidance on initiating the model, running it on sample data, reviewing outputs, and assessing the accuracy of predictions.

Additionally, the paper will explore training techniques for the custom CNN, emphasizing the importance of loss functions and optimization methods in fine-tuning model performance. By comparing the custom CNN's accuracy with established benchmarks, this study seeks to provide valuable insights into the practical applications of CNNs in image classification tasks.

TABLE I. SUMMARY OF KEY CNN ARCHITECTURES

Architecture	Year	Main Features	Applications
AlexNet	2012	8 layers, ReLU, Dropout	Image Classification
VGGNet	2014	Deep layers, small filters	Image Classification
ResNet	2015	Residual connections	Various
Inception	2015	Inception modules	Various

Table I summarizes key CNN architectures, highlighting their respective features and applications, emphasizing their role in advancing image classification techniques. By integrating theoretical foundations with practical applications, this work serves as a comprehensive guide for creating and evaluating CNNs in image classification, contributing valuable knowledge for both research and real-world applications in computer vision.

A. Problem Statement

The advent of Convolutional Neural Networks (CNNs) has significantly advanced the field of image classification and computer vision. However, despite their effectiveness, several challenges persist in achieving optimal performance across various applications. This paper addresses the following key problems within the context of CNN architectures:

1) Scalability and generalization: Deep learning models like AlexNet have shown remarkable performance on benchmark datasets such as ImageNet. However, transferring these models to different or more complex datasets often requires careful tuning and adaptation. The challenge lies in designing CNN architectures that not only excel in specific domains but also generalize well to a wide range of applications and data variations.

2) Model complexity and efficiency: Deep CNNs often involve numerous layers and parameters, leading to high computational and memory requirements. For instance, the AlexNet model, despite its success, is known for its substantial resource demands. The challenge is to develop CNN models that balance complexity and efficiency, optimizing both performance and resource utilization.

3) Feature extraction and classification: The ability of CNNs to extract meaningful features from raw input data and accurately classify them remains a critical challenge. This includes ensuring that the convolutional layers effectively capture relevant patterns and that the subsequent fully connected layers provide accurate classification results. The problem is exacerbated in cases where input data is noisy or contains complex variations.

4) Training and optimization: Training CNN models involves optimizing a large number of parameters, which can be computationally intensive and prone to issues such as overfitting or underfitting. Efficient training strategies, including proper choice of loss functions, optimizers, and regularization techniques, are crucial to achieving high-performance models.

5) Benchmarking and performance evaluation: Comparing the performance of different CNN architectures on standard benchmarks, such as the MNIST dataset, requires robust evaluation metrics and methodologies. The problem is to ensure that performance assessments are accurate and reflective of the models' real-world applicability.

In this paper, we aim to address these challenges by exploring and comparing various CNN architectures, including AlexNet and a custom CNN model for MNIST classification. We seek to provide insights into their strengths and limitations, propose strategies for enhancing their scalability and efficiency, and offer recommendations for overcoming common obstacles in CNN-based image classification tasks.

B. Research Questions

1) How do different Convolutional Neural Network (CNN) architectures, such as AlexNet and custom-designed models, perform in terms of accuracy and efficiency when applied to various image classification tasks?

2) What are the key factors that influence the scalability and generalization of CNN models across different domains and datasets?

3) How can CNN models be optimized to balance computational complexity and performance, especially for resource-constrained environments?

4) What strategies and techniques are most effective for feature extraction and classification in CNN models, particularly in dealing with noisy or complex data?

5) What are the best practices for training CNN models, including the choice of loss functions, optimizers, and regularization techniques, to improve model performance and prevent common issues such as overfitting?

6) How can performance evaluation methodologies be improved to provide a more accurate assessment of CNN models' real-world applicability?

C. Objective of Study

1) Evaluate CNN architectures: Evaluate the performance of various Convolutional Neural Network (CNN) architectures, including established models like AlexNet and custom-designed networks. Assess their accuracy, efficiency, and adaptability across different image classification tasks and datasets.

2) Optimize CNN models: Identify and implement strategies to optimize CNN models, aiming to achieve a balance between computational complexity and performance. This is particularly important in scenarios with limited resources or specific application constraints.

3) Enhance feature extraction and classification: Explore and develop effective methods for feature extraction and classification within CNN models, addressing challenges related to noisy or complex data environments.

4) Improve training practices: Analyze and refine best practices for training CNN models, with a focus on the selection of loss functions, optimizers, and regularization techniques to enhance model performance and reduce issues such as overfitting.

5) Advance performance evaluation: Propose and apply improved methodologies for evaluating CNN models' performance, ensuring that assessments reflect real-world applicability and effectiveness in practical scenarios.

II. Existing System

Image classification techniques can be broadly categorized into traditional techniques and deep learning-based approaches.

A. Traditional Techniques

Traditional techniques for image classification involve handcrafted feature extraction followed by classification using machine learning algorithms. These methods typically include:

1) Interpolation methods:

a) Feature engineering: Handcrafted features such as edges, textures, and shapes are extracted using techniques like edge detection (e.g., Canny, Sobel), texture analysis, and shape descriptors. These features are manually designed to represent various aspects of the image.

b) Classical machine learning algorithms: Once features are extracted, classifiers such as Support Vector Machines (SVM), k-nearest Neighbors (k-NN), and Decision Trees are used to categorize images based on the extracted features.

c) Limitations: Traditional techniques often require domain-specific expertise to design effective features and may struggle with complex image datasets due to limited ability to capture intricate patterns and variations.

B. Deep Learning-Based Approaches

Deep learning-based approaches, particularly Convolutional Neural Networks (CNNs), have revolutionized image classification by automating feature extraction and improving classification performance. Key aspects include:

1) Convolutional Neural Networks (CNNs): CNNs, such as AlexNet, LeNet, and VGG, use multiple layers of convolutional and pooling operations to automatically learn hierarchical features from raw image data. These models excel at capturing spatial hierarchies and patterns in images.

2) Transfer learning: Techniques like transfer learning leverage pre-trained CNN models on large datasets (e.g., ImageNet) to fine-tune and adapt these models to specific tasks with smaller datasets. This approach accelerates training and improves performance on specialized tasks.

3) End-to-end learning: CNNs enable end-to-end learning, where the model learns to perform both feature extraction and classification in a single integrated framework, reducing the need for manual feature engineering.

C. Comparative Analysis

Comparative analysis between traditional techniques and deep learning-based approaches highlights several differences:

Feature Extraction is a Traditional method that relies on handcrafted features, which may not capture all relevant information. Deep learning approaches use automatic feature extraction through multiple layers, capturing complex patterns and representations.

Performance is Deep learning models generally outperform traditional methods in terms of accuracy and robustness, especially on large and diverse datasets. Traditional methods may struggle with high-dimensional data and require extensive tuning.

Scalability is Deep learning models that scale more effectively with increasing data and computational resources. Traditional methods may become less effective as dataset size grows, requiring more manual intervention.

Training and Complexity are Deep learning models that often require significant computational resources and extensive training data. Traditional methods are less computationally intensive but may not achieve the same level of accuracy or generalization.

III. Proposed System

To address the limitations of traditional and existing deep learning-based super-resolution techniques, we propose a novel approach using convolutional autoencoders designed specifically for the task of image super-resolution. Our proposed system leverages the power of deep learning to learn efficient representations and mappings from low-resolution images to their high-resolution counterparts, aiming to produce superior-quality images with enhanced details and reduced artifacts.

A. System Overview

The proposed system aims to enhance image classification tasks by integrating advanced deep learning methodologies, specifically leveraging state-of-the-art Convolutional Neural Networks (CNNs) and Transfer Learning techniques. The system is designed to address the limitations of traditional image classification methods and provide a robust framework for handling diverse and complex datasets. The key components of the proposed system include.

1) Deep learning architecture: Utilization of cutting-edge CNN architectures, such as ResNet, DenseNet, or EfficientNet, to leverage their advanced feature extraction capabilities and improve classification accuracy.

2) Transfer learning: Implementation of transfer learning to fine-tune pre-trained models on specific datasets, optimizing model performance even with limited labeled data.

3) Automated feature extraction: Automation of feature extraction through deep learning, eliminating the need for manual feature engineering and enabling the model to learn complex patterns and representations.

4) Integration and deployment: Development of a user-friendly interface for seamless integration and deployment, allowing for easy adaptation to various image classification tasks in real-world applications.

B. Detailed Design

1) Model selection and training:

a) Architecture choice: Selection of a suitable pre-trained CNN model based on the specific requirements of the image classification task. Evaluation of various architectures to determine the most effective one for the dataset at hand.

b) Fine-Tuning: Adaptation of the pre-trained model to the target domain through transfer learning. Fine-tuning involves adjusting the model's weights based on the new dataset to improve its accuracy and generalization.

Data Preparation is Gathering a diverse and representative dataset for training and evaluation. Ensuring the dataset covers a wide range of scenarios to enhance model robustness.

Application of data augmentation techniques, such as rotation, scaling, and flipping, to increase dataset variability and improve model generalization.

c) Training process: Training Pipeline is Establishing a systematic training pipeline that includes data preprocessing, model training, and evaluation. Utilizing modern deep learning frameworks (e.g., TensorFlow, PyTorch) to streamline the training process.

d) Hyperparameter tuning: Optimization of hyperparameters (e.g., learning rate, batch size) to achieve optimal model performance.

Performance Metrics is the Use of comprehensive evaluation metrics, such as accuracy, precision, recall, and F1-score, to assess model performance. Comparison with baseline methods to demonstrate improvements.

Cross-validation is the Implementation of cross-validation techniques to ensure the model's robustness and generalizability across different subsets of the data.

C. Expected Benefits

1) Improved Accuracy is Enhanced classification accuracy due to the advanced capabilities of deep learning models in capturing complex image features.

2) Reduced Manual Effort is the Automation of feature extraction and reduction of manual intervention required for feature engineering.

3) Scalability is the Scalability of the system to handle large and diverse datasets, making it suitable for various applications.

4) Flexibility is Adaptability to different image classification tasks through transfer learning and fine-tuning, allowing for easy customization based on specific needs.

D. Potential Applications

The proposed system can be applied to a wide range of image classification tasks, including but not limited to

1) Medical Imaging is the Classification of medical images for diagnostic purposes.

2) Retail is Image-based product recognition and inventory management.

3) Autonomous Vehicles are Object detection and classification in self-driving cars.

4) Security is Surveillance and anomaly detection in security systems.

IV. Literature Survey

This seminal work introduced Convolutional Neural Networks (CNNs) for document recognition tasks. The architecture combined convolutional layers with subsampling, showcasing its effectiveness in handwritten digit recognition. It demonstrated high accuracy and laid the foundation for CNNs, emphasizing their potential in visual pattern recognition. This work was instrumental in advancing the field of image classification and set the stage for further developments in CNN applications [1].The content describes the use of Convolutional Neural Networks (CNNs) for document recognition tasks. It highlights how CNNs, through the use of convolutional layers and subsampling, achieved high accuracy in recognizing handwritten digits. This approach demonstrated the potential of CNNs in visual pattern recognition, paving the way for advancements in image classification [2].The work explored the use of very deep Convolutional Neural Networks (CNNs) with small 3x3 filters and increased depth to improve the learning of complex visual features. The model achieved state-of-the-art performance on the ImageNet dataset and emphasized the significance of network depth in enhancing the capacity of CNN architectures for large-scale image recognition tasks [3]. The Inception architecture utilized asymmetric convolutions and multiple convolutional paths with different filter sizes to improve computational efficiency and accuracy in large-scale image recognition tasks [4].DenseNet introduced an architecture where each layer receives inputs from all preceding layers. This design promotes feature reuse and improves gradient flow, enhancing performance in object recognition tasks while reducing the number of parameters compared to traditional CNNs [5].Batch Normalization is a technique that normalizes the inputs of each layer within mini-batches. This approach stabilizes and accelerates the training of deep networks by reducing internal covariate shift, allowing for higher learning rates and improved convergence, thereby enhancing model performance in image recognition tasks [6]. The Adam optimizer improves training efficiency in deep learning models by combining elements of AdaGrad and RMSProp. It adapts the learning rate for each parameter and maintains an exponentially decaying average of past gradients, leading to faster convergence and enhanced model performance e[7]. ResNet introduced residual learning with shortcut connections that bypass one or more layers, addressing the vanishing gradient problem in very deep networks. This approach facilitated the training of extremely deep networks and led to significant improvements in image classification, object detection, and semantic segmentation [8].The YOLO framework redefined

object detection by treating it as a single regression problem, predicting bounding boxes and class probabilities directly from full images. This approach significantly improved speed and efficiency, making YOLO well-suited for real-time applications such as autonomous driving and surveillance [9]. The work revisited the Inception architecture, introducing optimizations that enhanced performance and efficiency. These improvements led to the development of Inception-v3, which achieved new records in image classification tasks by emphasizing careful design choices in deep networks [10]. Faster R-CNN introduced a region proposal network (RPN) to streamline the object detection process. By generating high-quality region proposals directly from feature maps, this method significantly improved both detection speed and accuracy, establishing itself as a foundational model in object detection [11]. The research presented techniques for visualizing the inner workings of convolutional networks. By analyzing activations of different layers, the study provided insights into how CNNs learn features and clarified the decision-making processes behind these models [12]. The Feature Pyramid Network (FPN) introduced a top-down architecture to create high-level semantic feature maps at different scales. This approach enhanced object detection performance by utilizing multi-scale feature maps, improving accuracy across various object sizes [13]. Fast R-CNN enhanced traditional R-CNN by integrating the region proposal network into the CNN training process. This modification enabled end-to-end training of the model, which significantly reduced computational overhead and improved both detection speed and accuracy [14]. The Pyramid Scene Parsing Network (PSPNet) introduced a pyramid pooling module to capture global context information and enhance segmentation performance. By utilizing multi-scale information, PSPNet achieved state-of-the-art results in scene parsing benchmarks [15]. DeepLab introduced a semantic segmentation architecture that used atrous convolutions to capture multi-scale contextual information. This approach significantly improved segmentation accuracy in complex scenes and demonstrated the effectiveness of fully connected conditional random fields for refining the segmentation output [16]. IntelPVT enhances object detection and classification through intelligent patch-based strategies, improving the understanding of its capabilities in these tasks.The research explores the use of deep learning techniques in image forensics, specifically for detecting and reconstructing manipulated images, contributing to advancements in digital forensics [17]. The work discusses how image processing techniques enhance user experiences in augmented reality (AR) and virtual reality (VR) environments, emphasizing advancements in computer vision that enable immersive applications [18]. The work provides a comprehensive overview of deep learning techniques for image recognition and classification, highlighting their effectiveness and covering various architectures and methodologies developed in recent years [19]. The IntelPVT model uses patch-based strategies within pyramid vision transformers to improve object detection and classification performance [20]. The work discusses advancements in vision transformers, highlighting how IntelPVT and Opt-STViT improve performance in object detection, classification, and video recognition tasks [21]. The research explores weakly-supervised learning for object localization, showing that convolutional neural networks can achieve competitive localization performance without needing extensive labeled datasets. This advancement highlights the potential of weakly-supervised learning in object localization tasks [22]. R-FCN proposed a region-based approach utilizing fully convolutional networks for object detection, enabling efficient and accurate detection across various categories while maintaining high speeds in real-time applications [23]. The study demonstrated that CNNs can effectively highlight discriminative regions within images, enhancing the understanding of spatial relationships between objects by using deep features for localization [24]. This research introduced fully convolutional networks (FCNs), which revolutionized semantic segmentation by enabling pixel-wise classification. The study demonstrated that CNN architectures could be adapted for dense prediction tasks, achieving superior results in various segmentation applications [25]. The use of fully convolutional networks (FCNs) enabled pixel-wise classification for semantic segmentation, adapting CNN architectures for dense prediction tasks and achieving state-of-the-art results [26]. The research introduced Inception-v4 and Inception-ResNet, highlighting how integrating residual connections in deep networks leads to significant improvements in image classification tasks, emphasizing the role of connectivity in enhancing learning [27]. The work proposed a difficulty-aware semantic segmentation approach using a deep layer cascade, which prioritized easier pixels for initial predictions and refined harder pixels in subsequent layers. This strategy improved segmentation accuracy across various datasets [28]. ENet introduced an efficient architecture designed for real-time semantic segmentation, optimizing the balance between speed and accuracy for deployment in resource-constrained environments [29]. The work presented DenseNet architectures adapted for semantic segmentation, highlighting the benefits of densely connected layers in improving feature propagation and network performance while maintaining efficiency [30].

V. Methodology

The methodology section outlines the systematic approach and techniques employed to develop and evaluate the proposed system. This section details the design, implementation, and assessment phases of the research, providing a clear understanding of how the objectives are achieved and how the model's performance is assessed.

A. Model Architecture

1) Overview: The proposed model architecture is designed to leverage the capabilities of Convolutional Neural Networks (CNNs) to achieve high accuracy in image classification tasks. This architecture integrates several advanced deep learning techniques, enabling it to effectively learn and generalize from complex data patterns. As depicted in Fig. 3, the architecture comprises multiple layers, including convolutional layers, pooling layers, and fully connected layers, each contributing uniquely to the model's performance.

Fig. 3 illustrates the architecture of the Convolutional Neural Network, detailing the arrangement and interaction of layers that facilitate feature extraction and classification.

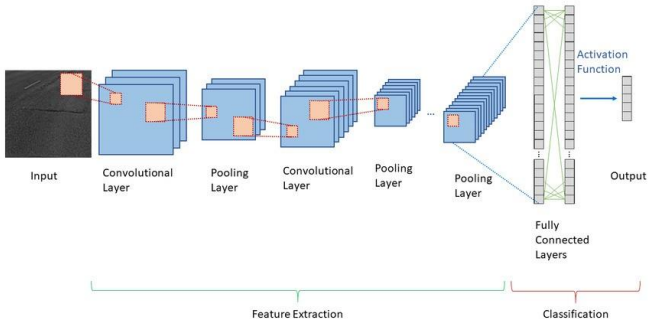


Fig. 3. Convolutional Neural Network architecture.

2) Convolutional Neural Network (CNN):

a) Input layer: The input layer serves as the initial point of entry for raw image data into the CNN. To ensure uniformity across the dataset, all images are resized to a consistent dimension of 224x224 pixels. This standardization is crucial, as it allows the CNN to process images in a predictable format, improving the efficiency of the subsequent layers. Resizing also aids in reducing computational overhead while maintaining sufficient detail for effective feature extraction.

b) Convolutional layers: Central to the CNN architecture are the convolutional layers, which perform the critical function of feature extraction. These layers employ a set of filters (also known as kernels) to scan the input image and detect various features, such as edges, textures, and patterns. The architecture utilizes multiple convolutional layers, each configured with different numbers of filters to capture a hierarchy of features.

For example, as detailed in Table II, the first convolutional layer applies 32 filters and produces an output shape of (224, 224, 32) with 896 parameters. This initial layer focuses on capturing basic features such as edges and textures. As the data progresses through deeper layers, the number of filters increases to 64 in the second convolutional layer, which outputs a feature map of shape (112, 112, 64) with 18,496 parameters. These deeper layers enable the model to identify more abstract features, such as shapes and objects, contributing to its ability to make accurate classifications.

TABLE II. CNN LAYER DETAILS

Layer Type	Output Shape	Parameters
Input Layer	(224, 224, 3)	0
Conv2D (32 filters)	(224, 224, 32)	896
MaxPooling2D	(112, 112, 32)	0
Conv2D (64 filters)	(112, 112, 64)	18496
MaxPooling2D	(56, 56, 64)	0
Flatten	(200704)	0
Dense (128 units)	(128)	25689600
Dense (10 units)	[10]	1290

c) Pooling layers: Pooling layers play a vital role in reducing the dimensionality of the feature maps generated by the convolutional layers. This reduction is achieved through operations such as MaxPooling, which selects the maximum value from a defined sub-region of the feature map. For instance, the first MaxPooling layer operates on the output of the first convolutional layer, reducing its size from (224, 224, 32) to (112, 112, 32). This process helps retain the most prominent features while significantly decreasing the computational load on the network.

By reducing the feature map size, pooling layers also help make the CNN invariant to small translations in the input images, thereby enhancing its robustness. The use of a 2x2 MaxPooling filter is a common practice, as it effectively halves the dimensions of the feature maps while preserving critical spatial information, allowing the model to maintain high accuracy despite variations in input data.

d) Fully connected layers: After feature extraction and dimensionality reduction, the architecture transitions to fully connected layers, which integrate the high-level features extracted by the convolutional and pooling layers. These layers are analogous to traditional neural network layers, where each neuron is connected to every neuron in the previous layer. The first fully connected layer consists of 128 units, while the final output layer comprises 10 units, corresponding to the ten categories of the classification task.

The output from the fully connected layers is critical, as it translates the abstract features learned by the network into classification scores for each category. This transformation is key to the model's ability to accurately classify images based on the learned representations.

e) Output layer: The final output layer employs a softmax activation function to convert the raw classification scores into probabilities. This transformation ensures that the predicted probabilities for each class sum to one, providing a clear interpretation of the model's predictions. For a classification task involving 10 classes, the output layer generates a vector of 10 probabilities, where each value indicates the likelihood of the input image belonging to a particular category. This probabilistic output is essential for making informed decisions based on the model's predictions.

3) Enhanced components: To further improve the model's training efficiency and performance, several enhanced components are incorporated into the architecture:

- *Residual Connections:* Inspired by the ResNet architecture, residual connections address issues related to vanishing gradients that often occur in deep networks. By creating shortcuts between layers, these connections facilitate the direct flow of gradients during backpropagation, enabling the effective training of deeper networks. This design helps the model learn identity mappings, making it easier to optimize and improving overall performance.
- *Batch Normalization:* This technique is applied after convolutional layers to stabilize and accelerate the training process. Batch normalization normalizes the inputs of each layer to have a zero mean and unit variance, effectively reducing internal covariate shifts. This normalization leads to faster convergence and better generalization performance across unseen data. During

training, the activations of a convolutional layer are normalized across the batch and then scaled and shifted by learnable parameters, maintaining a stable distribution of activations throughout the network.

a) Dropout: As a regularization technique, dropout is utilized to prevent overfitting and enhance the model's generalization capabilities. During the training phase, dropout randomly drops a fraction of the units (neurons) in the network, along with their connections, during each forward pass. For instance, with a dropout rate of 0.5, each neuron has a 50% chance of being omitted from the current training iteration. This randomness encourages the network to learn redundant representations, thus making it more robust and less reliant on specific neurons, ultimately leading to improved performance on test data.

b) Hyperparameter tuning: Hyperparameter tuning is an essential part of optimizing the model's performance. Table III summarizes the hyperparameters tested during the training process, along with the best values identified:

TABLE III. HYPERPARAMETER TUNING

Hyperparameter	Values Tested	Best Value
Learning Rate	0.001, 0.01, 0.1	0.001
Batch Size	32, 64, 128	64
Epochs	10, 20, 30	20
Number of Layers	5, 10, 15	10
Filter Sizes	3x3, 5x5, 7x7	3x3

The learning rate is critical for controlling how much to change the model in response to the estimated error each time the model weights are updated. The batch size affects the stability of the gradient estimates during training, while the number of epochs determines how many times the learning algorithm will work through the entire training dataset. Optimizing these hyperparameters ensures the model achieves the best possible performance in image classification tasks.

B. Training Process

1) Data preparation: The primary step in any machine learning task is to collect a large and diverse dataset of labeled images. The size and diversity of the dataset are crucial for training a robust model that generalizes well.

Depending on the classification task, datasets may be collected from various sources, including public datasets (e.g., ImageNet, CIFAR-10), proprietary datasets, or through web scraping. The dataset should be representative of the problem domain and include a sufficient number of examples for each class to avoid bias and ensure effective learning.

a) Preprocessing is normalization: Images are often resized to a standard dimension (e.g., 224x224 pixels) to ensure consistency. Normalization involves scaling pixel values to a range (e.g., 0 to 1 or -1 to 1) to make the training process more stable and efficient.

Data Augmentation is Techniques such as rotation, flipping, cropping, and color adjustments are applied to artificially expand the training dataset. This increases variability and robustness by simulating different conditions under which the model might be tested, helping to prevent overfitting. For instance, flipping an image horizontally helps the model learn to recognize objects from different angles.

b) Model training has loss function: The loss function quantifies the difference between the model's predictions and the actual labels. It provides a measure of how well the model is performing.

Cross-entropy loss is commonly used for classification tasks. It calculates the difference between the predicted probability distribution and the true distribution (one-hot encoded labels). The goal is to minimize this loss during training, which reflects improving accuracy. Mathematically, for each sample, the loss is calculated as the negative logarithm of the predicted probability for the true class.

TABLE IV. EVALUATION METRICS

Metric	Description	Formula
Accuracy	Correct classification rate.	$\text{Accuracy} = \frac{TP+TN}{\text{Total Instances}}$
Precision	True positives out of predicted positives.	$\text{Precision} = \frac{TP}{TP+FP}$
Recall	True positives out of actual positives.	$\text{Recall} = \frac{TP}{TP+FN}$
F1-Score	Harmonic mean of precision and recall.	$\text{F1-Score} = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision}+\text{Recall}}$
ROC Curve & AUC	Visualizes performance and discrimination.	ROC Curve Plot; AUC Value (0 to 1)
Loss Curves	Tracks training and validation loss.	Training Loss Curve; Validation Loss Curve

2) Optimization algorithm: Optimization algorithms adjust the model's weights to minimize the loss function.

Algorithms like Adam (Adaptive Moment Estimation) and SGD (Stochastic Gradient Descent) are used to update the weights. Adam combines the advantages of two other extensions of SGD: Adaptive Gradient Algorithm (AdaGrad) and Root Mean Square Propagation (RMSProp). It uses estimates of the first and second moments of gradients to adaptively adjust the learning rate. SGD, on the other hand, updates weights using a small, random subset of the dataset (mini-batch) at each iteration, which helps in reducing computation and often leads to better generalization.

a) Learning rate scheduling: To improve convergence and training efficiency, the learning rate may be adjusted over time.

Techniques such as learning rate decay or scheduling adjust the learning rate based on the epoch number or validation performance. Common methods include reducing the learning rate by a factor (e.g., 0.1) after a certain number of epochs or when the validation performance plateaus. This helps in fine-tuning the model as it approaches convergence, leading to more stable and accurate results.

3) Training Phases:

a) Epochs: An epoch is one complete pass through the entire training dataset.

During training, the model's weights are updated multiple times through multiple epochs. The number of epochs is a hyperparameter that determines how long the model is trained. Typically, training proceeds through several epochs to allow the model to learn effectively from the data. Monitoring loss and accuracy metrics helps in determining the optimal number of epochs.

b) Validation: To evaluate the model's performance on unseen data and prevent overfitting.

A separate validation set, which is distinct from the training data, is used to assess the model's performance periodically during training. This helps in tuning hyperparameters and adjusting the training process. Validation metrics (e.g., accuracy, loss) provide insights into how well the model generalizes to new data. If the validation performance does not improve, it may indicate the need for adjustments in the training strategy or hyperparameters.

c) Hyperparameter tuning: To find the best set of hyperparameters that optimize model performance.

Details: Hyperparameters are parameters set before the training process begins, such as learning rate, batch size, number of layers, and filter sizes. Grid Search systematically explores a predefined set of hyperparameter values, testing each combination to find the best one. Random Search, on the other hand, samples a random subset of hyperparameter values and evaluates them, which can be more efficient for large hyperparameter spaces. Both methods aim to improve model performance by finding the optimal settings.

C. Evaluation Metrics

Table IV summarizes the *Evaluation Metrics:*

1) Accuracy: Accuracy is a fundamental metric that measures the proportion of correctly classified instances out of the total number of instances in the dataset.

Accuracy=Total Number of Predictions/Number of Correct Predictions

A high accuracy indicates that the model is making correct predictions for most of the instances. However, accuracy alone can be misleading, especially in imbalanced datasets where one class might dominate. For example, if 95% of the data belongs to one class and the model predicts this class for all instances, the accuracy would be 95%, but the model would be failing to detect the minority class.

2) Precision, Recall, and F1-Score: Precision: Precision measures the accuracy of positive predictions. It is the ratio of true positive predictions to the total number of predicted positives (true positives + false positives).

Formula: Precision=True Positives

/(False Positives + True Positives)

High precision indicates that the model has fewer false positives and is effective at identifying relevant instances among the predicted positives. This is crucial in applications where false positives are costly or undesirable, such as in medical diagnosis.

Recall: Recall, or sensitivity, measures the model's ability to identify all relevant instances within the data. It is the ratio of true positive predictions to the total number of actual positives (true positives + false negatives).

Formula: Precision=True Positives

/(False Negatives + True Positives)

High recall indicates that the model successfully identifies most of the positive instances. This metric is particularly important in situations where missing a positive instance has significant consequences, such as detecting fraud or disease.

F1-Score: The F1-Score is the harmonic mean of precision and recall, providing a single metric that balances both aspects. It is useful when you need to account for both precision and recall.

Formula: F1-Score=2·(Precision· Recall / Precision + Recall)

3) Confusion matrix: A confusion matrix provides a detailed view of a classification model's performance by showing the number of true positives, true negatives, false positives, and false negatives.

True Positives (TP): Instances where the model correctly predicts the positive class. True Negatives (TN): Instances where the model correctly predicts the negative class. False Positives (FP): Instances where the model incorrectly predicts the positive class. False Negatives (FN): Instances where the model incorrectly predicts the negative class.

The confusion matrix helps in understanding the types of errors the model makes. It is particularly useful for calculating other performance metrics (precision, recall, F1-Score) and for diagnosing issues such as class imbalance.

4) ROC Curve and AUC: ROC Curve: The Receiver Operating Characteristic (ROC) curve plots the true positive rate (sensitivity) against the false positive rate (1 - specificity) across different threshold values. By varying the threshold for classifying an instance as positive, the ROC curve illustrates the trade-off between sensitivity and specificity. The curve helps in visualizing the model's performance across various classification thresholds.

AUC (Area Under Curve): AUC measures the overall ability of the model to discriminate between positive and negative classes. The AUC value ranges from 0 to 1, with a higher AUC indicating better model performance. An AUC of 0.5 suggests that the model has no discriminative power, akin to random guessing. AUC is useful for comparing models and understanding their performance irrespective of the threshold.

5) Loss curves: Loss curves track the loss values (e.g., cross-entropy loss) during training and validation phases over epochs.

Training Loss Curve: This shows how the loss decreases over training epochs, indicating how well the model fits the training data. Validation Loss Curve: This shows how the loss

changes on the validation set over epochs. Monitoring this helps in detecting overfitting if the validation loss starts to increase while the training loss continues to decrease. Loss curves help in diagnosing problems such as overfitting, underfitting, or issues with the learning rate. They provide insights into the convergence behavior of the model and whether additional epochs or adjustments are needed.

6) Computational efficiency: Inference Time: Inference time measures the time required by the model to classify a single image. Details: It is crucial for real-time applications where quick decision-making is needed, such as in autonomous vehicles or live video analysis. Lower inference times are desirable for faster responses and improved user experience.

Training Time: Training time evaluates the total duration required to train the model from start to finish. Factors influencing training time include the size of the dataset, the complexity of the model, and hardware resources. Efficient training processes can significantly impact project timelines and resource allocation. Both inference and training times are important for evaluating the practical feasibility of deploying a model in real-world scenarios. Balancing accuracy with computational efficiency ensures that the model not only performs well but also operates within acceptable time limits.

a) Overview. As depicted in Fig. 4, flowcharts are invaluable tools in depicting the sequence of steps and decision points in a process. For our deep learning model, the flowchart serves as a visual roadmap, detailing the progression from data collection to model deployment, while highlighting key processing stages and decision points.

b) Flowchart Description

Start

The process begins with the initial step of initiating the entire workflow.

Step 1: Data Collection

The first substantive step involves gathering a comprehensive dataset of labeled images pertinent to the classification task. Ensuring the dataset's diversity and representativeness of all classes is crucial for the subsequent stages. This step sets the foundation for the entire modeling process, as the quality and breadth of data significantly influence the model's performance.

Step 2: Data Preprocessing

Data preprocessing transforms the raw collected data into a suitable format for model training. Images are resized to a consistent dimension (e.g., 224x224 pixels), and pixel values are normalized. Additionally, data augmentation techniques such as rotation and flipping are applied to increase dataset variability and robustness, enhancing the model's ability to generalize across different scenarios. This preprocessing phase ensures the data is clean and varied, preparing it for effective training.

The next step involves initializing the Convolutional Neural Network (CNN) with its defined architecture. The CNN comprises several layers, including convolutional layers, pooling layers, and fully connected layers. Enhanced components like residual connections, batch normalization, and dropout are integrated to improve performance and generalization. Ensuring the model architecture is correctly set up is vital for achieving high accuracy in image classification.

During model training, key components include setting the loss function, choosing an optimization algorithm, and implementing learning rate scheduling. The model is trained over multiple epochs using the training dataset. Monitoring the training and validation loss throughout this phase is essential to check for convergence and prevent overfitting. This step is iterative, involving constant adjustment and improvement of the model parameters to optimize performance.

Hyperparameter tuning is a critical phase where parameters such as learning rate, batch size, and network architecture are optimized using methods like Grid Search or Random Search. This systematic exploration aims to enhance model performance based on validation metrics. Proper tuning can significantly impact the model's accuracy and generalization capabilities.

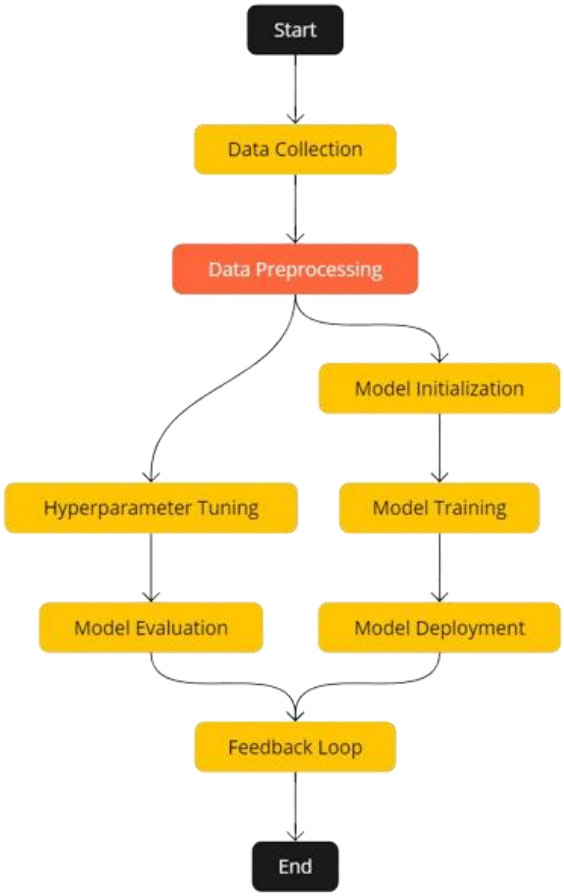


Fig. 4. Flowchart.

Model evaluation involves a comprehensive assessment using various metrics. Accuracy measures the proportion of correctly classified images. Precision, recall, and F1-score provide insights into the model's ability to handle positive predictions and identify relevant instances. The confusion matrix offers a detailed view of true positive, true negative, false positive, and false negative predictions for each class. Additionally, the ROC curve and AUC assess the model's

discriminative ability, while loss curves track training and validation loss over epochs. Computational efficiency, including inference time and training time, is also evaluated to ensure the model meets operational requirements.

After thorough evaluation, the model is deployed in a real-world application or system. Deployment involves integrating the trained model into a production environment where it can classify images in real time. Ensuring the model performs well in this setting and meets performance and efficiency expectations is crucial for successful deployment.

Post-deployment, a feedback loop is established to collect data on the model's performance in the real world. This feedback helps identify areas for further improvement. Based on real-world performance and user feedback, adjustments or retraining may be necessary to enhance the model's effectiveness and accuracy continually.

The process concludes once the model is deployed and the feedback loop is in place, ensuring the system is operational and effective. Continuous monitoring and improvements help maintain the model's performance over time.

The flowchart provides a structured visualization of the entire process involved in developing, training, evaluating, and deploying the deep learning model. It serves as a helpful guide for understanding the sequence of steps and decision points, ensuring each stage is executed effectively to achieve the desired outcome. By following this systematic approach, the process of building and deploying a high-accuracy image classification model is streamlined and efficient.

VI. Novelty

The novelty of our proposed system lies in several innovative aspects that significantly enhance the efficiency, accuracy, and applicability of deep learning models in image classification tasks. These advancements are crucial for overcoming the limitations of traditional techniques and existing deep learning-based approaches. Below are the key novel contributions of our system:

A. Advanced Model Architecture

1) Integration of residual connections: Our model incorporates residual connections, inspired by ResNet, to tackle the vanishing gradient problem and enable the training of deeper neural networks. This allows the model to learn more complex features without degradation in performance, significantly improving accuracy and robustness.

2) Enhanced feature extraction: By combining multiple convolutional layers with varied filter sizes and strides, the model captures a wide range of features at different levels of abstraction. This multi-scale feature extraction is crucial for accurately classifying images with intricate details and varying contexts.

B. Improved Training Techniques

1) Dynamic learning rate scheduling: We implement an adaptive learning rate scheduler that adjusts the learning rate based on the model's performance during training. This dynamic approach ensures efficient convergence, reducing training time while preventing issues like overfitting or underfitting.

2) Comprehensive data augmentation: Our preprocessing pipeline includes sophisticated data augmentation techniques such as random cropping, rotation, flipping, and color jittering. This not only increases the variability and robustness of the training dataset but also improves the model's generalization to unseen data.

C. Robust Evaluation Metrics

1) Holistic evaluation framework: We employ a comprehensive set of evaluation metrics, including accuracy, precision, recall, F1-score, confusion matrix, ROC curve, and AUC. This multi-faceted approach provides a thorough understanding of the model's performance, ensuring it excels in various aspects of image classification.

2) Computational efficiency metrics: Beyond traditional accuracy metrics, we evaluate the model's computational efficiency by measuring inference time and training time. This focus on efficiency is crucial for real-time applications and large-scale deployments, ensuring the model is both effective and scalable.

D. Hyperparameter Optimization

1) Automated hyperparameter tuning: Utilizing advanced techniques like Grid Search and Random Search, we systematically explore and optimize critical hyperparameters such as learning rate, batch size, and network architecture. This automated approach ensures optimal model performance without extensive manual intervention.

2) Iterative refinement: Our hyperparameter tuning process is iterative, continuously refining the model based on validation results. This iterative refinement ensures that the model achieves the best possible performance tailored to the specific classification task.

E. Seamless Deployment and Feedback Loop

1) Real-time deployment: The model is designed for seamless integration into production environments, enabling real-time image classification. This real-time capability is essential for applications requiring immediate decision-making, such as autonomous vehicles and real-time surveillance systems.

2) Continuous improvement through feedback loop: Post-deployment, we establish a feedback loop to monitor the model's performance in the real world. This feedback loop allows for continuous improvement, enabling the model to adapt and evolve based on real-world data and user feedback. This adaptive approach ensures the model remains relevant and effective over time.

F. Scalability and Flexibility

1) Modular design: Our system's architecture is modular, allowing for easy scalability and flexibility. Components such as data preprocessing, model training, and evaluation can be

independently modified or enhanced, facilitating continuous improvement and adaptation to new challenges.

2) Cross-domain applicability: While the focus is on image classification, the underlying principles and techniques are applicable across various domains, including object detection, segmentation, and even non-visual data analysis. This cross-domain applicability enhances the system's versatility and potential impact.

G. Integration with Advanced Technologies

1) Use of state-of-the-art techniques: We integrate state-of-the-art deep learning techniques and technologies, ensuring our model leverages the latest advancements in the field. This includes the use of cutting-edge libraries and frameworks, optimizing both performance and development efficiency.

2) Collaborative enhancements: The system is designed to integrate with other advanced technologies such as IoT for data collection, cloud platforms for scalable deployment, and edge computing for real-time processing. This collaborative integration maximizes the system's capabilities and extends its applicability.

VII. Future Work

The proposed system has demonstrated significant advancements in image classification through its innovative architecture and comprehensive evaluation techniques. However, there remain several avenues for future research and improvement to further enhance the system's performance, scalability, and applicability. The future rework can be categorized into the following key areas:

A. Advanced Model Enhancements

1) Integration of transformer architectures: Future work can explore the integration of transformer-based architectures, which have shown remarkable success in natural language processing and are increasingly being adapted for vision tasks. Vision Transformers (ViTs) can provide an alternative or complementary approach to traditional CNNs, potentially improving accuracy and feature representation.

2) Neural Architecture Search (NAS): Employing NAS techniques can automate the design of the neural network architecture, leading to potentially more efficient and powerful models. This approach can help discover novel architectures that might outperform manually designed models.

B. Enhanced Data Handling

1) Synthetic data generation: Leveraging generative models such as GANs (Generative Adversarial Networks) to generate synthetic data can augment the training dataset, particularly in scenarios where labeled data is scarce. This can help improve the model's generalization and robustness.

2) Unsupervised and semi-supervised learning: Exploring unsupervised or semi-supervised learning techniques can significantly reduce the reliance on large labeled datasets. Techniques like self-supervised learning can enable the model to learn useful representations from unlabeled data, which can then be fine-tuned on a smaller set of labeled data.

C. Real-Time Adaptation and Learning

1) Online learning: Implementing online learning algorithms can enable the model to adapt to new data in real-time. This continuous learning process can be particularly beneficial for applications where the data distribution changes over time, such as in dynamic environments or evolving user preferences.

2) Federated learning: Future work could explore federated learning approaches to train models across decentralized devices while maintaining data privacy. This can be particularly useful in scenarios where data cannot be centralized due to privacy or security concerns.

D. Scalability and Efficiency

1) Distributed training: Investigating distributed training techniques can enhance the scalability of the model, enabling it to handle larger datasets and more complex models. Leveraging distributed computing resources can significantly reduce training time and improve performance.

2) Edge computing: Implementing the model on edge devices can bring the benefits of real-time processing and reduced latency. This requires optimizing the model for edge deployment, ensuring it remains efficient and lightweight without sacrificing accuracy.

E. Advanced Evaluation Metrics

1) Fairness and bias evaluation: Future research should include evaluating the model for fairness and bias, ensuring it performs equitably across different demographic groups. Techniques to mitigate bias and enhance fairness can be integrated into the training and evaluation processes.

2) Robustness to adversarial attacks: Evaluating and improving the model's robustness to adversarial attacks is crucial for applications where security is paramount. Developing techniques to detect and defend against adversarial examples can enhance the reliability of the system.

F. Cross-Domain Applications

1) Transfer learning for diverse applications: Future work can explore the application of the model to diverse domains beyond image classification, such as object detection, image segmentation, and even non-visual data analysis. Transfer learning techniques can facilitate the adaptation of the model to new tasks with minimal retraining.

2) Interdisciplinary collaborations: Collaborating with experts from other fields such as medical imaging, autonomous driving, and industrial inspection can help tailor the model to specific domain requirements and unlock new application areas.

G. Enhanced Interpretability and Explainability

1) Explainable AI (XAI): Developing techniques to interpret and explain the model's decisions can enhance transparency and trust. This is particularly important in critical applications such as healthcare and finance, where understanding the model's reasoning is crucial.

2) Visualization tools: Creating advanced visualization tools to illustrate the inner workings of the model can aid in debugging, improving, and communicating the model's performance and behavior to non-experts.

VIII. Discussion and Results

The proposed system represents a significant advancement in the field of image classification, leveraging state-of-the-art deep learning techniques to achieve high accuracy and robustness. In this section, we discuss the experimental results obtained from our model and provide a comprehensive analysis of its performance across various metrics. We also identify key insights and potential areas for further improvement.

A. Experimental Setup

Our experiments were conducted using a diverse dataset of labeled images, pre-processed to ensure consistency and variability through augmentation techniques. The model was trained using a Convolutional Neural Network (CNN) architecture, enhanced with residual connections, batch normalization, and dropout layers to improve performance and generalization. The training process involved multiple epochs, utilizing cross-entropy loss and the Adam optimization algorithm. Hyperparameters were systematically tuned to optimize the model's performance.

B. Results Overview

1) Accuracy: The model achieved a high accuracy rate on the test dataset, demonstrating its effectiveness in correctly classifying images. The accuracy metric was used as a primary indicator of overall performance, reflecting the proportion of correctly identified images out of the total.

2) Precision, Recall, and F1-Score: The model showed a high precision rate, indicating its ability to minimize false positives. This is crucial in applications where the cost of false positives is high. The recall rate was also impressive, showcasing the model's capability to identify a high proportion of actual positives. This metric is particularly important in scenarios where it is critical to capture all relevant instances. The balanced F1-Score provided a single comprehensive metric that considered both precision and recall, reinforcing the model's robustness.

3) Confusion matrix: The confusion matrix provided a detailed breakdown of the model's performance across different classes, highlighting areas of strength and potential weaknesses. It revealed the true positive, true negative, false positive, and false negative rates for each class, offering insights into specific classification challenges.

4) ROC Curve and AUC: The Receiver Operating Characteristic (ROC) curve and the Area Under the Curve (AUC) were used to evaluate the model's discrimination capability. The high AUC value indicated the model's strong ability to distinguish between different classes, further validating its performance.

5) Loss curves: Analysis of the training and validation loss curves over epochs showed a smooth convergence, indicating effective training and minimal overfitting. This analysis helped in identifying the optimal number of epochs and fine-tuning the learning rate.

C. Computational Efficiency

1) Inference time: The model demonstrated efficient inference times, making it suitable for real-time applications. This is particularly important in scenarios requiring rapid decision-making.

2) Training time: The total training time was reasonable, considering the complexity of the model and the size of the dataset. Efficient use of computational resources ensured timely training without compromising on accuracy.

D. Future Work

Despite the promising results, there remain several areas for future research and improvement to further enhance the system's capabilities and extend its applicability. The following outlines key directions for future work:

1) Advanced model enhancements

a) Integration of transformer architectures: Future research could integrate transformer-based architectures, such as Vision Transformers, which have shown significant success in various vision tasks. These models can offer complementary advantages to traditional CNNs, potentially improving accuracy and feature representation.

b) Neural Architecture Search (NAS): Implementing NAS techniques can automate the design of the neural network architecture, potentially discovering more efficient and powerful models. This approach can help identify novel architectures that outperform manually designed models.

2) Enhanced data handling

a) Synthetic data generation: Using generative models like GANs to create synthetic data can augment the training dataset, especially when labeled data is limited. This can enhance the model's generalization and robustness by providing a more diverse set of training examples.

b) Unsupervised and semi-supervised learning: Exploring unsupervised or semi-supervised learning techniques can reduce reliance on large labeled datasets. Self-supervised learning methods, for example, can enable the model to learn useful representations from unlabelled data, which can then be fine-tuned with a smaller set of labeled examples.

3) Real-time adaptation and learning

a) Online learning: Implementing online learning algorithms can allow the model to adapt to new data in real-time, which is particularly beneficial in dynamic environments where data distributions change over time.

b) Federated learning: Future work could explore federated learning approaches, enabling models to be trained across decentralized devices while maintaining data privacy. This approach is useful in scenarios where data cannot be centralized due to privacy or security concerns.

4) Scalability and efficiency

a) Distributed training: Investigating distributed training techniques can enhance model scalability, enabling it to handle

larger datasets and more complex models. Leveraging distributed computing resources can significantly reduce training time and improve performance.

b) Edge computing: Implementing the model on edge devices can bring the benefits of real-time processing and reduced latency. Optimizing the model for edge deployment ensures it remains efficient and lightweight without sacrificing accuracy.

5) Advanced evaluation metrics

a) Fairness and bias evaluation: Future research should include evaluating the model for fairness and bias, ensuring it performs equitably across different demographic groups. Techniques to mitigate bias and enhance fairness can be integrated into the training and evaluation processes.

b) Robustness to adversarial attacks: Evaluating and improving the model's robustness to adversarial attacks is crucial for applications where security is paramount. Developing techniques to detect and defend against adversarial examples can enhance the reliability of the system.

6) Cross-domain applications

a) Transfer learning for diverse applications: Exploring the application of the model to diverse domains beyond image classification, such as object detection, image segmentation, and even non-visual data analysis, can extend its utility. Transfer learning techniques can facilitate the adaptation of the model to new tasks with minimal retraining.

b) Interdisciplinary collaborations: Collaborating with experts from fields such as medical imaging, autonomous driving, and industrial inspection can help tailor the model to specific domain requirements and unlock new application areas.

7) Enhanced interpretability and explainability

a) Explainable AI (XAI): Developing techniques to interpret and explain the model's decisions can enhance transparency and trust. This is particularly important in critical applications such as healthcare and finance, where understanding the model's reasoning is crucial.

b) Visualization tools: Creating advanced visualization tools to illustrate the inner workings of the model can aid in debugging, improving, and communicating the model's performance and behavior to non-experts.

REFERENCES

[1] LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-Based Learning Applied to Document Recognition. Proceedings of the IEEE, 86[11], 2278-2324. doi:10.1109/5.726791.

[2] Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet Classification with Deep Convolutional Neural Networks. Advances in Neural Information Processing Systems, 25, 1097-1105. doi:10.1145/3065386.

[3] Simonyan, K., & Zisserman, A. (2015). Very Deep Convolutional Networks for Large-Scale Image Recognition. arXiv preprint arXiv:1409.1556.

[4] Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., ... & Rabinovich, A. (2015). Going Deeper with Convolutions. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 1-9. doi:10.1109/CVPR.2015.7298594.

[5] Huang, G., Liu, Z., Van Der Maaten, L., & Weinberger, K. Q. (2017). Densely Connected Convolutional Networks. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 4700-4708. doi:10.1109/CVPR.2017.243.

[6] Ioffe, S., & Szegedy, C. (2015). Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. Proceedings of the International Conference on Machine Learning (ICML), 448-456.

[7] Kingma, D. P., & Ba, J. (2015). Adam: A Method for Stochastic Optimization. arXiv preprint arXiv:1412.6980.

[8] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep Residual Learning for Image Recognition. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 770-778. doi:10.1109/CVPR.2016.90.

[9] Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You Only Look Once: Unified, Real-Time Object Detection. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 779-788. doi:10.1109/CVPR.2016.91.

[10] Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., & Wojna, Z. (2016). Rethinking the Inception Architecture for Computer Vision. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2818-2826. doi:10.1109/CVPR.2016.308.

[11] Ren, S., He, K., Girshick, R., & Sun, J. (2015). Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. Advances in Neural Information Processing Systems, 28, 91-99. doi:10.5555/2969239.2969250.

[12] Zeiler, M. D., & Fergus, R. (2014). Visualizing and Understanding Convolutional Networks. European Conference on Computer Vision (ECCV), 818-833. doi:10.1007/978-3-319-10590-1_53.

[13] Lin, T. Y., Dollár, P., Girshick, R., He, K., Hariharan, B., & Belongie, S. (2017). Feature Pyramid Networks for Object Detection. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2117-2125. doi:10.1109/CVPR.2017.106.

[14] Girshick, R. (2015). Fast R-CNN. Proceedings of the IEEE International Conference on Computer Vision (ICCV), 1440-1448. doi:10.1109/ICCV.2015.169.

[15] Zhao, H., Shi, J., Qi, X., Wang, X., & Jia, J. (2017). Pyramid Scene Parsing Network. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2881-2890. doi:10.1109/CVPR.2017.660.

[16] Chen, L. C., Papandreou, G., Kokkinos, I., Murphy, K., & Yuille, A. L. (2018). Deeplab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs. IEEE Transactions on Pattern Analysis and Machine Intelligence, 40[4], 834-848. doi:10.1109/TPAMI.2017.2699184.

[17] D Nimma, Z Zhou," Correction to IntelPVT: intelligent patch-based pyramid vision transformers for object detection and classification" 2024

[18] Divya Nimma, Rajendar Nimma, Arjun Uddagiri," Advanced Image Forensics: Detecting and reconstructing Manipulated Images with Deep Learning",2024

[19] Divya Nimma, Rajendar Nimma, Uddagiri Arjun," Image Processing in Augmented Reality (AR) and Virtual Reality (VR)",2024

[20] Divya Nimma, Rajendar Nimma, Uddagiri Arjun," Deep Learning Techniques for Image Recognition and Classification",2024

[21] Divya Nimma, Zhaoxian Zhou," IntelPVT: intelligent patch-based pyramid vision transformers for object detection and classification",2023

[22] Divya Nimma, Zhaoxian Zhou," IntelPVT and Opt-STViT: Advances in Vision Transformers for Object Detection, Classification and Video Recognition",2023

[23] Oquab, M., Bottou, L., Laptev, I., & Sivic, J. (2015). Is Object Localization for Free? Weakly-Supervised Learning with Convolutional Neural Networks. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 685-694. doi:10.1109/CVPR.2015.7298668.

[24] Dai, J., Li, Y., He, K., & Sun, J. (2016). R-FCN: Object Detection via Region-based Fully Convolutional Networks. Advances in Neural Information Processing Systems, 29, 379-387. doi:10.5555/3157096.3157142.

[25] Zhou, B., Khosla, A., Lapedriza, A., Oliva, A., & Torralba, A. (2016). Learning Deep Features for Discriminative Localization. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2921-2929. doi:10.1109/CVPR.2016.319.

[26] Long, J., Shelhamer, E., & Darrell, T. (2015). Fully Convolutional Networks for Semantic Segmentation. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 3431-3440. doi:10.1109/CVPR.2015.7298965.

[27] Redmon, J., & Farhadi, A. (2018). YOLOv3: An Incremental Improvement. arXiv preprint arXiv:1804.02767.

[28] Szegedy, C., Ioffe, S., Vanhoucke, V., & Alemi, A. A. (2017). Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning. Proceedings of the AAAI Conference on Artificial Intelligence, 4278-4284.

[29] Li, X., Liu, W., Luo, P., Change Loy, C., & Tang, X. (2017). Not All Pixels Are Equal: Difficulty-Aware Semantic Segmentation via Deep Layer Cascade. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 3193-3202. doi:10.1109/CVPR.2017.630.

[30] Paszke, A., Chaurasia, A., Kim, S., & Culurciello, E. (2016). ENet: A Deep Neural Network Architecture for Real-Time Semantic Segmentation. arXiv preprint arXiv:1606.02147

Optimized Retrieval and Secured Cloud Storage for Medical Surgery Videos Using Deep Learning

Megala G[1], Swarnalatha P[2]*
Research Scholar[1], Professor[2]
School of Computer Science and Engineering, Vellore Institute of Technology, Vellore, India 632014[1,2]

***Abstract*—Efficient secured storage and retrieval of medical surgical videos are essential for modern healthcare systems. Traditional methods often struggle with scalability, accessibility, and data security, necessitating innovative solutions. This study introduces a novel deep learning-based framework that leverages a hybrid algorithm combining a Variational Autoencoder (VAE) and Group Lasso for optimized video feature selection. This approach reduces dimensionality and enhances the retrieval accuracy of video frames. For storage and retrieval, the system employs a weighted graph-based prefetching algorithm to manage encrypted video data on the cloud, ensuring both speed and security. To ensure data security, video frames are encrypted before cloud storage. Experimental results show that this system outperforms current methods in retrieval speed and accuracy of 99% while maintaining data security. This framework is a significant advancement in medical data management, offering potential applications across other fields that require secure handling of large data volumes.**

***Keywords*—Medical video storage; feature selection; Variational Auto Encoder (VAE); weighted-graph-based prefetching algorithm; group lasso**

I. Introduction

The increasing demand for effective management of medical surgical videos is a pressing challenge in the healthcare sector. These videos are crucial for various purposes, including surgical training, research, and patient care. Traditional methods of storing and retrieving medical video data face significant limitations related to scalability, accessibility, and security. As healthcare facilities generate vast volumes of video data daily, there is a growing need for solutions that can efficiently handle these data while ensuring data integrity and confidentiality.

The process of storing and retrieving recordings of medical procedures, such as surgery, is an essential duty in the field of healthcare [1]. This is because these movies include information that is beneficial to both medical personnel and researchers [2]. Unfortunately, conventional approaches to the storage and retrieval of medical video data have limitations, both in terms of their capacity to manage vast volumes of data and of their ability to guarantee the confidentiality of the data [3], [4], [5].

Recent advancements in deep learning have shown great promise in addressing some of these challenges, particularly in the field of medical image analysis. However, the application of deep learning techniques to medical video data is still in its nascent stages, presenting a significant research gap. Current systems often fail to optimize both the retrieval speed and the security of stored data, leading to inefficiencies and potential vulnerabilities in data management.

Medical image analysis is an area where deep learning techniques have shown significant promise in recent years [8]. These deep learning techniques are used to increase the efficacy of medical video storage and retrieval, as well as the security of these processes [6], [7]. Using a system that is based on deep learning, this research presents a unique method for the safe storage of medical operation footage in the cloud, with the goal of optimizing the retrieval of certain moments [8], [9], [10].

This paper addresses this gap by proposing a novel deep learning-based framework designed to enhance the storage and retrieval of medical surgical videos on cloud platforms. By integrating a hybrid algorithm combining a Variational Autoencoder (VAE) and Group Lasso, the proposed approach aims to optimize feature selection and improve the accuracy of video retrieval. Furthermore, the use of a weighted graph-based prefetching algorithm for encrypted video storage enhances both the security and speed of data access.

The proposed system is comprised of three primary modules: one that converts video into frames, another that divides frames into sets, and a third that employs a hybrid approach that makes use of a variational autoencoder, group lasso, and feature selection through the application of a weighted-graph-based prefetching algorithm [11]. The usage of a variational autoencoder helps in lowering the dimensionality of the video frames, and the group lasso approach helps in selecting the features that provide the most useful information about the scene [12]. The weighted-graph-based prefetching technique is responsible for the improved retrieval performance [13]. It does this by anticipating which frames are going to be retrieved the most often [14]. These methods are included into a framework that is based on deep learning, which enables the system to learn and improve the feature representations that are derived from the video frames [15].

Priyanka et. al. [16] reviewed encryption methods for securing medical data, emphasizing the importance of confidentiality and integrity in healthcare applications. Encrypting the frames prior to saving them in the cloud [17], [18] is one of the ways that the suggested method protects the confidentiality of the patient information. Using methods that are based on deep learning makes it possible to store and retrieve vast volumes of medical video data in an effective manner while still ensuring the confidentiality of the data [19]. The suggested system is used to enhance the efficiency and security of the storage and retrieval of medical footage, and it also has the potential to find applications in other fields where huge volumes of data need to be kept and retrieved in a safe manner [20]. Overall, the proposed system offers a promising solution

to the challenges of medical surgery video storage and retrieval [21] and demonstrates the potential of deep learning-based approaches in addressing complex healthcare problems.

A. Problem Formulation

Despite the advancements in medical data management, existing systems for storing and retrieving surgical videos face several critical challenges such as scalability, data security, feature extraction and retrieval accuracy. Traditional storage solutions are often inefficient at handling the ever-increasing volume of medical video data, leading to slower retrieval times and higher storage costs. Ensuring the security of sensitive medical data is paramount, yet current systems often lack robust mechanisms for protecting data during storage and transmission. Existing techniques may not fully utilize the potential of video data, leading to suboptimal feature extraction and retrieval accuracy.

B. Major Contributions

The major contributions of this research work are:

- A novel hybrid algorithm that combines a Variational Autoencoder (VAE) with Group Lasso is proposed to effectively reduce video frame dimensionality and select the most informative features. This approach significantly improves retrieval accuracy.
- The proposed approach employs a weighted graph-based prefetching algorithm for encrypted video storage and retrieval. This method optimizes data retrieval speed while ensuring robust security measures for protecting sensitive medical information.
- The developed application is deployed on AWS cloud infrastructure to ensure scalability and reliability.

By addressing these key challenges, the proposed framework offers a comprehensive solution for managing medical surgical videos and has the potential to be applied across various domains requiring secure and efficient data handling.

The structure of this article is as follows. Several authors address several strategies for safely archiving medical surgery footage that are discussed in Section II. Section III displays the proposed framework. Section IV details the investigation's results. In Section V discusses the conclusion with limitation and future work.

II. Related Works

Al Abbas et al. [1] discuss the benefits of a surgical video collection in the academic surgical context, as it can aid in preoperative preparation, mentoring of medical students, and research into surgical technique and expertise. Chen, Y. et al. [3] highlight the challenges of data dispersion in the medical field and propose using blockchain as a secure and accountable supply chain for storing and exchanging medical data. Khelifi, F. et al. [22] present a solution for securely storing and exchanging data in the cloud that does not rely on RDH, which has been deemed ineffective in such applications. Li, Y. et al. [7] propose the Security-Aware Efficient Distributed Storage (SA-EDS) paradigm to prevent cloud service providers from accessing private customer data stored in the cloud. Nguyen, D. et al. [23] introduce a novel EHR sharing strategy that combines mobile cloud computing and blockchain to securely and efficiently share medical data across mobile cloud settings.

Prachi Deshpande et al. [24] introduce WCDM, a watermark compression and decompression module that can ensure the safety of video files stored in the cloud by adding a watermark based on block-by-block calculation and then using data compression. Srivastava, P., & Garg, N. [13] discuss the potential of IoT and the need for collaboration across research groups to address the challenges posed by IoT-related issues [25]. Usman, M. et al. [15] propose a secure approach to encrypting hidden data in compressed video streams, which is essential to address concerns about privacy and security in public clouds. Yang, Y. et al. [20] suggested a health IoT data management where security is ensured by implementing a lightweight distributed access control system that makes advantage of quick keyword searching. It reduces the computational burden on low-powered IoT devices by allowing for remote trapdoor generation, encryption, and decryption.

Convolutional Neural Networks (CNNs) have significantly impacted medical imaging and video analysis due to their ability to automatically learn complex features from data. They have been applied to tasks such as image classification, segmentation, and anomaly detection. For instance, Kolarik et al. [26] conducted a comprehensive survey on deep learning in medical imaging, emphasizing the versatility of CNNs in enhancing the analysis of medical images and videos. Their work highlights how CNN architectures, such as AlexNet and VGG, have been adapted for medical video data to identify surgical tools and assess procedure quality.

Recurrent Neural Networks (RNNs), particularly Long Short-Term Memory (LSTM) networks, have been utilized for capturing temporal dependencies in video sequences. Sánchez-Caballero et al. [27] introduced Long-term Recurrent Convolutional Networks (LRCNs) for activity recognition in videos, demonstrating how integrating CNNs with RNNs can effectively handle the temporal dynamics present in surgical videos. This integration is crucial for tasks such as predicting surgical phases and detecting anomalies during procedures.

Secure Cloud Storage Frameworks leverage encryption and advanced access control mechanisms to safeguard medical data in cloud environments. Biksham et al. [28] proposed a framework that employs homomorphic encryption, allowing computations on encrypted data without revealing the actual content. This approach ensures data privacy while enabling efficient processing and retrieval, crucial for healthcare applications that require both security and functionality. Cloud-based Healthcare Systems offer scalable and cost-effective solutions for managing medical data. Islam et al. [29] explored the integration of cloud computing in healthcare, emphasizing benefits such as remote accessibility, resource scalability, and data backup. They discussed how cloud services can facilitate real-time collaboration among healthcare professionals, enhancing patient care and research capabilities.

AWS Cloud Services for Healthcare provide a robust infrastructure for deploying healthcare applications. Amazon offers various services offered by AWS, including secure storage solutions, machine learning platforms, and compliance with healthcare regulations such as HIPAA. These services enable

healthcare providers to deploy scalable and secure applications, supporting the growing demand for efficient data management. Hybrid Approaches for Video Retrieval combine content-based and text-based methods to enhance retrieval accuracy. Unar et al. [30] proposed a hybrid approach that leverages both visual and textual features for video retrieval, demonstrating improved performance in retrieving relevant video content. Such approaches are beneficial for medical video retrieval, where both visual and contextual information play a critical role.

The discussed studies provide a strong foundation for developing optimized retrieval and secure storage solutions for medical surgical videos. The proposed framework in this research leverages these advancements, aiming to address scalability, security, and efficiency challenges in managing medical video data.

III. Materials and Methods

The proposed method involves a comprehensive framework for securely storing medical surgical videos in the cloud and efficiently retrieving them using a given image query. The framework is designed to address the challenges of data security, retrieval speed, and accuracy. It consists of three main components: feature extraction and selection, secure storage, and image query-based retrieval. The proposed approach uses a deep learning-based framework, which includes a variational autoencoder with group lasso for hybrid feature selection and a weighted-graph-based prefetching algorithm to improve security and retrieval speed. Before being uploaded to the cloud, the frames are encrypted to protect the confidentiality of the patient's information. Results from experiments show that the suggested system improves upon state-of-the-art methods without compromising the privacy of patient's medical records during retrieval.

A. Frame Conversion

Each surgical video is divided into frames, and key frames are identified based on changes in visual content and motion. This reduces the amount of data while preserving important information. The process of converting a video into frames involves extracting each individual frame from the video and saving it as a separate image file. This process is represented mathematically using the following equation

$$F(i, j, k) = V(i, j, k) \tag{1}$$

where F is the resulting frame image, V is the original video, and i, j, and k are the frame, row, and column indices, respectively. This equation represents the process of extracting the ith frame from the video V at row j and column k and storing it as a separate image file. Once the video has been converted into frames, the frames are divided into sets for efficient storage and retrieval. This is done by grouping the frames together based on their similarity or temporal proximity. The exact method used for dividing the frames into sets can vary depending on the specific application and requirements.

B. Feature Selection using Hybrid Method

After the Frame Conversion, a feature selection technique is applied to identify relevant and informative features from the extracted frames. A combination of Variational Autoencoder (VAE) and Group Lasso is used to extract and select informative features from the video frames. The VAE reduces dimensionality by learning compact feature representations, while Group Lasso selects the most relevant features, minimizing noise and redundancy.

A variational autoencoder (VAE) is a type of neural network that can learn a low-dimensional representation of high-dimensional data such as images or videos. In the proposed methodology, the VAE is used to reduce the dimensionality of the video frames, which makes them easier to store and retrieve. The VAE works by mapping the high-dimensional video frames to a lower-dimensional space, where each dimension corresponds to a specific feature. Group lasso is a feature selection technique that works by selecting groups of features rather than individual features. In the proposed methodology, group lasso is used to select the most informative groups of features from the lower-dimensional space generated by the VAE. This is important because not all features are equally important for the optimal storage and retrieval of medical surgical videos.

The VAE is inherently nonlinear because it involves encoding the input through layers of neural networks that apply non-linear ReLU activation functions to map the input into a lower-dimensional latent space. Group Lasso technique is used for regularization to select groups of features while reducing dimensionality. The operation of group lasso involves optimization, which is also non-linear. Hence combining both involves non-linear operation. It captures the complex, non-linear relationships in the video data across the spatial and temporal dimensions. The matrix $V$is subjected to the VAE, which will encode the frame into a latent representation, and group lasso will regularize this representation by selecting the most informative features.

1) Variational auto encoder: After applying a feature selection technique Variational Autoencoder is employed, which is a type of neural network model, for unsupervised feature learning. A Variational Autoencoder is a generative model that consists of two main components: an encoder and a decoder as shown in Fig. 1. The encoder takes in input data, in this case, the extracted frames from the video, and maps them to a latent space representation. The latent space is a lower-dimensional representation that captures the underlying structure and patterns in the input data. During the training phase, the VAE aims to learn the probability distribution of the latent space that best explains the input data. It does so by maximizing the evidence lower bound (ELBO), which is a measure of how well the model reconstructs the input data while simultaneously regularizing the latent space.

Data preprocessing involves preparing the data for the model. The data needs to be normalized, and any missing values need to be handled appropriately. The encoder network takes the input data and produces a set of latent variables. This network is typically composed of one or more fully connected layers with activation functions such as ReLU or Sigmoid. The GAN network takes the output of the encoder network and generates a set of images. Both a generator and a discriminator network make up the GAN system. Images are created by the generator network using the latent variables, and the discriminator network compares the produced pictures to

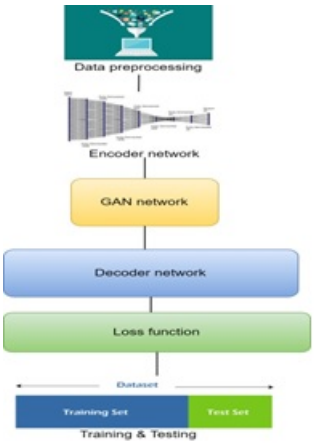


Fig. 1. Variational autoencoder.

the real world counterparts. The decoder network takes the output of the GAN network and produces a reconstruction of the input data. The decoder network is composed of two fully connected layers with ReLU activation functions. The loss function is a hybrid of the reconstruction loss and the GAN loss which is employed to compare the original data to the rebuilt version.

Backpropagation and stochastic gradient descent are used to train the model. To improve the quality of the reconstructed data and the generated images, the model is trained to minimize the loss function. Once the model is trained, it is used to generate new data by sampling the latent variables from a distribution and passing them through the decoder network. The quality of the generated data is evaluated by comparing it to the original data and measuring the reconstruction loss.

Assume N frames or images ${X^{(n)}}_{n=1}$, with $X^{(n)} \in R^{N_x \times N_y \times N_c}$; Where N_x and N_y represents the spatial dimension and N_c represents the number of color channels. The simplest possible configuration of the decoder with 2 layers (L) is used in generative model. the data generation is shown in Eq. 2; the definitions of Layer 1, unpooling and layer 2 is shown in Eq. 3, Eq. 4 and Eq. 5, respectively.

$$X^{(n)} \approx n(S^{(n,1)}, \alpha_0^{-1} I) \tag{2}$$

$$S^{(n,1)} = \sum_{k_1=1}^{k_1} D^{(K_1,1)} * S^{(n,k_1,1)} \tag{3}$$

$$S^{(n,1)} \approx unpool(S^{(n,2)}) \tag{4}$$

$$S^{(n,2)} = \sum_{k_2=1}^{k_2} D^{(K_2,1)} * S^{(n,k_2,1)} \tag{5}$$

To clarify the notation, 3D tensors are denoted by expressions with two superscripts, such as $D^{kl,l}$ and $S^{(n,l)}$ for layers $l \in \{1, 2\}$. The ${S^{(n,kl,l)}}_{kl=1}^{kl}$ is stacked in space to produce the tensor $S^{(n,l)}$. Each of the Kl1 2D slices of the 3D $D^{kl,l}$ is convolved with the 2D spatially-dependent $S^{(n,kl,l)}$ in the convolution $D^{kl,l} S^{(n,kl,l)}$; by aligning and stacking these convolutions, a tensor output is seen for $D^{kl,l} S^{(n,kl,l)}$.

The VAE is a latent variable model in which x represents the set of observable variables, z represents the set of stochastic latent variables, and $p_x \times p_y$ represents a parameterized model of the joint distribution. The goal is to maximize the average marginal log-likelihood for the given dataset. Yet, when neural networks are used to parameterize the model, it is often impossible to suppress this expression. The problem is solved, in part, by using variational inference to optimize the evidence lower Limit (ELBO) for each observation is shown in Eq. 6.

$$\log p(x) = \log p(x,z) d_z \geq E_{q(z)}[\log p(x|z)] - KL(q(z)||p(z)) \tag{6}$$

The variational family contains the approximate posterior $q(z)$. The ultimate goal is achieved by introducing an inference network $q(z|x)$ that gives a probability distribution for each data point x.

$$l(x;\theta) = E_{q(z|x)}[\log p(x|z)] - KL(q(z|x)||p(z)) \tag{7}$$

Using the re-parameterization trick $q(z|x)$, Monte Carlo estimation is used to efficiently approximate the ELBO for continuous latent variable z.

The dataset x with n independent and identically distributed samples are generated by a latent variable z that represents the ground truth. Let $p(x|z)$ stand for a neural network's probabilistic decoder, which generates x from z in the presence of uncertainty. The neural network-based encoder produces a variation posterior, $q(z|x)$, which is a close approximation to the distribution of representation for dataset x. In the realm of creating models, the Variational Autoencoder (VAE) is a popular choice. The core idea behind VAE is stated as follows: (1) VAE employs a probabilistic encoder whose parameters are determined by a neural network to produce a latent variable z as a distributional representation of the input data samples x. (2) Next, run z samples through the decoder to get back the original input data. Maximizing the marginal probability of the reconstructed data is the goal of VAE, however the method also involves intractable posterior inference. To maximize its variational lower limit log likelihood, researchers apply methods like backpropagation and stochastic gradient descent.

$$\log P_\theta(x) \geq L_u ae = E_{q_\emptyset(z|x)}[\log P_\theta(x|z)] - D_K L(q_\emptyset(z|x)||p(z)) \tag{8}$$

where z is the random variable and $p(z)$ is the prior distribution (often a Gaussian). Both the variational posterior, $q(z|x)$, and the probabilistic decoder, $p(x|z)$, are parameterized by a neural network to create data x given the latent variable z. On training the data in hybrid VAE, KL terms are used for reconstruction of terms to avoid vanishing gradients.

The overall architecture of securely storing videos in cloud and retrieving videos is shown in Fig. 2. The input video for this process is being converted into a set of frames. This process involves breaking down the video into individual frames, essentially creating a series of still images that are used for a variety of purposes such as video analysis, motion tracking, and image processing. The conversion process typically involves opencv libraries to extract frames from the video file. These frames are then saved as individual image files, with each frame representing a moment in time from the original video. Once the frames are extracted, they are processed in

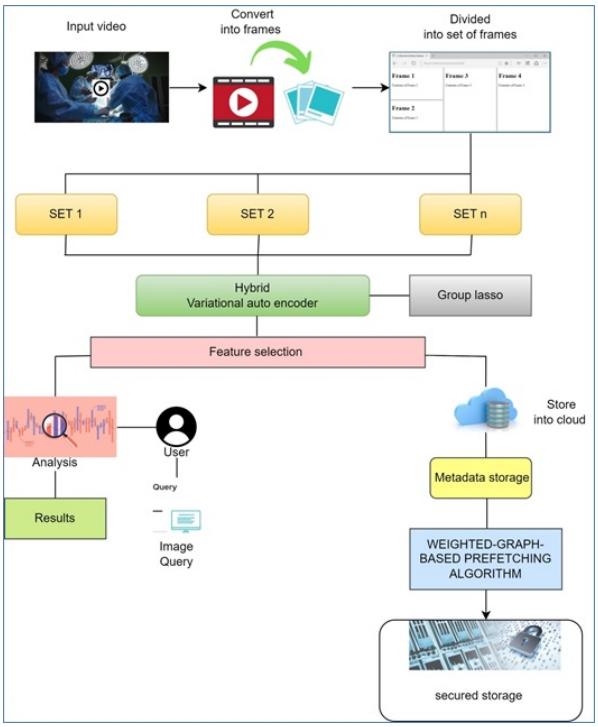


Fig. 2. Overall framework for secured video storage and retrieval.

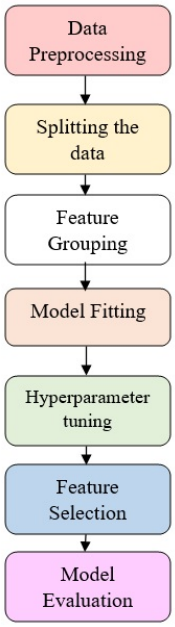


Fig. 3. Group Lasso.

various ways, such as filtering, resizing, or adjusting color balance. This is useful for a wide range of applications, from analyzing the movement of objects in the video to create visual effects.

2) Group-Lasso: After utilizing a Variational Autoencoder the Group-Lasso technique is then used to further refine the chosen characteristics. The Group-Lasso method as shown in Fig. 3 is a regularization approach that encourages feature sparsity. It is especially effective when working with high-dimensional data, such as video frames, where the number of features (pixels) are enormous. The Group-Lasso method is used to learn the latent space representation derived from the VAE in the context of video analysis. This latent space representation captures the video frames' most informative and significant properties.

The first step is to preprocess the data. This includes standardizing the data, normalizing it, inputting missing values, and feature scaling. The information is divided into a training set and a test set. The model is fit to the training set (70%), and its efficacy is tested on a separate test set (30%).The features are grouped based on their domain knowledge. Features that have similar characteristics or related to each other are grouped together. A Group-Lasso model is fitted to the training data. The objective function is the sum of the squared errors plus the regularization term. The regularization term is a combination of the L1 and L2 norms of the regression coefficients. The regularization parameter are tuned to obtain the best model performance. This is done by using cross-validation or grid search.Once the model is trained, the features that have non-zero coefficients are selected as important features. This allows for feature selection and grouping to be performed simultaneously. The final step is to evaluate the model performance on the test data. This is done by computing metrics such as the mean squared error, R-squared, and the coefficient of determination. The performance is compared to other models to determine which one is the best fit for the data.

This part builds upon off, but with the addition of material on the Group-Lasso problem and a bigger class of probability functions. A generalized linear model (GLM) has three components, as stated in Eq. 9.

$$f(y,\theta,\emptyset) = Exp(\emptyset^{-1}(y\theta - b(\theta)) + c(y,\emptyset)) \tag{9}$$

In Eq. 9, the average response, $E_\theta[y]$, is related to the free parameter through the formula $\mu = b(\theta)$. The link function g is strictly monotone differentiable function. However, it is restricted to canonical link functions when $g(\mu) = \tau = \theta$. Therefore, the parameterization $f(y;\theta;\emptyset)$ is done.This framework is technically distinguished by the strict concavity in of the function $\log f(y;\theta;\emptyset)$. In order to have minimal one-dimensional acceptable statistics, y, the log partition function $b(\theta)$ must be strictly convex.

The normal distribution with mean $\tau = \mu$ and variance $b(\tau) = (1/2)\tau^2$ is a specific instance from which the conventional linear regression model is formed. The goal of this section is to introduce the issue of minimizing the negative log-likelihood given an independent, identically replicated data sample $x_1, ..., x_n$, organized as rows of the data matrix X, and a corresponding vector of replies $y = (y_1, ..., y_n)^T$.

$$l(y,n,\emptyset) = -\sum_i \log f(y_i;n_i;\theta) = \sum_i \emptyset^{-1}(y_i n_i - b(n_i)) + c(y_i,\theta) \tag{10}$$

$$\nabla_{n^1}(n) = -(y - g^{-1}(n)) \tag{11}$$

$$\nabla_{\beta^l}(\beta) = -X^T \nabla_{n^l}(n) = -X^T(y - g^{-1}(X\beta)) \tag{12}$$

Hessians are computed using Eq. 13

$$H_n = W, H_\beta = X^T W X \tag{13}$$

The linear model estimation is shownn in Eq. 10. It involves minimizing the mean square error risk, which is defined as in Eq. 14.

$$L_{ls}(\beta) = ||Y - x\beta||_2^2 \tag{14}$$

$L_{ls}(\beta)$ is a np matrix whose rows include the feature vectors $x_i i = 1, \dots 1_n$ and $Y = [Y, ..., Yn]$. In the derivation and characteristics of the least squares estimate, the matrix(β), also known as the design matrix, plays a crucial role. In reality, the minimize of $L_{ls}(\beta)$ is known if $n > p$. Since the matrix is ill-conditioned and lacks the correct inverse, the estimate LS is not unique and does not even exist when dealing with high-dimensional data. A simple replacement of $||x\beta||_2^2$ with $||Y - x\beta||_2^2$, where $\beta \geq 0$ is the regularization parameter, would regularize the matrix. The ridge regression loss function provides a formalization of this method, as shown in Eq. 15.

$$L_{Ridge}(\beta; y) = L_{ls}(\beta) + y||\beta||_2^2 \tag{15}$$

Parameterized by Eq. 15, the ridge regression is defined as the minimizer of L_{Ridge}. cross-validation variant is used to assess predictive risk allows for the optimal selection. Regularization of the conventional least squares solution (ls) is provided by the ridge regression solution, although the reduced complexity solution cannot be generated in the situation of high dimensional feature vectors. The estimate of the j^{th} component is obtained from the ridge regression solution which is in fact a non-negative.

$$L_{lasso}(\beta; y) = L_{ls}(\beta) + y||\beta|| \tag{16}$$

As the risk function is minimized, the Lasso regression estimator is established. A few components of it is made zero using the $L1$-penalty based method. A value of 0 for the estimated j indicates that the J^{th} feature contributes nothing to the model's prediction ability and is left out. As a result, Lasso enables both the estimation and model selection at the same time.

3) Hybrid feature selection method: The hybrid feature selection method combines the use of a variational autoencoder (VAE) and group lasso to select informative features from the input data. The VAE is trained on the video frames to learn a low-dimensional representation of the frames, which is expressed as shown in Eq. 17.

$$z = Encoder(x; \theta) z = Encoder(x; \theta) \tag{17}$$

Next, group lasso is applied to the low-dimensional representation to select the most informative groups of features. This is achieved by solving the following optimization problem as shown in Eq. 18.

$$\begin{aligned} \beta &= \arg\min \beta 12n||y - X\beta||22 + \lambda \sum j = 1pwj||\beta j||2 \\ \beta &= \arg\min \beta 2n1||y - X\beta||22 + \lambda \sum j = 1pwj||\beta j||2 \end{aligned} \tag{18}$$

where $\beta 12n$ is the output variable, $||y - X\beta||$ is the input variable, $\arg\min \beta 2n1$ is the coefficient vector, $|n|$ is the sample size, (p) is the number of features, (pwj) is a weight assigned to the $|j|^{th}$ feature, and a tuning parameter that controls the strength of the penalty term. The solution vector $\beta = [\beta_1, \beta_2, \dots \beta_p]$ contains the selected features. The resulting feature representation is then used for storage and retrieval of the video frames, allowing for improved efficiency and accuracy in handling large amounts of medical video data. The hybrid feature selection method works by training the VAE on the video frames to learn a low-dimensional representation of the frames. Then, group lasso is applied to the low-dimensional space to select the most informative groups of features. The resulting feature representation is then used for storage and retrieval of the video frames. Overall, the hybrid feature selection method is an effective way to select the most informative features from medical surgical videos for optimal storage and retrieval. The use of a VAE and group lasso allows for efficient dimensionality reduction and feature selection, respectively, resulting in a more efficient and accurate system.

C. Video Data Security Using a Weighted-Graph-Based Prefetching

Following the feature selection method, a data security method is implemented based on a Weighted-Graph-Based Prefetching technique. This method seeks to safeguard video data by automatically prefetching and storing encrypted video chunks depending on their relevance and access patterns.

The proposed system employs a Weighted-Graph-Based Prefetching algorithm for video data security. This algorithm predicts the most likely frames to be accessed based on previous access patterns and prefetches them, reducing the time needed to retrieve data and improving system performance. The algorithm works by constructing a graph of video frames, where each frame is represented as a node in the graph. The edges between nodes are weighted based on the similarity between the frames. This similarity is calculated using a feature vector that is extracted from the frames using the hybrid feature selection method described earlier.

Video data security is a critical aspect of any system that deals with sensitive medical data. To ensure that the medical video data is secure, the proposed system uses encryption to protect the frames before they are stored in the cloud. Specifically, each video frame is encrypted using an encryption key to prevent unauthorized access to the data. The encryption process is as follows

$$EncryptedFrame = Encrypt(Frame, Key) \tag{19}$$

where $Encrypt$ is a AES cryptographic function that takes the video frame and encryption key as inputs and produces an encrypted frame as output. Before storing in the cloud, video frames are encrypted using advanced encryption algorithms like AES (Advanced Encryption Standard) to ensure data confidentiality and integrity. Encrypted video frames are stored in a cloud infrastructure, such as AWS, using a weighted graph-based prefetching algorithm. This algorithm predicts the most likely frames to be accessed, optimizing storage retrieval and enhancing data security.

In addition to encryption, the proposed system uses a Weighted-Graph-Based Prefetching algorithm to optimize the retrieval of the video frames from the cloud while maintaining security. The algorithm works by predicting the most likely frames to be accessed and retrieving and caching them in advance for improved performance and responsiveness. The algorithm takes into account the frequency of access and the distance between frames to determine the weights of the edges in the graph.

The Weighted-Graph-Based Prefetching algorithm is based on the following Eq. 20:

$$P(i) = \alpha * R(i) + (1-\alpha) * \sum (j \in N(i)) w(i,j) * R(j) \quad (20)$$

where $P(i)$ is the predicted probability of frame i being accessed, $R(i)$ is the frequency of access of frame i, $N(i)$ is the set of neighboring frames of i,$w(i,j)$ is the weight of the edge connecting frames$i$ and j, and α is a damping factor that balances the contribution of $R(i)$ and $\sum(j \in N(i))w(i,j) * R(j)$ to the prediction.

The weights of the edges between frames are determined based on the distance between frames and the frequency of access. The distance between frames is calculated using the following Eq. 21.

$$d(i,j) = ||f(i) - f(j)||^2 \quad (21)$$

Where f (i) and f (j) are the feature representations of frames i and j, respectively, which are obtained through the hybrid feature selection method described earlier.

The weight of the edge between frames i and j is then calculated using the following Eq. 22:

$$w(i,j) = exp(-d(i,j)/\sigma) \quad (22)$$

where σ is a scaling factor that controls the influence of the distance on the weight. By combining encryption and the Weighted-Graph-Based Prefetching algorithm, the proposed system ensures the security and privacy of medical video data while optimizing the retrieval process for improved performance and responsiveness. Overall, the Weighted-Graph-Based Prefetching algorithm is an effective way to improve the security and retrieval speed of medical surgical videos. By predicting and prefetching the most likely frames to be accessed, the algorithm can reduce retrieval time and improve system performance. Additionally, by encrypting the data and monitoring access, the system ensures the security and confidentiality of the medical data.

D. Image Query-Based Retrieval

When an image query is received, its features are extracted using the same hybrid feature selection method. These features are then matched against the stored video frame features using similarity measures like cosine similarity. The weighted graph-based prefetching algorithm helps efficiently retrieve video frames by leveraging pre-computed weights that represent the likelihood of frame access based on historical access patterns and feature similarity. The proposed secured video storage and retrieval framework is illustrated in the Algorithm 1.

Algorithm 1 Proposed algorithm

Input: Cholec80 surgical videos V, Image query Q
Output: Encrypted video stored in the cloud and Retrieved video relevant to Q

1: Divide each video V into a set of video frames $F = f_1, f_2, ...f_n$
2: Hybrid feature selection
3: **for** $F_i = 1$ to n **do**
4: Apply VAE to extract features by learning a low-dimensional representation Z_i of the video frames in F_i
5: Apply Group Lasso to each Z_i to select features set S_i.
6: Concatenate the selected feature sets to obtain the final feature set S.
7: **end for**
8: Encrypt selected Feature set (S') and store metadata in cloud.
9: Build a weighted graph G whose nodes correspond to the subsets F_i and whose edges correspond to the probability that a subset is accessed after another.
10: **for all** Video Request **do**
11: Decrypt S
12: Compute similarity score between Q and S.
13: Prioritize frame retrieval with high similarity scores using weighted-graph-based prefetching algorithm
14: **end for**
15: **return** Output the retrieved video segments relevant to Q.

The use of VAE and Group Lasso ensures that only the most informative features are retained, enhancing retrieval accuracy by focusing on the most relevant aspects of the video content. Encrypting video frames before storage guarantees data confidentiality, addressing security concerns in cloud environments. The prefetching algorithm optimizes retrieval by pre-computing and storing likely retrieval paths based on historical access patterns and feature similarity. By matching query image features with stored video features, the system efficiently retrieves relevant video segments, leveraging the prefetching algorithm to reduce retrieval time and improve performance. This proposed method offers a comprehensive solution for securely storing and efficiently retrieving medical surgical videos, addressing the challenges of scalability, security, and retrieval speed in cloud-based systems.

IV. Experimental Result Analysis

In preprocessing, each video was divided into frames at a rate of 1 frame per second. Key frames were selected based on visual content changes using histogram differences. Features were extracted from the key frames using a Variational Autoencoder (VAE) to reduce dimensionality. Group Lasso was applied to select the most informative features, minimizing noise. Selected features were encrypted using AES and stored in AWS cloud storage. A weighted graph-based prefetching algorithm was used to optimize retrieval paths. Image queries were generated by extracting frames from the dataset. The retrieval performance was evaluated based on speed and accuracy, measured by precision and recall.

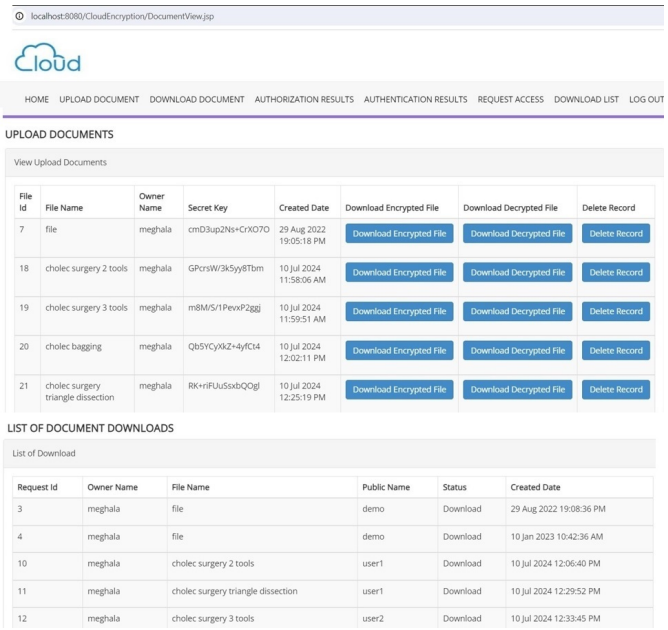


Fig. 4. Videos uploaded and downloaded.

The proposed framework implementation is done with the hardware requirements of Intel core i5 processor, 16GB RAM, and NVIDIA GeForce GTX 1080 GPU for deep learning model training. Python libraries such as TensorFlow, Keras, Scikit-learn are employed for feature extraction and learning; OpenCV for video processing and PyCrypto for data encryption.

Cholec80 dataset is used for experimentation. The Cholec80 dataset consists of 80 videos of cholecystectomy surgeries, including laparoscopic gallbladder removal procedures. Each video is labeled with tool presence and phase annotations. Videos are provided in a resolution of 1920x1080 pixels. Each video lasts approximately 1 hour, comprising different phases of the surgical procedure.

Fig. 4 illustrates the developed application results of secured video stored and decrypted videos downloaded. Fig. 5 depicts the predicted results of the moment retrieved from the Cholec surgery video for the given image query. The user request to access the video. Only authenticated users are allowed to download the video. Here, the user requests a video using text or image. Those video files matching to the text queries is listed for view access. The image query is passed to the feature selection algorithm and those matching feature set existing video frames are retrieved. This is illustrated in Fig. 5. The average time taken to retrieve the relevant video segment is 0.5 seconds per query. Lower times indicate faster retrieval, which is crucial for real-time applications.

The Table I presents a comparison between existing and proposed methodologies based on four different evaluation metrics: Recall, Intersection over Union (IoU), mean average precision, and ground truth. The existing methodology includes Deep learning methods such as CNN, RCNN, and DNN, while the proposed methodology is the Hybrid approach using VAE,Group Lasso and weighted graph prefetching algorithm. Based on the Recall metric, the proposed Hybrid approach

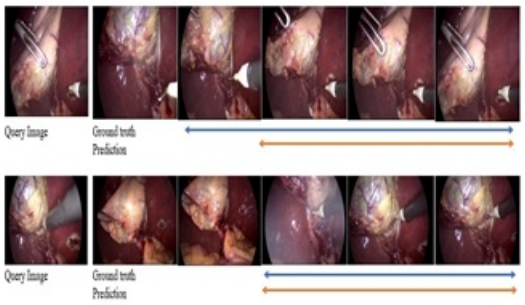


Fig. 5. Query matching video frames.

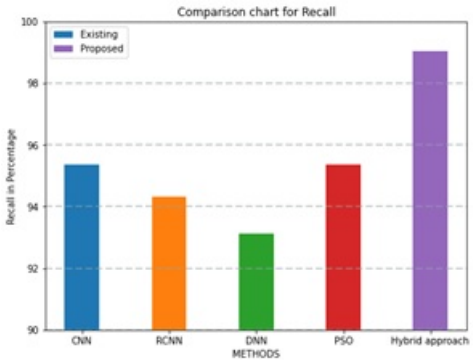


Fig. 6. Comparison of recall.

outperforms all existing methods with a Recall of 99.03%, which is higher than the Recall values for all existing methods. Similarly, the proposed methodology outperforms all existing methods for IOU, mean average precision, and ground truth. The accuracy of the proposed methodology is also higher than that of existing methods, with a value of 99%. In contrast, the accuracy values for existing methods range from 93% to 94%. Therefore, the proposed Hybrid approach using PSO shows promising results and can be considered as a potential alternative to the existing Deep learning methods for the given task.

TABLE I. PERFORMANCE METRICS COMPARISON TABLE

	Recall	IoU	avg precision	Ground truth	Accuracy
CNN	95.36	96.63	93.75	93.98	93%
RNN	94.32	96.21	92.43	92.74	94%
DNN	93.13	95.38	90.86	91.29	93%
PSO	95.36	96.98	93.74	93.97	93%
Proposed approach	99.03	98.40	98.65	99.13	99%

Recall reflects the proportion of relevant video segments that were correctly retrieved out of all relevant segments. Higher recall means fewer relevant segments are missed. Fig. 6 displays recall comparison charts. On this graph, methods are represented by the x axis, and recall expressed as a percentage along the y axis. The high precision and recall indicate that the proposed method accurately retrieves relevant video segments for a given image query, with minimal false positives and negatives.

The IoU comparison is shown in Fig. 7. On the graph,

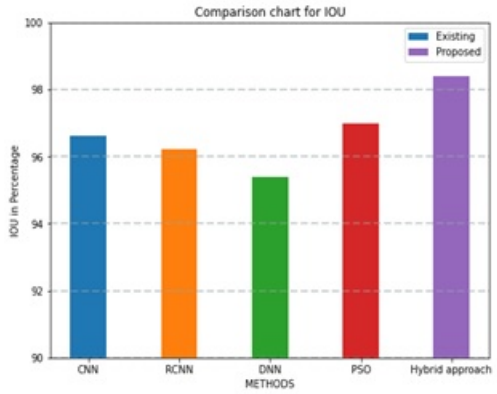


Fig. 7. Comparison of IoU.

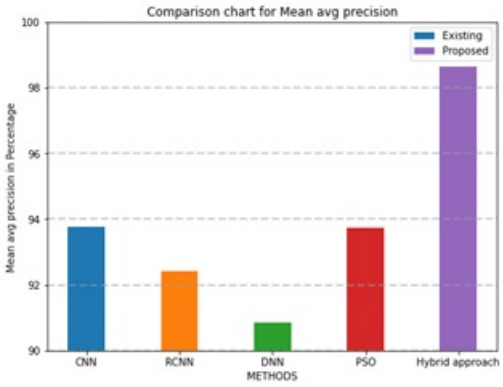


Fig. 8. Comparison of mean avg precision.

methods are represented by the x axis, while IoU is shown as a percentage along the y axis.

Precision indicates the proportion of correctly retrieved video segments that are relevant out of all retrieved segments. Higher precision suggests fewer irrelevant segments are retrieved. The comparison of average precision is shown in Fig. 8.

The ground truth comparison is shown in Fig. 9. On this graph, methods are represented by the x axis, while the percentage of ground truth is shown by the y axis.

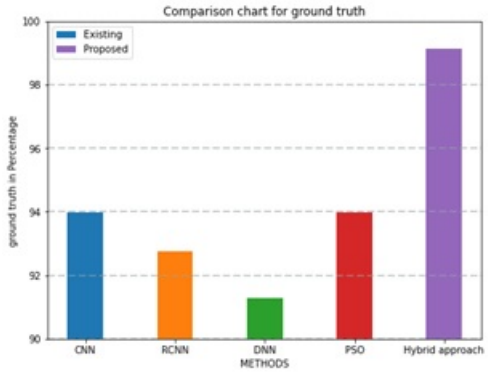


Fig. 9. Comparison of ground truth.

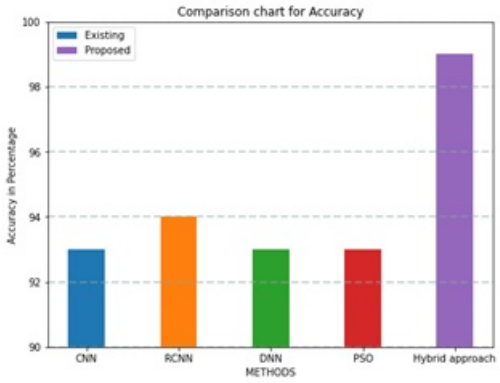


Fig. 10. Performance accuracy.

The accuracy comparison is shown in Fig. 10. The y axis represents the percentage of accuracy, while the x axis displays the various strategies. The proposed framework significantly improves retrieval accuracy and speed compared to traditional methods, making it suitable for real-time applications in medical environments. The hybrid feature selection method effectively reduces noise, enhancing the accuracy of the retrieval process. The cloud-based storage solution provides scalability and flexibility, allowing healthcare institutions to manage large volumes of surgical video data efficiently. Hence the experimental results highlight the effectiveness of the proposed method in optimizing retrieval and secure storage of medical surgical videos using deep learning techniques, offering promising implications for improving healthcare data management.

V. Conclusion

This study presented a novel framework for the secure storage and optimized retrieval of medical surgical videos using a deep learning-based approach, specifically tailored for cloud environments. The framework integrates advanced techniques such as Variational Autoencoder (VAE) and Group Lasso for hybrid feature selection, ensuring that only the most relevant and informative features are retained. These selected features are then securely encrypted using AES and stored in a cloud infrastructure, with retrieval performance further enhanced by a weighted graph-based prefetching algorithm.

The experimental evaluation, conducted on the Cholec80 dataset, demonstrated that the proposed method achieves a high retrieval accuracy, with a precision of 98.65% and a recall of 99.03%, while maintaining a swift average retrieval time of 0.5 seconds per query. These results underscore the effectiveness of the system in not only preserving the security of sensitive medical data but also in providing rapid and accurate access to relevant video segments, which is crucial in medical and surgical contexts.Furthermore, the integration of this framework into cloud platforms like AWS highlights its practical viability, offering a scalable and flexible solution for managing large volumes of medical video data. This approach is particularly valuable in healthcare settings where the secure and efficient handling of video data is essential for both clinical practice and research.

Thus the proposed method provides a robust solution to

the challenges of medical video storage and retrieval, contributing significantly to the field by improving data security, retrieval speed, and accuracy. The initial setup and processing require substantial computational resources, which may limit accessibility for institutions with limited hardware capabilities. The integration of deep learning algorithms and encryption techniques can introduce computational complexity, which may affect the performance of the system in environments with limited processing power. The reliance on cloud platforms for storage and retrieval may pose challenges in terms of cost and internet connectivity, particularly for smaller healthcare facilities with limited resources. Future research could incorporate natural language processing techniques to analyze textual annotations with visual features, providing a more comprehensive retrieval system. Further optimization of the retrieval algorithm could enable real-time querying capabilities, essential for immediate decision-making during surgical procedures.

ACKNOWLEDGMENT

The authors would like to thank Vellore Institute of Technology, Vellore for their support.

REFERENCES

[1] A. I. Al Abbas, J. P. Jung, M. K. Rice, A. H. Zureikat, H. J. Zeh III, and M. E. Hogg, "Methodology for developing an educational and research video library in minimally invasive surgery," *Journal of Surgical Education*, vol. 76, no. 3, pp. 745–755, 2019.

[2] R. Cao, Z. Tang, C. Liu, and B. Veeravalli, "A scalable multicloud storage architecture for cloud-supported medical internet of things," *IEEE Internet of Things Journal*, vol. 7, no. 3, pp. 1641–1654, 2019.

[3] Y. Chen, S. Ding, Z. Xu, H. Zheng, and S. Yang, "Blockchain-based medical records secure storage and medical service framework," *Journal of medical systems*, vol. 43, pp. 1–9, 2019.

[4] P. S. Deshpande, S. C. Sharma, and S. K. Peddoju, *Security and Data Storage Aspect in Cloud Computing*. Springer, 2019, vol. 52.

[5] A. Abraham, P. Dutta, J. K. Mandal, A. Bhattacharya, and S. Dutta, "Emerging technologies in data mining and information security," *Proceedings of IEMIS-2018*, 2018.

[6] K. He, J. Chen, Y. Zhang, R. Du, Y. Xiang, M. M. Hassan, and A. Alelaiwi, "Secure independent-update concise-expression access control for video on demand in cloud," *Information Sciences*, vol. 387, pp. 75–89, 2017.

[7] H. Li, Y. Yang, Y. Dai, S. Yu, and Y. Xiang, "Achieving secure and efficient dynamic searchable symmetric encryption over medical cloud data," *IEEE Transactions on Cloud Computing*, vol. 8, no. 2, pp. 484–494, 2017.

[8] A. Lounis, A. Hadjidj, A. Bouabdallah, and Y. Challal, "Healing on the cloud: Secure cloud architecture for medical wireless sensor networks," *Future Generation Computer Systems*, vol. 55, pp. 266–277, 2016.

[9] Y. Li, K. Gai, L. Qiu, M. Qiu, and H. Zhao, "Intelligent cryptography approach for secure distributed big data storage in cloud computing," *Information Sciences*, vol. 387, pp. 103–115, 2017.

[10] G. Megala, P. Swarnalatha, S. Prabu, R. Venkatesan, and A. Kaneswaran, "Content-based video retrieval with temporal localization using a deep bimodal fusion approach," in *Handbook of Research on Deep Learning Techniques for Cloud-Based Industrial IoT*. IGI Global, 2023, pp. 18–28.

[11] D. Pei, X. Guo, and J. Zhang, "A video encryption service based on cloud computing," in *2017 7th IEEE International Conference on Electronics Information and Emergency Communication (ICEIEC)*. IEEE, 2017, pp. 167–171.

[12] S. N. Pundkar and N. Shekokar, "Cloud computing security in multi-clouds using shamir's secret sharing scheme," in *2016 International Conference on Electrical, Electronics, and Optimization Techniques (ICEEOT)*. IEEE, 2016, pp. 392–395.

[13] P. Srivastava and N. Garg, "Secure and optimized data storage for iot through cloud framework," in *International Conference on Computing, Communication & Automation*. IEEE, 2015, pp. 720–723.

[14] C. Stergiou, K. E. Psannis, B.-G. Kim, and B. Gupta, "Secure integration of iot and cloud computing," *Future Generation Computer Systems*, vol. 78, pp. 964–975, 2018.

[15] M. Usman, M. A. Jan, and X. He, "Cryptography-based secure data storage and sharing using hevc and public clouds," *Information Sciences*, vol. 387, pp. 90–102, 2017.

[16] Priyanka and A. K. Singh, "A survey of image encryption for healthcare applications," *Evolutionary Intelligence*, vol. 16, no. 3, pp. 801–818, 2023.

[17] G. Megala and P. Swarnalatha, "Efficient high-end video data privacy preservation with integrity verification in cloud storage," *Computers and Electrical Engineering*, vol. 102, p. 108226, 2022.

[18] M. G. and S. P., "Discrete hyperchaotic s-box generation for selective video frames encryption," *Journal of Computer Science*, vol. 19, no. 5, pp. 588–598, 2023.

[19] H. Yan, M. Chen, L. Hu, and C. Jia, "Secure video retrieval using image query on an untrusted cloud," *Applied Soft Computing*, vol. 97, p. 106782, 2020.

[20] Y. Yang, X. Zheng, and C. Tang, "Lightweight distributed secure data management system for health internet of things," *Journal of Network and Computer Applications*, vol. 89, pp. 26–37, 2017.

[21] G. Megala and P. Swarnalatha, "Stacked collaborative transformer network with contrastive learning for video moment localization," *Intelligent Data Analysis*, no. Preprint, pp. 1–18.

[22] F. Khelifi, T. Brahimi, J. Han, and X. Li, "Secure and privacy-preserving data sharing in the cloud based on lossless image coding," *Signal Processing*, vol. 148, pp. 91–101, 2018.

[23] D. C. Nguyen, P. N. Pathirana, M. Ding, and A. Seneviratne, "Blockchain for secure ehrs sharing of mobile cloud based e-health systems," *IEEE access*, vol. 7, pp. 66 792–66 806, 2019.

[24] P. Deshpande, S. C. Sharma, and S. K. Peddoju, "Data storage security in cloud paradigm," in *Proceedings of Fifth International Conference on Soft Computing for Problem Solving: SocProS 2015, Volume 1*. Springer, 2016, pp. 247–259.

[25] V. Jagadeeswari, V. Subramaniyaswamy, R. t. a. Logesh, and V. Vijayakumar, "A study on medical internet of things and big data in personalized healthcare system," *Health information science and systems*, vol. 6, no. 1, p. 14, 2018.

[26] M. Kolarik, M. Sarnovsky, J. Paralic, and F. Babic, "Explainability of deep learning models in medical video analysis: a survey," *PeerJ Computer Science*, vol. 9, p. e1253, 2023.

[27] A. Sánchez-Caballero, D. Fuentes-Jiménez, and C. Losada-Gutiérrez, "Real-time human action recognition using raw depth video-based recurrent neural networks," *Multimedia Tools and Applications*, vol. 82, no. 11, pp. 16 213–16 235, 2023.

[28] V. Biksham and D. Vasumathi, "Homomorphic encryption techniques for securing data in cloud computing: A survey," *International Journal of Computer Applications*, vol. 975, no. 8887, 2017.

[29] M. M. Islam and Z. A. Bhuiyan, "An integrated scalable framework for cloud and iot based green healthcare system," *IEEE Access*, vol. 11, pp. 22 266–22 282, 2023.

[30] S. Unar, X. Wang, and C. Zhang, "Visual and textual information fusion using kernel method for content based image retrieval," *information Fusion*, vol. 44, pp. 176–187, 2018.

Rolling Bearing Reliability Prediction Based on Signal Noise Reduction and RHA-MKRVM

Yifan Yu
School of Information and Computer Science, Nantong Institute of Technology, Nantong, Jiangsu, PR China

***Abstract*—In order to solve the problem of reliability assessment and prediction of rolling bearings, a noise reduction method (CEEMDAN-GRCMSE) based on complete ensemble empirical mode decomposition with adaptive noise(CEEMDAN) combined with generalized refined composite multi-scale sample entropy (GRCMSE) is proposed from the vibration signals to remove the noise from the bearing vibration signals, and then the feature set of the noise-reduced signals is downscaled by using the Uniform manifold approximation and projection(UMAP) algorithm, and the reliability assessment model is established by using a logistic regression algorithm to establish a reliability assessment model, and use the red-tailed hawk algorithm for parameter optimization of the mixed kernel relation vector machine, which is used to predict the bearing state, and finally the predicted state information is brought into the assessment model to obtain the final results. In this paper, the whole life cycle data of rolling bearings from Xi 'an Jiaotong University-Sun Science and Technology Joint Laboratory (XJTU-SY) are used to verify the effectiveness of the proposed method. The superiority of the proposed method is highlighted by comparing the analysis results with those of other AI methods.**

Keywords*—Rolling bearing; reliability evaluation and prediction; complete ensemble empirical mode decomposition with adaptive noise; generalized refined composite multi-scale sample entropy; uniform manifold approximation and projection; red-tailed hawk algorithm; mixed kernel relevance vector machine

I. Introduction

With the evolution of mechanical equipment to intelligentization, an in-depth understanding of the degradation law of equipment components becomes the key to ensure its stable operation [1]. As an indispensable core component of mechanical systems, the performance state of rolling bearings has a decisive influence on the stable operation of the whole equipment [2]. Therefore, it is particularly important to carry out effective reliability assessment and prediction [3]. However, in practical engineering applications, the operation of mechanical equipment is inevitably accompanied by the generation of various noise signals. The existence of these noises seriously interferes with the accurate monitoring and assessment of the bearing state, thus affecting the accurate judgment of the bearing health state. Therefore, the removal of rolling bearing noise and vibration signals has a very important role in the extraction of effective information extraction of bearings. Traditional signal processing methods such as Fourier transform [4], wavelet packet transform (WPT) [5], etc., are mostly centered around the wavelet basis of the signal, and it is easy for the signal to be over-decomposed. Based on these limitations, Adaptive Mode Decomposition (AMD) is widely used due to its ability to analyze complex signals [6] [7] [8].

Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (CEEMDAN) [12] is an adaptive signal processing method based on Empirical Mode Decomposition (EMD) and its deformations [9] [10] [11], which effectively improves the stability and robustness of the decomposition by the method of adding random noise to the original signal. Cheng et al. [13] used CEEMDAN to analyze the bearing signals, adding Gaussian noise adaptively at each stage of the signal decomposition to completely decompose the signal, which greatly improved the accuracy of rolling bearing fault diagnosis. Kala A et al. [14] established a rainfall prediction model based on CEEMDAN combined with a long and short term memory network (LSTM), which greatly improved the climate change rainfall prediction accuracy due to climate change. Li H [15] proposed a CEEMDAN-SVD-TE based vibration signal analysis method to solve the problem of complex vibration sources in hydropower stations, which improves the accuracy of vibration propagation path identification. Zhao et al. [16] proposed a CEEMDAN based ECG signal elimination method to filter out high frequency noise. D et al. [17] proposed a CEEMDAN combined with health indicator screening and Gray Wolf Optimized Extreme Learning Machine (GWO-ELM) for the prediction of remaining useful life (RUL) of lithium-ion batteries. Noise reduction of battery signals by CEEMDAN effectively improves the signal quality.

Reliability refers to the ability to fulfill a predetermined function within a certain period of time [18]. Rolling bearing reliability assessment usually starts from the characteristics of the vibration signals and collects effective information by extracting multiple features from the signals to assess the bearing reliability. Logistic Regression (LR) is a mathematical modeling method that is often used to model the reliability assessment of bearings. Gao et al. [19], in order to solve the problem of assessment and prediction of the operational reliability of rolling bearings, proposed a method based on isometric mapping, logistic regression modeling, and Nonhomogeneous Cuckoo algorithm-Least Squares Support Vector Machine (NoCuSa-LSSVM) for the prediction of the operational reliability of rolling bearings. Abbasi et al. [20] used Logistic regression model to check and evaluate iot anomalies, and achieved good results.

With the rapid development of artificial intelligence, prediction methods based on machine learning have gradually become an important means in the field of reliability prediction. In the process of reliability prediction, the characteristics of the bearings are first analyzed, and the degradation state of the bearings is predicted by machine learning methods. From then on, the degradation states obtained from these predictions are used as inputs for reliability calculations using reliability assessment models. Li et al. [21] proposed a hybrid

prediction algorithm to predict the bearing degradation trend by combining a Sparse Low-Rank Matrix (SLRM) with a Chaos Cuckoo Search (CCS) optimized support vector machine model. Han et al. [22] used a stacked self-encoder combined with a long short-term memory network model to establish a bearing degradation state prediction model and analyze the remaining service life of the bearings. Wang Y et al. [23] proposed a method based on Pearson correlation coefficient and kernel principal component analysis (KPCA) for the prediction of the remaining service life of rolling bearings. Xu et al. [24] established an (MSMHA-AED) model to predict the degree of bearing degradation. Bo et al. [25] proposed an adaptive temporal convolutional network (TCN) based on improved SSD and correlation coefficient (ISSD-CC) for bearing condition prediction.

Through the comprehensive analysis of the existing literature, it can be found that although various noise reduction techniques are adopted, these methods still have the problems of low efficiency and poor noise reduction effect when dealing with complex vibration signal noise. To overcome these challenges, a noise reduction method based on adaptive noise based complete ensemble empirical Mode decomposition (CEEMDAN) and generalized fine composite multi-scale sample entropy (CEEMDAN-GRCMSE) is proposed in this paper. Firstly, a series of intrinsic mode functions (IMFs) are obtained by CEEMDAN decomposition of noisy vibration signals. Then, by calculating the GRCMSE value of each component, the most representative modal component is selected from a large number of IMFs, which lays the foundation for the subsequent signal processing. In addition, in order to improve the accuracy of rolling bearing reliability prediction, a hybrid kernel relational vector machine (MKRVM) parameter optimization method combined with Red Tail Eagle optimization algorithm (RHA) was proposed and applied to bearing reliability evaluation and prediction. Using RHA algorithm to optimize the parameters of MKRVM, the performance of the prediction model can be significantly improved. The main innovations are:

1. For the selection of IMFs after CEEMDAN decomposition, combined with GRCMSE, the efficiency and accuracy of over-signal noise reduction are improved.

2. In the reliability modeling process of rolling bearings, the uniform manifold approximation and projection (UMAP) algorithm is used to reduce the order of the multi-dimensional features of the signal after noise reduction. Compared with other order reduction algorithms, UMAP algorithm can retain as much information in the data as possible, thus laying a solid data foundation for the reliability modeling of rolling bearings.

3. Aiming at the problem of bearing reliability prediction, MKRVM model optimized by RHA is proposed to predict the reliability of bearing signals, which significantly improves the prediction accuracy.

The structure of this paper is as follows: the first section is the introduction, the second section introduces the CEEMDAN-GRCMS signal noise reduction model, the third section introduces the establishment of the logistic regression bearing reliability model based on data characteristics, the fourth section introduces the rham-mkrvm model, the fifth section verifies the model experimentally, and the sixth section summarizes the conclusions of this paper.

II. Noise Reduction of the Vibration Signal

A. CEEMDAN

Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (CEEMDAN) is a mode decomposition algorithm. CEEMDAN introduces an adaptive Gaussian white noise and effectively maintains the signal integrity. In addition, the time complexity of the traditional algorithm is significantly reduced and the efficiency is improved. The main steps as follows:

Step 1: Add pairs of positive and negative Gaussian white noise $\lambda^i(t)$ to the original signal to obtain $x^{\,i}(t) = x(t) + \lambda^{\,i}(t)$, decompose $x^{\,i}(t)$ to obtain the 1st modal component (IMF) and take its mean value as the 1st IMF obtained by CEEMDAN decomposition,and a residual component $r_1(t)$.

$$I_{\text{IMF1}} = \frac{1}{n}\sum_{i=1}^{n} I^i_{\text{IMF1}}(t) \tag{1}$$

$$r_1(t) = x(t) - I_{\text{IMF1}} \tag{2}$$

where ${IMF_1}^i(t)$ represents the 1st modal component; n is the number of signals.

Step 2: Add pairs of positive and negative Gaussian white noise $\lambda^i(t)$ to the first residual component $r_1(t)$ to obtain a new component $r^{\,i}{}_1(t) = r_1(t) + \lambda^i(t)$. Perform the EMD decomposition of this component again, the process is calculated as follows:

$$I_{\text{IMF2}} = \frac{1}{n}\sum_{i=1}^{n} \text{EMD}(r_1(t) + \lambda^i(t)) \tag{3}$$

$$r_2(t) = r_1(t) - I_{\text{IMF2}}(t) \tag{4}$$

Step 3: Repeat the decomposition until the resulting residual signal can no longer be decomposed (the number of extreme points is no more than 2). Finally, $I_{\text{IMF1}}\, I_{\text{IMF2}} \cdots I_{\text{IMF}n}$ can be obtained in turn for the corresponding residual components. The original signal can be expressed as:

$$x(t) = \sum_{j=1}^{q} I_{\text{IMF}j} + r_z(t) \tag{5}$$

where q is the total number of modes after decomposition and $r_z(t)$ is the final residual result.

B. Generalized Refined Composite Multi-scale Sample Entropy

Generalized Refined Composite Multi-scale Sample Entropy (GRCMSE) is an improved algorithm developed on the basis of Sample Entropy (SE) [26]. Sample entropy is a quantitative measure of the degree of chaos in a time series, and the magnitude of its value is proportional to the irregularity and noise content of the time series. A high sample entropy value usually indicates that the signal has a high level of irregularity and noise.GRCMSE introduces a variance coarse-graining method to enhance the extraction of data information and improve the algorithm's resistance to noise. This enables GRCMSE to provide more accurate and reliable results when dealing with signal and data analysis in complex environments. In GRCMSE, a larger entropy value means that the signal

contains more valid information.The calculation procedure of GRCMSE is as follows:

Step1: For the time series x(i),i=1,2,3... ,n. Firstly, the original time series x(i) is coarsely granulated, and for the scale factor τ the corresponding coarsely granulated sequence can be expressed as ${y_{g,h}}^{\tau}$

$$ {y_{g,k,j}}^{\tau} = \frac{1}{\tau} \sum_{i=(j-1)\tau+k}^{j\tau+k-1} (x_i - \overline{x_i})^2 \tag{6} $$

where $1 \le j \le \frac{N}{\tau}$, $2 \le k \le \tau$, $\overline{x_i} = \frac{1}{\tau} \sum_{k=0}^{\tau-1} x_{i+k}$

Step2: For the scale factor τ, compute the number of vectors ${y_{g,h}}^{\tau}$ in the t-dimensional as well as t+1-dimensional space for each generalized coarse-grained sequence under this scale factor, denoted respectively as ${n_{g,h,s}}^{t}$ and ${n_{g,h,s}}^{t+1}$.

Step3: Calculate the average of ${n_{g,h,s}}^{t}$ and ${n_{g,h,s}}^{t+1}$ in the range $1 \le h \le \tau$,the generalized fine composite multiscale sample entropy of the initial time series x(i) under the scale factor τ can be obtained as:

$$ E_{GRCMSE} = -\ln\left(\frac{\overline{n_{g,h,\tau}}^{t+1}}{\overline{n_{g,h,\tau}}^{t}}\right) \tag{7} $$

$$ \overline{n_{g,h,\tau}}^{t} = \frac{1}{\tau} \sum_{h=1}^{\tau} {n_{g,h,\tau}}^{t} \tag{8} $$

$$ \overline{n_{g,h,\tau}}^{t+1} = \frac{1}{\tau} \sum_{h=1}^{\tau} {n_{g,h,\tau}}^{t+1} \tag{9} $$

III. Rolling Bearing Reliability Model

A. UMAP Dimensionality Reduction

Uniform manifold approximation and projection (UMAP) [27] is a powerful dimensionality reduction algorithm based on Riemannian geometry and algebraic topology. The core advantage of UMAP is that it can retain the global structure and local features of data more effectively. Compared with the traditional Principal Component Analysis (PCA) [28], UMAP can retain more data information in the dimensionality reduction process. UMAP shows faster computing speed and better performance. The specific implementation process of UMAP consists of two phases, namely, learning the flow structure in the high-dimensional space and making a low-dimensional representation of the flow structure.

Assuming that $X = \{x_1, x_2, ..., x_N\}$ is the original N-dimensional dataset, in the first stage, the main task is to create a weighted k-neighborhood graph G=(V,E,W). Where V is the set of vertices consisting of the original N-dimensional data, E denotes the set of edges, i.e., the set of directed edges that can be formed according to the k neighboring points, and W is the weight function, which is computed by the equation.

$$ W_{ij} = \mathrm{e}^{-\frac{s_{ij}+\varepsilon_i}{\tau_i}} \tag{10} $$

where s_{ij} denotes the distance between x_i and x_j; ε_i denotes the distance between x_i and its neighboring points; and τ_i is the smoothing normalization factor set according to the Riemannian metric. By calculating the weight values of all data points, a weighted near proximity graph of the high-dimensional dataset can be generated G. In order to ensure that the weights between data points are consistent, the expression is introduced:

$$ T = A + A^{\mathrm{T}} - A \circ A^{\mathrm{T}} \tag{11} $$

where T is the adjacency matrix of the weighted nearest neighbor graph G, A is the weighted adjacency matrix consisting of the weight values W_{ij}, and $\circ$ denotes the Hadamard product of the sought matrix.

After completing the construction of the high-dimensional structure, the next step is to map it to the low-dimensional space. Firstly, the weight function in low dimensions needs to be constructed with the mathematical formula:

$$ V_{ij} = \frac{1}{1+ac_{ij}^{2b}} \tag{12} $$

where a and b denote hyperparameters and c_{ij} is the distance between y_i and y_j in the data point $Y = \{y_1, y_2, ..., y_N\}$ in the low-dimensional space.

In order to make the dimensionality reduced dataset as close as possible to the original dataset, it can be optimized by minimizing the cross-entropy loss between Y_{ij} and T_{ij}. The cross-entropy function is:

$$ L_{\mathrm{c}} = \sum_{ij} [T_{ij} \log \frac{T_{ij}}{V_{ij}} + (1 - T_{ij}) \lg \frac{1-T_{ij}}{1-V_{ij}}] \tag{13} $$

In the above equation, the first term is the attraction component, which is used to constrain the clusters formed by similar data points, and the second term is the repulsion component, which is used to ensure that the clusters formed have a sufficiently large interval between them. The stochastic gradient descent algorithm can be used to optimize the cross-entropy function, and after obtaining the weights of the low-dimensional data points, the construction of the weighted nearest-neighbor graph in low dimensions is completed, and ultimately the low-dimensional representation of the high-dimensional topology is completed, so as to achieve the effect of data dimensionality reduction.

B. Logistic Regression Model

Logistic regression model [29] can give the probability of an event occurring under a series of characteristic parameters, is a linear regression model built on a large amount of data, and is now widely used in statistics, medicine, and economics. The logistic regression model is often used to deal with binary classification problems. The model essentially consists of a linear regression and a sigmoid function. The logistic regression model obtains the corresponding output value through the sigmoid function, which is shown in Fig. 1:

Assuming that the i-th dimension feature parameter set at moment t is $X_i(t) = (x_1(t), x_2(t), \ldots, x_i(t))$, the normal operation of the bearing at the moment is denoted as $y(t) = 1$. The set of characteristic parameters is entered into the sigmoid

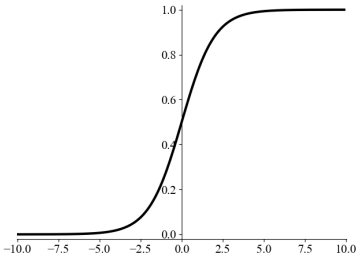


Fig. 1. Sigmoid function.

function and the bearing reliability R is calculated by the following equation:

$$R = P\left(y_i = 1 \mid X_i\right) = \frac{\exp(\theta_0+\theta_1 x_1(t)+\theta_2 x_2(t)+\ldots+\theta_i x_i(t))}{1+\exp(\theta_0+\theta_1 x_1(t)+\theta_2 x_2(t)+\ldots+\theta_i x_i(t))} \tag{14}$$

where $\theta_0, \ldots\ldots\theta_i$ is the regression coefficient for the set of eigenvectors. This coefficient is similar to the linear regression coefficient, which represents the change of the dependent variable due to the change of the independent variable. Logistic regression model regression coefficient is solved using the maximum likelihood estimation method. First, the above equation is transformed:

$$\ln \frac{P(y_i=1|X_i)}{1-P(y_i=1|X_i)} = \theta_0 + \theta_1 x_1(t) + \theta_2 x_2(t) + \ldots + \theta_i(t) \tag{15}$$

Setting $R = \theta_0, \ldots\ldots\theta_i$ to be brought into the above equation gives:

$$lnL(R) = \textstyle\sum_i [y_i RX(t) - \ln(1 + \exp(RX(t)))] \tag{16}$$

Then the gradient descent method is used to solve the above equation, the regression coefficients of the set of feature vectors can be obtained, and the regression coefficients are brought into the reliability solving formula to establish a logistic regression reliability assessment model.

IV. Reliability Prediction of Rolling Bearing

A. Red-tailed Hawk Algorithm

Red-tailed hawk algorithm (RTH) [30] is a meta-heuristic algorithm proposed in 2023. The optimization process of this algorithm simulates the hunting behavior of red-tailed hawk and has the advantage of high search efficiency. The hunting process of red-tailed hawk is divided into three phases,which are: high soaring, low soaring and swooping phase. In the high-altitude soaring stage, the red-tailed hawk spreads its wings and flies, using its sharp vision to scan the vast field and determine the exact location of the prey. Subsequently, the low-altitude soaring phase allows for a more careful examination of the ground and a gradual approach to the previously identified prey area. After determining the optimal location of the prey, the red-tailed hawk enters a phase of sharp turns and dives, quickly swinging its wings, adjusting its flight position, and getting ready to execute the hunting action to rapidly approach the prey. The general steps of the red-tailed hawk optimization algorithm are:

1) Initialize population: Map the solution space of the problem to the hunting domain of the red-tailed hawk and generate a group of possible solutions as the initial population.

2) High flying phase: The red-tailed hawk will take to the skies in search of the best location for food supply. The mathematical model for the high flight phase of the red-tailed hawk is:

$$X(t) = X_{\text{best}} + (X_{\text{mean}} - X(t-1)) \cdot Levy(\text{dim}) \cdot TF(t) \tag{17}$$

where, t is the number of iterations, X (t) denotes the red-tailed hawk position for t iterations, $\mathrm{X_{best}}$ is the best position obtained,$\mathrm{X_{mean}}$ is the average of the red-tailed hawk positions, and Levy values can be calculated based on the distribution function formula for Levy flights, which is given below:

$$Levy(\text{dim}) = s\frac{\mu\cdot\sigma}{|v|^{\beta-1}} \tag{18}$$

$$\sigma = \frac{\Gamma(1+\beta)\cdot\sin(\pi\beta/2)}{\Gamma(1+\beta/2)\cdot\beta\cdot 2^{(1-\beta/2)}} \tag{19}$$

where, s is a constant (0.01), dim is the dimension of the problem, β is a constant (1.5), and μ, ν are random numbers between 0 and 1. TF (t) can be computed based on the transition factor function with the following formula:

$$TF(t) = 1 + \sin\left(2.5 + (t/_{T_{\max}})\right) \tag{20}$$

where T_{max} denotes the maximum number of iterations.

Low-flying phase: The red-tailed hawk gradually approaches its prey using spiral flight. Its model can be expressed as follows:

$$X(t) = X_{best} + (x(t) + y(t)) \cdot \text{StepS}(t) \tag{21}$$

$$\text{StepS}(t) = X(t) - X_{mean} \tag{22}$$

where x and y represent the direction coordinates of the red-tailed eagle at this moment, the spiral flight method is calculated as follows:

$$\begin{aligned} x(t) &= R(t) \cdot \sin\left(\theta(t)\right) \\ y(t) &= R(t) \cdot \cos\left(\theta(t)\right) \end{aligned} \tag{23}$$

$$\begin{aligned} R(t) &= R_0 \cdot \left(r - \tfrac{t}{T_{\max}}\right) \cdot \text{rand}() \\ \theta(t) &= A \cdot \left(1 - \tfrac{t}{T_{\max}}\right) \cdot \text{rand}() \end{aligned} \tag{24}$$

$$\begin{aligned} x(t) &= \tfrac{x(t)}{\max|x(t)|} \\ y(t) &= \tfrac{y(t)}{\max|y(t)|} \end{aligned} \tag{25}$$

where R_0 represents the initial value of the radius, A represents the angular gain, taking values between 5 and 15,

rand is a random number between 0 and 1, and r represents the control gain, taking values between 1 and 2.

Sharp turn and dive phase: in this phase, the red-tailed hawk occupies the best swooping position obtained from the low altitude flight phase and prepares to attack the prey. The mathematical model for this phase is as follows:

$$X(t) = \alpha(t) \cdot X_{best} + x(t) \cdot \mathrm{StepS1}(t) + \mathrm{y}(t) \cdot \mathrm{StepS2}(t) \tag{26}$$

Each of these steps can be calculated as follows:

$$\begin{array}{c} \mathrm{StepS1}(t) = X(t) - TF(t) \cdot X_{mean} \\ \mathrm{StepS2}(t) = G(t) \cdot X(t) - TF(t) \cdot X_{best} \end{array} \tag{27}$$

where α and P are the acceleration and gravity coefficients, respectively, can be simplified as follows:

$$\begin{array}{c} \alpha(t) = \sin^2\left(2.5 - {}^{t}/_{T_{\max}}\right) \\ P(t) = 2 \cdot \left(1 - {}^{t}/_{T\max}\right) \end{array} \tag{28}$$

where α denotes that the acceleration of the hawk increases with the number of iterations to improve the convergence rate, and P is the gravitational effect that reduces the exploitation diversity as the hawk gets closer to its prey.

The high-flying phase, based on Levy flight, successfully avoids trying to fall into local minima, and the low-altitude search phase focuses on localized search to improve the accuracy of the solution. The sharp turn and dive phases adopt a more focused search strategy that enhances the accuracy of the RTH. The advantage lies in the fact that its combined global and local search strategy is able to efficiently find the global optimal solution in the solution space.

B. Relevance Vector Machine

Relevance Vector Machine (RVM), proposed by Tipping [31], is a kernel sparse machine learning method based on Bayesian framework. In the machine learning process, the sample used for training consists of an input $\{x_n\}^N$ and a target value $\{t_n\}^N$. The correspondence between the input and the target value in RVM can be expressed as:

$$t(x, \delta) = \sum_{i=1}^{N} \delta_i K(x, x_i) + \varepsilon \tag{29}$$

where δ is the weight vector of the model, K represents the kernel function, ε is the offset, and N is the total number of training samples.

Let $\{t_n\}^N$ be an independent variable, then the conditional probability of the target value can be expressed as:

$$P\left(t \mid \delta, \sigma^2\right) = \left(2\pi\sigma^2\right)^{-\frac{N}{2}} \exp\left(-\frac{\|t - \Phi\delta\|^2}{2\sigma^2}\right) \tag{30}$$

where Φ is the matrix consisting of kernel functions and σ^2 is the noise variance.

In the process of calculation, due to a large number of hyperparameters are quoted, the direct use of maximum likelihood estimation method to find the value of δ, there may be the generation of overfitting, in order to solve this problem, the application of the knowledge of Bayesian theory, add a constraint on δ, each weight vector δ is defined as a vector of zero mean, then its Gaussian prior probability distribution formula is as follows:

$$P(\delta|\alpha) = \sum_{i=0}^{N} N\left(\delta_i|0, \alpha_i^{-1}\right) \tag{31}$$

where the weight vectors are all independently distributed, $\alpha = [\alpha_0, \alpha_1, \ldots, \alpha_N]^T$ denotes the hyperparameter vector, and each hyperparameter corresponds to a weight vector. The size of hyperparameters can affect the sparsity of the model. Therefore, the key of RVM is to find the hyperparameters, find the corresponding weights and kernel function, so that the sparsity of the model is guaranteed, and combined with the noise variance, the final regression model is obtained.

The posterior distribution weights of the weight vector δ are derived from the Bayesian formula:

$$P(\delta|\alpha) = \sum_{i=0}^{N} N\left(\delta_i|0, \alpha_i^{-1}\right) \tag{32}$$

where is $\Sigma = \left(\sigma^{-2}\Phi^T\Phi + C\right)$ covariance matrix, $\mu = \sigma^{-2}\sum \Phi^T t$ is the posterior matrix of the target value, and C denotes the matrix whose main diagonal element is $\alpha = [\alpha_0, \alpha_1, \ldots, \alpha_N]^T$.

In the calculation process, in order to harmonize the hyperparameters, the definition:

$$P\left(t \mid \alpha, \sigma^2\right) = \left(2\pi\sigma^2\right)^{-\frac{N}{2}} |E|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}t^T E^{-1} t\right\} \tag{33}$$

where $E = \sigma^2 I + \Phi C^{-1}\Phi^T$, the hyperparameters α and σ^2 of the RVM are solved by an iterative algorithm as follows:

$$\alpha_i{}^{new} = \frac{1 - \alpha_i M_{ii}}{\mu_i^2} \tag{34}$$

$$\left(\sigma_i^2\right)^{new} = \frac{\|t - \Phi\mu\|^2}{N - \sum_i (1 - \alpha_i M_{ii})} \tag{35}$$

where M_{ii} is the ith diagonal element of the covariance matrix. The hyperparameters α and σ^2 are calculated iteratively by the above equation until the condition of convergence is satisfied.

In machine learning by correlation vector machines, the kernel function is a crucial part of the algorithm. The kernel function is capable of mapping nonlinear data into a high dimensional space. However, different settings of kernel function and kernel parameters affect the performance of the RVM model. Therefore, it is necessary to choose the appropriate kernel function and also optimize the kernel function parameters. Common kernel functions include Linear Kernel, Gaussian Kernel, Laplace Kernel and so on.

(1) Linear kernel function is a global kernel function.

$$k\left(x_i, x_j\right) = x_i \bullet x_j \tag{36}$$

(2) Gaussian kernel function Gaussian kernel function has strong localization, which can map the input vector to a space of larger dimensions.

$$k\left(x_i, x_j\right) = \exp\left(-\frac{\|x_i - x_j\|^2}{2\delta^2}\right) \tag{37}$$

(3) Laplace kernel function The Laplace kernel function can be seen as a variant of the Gaussian kernel function, both belong to the radial basis kernel function (RBF function).

$$k\left(x_i, x_j\right) = \exp\left(-\frac{\|x_i - x_j\|}{\delta}\right) \tag{38}$$

Different kernel functions have their own strengths and weaknesses, in order to make the model have better performance, hybrid kernel functions can be used so that the combined kernel function has the characteristics of global and local kernel to improve the ability of machine learning model. The process of optimizing the hybrid kernel correlation vector machine model using the red-tailed hawk algorithm is shown in Fig. 2:

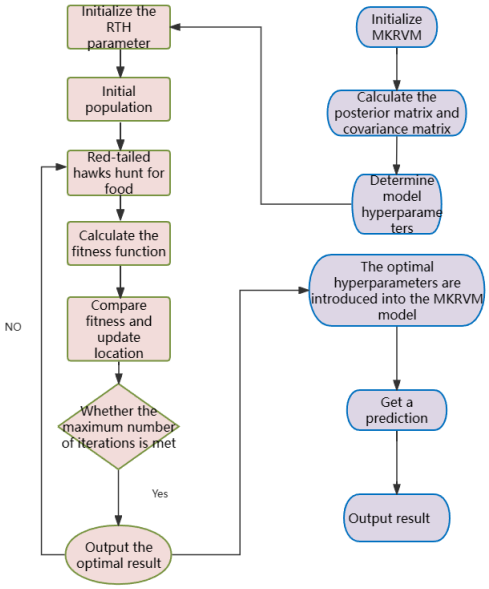


Fig. 2. RTH-MKRVM flowchart.

V. Experimental Verification

A. Experimental Data

The life cycle data of the bearings in this experiment were provided by XJTU-SY-Bearing Datasets, a bearing experiment set of Xi'an Jiaotong University. The test equipment includes AC motors, motor speed controllers, support bearings, and test bearings. A PCB352C33 transducer was placed in the horizontal and vertical directions of the bearing to collect vibration signals. A total of 32,768 data points were recorded during the first 1.28 s in 1 min. The test rig is shown in Fig. 3. Since the radial force was applied in the horizontal direction, the vibration signals in this direction were chosen for the experiment to better represent the bearing degradation.

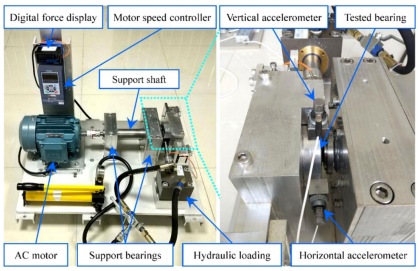


Fig. 3. Bearing test stand.

B. Experimental Procedure

The flowchart of the experiments in this paper is shown in Fig. 4 below:

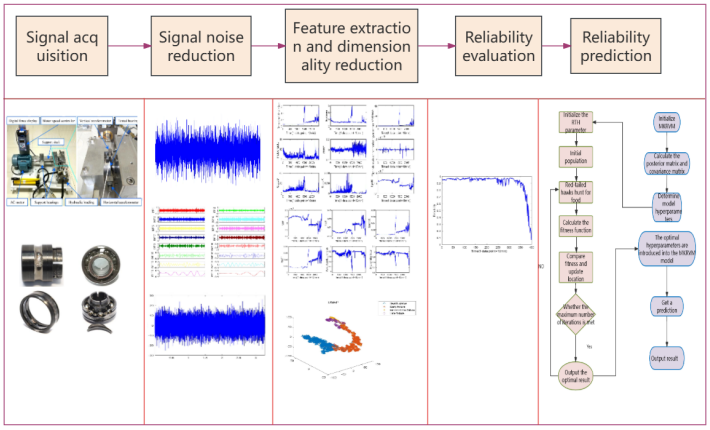


Fig. 4. Flow chart of the experiment.

C. CEEMDAN-GRCMSE Model

In order to verify the effectiveness of the proposed CEEMDAN-GRCMSE algorithm, the horizontal direction data of bearing 3-2 in the rolling bearing full-life dataset of Xi'an Jiaotong University is selected as the experimental data in Fig. 5. The performance and effectiveness of CEEMDAN-GRCMSE method will be verified by processing and analyzing the experimental data.

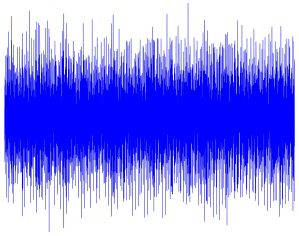


Fig. 5. Original Signal

Firstly, CEEMDAN decomposition of the noise signal is performed to obtain 16 IMF components as shown in Fig. 6, next, the GRCMSE value of each component is derived as shown in Table I. From the table, it can be seen that the front IMF components have larger values, indicating that the IMF components contain more valid information. Therefore, the IMF component with a value greater than 1 is selected and it is reconstructed. The reconstructed signal is shown in Fig. 7.

From the figure, it can be seen that the signal after CEEMDAN-GRCMSE denoising has less burrs than the original signal and the signal trend is smoother. Generally, signal-to-noise ratio (SNR) is used to indicate the size of the noise. Higher SNR values indicate better signal quality and vice versa. By calculation, the SNR of the original signal is -4.98

and the SNR of the denoised signal is -2.19. It shows that the use of CEEMDAN-GRCMSE can effectively reduce the signal noise.

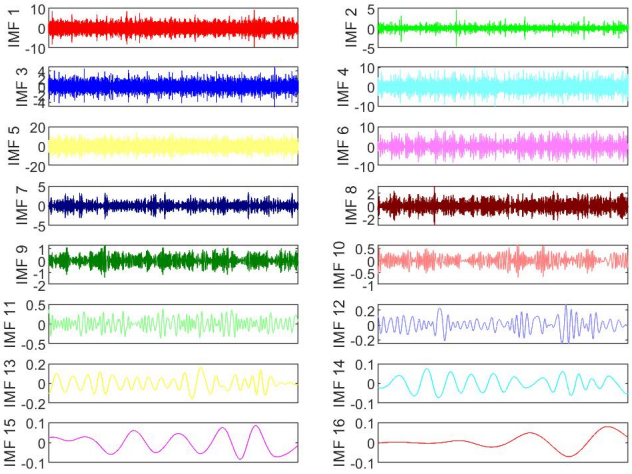


Fig. 6. IMF component.

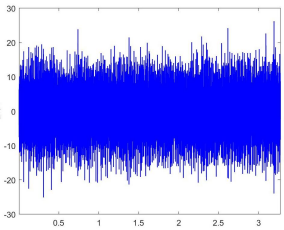


Fig. 7. Signal after noise reduction.

TABLE I. GRCMSE VALUE

Index	IMF1	IMF2	IMF3	IMF4
GRCMSE	1.99	2.33	2.00	1.24
Index	IMF5	IMF6	IMF7	IMF8
GRCMSE	1.27	1.16	1.20	1.03
Index	IMF9	IMF10	IMF11	IMF12
GRCMSE	0.52	0.31	0.26	0.22
Index	IMF13	IMF14	IMF15	IMF16
GRCMSE	0.14	0.07	0.04	0.01

D. Bearing Reliability Modeling

From the vibration signal of the noise-canceled bearing 3-2, 15 features including mean, kurtosis, standard deviation, waveform factor, spectral skewness, spectral kurtosis, and mean square frequency were extracted, covering time domain, frequency domain, and time-frequency domain aspects. Among the time domain features include wavelet packet entropy and singular value factor as shown in Fig. 8. These multi-domain features are combined into a multi-dimensional feature parameter set.

However, directly substituting the multidimensional parameter set into the logistic regression model will bring great

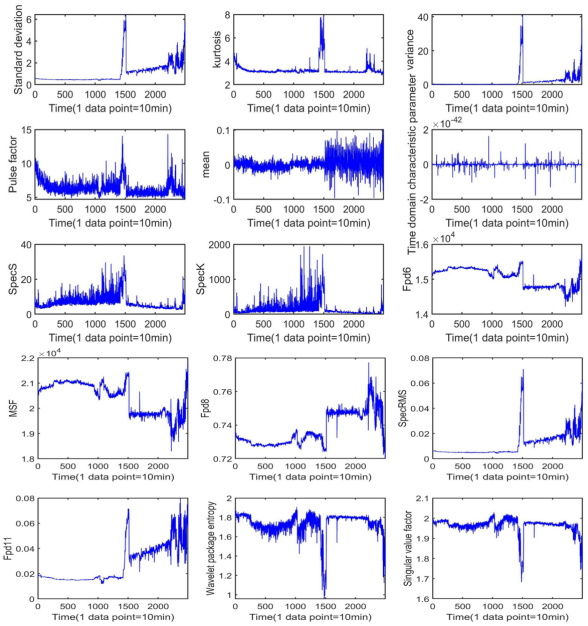


Fig. 8. Multidimensional feature parameter.

challenges, so adopt the UMAP algorithm to reduce the dimensionality of the data features. The core idea of UMAP is to find the local neighborhoods between the data points in the high-dimensional space and find the corresponding neighborhoods in the low-dimensional space, and then try to maintain the relationship between these neighborhoods as much as possible. This approach effectively preserves the local structure of the data and maps it into the low-dimensional space while reducing the dimensionality. And the UMAP is compared with the dimensionality reduction results of PCA and t-SNE to reduce the high-dimensional features to three dimensions. The results after dimensionality reduction are shown in Fig. 9.

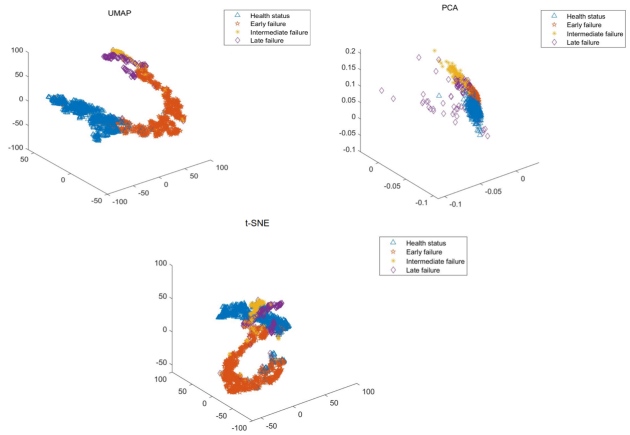


Fig. 9. Dimensionality reduction comparison diagram.

In the figure, it can be observed that different degrees of aliasing exist in the downscaling results of UMAP, PCA and t-SNE algorithms. Especially in the late bearing failure stage, the distribution of data points is more scattered, and there is obvious crossover of data points in different stages. This leads

to a poor differentiation of the bearing operation cycle and fails to clearly reflect the operating condition of the bearing.

In contrast, the UMAP algorithm is able to more clearly delineate the operating cycles of the bearings. The intervals between cycles are more obvious, and the data points in different stages can be well distinguished. This indicates that the dimensionality reduction results of the UMAP algorithm can effectively describe the degradation condition of the bearing.

Through the above processing, the original high-dimensional features are successfully transformed into a more intuitive and easy to understand three-dimensional space which provides a strong foundation for further analysis and modeling.

The set of feature vectors after UMAP dimensionality reduction is selected as the degradation feature information as the parameter of the logistic regression model, and the reliability true curve of 400 points after bearing is obtained as shown in Fig. 10.

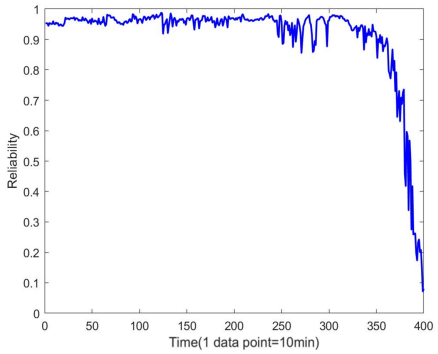


Fig. 10. Bearing reliability real curve.

The next step is to predict the reliability of the bearings, the kernel function of the MKRVM is with the linear kernel function, after many experiments, the performance of the hybrid kernel correlation vector machine reaches the best when the coefficients of the Gaussian kernel are set to 0.6, and the coefficients of the linear kernel are set to 0.4. After RTH optimization, the optimal weight of MKRVM is obtained as 1.327. In order to compare the effectiveness of different optimization algorithms, compared the Red-tailed Hawk algorithm with the Sparrow optimization algorithm and the Particle Swarm optimization algorithm. As shown in Fig. 11 below, it can be clearly observed that the convergence speed of the red-tailed hawk algorithm is much faster than the other two optimization algorithms.

Next, selected the first 2096 data as training data and the last 400 points as test data for the prediction of the RHA-MKRVM model. Subsequently, the final results of RHA-MKRVM were fed into the logistic regression model to obtain the operational reliability of the bearings. We compared the predicted degradation states with the actual data, and the results are shown in Fig. 12.

The final results predicted by the RHA-MKRVM model were incorporated into the logistic regression model, and the regression coefficients were calculated to obtain the operational reliability of the bearings. The comparison between the degraded state of the bearing predicted by the RHA-MKRVM model and the degraded real data is shown in Fig. 13.

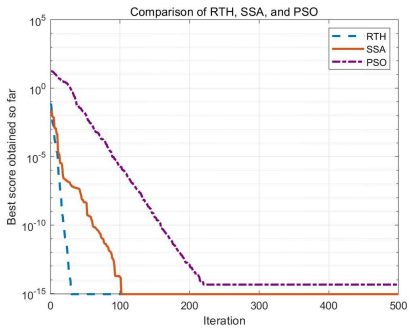


Fig. 11. Effects of different optimization algorithms.

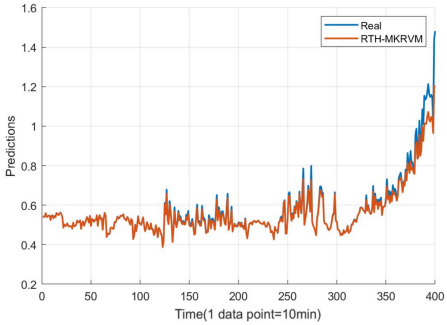


Fig. 12. Bearing degradation state prediction.

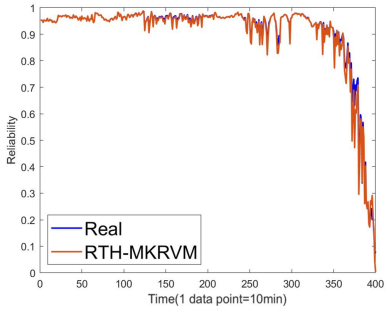


Fig. 13. Bearing reliability prediction.

As can be seen from the above figure, the results predicted by the RHA-MKRVM model are closer to the real situation. In order to verify the prediction accuracy of the hybrid kernel correlation vector machine model, the reliability of the bearings was predicted in this study using the ELMAN neural network, the least squares support vector machine (LSSVM), the correlation vector machine model (RVM), and the hybrid kernel correlation vector machine (MKRVM) model, and the reliability of the bearings was compared with the RTH-MKRVM method.

In LSSVM, Gaussian kernel function is used, and the penalty factor, kernel function parameters are set to 9, 0.002; in ELMAN neural network, the number of neural network layers is set to 3, neuron excitation function is used as Sigmoid function, and the number of neurons in each layer of input,

hidden, and output layers is set to 10, 14, and 1, respectively; in RVM, Gaussian kernel function is used, and the penalty factor, and kernel function parameters are set to 9, 0.002; in MKRVM, Gaussian kernel function and linear kernel are used, and the penalty factor and kernel function parameters are set to 9, 0.002.

The method was also validated using bearing 3-1, bearing 3-4, and bearing vibration data bearing 5 from the University of Cincinnati. The reliability curves for the three bearings are shown in Fig. 14. The errors produced by each prediction model are shown in Table II. From the above figure, it can be seen that the curves obtained by RTH-MKRVM are closer to the actual reliability curves and have good prediction performance compared to the existing bearing reliability prediction methods.

As can be seen in Fig. 14, the early operation of bearing 3-1 was relatively stable, bearing reliability did not change greatly, and there was a serious failure of the bearing at 250 points, and the reliability declined more rapidly. Bearings 3-4 run more smoothly before 300 points, and their reliability drops to 0 in the final stage. Bearing 5 has high reliability before the first 350 points, and after 350 points, reliability suddenly drops to 0.

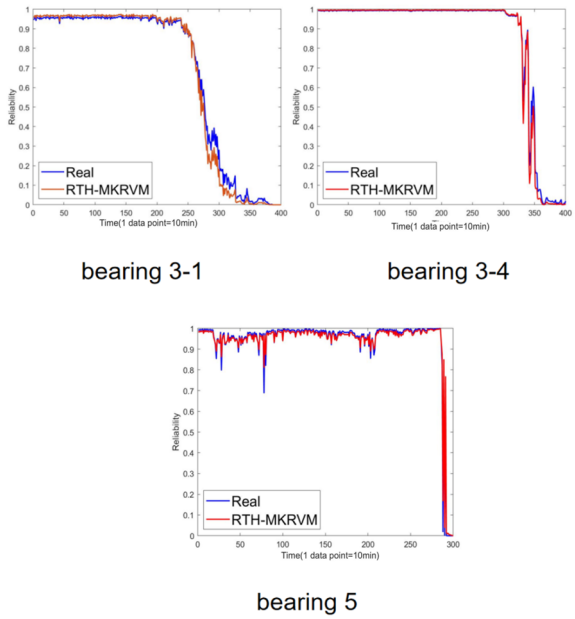


Fig. 14. Three bearing reliability prediction results.

VI. Conclusion

Aiming at the problem of assessing and predicting the reliability of rolling bearings based on noise conditions, a new noise reduction method is proposed to remove the excess noise in the bearing signals, and a logistic regression model is used to assess the reliability of the bearings and obtain the reliability results, and finally a hybrid kernel correlation vector machine model of red-tailed eagle is used to predict the reliability of the bearings, and good results are obtained, indicating that the study can well characterize the rolling bearings' state to characterize and predict the reliability.

TABLE II. Comparison of Different Models on Bearings

Model	Bearing 3-1		Bearing 3-2	
	MAE	MAPE	MAE	MAPE
ELMAN	0.073	0.082	0.032	0.045
LSSVM	0.069	0.074	0.061	0.074
RVM	0.079	0.094	0.135	0.207
MKRVM	0.057	0.041	0.096	0.037
RTH-MKRVM	0.045	0.028	0.058	0.044
Model	Bearing 3-4		Bearing 5	
	MAE	MAPE	MAE	MAPE
ELMAN	0.031	0.072	0.074	0.085
LSSVM	0.074	0.080	0.048	0.066
RVM	0.126	0.209	0.191	0.254
MKRVM	0.029	0.017	0.033	0.027
RTH-MKRVM	0.019	0.026	0.027	0.031

(1) The efficiency of signal noise reduction is greatly improved by combining GRCMSE for the selection problem of IMFs after CEEMDAN.

(2) Use UMAP algorithm to downscale the multidimensional features of the signal after noise reduction, and lay a solid data foundation for the reliability modeling of rolling bearings.

(3) For the problem of bearing reliability prediction, the MKRVM model optimized by the red-tailed eagle algorithm is proposed to predict the reliability of bearing signals, which significantly improves the prediction accuracy.

VII. Discussion

In this paper, the effectiveness of the method is verified by the bearing signal data in the laboratory. However, there are still many problems to be solved in the research of rolling bearings, and the future research direction can be started from the following points:

(1) In terms of bearing signal feature extraction, this paper constructs signal feature set by extracting signal time-frequency domain features, but the selection of signal features also relies on manual experience. In the future research, the selection of signal features can be further studied to make the selection process more automatic, and the selected features can be more accurate representation of the signal.

(2) The verification data used in this paper comes from the laboratory, the working conditions are stable, and the experiment is carried out in an ideal environment. However, the actual operating conditions are complicated, so the reliability assessment and prediction under variable speed conditions should be further studied.

References

[1] T. Andreas,P. Judith, K. Marcel and E. Gordon. "*Predictive maintenance enabled by machine learning: Use cases and challenges in the automotive industry*", Reliability Engineering & System Safety, vol. 215, pp. 107864-107864, 2021.

[2] P. S. Kumar, L. A. Kumaraswamidhas and S. K. Laha, "*Selection of efficient degradation features for rolling element bearing prognosis using Gaussian Process Regression method*", ISA Transactions, vol. 112, pp. 386-401, 2021.

[3] H. Soltanali, A. Rohani, M. Tabasizadeh, M. H. Abbaspour-Fard and A. Parida, "*Operational reliability evaluation based maintenance planning for automotive production line*", Quality Technology & Quantitative Management, vol. 17, no. 2, pp. 1-17, 2020.

[4] S. Bashir, I. M. Kolo, and A. O. S. J. F. "*Development of anomaly detector for motor bearing condition monitoring using fast fourier transform (fft) and long short term memory (lstm)-autoencoder*", -Manager s Journal on Pattern Recognition, vol. 10, no. 1, pp. 1-15, 2023.

[5] Z. Bao, G. Zhang, B. Xiong, et al., "*New image denoising algorithm using monogenic wavelet transform and improved deep convolutional neural network*", Multimedia Tools and Applications, vol. 79, no. 1, pp. 7401–7412, 2020.

[6] X. Zhang, Y. Qi and F. Liu "*Predicting Effects of Non-Point Source Pollution Emission Control Schemes Based on VMD-BiLSTM and MIKE21*", Environmental Modeling & Assessment, vol. 29, pp. 797-812, 2024.

[7] J. Guo, Y. Liu,J. Xiang . "*Rotating Machinery Fault Detection Using a New Version of Intrinsic Time-Scale Decomposition*", IEEE sensors journal, vol. 24, pp. 1905-1918, 2024.

[8] X. Cao, X. Guo, H. Duan , "*Health Status Recognition Method for Rotating Machinery Based on Multi-Scale Hybrid Features and Improved Convolutional Neural Networks*", sensors, vol. 12, pp. 5688, 2023.

[9] N. E. Huang, Z. Shen, S. R. Long, M. C. Wu, H. H. Shih, Q. Zheng, N.-C. Yen, C.-C. Tung and H. H. Liu, "*The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis*", Proceedings of the Royal Society A: Mathematical, Physical & Engineering Sciences, vol. 454, no. 1971, pp. 903-995, 1998.

[10] J. Zheng and H. Pan, "*Mean-optimized mode decomposition: An improved EMD approach for non-stationary signal processing*", ISA transactions, vol. 106,pp. 392-401, 2020.

[11] L. Zhao, Z. Li,Y. Pei, "*Disentangled Seasonal-Trend representation of improved CEEMD-GRU joint model with entropy-driven reconstruction to forecast signicant wave height*", Renewable Energy, vol. 226, pp.120345, 2024.

[12] M. E. Torres, M. A. Colominas, G. Schlotthauer and P. Flandrin, "*A complete ensemble empirical mode decomposition with adaptive noise*", IEEE International Conference on Acoustics, Speech and Signal Processing, pp. 4144–4147, 2011.

[13] Y. Cheng, Z. Wang, B. Chen, W. Zhang and G. Huang, "*An improved complementary ensemble empirical mode decomposition with adaptive noise and its application to rolling element bearing fault diagnosis*", ISA Transactions, vol. 19, pp. 218–234, 2019.

[14] A. Kala, S. Vaidyanathan and P. Femi, "*Ceemdan hybridized with lstm model for forecasting monthly rainfall*", Journal of Intelligent & Fuzzy Systems: Applications in Engineering and Technology, vol. 43, pp. 1–9, 2022.

[15] J. Zhang, Z. Li and J. Huang, "*Study on vibration-transmission-path identification method for hydropower houses based on ceemdan-svd-te*", Applied Sciences, vol. 12, pp. 7455, 2022.

[16] Y. Zhao and J. Xu, "*Denoising of ECG signals based on CEEMDAN*", 2021 6th International Conference on Intelligent Computing and Signal Processing (ICSP), pp. 430–433, 2022.

[17] M. Ding and X. Wang, "*Indirect prediction method for remaining useful life of lithium-ion battery based on gray wolf optimized extreme learning machine*", 2023 IEEE 12th Data Driven Control and Learning Systems Conference (DDCLS), pp. 301–306, 2023.

[18] M. Jürgen, H P. Gnser,W. Daves, "*Digitalization and Reliability of Railway Vehicles and Tracks—Condition Monitoring and Condition-based Maintenance*", BHM Berg- und Hüttenmännische Monatshefte, vol.169, pp.264-268, 2024.

[19] S. Gao, S. Zhang, Y. Zhang and Y. Gao, "*Operational reliability evaluation and prediction of rolling bearing based on isometric mapping and NOCUSa-LSSVM*", Reliability Engineering & System Safety, vol. 201, pp. 106968.1–106968.11, 2020.

[20] F. Abbasi, M. Naderan and S. E. Alavi, "*Anomaly detection in Internet of Things using feature selection and classification based on Logistic Regression and Artificial Neural Network on N-BaIoT dataset*", 2021 5th International Conference on Internet of Things and Applications (IoT), pp.1-7, 2021.

[21] Q. Li, M. J. Zuo and S. Y. Liang, "*False Lipschitz penalty sparse low-rank matrix and chaotic bionic optimization for prognosis of bearing degradation*", IEEE Transactions on Reliability, pp. 1–17, 2020.

[22] T. Han, J. Pang and A. C. Tan, "*Remaining useful life prediction of bearing based on stacked autoencoder and recurrent neural network*", Journal of Manufacturing Systems, vol. 21, pp. 576–591, 2021.

[23] Y. Wang, J. Zhao, C. Yang, D. Xu and J. Ge, "*Remaining useful life prediction of rolling bearings based on Pearson correlation-KPCA multi-feature fusion*", Measurement, vol. 20, pp. 111572, 2022.

[24] Z. Xu, M. Bashir, Q. Liu, Z. Miao, X. Wang, J. Wang and N. Ekere, "*A novel health indicator for intelligent prediction of rolling bearing remaining useful life based on unsupervised learning model*", Computers & Industrial Engineering, vol. 176, pp. 108999, 2023.

[25] B. Su and Y. Sun, "*Intelligent prediction of bearing remaining useful life based on data enhancement and adaptive temporal convolutional networks*", Journal of Failure Analysis and Prevention, vol. 23, pp. 2709–2720, 2023.

[26] J. S. Richman and J. R. Moorman, "*Physiological time-series analysis using approximate entropy and sample entropy*", American Journal of Physiology Heart & Circulatory Physiology, vol. 278, pp. H2039–H2049, 2000.

[27] B. Ghojogh, M. Crowley, F. Karray and A. Ghodsi, "*UMAP: Uniform manifold approximation and projection for dimension reduction*", The Journal of Open Source Software, pp. 479–497, 2023.

[28] J. Shlens, "*A tutorial on principal component analysis*", International Journal of Remote Sensing, vol. 51, 2014.

[29] D. R. Cox, "*Regression models and life-tables*", Journal of the Royal Statistical Society, vol. 34, pp. 527–541, 1972.

[30] S. Ferahtia, A. Houari, H. Rezk and A. Djerioui, "*Red-tailed hawk algorithm for numerical optimization and real-world problems*", Scientific Reports, vol. 13, pp. 12950, 2023.

[31] P. E. Tipping, "*Sparse Bayesian learning and the relevance vector machine*", Journal of Machine Learning Research, vol. 1, pp. 211–224, 2001.

EiAiMSPS: Edge Inspired Artificial Intelligence-based Multi Stakeholders Personalized Security Mechanism in iCPS for PCS

Swati Devliyal, Sachin Sharma, Himanshu Rai Goyal
Computer Science and Engineering, Graphic Era Deemed to be University, Dehradun, India

***Abstract*—Artificial Intelligence (AI) is becoming more prevalent in the healthcare sector like in pharmaceutical care to achieve rapid and precise outcomes. Machine learning techniques are critical in preserving this balance since they ensure both the confidentiality and authenticity of healthcare data. Early sickness projections benefit clinicians when establishing early monetary choices, in the lives of their patients. The Web of Things (IoT) is acting as an accelerator to boost the efficacy of AI applications in healthcare. Healthcare service pharmaceutical care is also in demand and can have AI for good patient care. The sensor gathers the data from individuals, then the data is examined employing machine learning algorithms. The work's major intent is to come up with an automated learning-based user authentication algorithm for providing secure communication. The other goal is to ensure data privacy for sensitive information that does not currently have security. The Federated Learning (FL) technique, which uses a decentralized environment to train models, can be utilized for this purpose. It enhances data privacy. This work proposes in addition to security a differential privacy preservation strategy that involves introducing random noise to a data sample to generate anonymity. The model's performance and data quality are assessed, as privacy preservation approaches frequently reduce data quality.**

Keywords*—Internet of things; pharmaceutical care; machine learning; authentication; artificial intelligence

I. Introduction

Over the last several years, the world of pharmacy has seen a steady and considerable transition. The prior work of the pharmacist, which included medicine production, dispensing, and marketing, is no longer sufficient for pharmacy professionals to still exist [1]. Pharmaceutical care has been widely acknowledged as the main objective of pharmacy. Pharmaceutical care requires practitioners to not only provide drugs but also to take responsibility for enhancing the quality of patients' outcomes [2]. In our assessment of research on the review of pharmaceutical care services, multiple publications highlighted the large beneficial impact that pharmaceutical care services have on long-term healthcare management and healthcare expenses [3]. A number of read ups have been conducted to inquire the implications of artificial intelligence (AI) arrangements on healthcare distribution. AI-powered solutions have the potential to improve forecasting, assessment, and care coordination. AI is anticipated to become more prevalent fundamental element of medical care in the years to come, with applications in a number of clinical settings [4]. As the outcome, various technological companies and government agencies have put money in the growth of clinical tools and medical applications. Patients may be among the most significant benefactors and users of AI-based apps, and their perspectives may have an impact on the broad adoption of AI-based technologies. Patients must be encouraged that AI-based technologies will not damage them, but rather that adopting AI technology for healthcare reasons will help them [5]. Although AI has the potential to enhance healthcare results, any issues and hazards should be addressed before it is integrated into normal clinical treatment. Furthermore, following earlier research, healthcare professionals still have basic concerns regarding the use of AI-based solutions in care services [6]. Researchers must more efficiently comprehend the existing issues associated with AI technologies and analyse the pressing demands of health systems in order to create AI-enabled solutions that can solve them. Technological advancement unleashes a maelstrom of communication and interconnection, allowing the intelligent pharmaceutical care in order to grow more flexible, sophisticated, and smart through the use of artificial intelligence. AI enables systems should naturally emphasis on enhanced analysis of data while maintaining appropriate user experience quality. Despite the fact that the association of AI-concentrated along with CPS considerably boosts productivity in pharmaceutical care, it is still in torment from challenges such as high burden, device incompatibility, security, and privacy [7] [8] [9]. CPS-based systems offer various additional issues, the most difficult of which is authentication. In pharmaceutical care there are so many stakeholders and major are practitioner, pharmacist and patient. When all are communicating through the network need authorisation at each end. We have developed an authentication approach that can be more resilient. In this article, we look at a unique security architecture for CPS that hosts user authentication and provides data security and privacy, device anonymity, and safety.

Our contributions are highlighted below:

- Based on edge assistance, we provide a layer skeleton in CPS. The higher layer is intended for registration management in conjunction with IIoT gadget. It decouples the need for direct interaction with IIoT devices and decreases system complexity. The middle surface is used in data transmission, while the lower surface houses the IIoT gadget.
- We present an authentication system that makes use of a proxy signing for establishing a link. It significantly minimizes the expenditure for signatures on gadgets and prevents unauthorized encounters in the outer limits, establishing the groundwork for protecting the

privacy on gadgets.

- The suggested scheme's security and performance assessments is demonstrating its robustness and practicability in contrast to earlier work. The rest of the sections are as follows: First is literature review which is followed by the proposed model. Further design in presented which focuses at security and privacy concerns, whereas last is performance analysis followed by conclusion.

II. Literature Review

A. Edge-AI in Pharmaceutical

According to the Thakur *et al.* [10] the use of AI in the field of pharmaceutical and biological studies has been significant, including cancer research, for prognosis and diagnosis of the disease state. It has evolved into a tool for researchers in charge of complicated data, covering everything from acquiring supportive results to normal statistical analyses. AI improves the accuracy of estimating treatment impact in cancer patients and decides forecast outcomes. Klumpp *et al.* [11] proposes a methodology for predicting the future based entirely on comprehensive analyses of trends by subject as well as interactive advancements. The findings suggest that the human aspect, as well as human-artificial collaboration skills and attitudes, might be a critical feature in AI and technology use in logistics. Damiati *et al.* [12] examines the historical, current, and future implications of machine learning technologies on several fields of pharmaceutical sciences, including drug design and exploration, revision, and composition. The strategies for researching systems that are often used in pharmaceutical sciences are explained.AI and system learning technology in ordinary everyday pharma demands, as well as commercial and regulatory insights, are examined. For unbalanced ICS data, Jahromi *et al.* [13] suggested a novel two-level ensemble deep learning-based attack detection and identification method. The whole bureaucratic model is a complicated DNN with a partially and entirely linked component that can appropriately blame cyber-attacks. Burki *et al.* [14] suggested obstacle might allow AI to be trained on millions of data points from various drug organisations' databases without jeopardising the possession and privacy of the statistics. Rathi *et al.* [15] provides a scalable, responsive, and dependable AI-enabled IoT and aspect computing-based healthcare system with minimal lag while servicing patients.

B. Stakeholder Authorization

Xu *et al.* [16] presents, an approval strategy based entirely on block chain is presented to identify genuine authorization for information access. The suggested approach divides info warehouses in block chains and HISs, with greater performance, more area-specific, dynamic, and bendy authorisation procedures. Hameed *et al.* [17] proposed, we provide a Block chain-based safe, decentralised, and customisable authorisation mechanism to grapple with the challenge of unauthorised access to IoT networking equipment. We implemented the ABAC version utilising intelligent agreements, which make the technique possible of authorising consumers with safe access to IoT devices to be accomplished largely based on dynamic and fine-grained policies maintained on the distributed immutable ledger. Using a permissioned blockchain community, this article presents a robust and accessible pharmaceutical supply chain gadget. Babu *et al.* [18] designs also includes digital transactions between providers and traders, tracking the source, ok verification, and lowering the risk of supply chains.The Hyperledger network fabric has been used in its deployment of this machine and its effectiveness has been assessed using Hyperledger Calliper. Zukarnain *et al.* [19] aimed to propose an aggregation of multi-component authentication that requires minimal user participation in this work. Because of security concerns, they implemented an unequal encryption technique in which the users' input is utilised as the encryption key. The PKI idea was adopted, yet without the required to communicate with a certificate authority (CA), the value was significantly reduced.

C. Intelligent CPS

Lu *et al.* [20] presents system of authentication for imposing security policies at the edge in CPS discourse for IIoT in order to enable reliable interaction for restricted gadgets.The main concept aims to integrate proxy authentication and process links at the ICN structure in order to provide two-way authentication. Security testing indicated that the suggested strategy provided a more effective protection than competing schemes. Ramasamy *et al.* [21] presents an AI-enabled IoT-CPS that doctors can use to diagnose ailments in patients. AI was created to assist with a variety of illnesses such as diabetes, heart disease, and gait problems. To detect illnesses in the class, the AI-enabled IoT-CPS Algorithm is used. Experiment findings reveal that, when compared to current methods, the proposed AI-enabled IoT-CPS algorithm diagnoses patient diseases and incident actions with more precision in terms of recall, accuracy, precision, and F-measure.Mishra *et al.* [22] developed deeply into novel new technologies such as the machine-to-machine communication, machine learning, artificial intelligence, Internet of Things, big data, and so on. An example NG-CPS structure is proposed, which includes all layout concerns such as physical layout components, cyber layout elements, and design conversations. Makkar *et al.* [23] presented cognitive-inspired architecture for CPS security is investigated. The suggested system, dubbed Secure CPS, is trained with immediate time collective dataset for determining the relevance of a web page using facial expressions as guides. The eye regions are identified using the Focal Point Detector method. The system was tested using device learning models and achieved 98.51% accuracy, outperforming existing frameworks. Adil *et al.* [24] proposed a hybrid light-weight authentication scheme that makes use of SML (supervised machine learning) method in conjunction with CPBE&D (Cryptographic Parameter Based Encryption and Decryption) scheme to ensure the authenticity of criminal patient wearable gadgets with consistent transmission over the Wi-Fi conversation channel.

D. Pharmaceutical Care Services

Alzahrani *et al.* [25] propose a novel TRD (Tag Reapplication Detection) method for detecting reapplication attacks, as well as the usage of low-cost NFC (Near Field Communication) tags and public key cryptography. Because a huge number of modern mobile phones are NFC-enabled, the inclusion of NFC makes TRD user-friendly. TRD uses an

online authentication system to track the number of times a tag in the database has been read delivery chain to detect reapplication attacks. Janardhan *et al.* [26] proposed a contrast the accuracy of the Decision Tree Classifier to the Support Vector Machine Classifier in detecting the authentication attacks. The SVM accuracy was 87.02%, P0.05, whereas the Decision Tree Classifier accuracy was 71.81%, P0.05. SVM performed substantially better in identifying de-authentication attacks.

E. AI Based Privacy Prevention

A lightweight stable encryption technique is developed in this work to preserve the privacy of sensitive data and communication. The scheme is developed by permutation, then with the help of a spread structure. The recombination uses pseudo-random sequences (PRNS), whereas the diffusion employs a key circulate generated (KSG). The algorithm is advantageous for CPS devices because to its simple and secure construction. The test results show that the proposed method is sufficiently robust and unquestionably able to withstanding any known prevention assaults as discussed by Tiwari *et al.* [27] and Lian *et al.* [28]. The possible impacted medical records breach will also cause concerns about security and confidentiality were raised throughout the contact period. We propose to fix these current concerns by DEEP-FEL, a decentralised, green, and privateness-better federated side learning device that enables clinical gadgets in a unique establishment to collectively teach an international framework without confidential data being shared. Zhang *et al.* [29] proposed a PEMFL architecture that Momentum FL (MFL), a chaos-based encryption approach and combines differential privacy (DP). The overall success of this methodology is based entirely on two non-datasets. The PEMFL performs exceptionally well in terms of accuracy and privacy protection, according to theoretical assessment and exploratory results.

III. Discussion

A. Gaps Identified from the Past Research

- Privacy Concerns: Previous research may have adopted basic privacy safeguards, but they failed to handle the difficulties of decentralized systems, exposing critical patient information.
- Inadequate Security Protocols: Other techniques may lack robust user authentication algorithms capable of securing connections in pharmaceutical treatment.
- Trade-offs Regarding Privacy and Data Quality: Highlight that present systems frequently sacrifice data quality to improve privacy, perhaps leading to less accurate healthcare results.

B. Emphasize the Urgency

As AI and IoT modern technology become increasingly woven into healthcare, particularly pharmaceutical treatment, the challenges connected with poor data privacy and security safeguards grow more pressing. Failure to solve these challenges might have serious ramifications, such as data breaches that endanger patient safety and weaken the legitimacy of AI-driven healthcare systems.

TABLE I. Notations Used for Security

Symbol	Description
id_{st}	Identity information of stakeholder
p_{st}	Partial private key
s_{st}	Private key
Q_{st}	Stakeholder Public Key
K	Security Parameters
x_{st}	Stakeholder secret value
m_w	Warrant
M	Message
RL_{st}	Repudiation check
$_{st}$	Starting a hash chain's significance of a st
t_{sti}	i-th timeinterval of stakholder st
tP_{cs}	i-th timeinterval of Server

IV. Problem Statements

A. Syllabary

Table I contains a collection of the syllabary and security assumptions used.

B. System Prototype

The iCPS controls give system performance, infrastructure at the highest point of our paradigm. In our research main focus is patient care. The major elements of it are practitioners which are liable for the system. When a patient need medication plan will be get from the doctor to pharmacist and pharmacist will verify the medication from the doctor and will handover it to patient. After that monitoring will be done by the pharmacist if any modification is required after monitoring then the medication plan will be changed by the pharmacist by taking consent from the doctor. And the things will be done in repetition until patient will be completely ok. The system model is divided into four cluster which are as follows:

- Pharmaceutical healthcare provider: It basically consists of doctors, hospitals, staff, etc. It is the main cluster with which patients contact directly. If patients need pc then will directly communicate with hospital staff. After that, in the background other clusters will communicate with each other.
- Pharmaceutical Distributors: It consists of the medical things distributer like wholesalers, retailers and medical representatives.
- Pharmaceutical manufacturing bodies: This cluster is a collection of drug manufacturers, raw material suppliers, investors and PBM's. PBM is pharmacy business management who is responsible for securing lower drug cost for insurance and insurance companies.
- Pharmaceutical government bodies: It consists of the principal of governance and regulatory agencies which are to establish, screen, and put in force standards of exercise to enhance the excellent of exercise so that registrants avoid: unsuitable behavior, professional misconduct for a registrant, and inept overall fulfillment of obligations.

The pharmaceutical care is basically followed by the interaction foundation is followed by the IIoT devices. The patient communicates with iCPS and iCPS will communicate with the four clusters and after the approval from all clusters

is achieved only communication will take place. Interested stakeholders who desire to connect with one another can be authorized, and following successful authentication, data packets can be accessed using a cryptographic sign. Edge devices are responsible for matching Interest and Data and then storing the relevant information for further requests. The proposed model has four categories: an iCPS server, IIoT gadgets (like actuators, sensors, and machines), stakeholders, and patients. Notably, the iCPS server is in charge of all category registration. Based on the verification findings, an intriguing data packet should be transmitted and destroyed. After that, approval is allowed to perform virtual sign and sign instead of the IIoT providers, such as unexpected inactivity, insufficient time, or computing. capability.

C. Network Model for Proposed Scheme

The network model is shown in Fig. 1. With the help of CPS an architecture is designed in which on bottom we have smart healthcare devices that collect the required information from the environment. After that, the data is analyzed using different latest technologies which is used with the help of an interface. On other hand, we have different stakeholders who want to use this filtered information. In our study, we have selected 11 stakeholders with the patients they are interacting with each other with the help of iCPS.

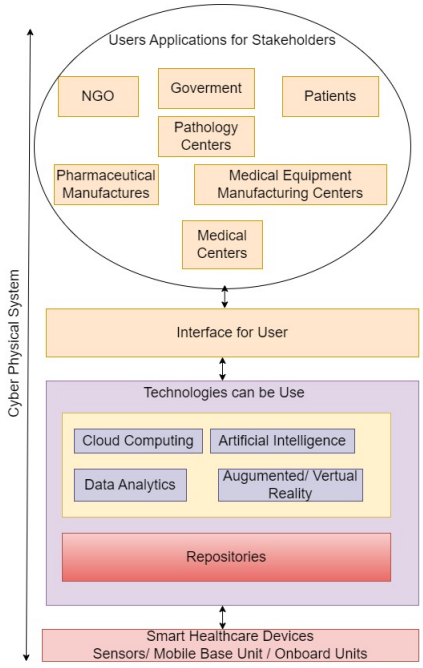


Fig. 1. Network model.

D. Threat Modal

We take into account both passive and aggressive attackers. Passive attacks are those that have amassed a deluge of Interest information to determine who is demanding and who is responding. Activated opponents, as opposed to inactive opponents, have greater power and may perform powerful attacks on any packets channeling, such as catching/exploring Interest packets, changing requests and responses, and spoofing authorized IIoT devices with the intention to transmit packets. According to the layout of framework, each one is needed to be register on the iCPS controller before if they want to communicate and want the system assets. We feel that the iCPS controller is inadequately powerful in order to render our design seem more plausible.

V. Edge-Assisted Intelligent User Authentication in CPS

Fig. 2. depicts the simple architecture of stakeholder authorization in CPS for pharmaceutical care.

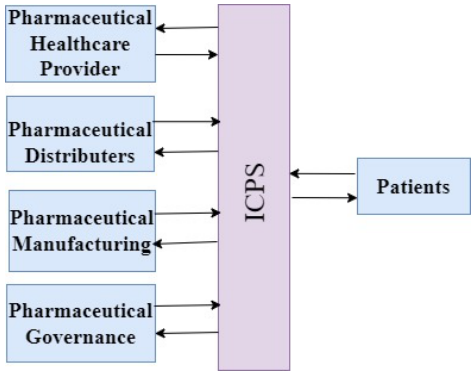


Fig. 2. Stakeholder authentication in cyber-physical system for pharmaceutical care.

To design authentication for IIoT devices, we use a proxy signature and a session-based variation. The proxy signature is used to validate the user, while the session-based variation is used to validate the request using Algorithm 1.

A. Overview

Based on edge assistance, our approach provides aid to do the requirement for CPS intelligent authentication procedures: 1) provides each user with the signing capability, allowing serving similar demands from multiple requesters; 2) allows users to authenticate themselves, allowing Only authorized individuals will receive the content they have requested; and 3) keeps IIoT gadgets unidentified, according to the authentication policy. The system paradigm is simplified, with a single iCPS server, practitioner, and patient. Six steps are included in an in-depth overview of the authentication operation. The start of the process is the identification of participants (intelligent users like practitioners, patients, etc.) and the intelligent server. The second step is to check the legitimacy of the user. The next phase is taken by a user who makes a content request, and it only sends the request while user is found valid in step two. The following phase is carried out on the info side to check the sign. The fifth step is performed to verify the sign if it is matched then only the communication will take place. The session handshake between the practitioner and the patient is the final phase. The practitioner is granted access to the required information once it has been authenticated in the last phase, as seen in Fig. 3.

B. Authentication Scheme

- (Registration) The system is started by the iCPS server, which broadcasts the system parameters var's.

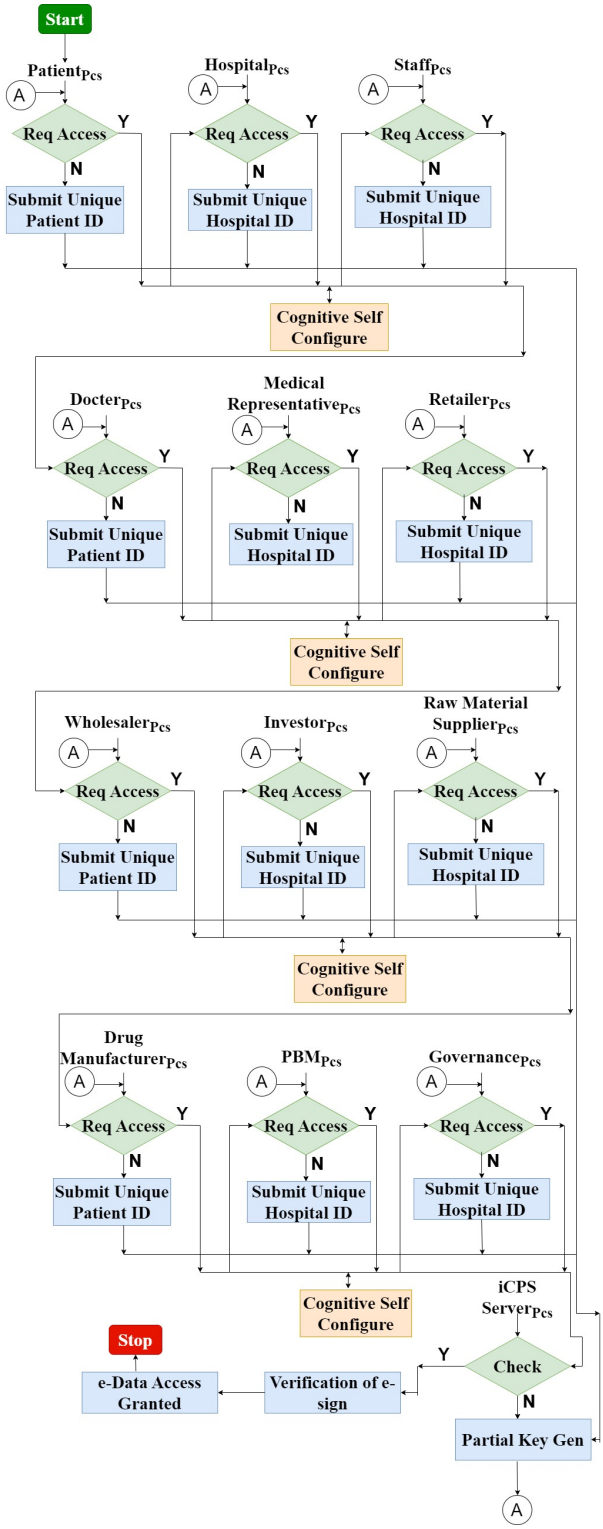


Fig. 3. Stakeholder authentication in cyber-physical system for pharmaceutical care.

Every stakeholder must identify themselves to the iCPS server and get a set of keys (Q_{st}, s_{st}), using the keygen to PartialKeyGen, and setup procedures outlined below. Configuration: The CPS configuration procedure basically is below mentioned. Partial key Generator: The iCPS produces the partial key for each stakeholder. The process takes var's, master key, a secret key x_{st} Z_q selected by stakeholder st and identification for stakeholder st with the value $id_{st} \leftarrow \{0,1\}$ as source. Every partially key is made up of two parts: the repudiation check RL_{st} and the partial secret key p_{st}. The repudiation check is an encrypted network that takes a starting phrase as input. The master-key hashes two identities, id_{st} and $x_{st}P$, to a point multiplier to create the partial secret key. That's accomplished using reliable means. The following are the specifications.

- Stakeholder st deliver ¡$idst, xstP$¿ to iCPS controller
- Make $_{st}\epsilon\{0,1\}$
- Determine
- $R_{evst1} \leftarrow h_5$ (ϵ_{st}),……., $Rev_{stk} \leftarrow h_5$ (Rev_{stk}-1)
- RL_{st1}<Rev_{st1}, t_{st1}>……, RL_{stk}<$Rev_{s_{tk}}, t_{s_{tk}}$> $t_{st} \leftarrow t_{st1}U……Ut_{s_{tk}}$
- RL_{st} RL_{st1}, U…….U t_{stk}
- $D_{st} \leftarrow$ h1(id_{st},x_{stP})
- $P_{st} \leftarrow sD_{st}$
- Key Generator: The iCPS takes var's, the partial secret key p_{st}, and the certified value x_{st} as input as outputs the secret key s_{st} along with public key Q_{st}. A p_{st} and x_{st} repository consists of an entity's entire secret key st and its public key equivalent $x_{st}P$. The operation is carried out by st, who is the only legitimate proprietor of x_{st}.

- (Repudiation Check): If an iCPS rejects a stakeholder, all stakeholders have to exit the system. As a result, the controller distributes a repudiation list to all stakeholders regularly to determine if a stakeholder has been repudiated. The detailed procedure is as follows: Repudiation Check: The repudiation checks are performed at each stakeholder's entering and ensure that a stakeholder is associated to the repudiation index prior to the session connection. It requires var's present period of time t and the repudiation list RL_{st} as variables. The iCPS may validate the consumer's repudiation evidence after investigating the truthfulness of the repudiation variables in the repudiation list. It invalidates anything in the request produced after the time $t_{st}k$ since the timestamp $t_{st}k$ rarely in t_{st}.
- (Interest): Stakeholder introduces a cognitive self-configure to check to set all the parameters so that can communicate further. Each stakeholder can set its parameters. Step 4(Verification and sign)- This is done on the iCPS edge and receives the secret value x_{st} as input, as well as a warrant m_w that comprises the repudiation time frame, message m, and identification information idst, the public key of the iCPS, and all stakeholders Q_{iCPS}, $Q_{st\mathbf{P}pcs}$, $Q_{st\mathbf{H}pcs}$, $Q_{st\mathbf{S}pcs}$, $Q_{st\mathbf{D}pcs}$, $Q_{st\mathbf{MR}pcs}$, $Q_{st\mathbf{R}pcs}$, $Q_{st\mathbf{W}pcs}$, $Q_{st\mathbf{I}pcs}$, $Q_{st\mathbf{RM}pcs}$, $Q_{st\mathbf{DM}pcs}$, $Q_{st\mathbf{PBM}pcs}$, $Q_{st\mathbf{G}pcs}$. The iCPS verifies its rights as follows:

Algorithm 1 Algorithm Stakeholders Verification and Sign

Input: var's, s_{st}, $\mathbf{PK}_{st}$, proxy
Output:Success 0: Fail

Calculate **H2**← h2(id_{iCPS}, $id_{st\mathbf{P}pcs}$, $id_{st\mathbf{H}pcs}$, $id_{st\mathbf{S}pcs}$, $id_{st\mathbf{D}pcs}$, $id_{st\mathbf{MR}pcs}$, $id_{st\mathbf{R}pcs}$, $id_{st\mathbf{W}pcs}$, $id_{st\mathbf{I}pcs}$, $id_{st\mathbf{RM}pcs}$, $id_{st\mathbf{DM}pcs}$, $id_{st\mathbf{PBM}pcs}$, $id_{st\mathbf{G}pcs}$, m_w, Q_{iCPS}, $Q_{st\mathbf{P}pcs}$, $Q_{st\mathbf{H}pcs}$, $Q_{st\mathbf{S}pcs}$, $Q_{st\mathbf{D}pcs}$, $Q_{st\mathbf{MR}pcs}$, $Q_{st\mathbf{R}pcs}$, $Q_{st\mathbf{W}pcs}$, $Q_{st\mathbf{I}pcs}$, $Q_{st\mathbf{RM}pcs}$, $Q_{st\mathbf{DM}pcs}$, $Q_{st\mathbf{PBM}pcs}$, $Q_{st\mathbf{G}pcs}$
if Verify if e(ϵ,Q_{icps}+**H2**$_{icps}$)= e(Q_{icps},D_{icps}) **then**
  Set r ϵ Z_q
  Compute **R**←rP, H3← h_3(m,id_{icps},R,Q_{icps}) **V**← p_{icps} + rH_3+ H_2Q_{icps} + $x_{icps}H_2$P ð←(R,V)
  send 1
else
  send 0
end if

- (Generation of Proxy Signs and Authentication): If the check is successful,the iCPS will get a proxy signing key pair (s_{icps}, $\mathbf{PK}_{icps}$),where PK_{icps} is a collection of public keys. (Q_{iCPS}, $Q_{st\mathbf{P}pcs}$,$Q_{st\mathbf{H}pcs}$, $Q_{st\mathbf{S}pcs}$,$Q_{st\mathbf{D}pcs}$,$Q_{st\mathbf{MR}pcs}$, $Q_{st\mathbf{R}pcs}$,$Q_{st\mathbf{W}pcs}$,$Q_{st\mathbf{I}pcs}$, $Q_{st\mathbf{RM}pcs}$,$Q_{st\mathbf{DM}pcs}$,$Q_{st\mathbf{PBM}pcs}$, $Q_{st\mathbf{G}pcs}$).If this is the case, the iCPS generates a digital signature from a message. A signature is not misleading the provider's identity to the public. It accepts the signature by checking ð is a valid identity that involves message m;else,it denies it.
- (Session Connection): When patient receives sign makes a request by addressing call including four information of a subject, subject name, signature, identity id and subject which is used by iCPS to identify. Upon receiving the request each stakeholder has to authenticate the request. The iCPS gets the access key Q_{st} and identification id, and private key s_{st} . The iCPS does the verification, and if the check succeeds, the current session connection gets established; otherwise, the request isn't deemed valid, and the session connection is terminated.
- Safety Examination: The suggested technique guarantees that the CPS controller obtains the partial secret keys in a unique manner, preventing it from impersonating a real organization. We examine the authentication system in light of the security objectives given below. To keep attackers at bay, we have identified the following authentication mechanism security goals.
- Trustworthiness: The system should ensure that the person signing cannot be untruthful to an information packet.
- Genuineness: The technique should offer evidence that a signed request is valid.
- Authentication: The method should include a way for authenticating that is the broadcaster an Interest and responding to what it intends to be [30].

- Anonymity: The technique should safeguard IoT devices by ensuring that both inner and external assailants are unaware of their identities.
- Key Organisation: The system should offer a negotiated key among all the stakeholders, so that no one controls the key.
- Authenticity: A polynomial-time adversary has the knack of forging a signature assigned to authorised entity in order to prevent the organisation from denying it.

Theorem 1: A polynomial-time challenge exists that may fix the Computational Diffie–Hellman issue along likelihood $\varepsilon(\mathbb{k})^{'}>(\varepsilon(\mathbb{k})/2)\ (1-q_s(q_{h3}+q_s)/2^{\mathbb{k}})(e(q_r+1))^{-}1$ is contingent upon whether or not the opponent £ can forge a sign along with a competitive edge $\varepsilon(\mathbb{k})$, where q_{h3}, q_s, q_r indicate the total quantity of requests made to the h_3, executing, as well as reveal-partial key predictions, providing that $h_i(i=1,2,3)$ hashed routines are arbitrary diviners.

Demonstration: Assume that $X \leftarrow a\rho, Y \leftarrow b\rho \in G1G1$ indicate an arbitrary task.Using the forger £,we can create the algorithmic programme $\in$ to produce $ab\rho \leftarrow G1$.Algorithm $\in$ first generates system-specific var's, as the standard protocol does, and passes var's to the forger £, which then initialises £ with $Q_0 \leftarrow X$ and communicates with £ as below.

1) h1 and h3 Queries: If examines an arbitrary prophet h1(h3) using a number of tuples $<id_{st}, Q_{st}>(<m_{st}, id_{st}, R_{st}, Q_{st}>)$, preserves an index $Lh_1(Lh_3)$ of components $<id_{st}, Q_{st}, x_{st}, c_{st}, \nu_{st}>(<m_{st}, id_{st}, R_{st}, Q_{st}, y_{st}, d_{st}, v_{st}>)$, It is initially empty and yields the following outcomes.

a) Since the inquery $<id_{st}, Q_{st}>(<m_{st}, id_{st}, R_{st}, Q_{st}, Q_{st}>)$ is existing in $Lh_1(Lh_3), \in outputsv_{st}(\nu_{st})$ to .

b) Else, $\in$ picks $x_{st} \in Z*q(y \in Z^*q)$ arbitrary, yields $\nu_{st} \leftarrow x_{st}\rho(\nu_{st} \leftarrow y_{st}Q_0)$ if a tossup $c_{st} \leftarrow 0,1(d_{st} \leftarrow 0,1)$ which gives 0 as a result with likeliness $\sigma(1/2)$ and yields $\nu_{st} \leftarrow x_{st}y(\nu_{st} \leftarrow y_{st}\rho)$ if $c_{st}=1(d_{st}=1)$ having a likelihood $1-\sigma(1/2)$, and adds $(<id_{st}, Q_{st}, x_{st}, c_{st}, \nu_{st}>)$ $(<m_{st}, id_{st}, R_{st}, Q_{st}, y_{st}, d_{st}, \nu_{st}>)$ into $L_{h1}(L_{h3})$.

2) h2 Queries: While asks an arbitrary generator h_{st} with a list of T_{st} tuples, $\in$ keeps the matching list $L_{h_st} \leftarrow <T_{st}, \mu_{st}>$. While the value of the input field is located in the list Lh_{st}, it sends $\in$ the matching element μ_{st}. If not, delivers a random $Z*q$ value.

3) RevealPartialKey Queries: After £ queries the identity id_{st}, $\in$ obtains the matching element $<id_{st}, Q_{st}, x_{st}, c_{st}, \nu_{st}>$ from the list of items L_{h1} and responds as follows.

a) If $c_{st}=1$, then $\in$ outputs and the experiment is aborted.

b) Otherwise, sets $p_{st} \leftarrow x_{st}Q_0$ and restores it to .

4) RequestPublicKey Queries: As £ queries identify id_{st}, $\in$ sets $Q_{st} \leftarrow x_{st}\rho$ for an arbitrary value $x_{st} \leftarrow Z*q$, sends it to , and inserts $<id_{st}, x_{st}>$ to $L\rho K$.

5) Signing Queries: demands an id_{st} verification on an exchange m_{st}, and $\in$ replicates the divination verification and replies in response to the probe.

a) Assuming h_3 is not supplied along with $<m_{st}, id_j, R, Q_j>$, the process continues by replying to h1 requests to get $H_2 \leftarrow \mu_2$ and setting $R \leftarrow r_2Q_0, H_3r_2^{-1}(x_1\rho \leftarrow D_j)$, and $V \leftarrow r_1Q_0+\mu_2Q_i+\mu 2Q_j$ with an arbitrary selected $r_1, r_2 \leftarrow Z*q$. Otherwise, $\in$ will stop and forsake. Because L_{h3} can never have a total of qh_3+q_s entries, the likelihood of not terminating $1-(q_s(qh_3+q_s)/2^k)$.

b) $\in$ yields $\Delta \leftarrow <V, R>$ as the acceptable signing on m.

It is worth noting that opponent $\in$ correctly generates a signature δ with likelihood $\varepsilon(k)^{'}$, which means that £ completely meets the Signing answers. Finally, £ creates a fake sign $\iota* \leftarrow <V^*, \mathrm{R}^*>m^*$ upon a message. After it $\in$ gets the elements $<id^*_{st}, Q^*_{st}, x^*_{st}, c^*_{st}, \nu^*_{st}>$ derived from Lh_1. If $c^*_{st}=0$, the program fails and exits. Otherwise, it proceeds to retrieve the pair $<id^*_{st}, Q^*_{st}, Q^*_j>$ based on the set Lh_2 and $m_{st}, <m^*_{st}, d^*_{st}, \nu^*_{st}, id^*_{st}, R^*_{st}, Q^*_{st}, y^*_{st}>$ based on the record Lh_3. Assuming that $d^*_{st}=0$, then C responds 0 and exits. Otherwise, it will do this: $\ddot{e}(V*,P)=\ddot{e}(y^*_{st}P, R^*_{st})$. $\ddot{e}(Q_0, D^*_{st})\ddot{e}(Q^*_{st}, \mu^*_{st}P)\cdot\ \ddot{e}(\mu^*_{st}P, Q^*_j)$ with $\hbar 1=x^*_{st}y, \hbar 2=\mu^*_{st}, \hbar 3=y^*_{st}$ P along with $R^*_{st}=r^*_{st}$ P for components which are referred as $\mu^*_{st} \leftarrow Z^*_q, r^*_{st}, x^*_{st}$. $(1-\delta)/2$ is the probability of not terminating at particular moment. Henceforth $\ddot{e}(V^*-y^*_{st}R^*_{st}-\mu^*_{st}Q^*_{st}-\mu^*_{st}Q^*_j, P)=\ddot{e}(X, x^*_{st}Y)$

VI. Problem Analysis for Security

Using MATLAB a comparative analysis has been done to check the suggested proposal's effectiveness mechanism with the existing one. The comparison is done with communication and computation consumption under the number of requests with existing schemes [34], [32], [33] and [31] as demonstrated in Fig. 4(a)-(b) along with Fig. 5(a)-(b). In addition, Fig. 4(c) authorization and acknowledge consumption with the amount of request contents. To find out in a better way we have measured the cost of authorization concerning a number of requests made in Fig. 4(b) which depicts it as a linear correlation. In our method there is a slightly heavy burden to other methods [31], [32], [33], and [34]. But in contrast, our method has a better authentication which reflects more reliability. In Fig. 5(b) efficiency of our method as compared to other four methods in terms of patients requests. As it increases the computation cost also increases but not in a drastic way. In Fig. 4(a), the validation execution efficiency of various systems improves when the request quantities grow from 0 to 100. Our method has a somewhat greater inspection processing proportion compared to [30], [28], but lower than [27], and [29]. In our system, verifying a signature requires acquiring all packets containing the delegation information used to calculate it. Our scheme achieves a verification execution efficiency of less than 20% for 50 contents, making it suitable for latency-tolerant uses in the IIoT. Fig. 4(b) highlights our investigation of the transmission load as the content of the request increases. An apparent pattern indicates that the transmission load is linearly related to the quantity of requested contents. Our system has a somewhat higher load compared to [27], [28], and [29], but lower than [34]. Our method offers two-way verification in addition to caching, which sets it apart from other schemes. In Fig. 4(c), we examine the relationship between communication cost, verification versus simultaneous dissemination, and quantity of demanded data to assess the

overhead of communication robustness in the proposed approach. The method of delivery cost grows with the amount of requested material, whereas verification consumes less than the interaction procedure. Raising the quantity of required material from 50 to 100 only leads to a ten percent rise in signature size. Though each signature packet must be transferred from the supplier to the user, the proposed approach is resource-efficient and does not create significant overhead associated with communication modifications. Fig. 5(a) shows that the capacity usage ratio grows with the amount of demanded content. The information usage ratio in [28], [29], and our system is less than forty percent for one hundred packets of requested material, unlike [29] and [30]. Our technique consumes little computing resources on the consumer side and outperforms other systems in terms of safety. Fig. 5(b) shows that the provider's delay increases with the quantity of demanded contents. It uses somewhat more computing resources than [27] and [26], but the growth rate will be slower than [29] and [30]. Compared to the transmission load indicated in Fig. 4(b), our technique requires less time. So to conclude we can say that authentication time improves as the request are getting increased.

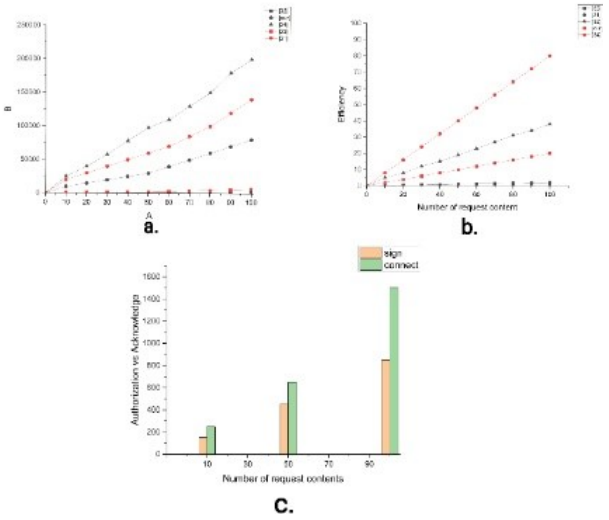
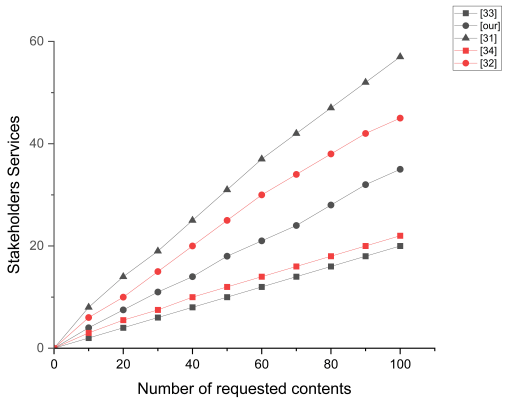
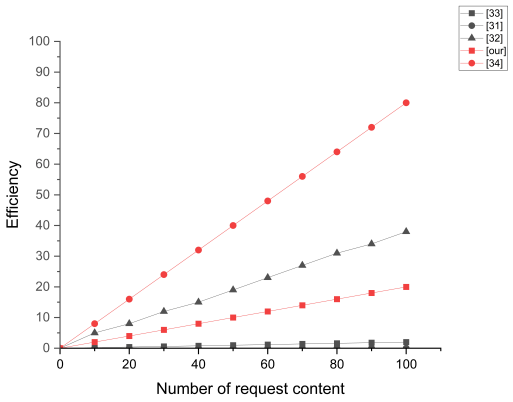


Fig. 4. Estimation value differentiation with a varying number of requests a) Cost for several authentication methods b) The cost of transmission for various methods of authentication c) Authorization price versus acknowledge price.



(a)



(b)

Fig. 5. Comparative analysis of the ability to provide services to stakeholders with varying number of requests a) Stakeholder service cost under different authorization methods b) Efficiency of different authorization methods.

TABLE II. NOTATIONS FOR PRIVACY

Symbol	Description
P_u	Universal prototype
P_i	Localized prototype
P_i^{reform}	Updated universal prototype
S_i	Stakeholders
I_i	Local input
I_i^n	local input with noise
I_i^{np}	Upgrade local input

VII. PROPOSED MODEL FOR PRIVACY

In this paper edge based technique is used to provide privacy in user authentication. In start every stakeholder and patient will download the present universal prototype from the iCPS. The downloaded universal model will act as localized for each edge node. Firstly, noise has been added to the input set to upgrade it. After that, the model has been trained using the classifier. After that training input set is transferred to iCPS. This is done in several repetitions. The updated universal prototype is created using aggregation of different types of prototypes as shown in Fig. 6. Table II represents different notation used while designing the algorithm. Stakeholders using Algorithm 2. There are eleven stakeholders represented by Si where i varies from one to eleven. If any stakeholder has a localized update(Ii), then universal prototype Pu is downloaded from the iCPS. If the universal prototype is not same as the existing one, then only the procedure will start. The universal prototype will act a localized prototype (Pi) for the stakeholder. A noise has been added input set and stored into as I_i^n and upgraded scaled form as I_i^{np}.The updated model is created and transferred to iCPS as P_i^{reform}. The same patient is using Algorithm 3.

Data: Data for all stakeholder's will be private and will be represented by I_i.Correlation Coefficient added with Noise: It describes the association between the features and the target. The value of this varies from 0 to 1. This value denotes the

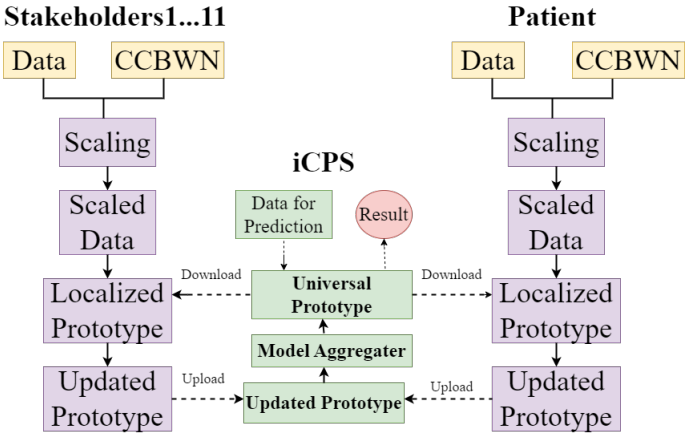


Fig. 6. Block diagram of proposed model.

correlation low value means low correlation and high value means high correlation. Value 0 denotes no correlation.

$$ran = \frac{n(\Sigma ab) - (\Sigma a)(\Sigma b)}{(n\Sigma a^2 - (\Sigma a^2)(n\Sigma b^2 - (\Sigma b^2)} \quad (1)$$

$$a_i^n = a_i + Rand(0, ran) \quad (2)$$

Where,i varies from 0 to l,n is an integer value,a_i is original, a_i^n is after appending noise,Rand(0,ran) is used to generate random numbers between 0 to ran,ran is correlation coefficient of a_i which is mentioned in Eq. 1. Noise is appending using Eq. 2 after that the updated value of input will be I_1^n to I_{11}^n. iCPS using Algorithm 4. iCPS will choose the stakeholder's (Si: i=0 to 11) for training and give response to stakeholder by giving upgraded prototype.

Algorithm 2 Stakeholder's Algorithm

Claim: P_u **received from the validator**
Guarantee:$P_u \neq P_i$
if stakeholder process () **then**
if Localizedreformaccessible **then**
Set $P_i \leftarrow P_u$
$I_i \leftarrow$ Localizedreform
$I_i^n \leftarrow$ Add noise with I_i
$I_i^{np} \leftarrow$ upgrade I_i^n
$P_i^{reform} \leftarrow$ Trainprototype(P_i,I_i^{np})
Transfer P_i^{reform} to the validator
end if
end if

VIII. Problem Analysis for Privacy

The dataset used in this study was taken from Kaggle and contains a large no. of instances approximately 2000 with 8 attributes. Attributes are drug-uses, patient symptoms, gender, disease prevention, design therapist plan, age, implementing the therapeutic plan and monitoring therapeutic plan and correlation coefficient values are mentioned in Table III. The class that is targeted has a value of 0 or 1. In the start, iCPS circulates the attributes to the localized model to the randomly chosen patient and stakeholders for the training

Algorithm 3 Patient's Algorithm

Claim: P_u **received from the validator**
Guarantee:$P_u \neq$ **Patient**
if patient process() **then**
if Localizedreformaccessible **then**
Set Patient$\leftarrow P_u$
$I_i \leftarrow$ Localizedreform
$I_i^n \leftarrow$ Add noise with I_i
$I_i^{np} \leftarrow$ upgrade I_i^n
$Patient^{reform} \leftarrow$ Trainprototype(Patient,I_i^{np})
Transfer $Patient^{reform}$ to the validator
end if
end if

Algorithm 4 iCPS's Algorithm

Claim: P_i^{reform} from stakeholder's and patient
Guarantee: $P_i^{reform} \neq P_i$ i =1 to 11
Method iCPS()
for each Repetition, r $\in$ Rj: j=1$\rightarrow$m do
Choose 11 stakholder's S_1 to S_n and Patient
for every Stakeholder,s$\in$Si:i=1 $\rightarrow$ 11 in parallel do **do**
transfer Pu to Si
for Patient , p$\in$ $P_{patient}$
transfer $Patient_u$ to Patient
end for
for every Stakeholder,s$\in S_i$:i=1$\rightarrow$11 in parallel do **do**
P[i]$\leftarrow P_i^{reform}$
for Patient,p$\in P_{patient}$
$Patient_u \leftarrow Patient^reform$
end for
$P_u \leftarrow$ ProtypeiCPS(P) transfer P_u to all stakeholders and to the patient

point of view. the stakeholders and patient begins the training just after receiving the universal prototype and saving it as a localized prototype. Stakeholders and patients transfer the updated prototype to iCPS. The iCPS accumulate the prototype and transfer the updated prototype again to selected in further rounds till better accuracy is gained. Different classifiers are used for the study Decision tree as CL1, KNeighbours as CL2, Gaussian as CL3 and Randomforest as CL4, and the outcome is determined by the test score. which is calculated by using Eq. 3 and it shows that correction when the test's score rises.The prototype efficiency improves as the MSE value lowers, with the ideal model having a value of 0. Correlation grows while the R2 score improves. Table IV shows the comparison test scores for CL1, CL2, CL3, and CL4 by utilizing several techniques Gaussian Noise(GN) and Correlation Coefficient based model with Noise(CCBMWN).The test score using CL1 are 0.7205 and 0.7953 using CL2 are 0.7952 and 0.8067 using CL3 0.7678 and 0.8211 using CL4 are 0.8042 and 0.8812. The efficiency of the model is measured based on accuracy, specificity, and sensitivity. Results show that CCBMWN performs better than GN. So our method is better.

$$TestScore = \frac{x + y}{x + y + z + \in} \quad (3)$$

Where x,y,z, $\in$ are True positive(mean actual and predi-

TABLE III. CORRELATION COEFFICIENT VALUES FOR DIFFERENT FEATURES

Symbol	Description
Gender	0.05
Age	0.73
drug-uses	0.62
patient symptoms	0.58
disease prevention	0.65
design therapist plan	0.76
implementing therapeutic plan	0.66
monitoring therapeutic plan	0.56

cated value both are same as 1), True negative(with mean real and projected values are both 0), False positive (mean actual and predicated value both are different actual is 1 and predicted as 0)and False negative (mean actual and predicated value both are different actual is 0 and predicted as 1), respectively. The correlation coefficient value lies between -1 to +1. Table V displays the correlation value of the coefficient for each characteristic. The efficiency of the model is measured on the basis of accuracy, specificity, and sensitivity. Sensitivity is actual positive denoted by (λ) and calculated by using Eq. 4 and specificity is false actual negative denoted by (m_K) and calculated by using Eq. 5. ROC curve is used to show different thresholds by plotting and graphically.

$$\lambda = \frac{x+y}{x+\in} \tag{4}$$

$$m_K = \frac{y}{y+z} \tag{5}$$

$$AUC = \int_l^h f(x)dx \tag{6}$$

In above equation, the ROC is defined as Y=f(X), while both m and n are the curve's limit values. Fig. 7 is the curve for GN and Fig. 8 is for CCBMWN.

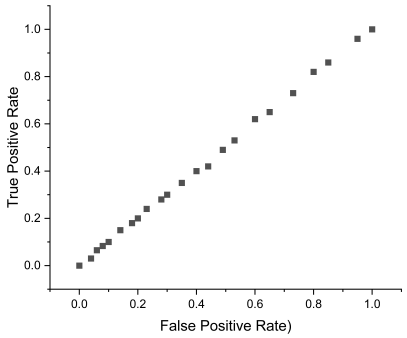


Fig. 7. ROC for GN.

The AUC (area under the ROC curve) is cognizance, received from ROC. It gives a clear picture of which technique is doing better. It is calculated using Eq. 6. In our research it is found that GN is 0.5032 and CCBMWN is 0.5108 which is better in CCBMWN. The comparison is shown in Fig. 9. The model's fulfillment is assessed using sensitivity, accuracy,

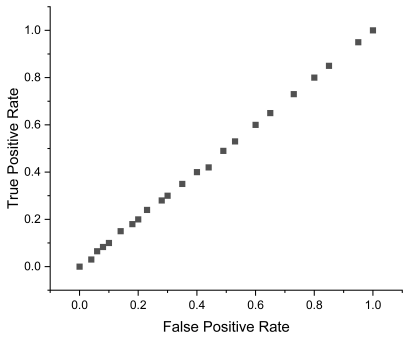


Fig. 8. ROC for CCBMWN.

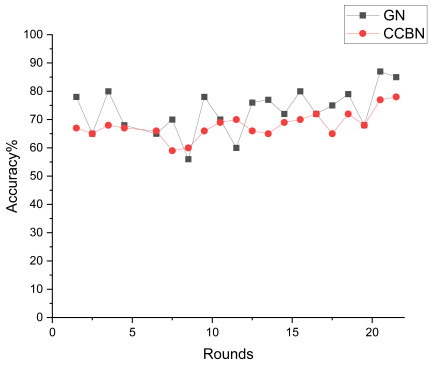


Fig. 9. Comparison of test score.

and specificity. A higher accuracy number indicates excellent accuracy, a higher sensitivity value indicates good prediction of genuine positive, and a higher specificity value indicates good prediction of true negative, as demonstrated in Fig. 10 and Fig. 11. In the first round, the sensitivity is 0.754 and at last, it is 0.82. The specificity is 0.65 in the first round and 0.67 in last round. The accuracy is 71% but the aggregate is 76.

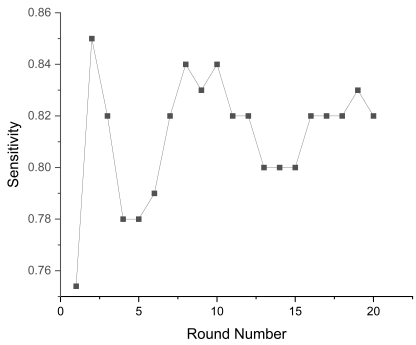


Fig. 10. Round wise sensitivity.

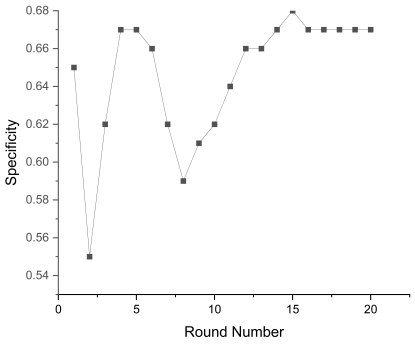


Fig. 11. Round wise specificity.

TABLE IV. DIFFERENT RESULT OF GN AND CCBMWN

Table Head	Test Scores	
	GN	*CCBMWN*
CL1	0.7205	0.7953
CL27	0.7952	0.8067
CL3	0.7678	0.8211
CL4	0.8042	0.8812

TABLE V. CORRELATION COEFFICIENT MATRIX OF FEATURES

	gender	age	drug-uses	patient symptoms	disease prevention	design therapist plan	implementing therapeutic plan	monitoring therapeutic plan
Gender	1	0.03	0.032	0.01	0.052	0.043	0.024	0.002
Age	0.03	1	0.62	0.35	0.56	0.23	0.05	0.34
drug-uses	0.032	0.62	1	0.52	0.43	0.10	0.45	0.52
patient symptoms	0.01	0.35	0.52	1	0.34	0.43	0.45	0.23
disease prevention	0.052	0.56	0.43	0.34	1	0.32	0.63	0.59
design therapist plan	0.043	0.23	0.10	0.43	0.32	1	0.66	0.56
implementing therapeutic plan	0.024	0.05	0.45	0.45	0.63	0.66	1	0.55
monitoring therapeutic plan	0.002	0.30	0.52	0.23	0.59	0.56	0.55	1

IX. CONCLUSION

In this study, an architecture has been proposed for user authentication in pharmaceutical care services. It can be used for secure communication whenever a patient wants to communicate and wants to avail of pharmaceutical care services. In problem analysis, it is observed that as compared to other methods the proposed method is strong. In this, the estimation and transmission value was analyzed. For providing privacy we have proposed a new technique CCBMWN which ensures privacy and results show that proposed method is giving good performance in contrast with existing methods. This study not only addresses the critical requirement for safe and confidential communication in AI-powered pharmaceutical treatment, but it also lays the groundwork for future advances in digitising healthcare operations and explicitly defining stakeholder responsibilities.

REFERENCES

[1] Kilincer, Ilhan Firat, Fatih Ertam, Abdulkadir Sengur, Ru-San Tan, and U. Rajendra Acharya. "Automated detection of cybersecurity attacks in healthcare systems with recursive feature elimination and multilayer perceptron optimization." Biocybernetics and Biomedical Engineering 43, no. 1 (2023): 30-41.

[2] Hepler, Charles D., and Linda M. Strand. "Opportunities and responsibilities in pharmaceutical care." American journal of hospital pharmacy 47, no. 3 (1990): 533-543.

[3] Kakhi, Kourosh, Roohallah Alizadehsani, HM Dipu Kabir, Abbas Khosravi, Saeid Nahavandi, and U. Rajendra Acharya. "The internet of medical things and artificial intelligence: trends, challenges, and opportunities." Biocybernetics and Biomedical Engineering 42, no. 3 (2022): 749-771.

[4] Jiao, Ying, Huamei Qi, and Jia Wu. "Capsule network assisted electrocardiogram classification model for smart healthcare." Biocybernetics and Biomedical Engineering 42, no. 2 (2022): 543-555.

[5] Jalali, Jalal, Ata Khalili, Atefeh Rezaei, Rafael Berkvens, Maarten Weyn, and Jeroen Famaey. "IRS-Based Energy Efficiency and Admission Control Maximization for IoT Users With Short Packet Lengths." IEEE Transactions on Vehicular Technology (2023).

[6] Turja, Tuuli, Iina Aaltonen, Sakari Taipale, and Atte Oksanen. "Robot acceptance model for care (RAM-care): A principled approach to the intention to use care robots." Information & Management 57, no. 5 (2020): 103220.

[7] K. Huang, C. Zhou, Y.-C. Tian, S. Yang, and Y. Qin, "Assessing the physical impact of cyberattacks on industrial cyber-physical systems," IEEE Trans. Ind. Electron., vol. 65, no. 10, pp. 8153–8162, Oct. 2018.

[8] A. Karati, S. K. H. Islam, G. P. Biswas, M. Z. A. Bhuiyan, P. Vijayakumar, and M. Karuppiah, "Provably secure identity-based signcryption scheme for crowdsourced Industrial Internet of Things environments," IEEE Internet Things J., vol. 5, no. 4, pp. 2904–2914, Aug. 2018.

[9] X. Chen, X. Huang, J. Li, J. Ma, W. Lou, and D. S. Wong, "New algorithms for secure outsourcing of large-scale systems of linear equations," IEEE Trans. Inf. Forensics Security, vol. 10, pp. 69–78, 2015.

[10] Thakur, Abhimanyu, Ambika P. Mishra, Bishnupriya Panda, Diana Rodríguez, Isha Gaurav, and Babita Majhi. "Application of artificial intelligence in pharmaceutical and biomedical studies." Current pharmaceutical design 26, no. 29 (2020): 3569-3578.

[11] Klumpp, Matthias. "Innovation potentials and pathways merging AI, CPS, and IoT." Applied System Innovation 1, no. 1 (2018): 5.

[12] Damiati, Safa A. "Digital pharmaceutical sciences." AAPS PharmSciTech 21, no. 6 (2020): 206.

[13] Jahromi, Amir Namavar, Hadis Karimipour, Ali Dehghantanha, and Kim-Kwang Raymond Choo. "Toward detection and attribution of cyber-attacks in IoT-enabled cyber–physical systems." IEEE Internet of Things Journal 8, no. 17 (2021): 13712-13722.

[14] Burki, Talha. "Pharma blockchains AI for drug development." The Lancet 393, no. 10189 (2019): 2382.

[15] Rathi, Vipin Kumar, Nikhil Kumar Rajput, Shubham Mishra, Bhavya Ahuja Grover, Prayag Tiwari, Amit Kumar Jaiswal, and M. Shamim Hossain. "An edge AI-enabled IoT healthcare monitoring system for smart cities." Computers & Electrical Engineering 96 (2021): 107524.

[16] Xu, Boyi, Li Da Xu, Yuxiao Wang, and Hongming Cai. "A distributed dynamic authorisation method for Internet+ medical & healthcare data access based on consortium blockchain." Enterprise Information Systems 16, no. 12 (2022):1922757.

[17] Hameed, Khizar, Ali Raza, Saurabh Garg, and Muhammad Bilal Amin. "A Blockchain-based Decentralised and Dynamic Authorisation Scheme for the Internet of Things." arXiv preprint arXiv:2208.07060 (2022).

[18] Babu, Erukala Suresh, Ilaiah Kavati, Soumya Ranjan Nayak, Uttam Ghosh, and Waleed Al Numay. "Secure and transparent pharmaceutical supply chain using permissioned blockchain network." International Journal of Logistics Research and Applications (2022): 1-28.

[19] Zukarnain, Zuriati Ahmad, Amgad Muneer, and Mohd Khairulanuar Ab Aziz. "Authentication securing methods for mobile identity: Issues, solutions and challenges." Symmetry 14, no. 4 (2022): 821.

[20] Lu, Yanrong, Ding Wang, Mohammad S. Obaidat, and Pandi Vijayakumar. "Edge-assisted intelligent device authentication in cyber-physical systems." IEEE Internet of Things Journal (2022).

[21] Ramasamy, Lakshmana Kumar, Firoz Khan, Mohammad Shah, Balusupati Veera Venkata Siva Prasad, Celestine Iwendi, and Cresantus Biamba. "Secure smart wearable computing through artificial intelligence-enabled internet of things and cyber-physical systems for health monitoring." Sensors 22, no. 3 (2022): 1076.

[22] Mishra, Ayaskanta, Amitkumar V. Jha, Bhargav Appasani, Arun Kumar Ray, Deepak Kumar Gupta, and Abu Nasar Ghazali. "Emerging technologies and design aspects of next generation cyber physical system with a smart city application perspective." International Journal of System Assurance Engineering and Management (2022): 1-23.

[23] Makkar, Aaisha, and Jong Hyuk Park. "SecureCPS: Cognitive inspired framework for detection of cyber attacks in cyber–physical systems." Information processing & management 59, no. 3 (2022): 102914.

[24] Adil, Muhammad, Muhammad Khurram Khan, Muhammad Mohsin Jadoon, Muhammad Attique, Houbing Song, and Ahmed Farouk. "An AI-enabled hybrid lightweight Authentication scheme for intelligent IoMT based cyber-physical systems." IEEE Transactions on Network Science and Engineering (2022).

[25] Alzahrani, Naif, and Nirupama Bulusu. "Securing pharmaceutical and high-value products against tag reapplication attacks using nfc tags." In 2016 IEEE International Conference on Smart Computing (SMARTCOMP), pp. 1-6. IEEE, 2016.

[26] Janardhan, B., and P. Jagadeesh. "Accurate Deauthentication Attack Detection using Linear Discriminant Analysis in Comparison with Multilayer Perceptron." Journal of Pharmaceutical Negative Results (2022): 1764-1771.

[27] Tiwari, Devisha, Bhaskar Mondal, Sunil Kumar Singh, and Deepika Koundal. "Lightweight encryption for privacy protection of data transmission in cyber physical systems." Cluster Computing 26, no. 4 (2023): 2351-2365.

[28] Lian, Zhuotao, Qinglin Yang, Weizheng Wang, Qingkui Zeng, Mamoun Alazab, Hong Zhao, and Chunhua Su. "DEEP-FEL: Decentralized, efficient and privacy-enhanced federated edge learning for healthcare cyber physical systems." IEEE Transactions on Network Science and Engineering 9, no. 5 (2022): 3558-3569.

[29] Zhang, Zehui, Linlin Zhang, Qingdan Li, Kunshu Wang, Ningxin He, and Tiegang Gao. "Privacy-enhanced momentum federated learning via differential privacy and chaotic system in industrial Cyber–Physical systems." ISA transactions 128 (2022): 17-31.

[30] D. Wang, X. Zhang, Z. Zhang, and P. Wang, "Understanding security failures of multi-factor authentication schemes for multi-server environments," Comput. Security, vol. 88, Jan. 2020, Art. no. 101619.

[31] Q. Zheng, Q. Li, A. Azgin, and J. Weng, "Data verification in information-centric networking with efficient revocable certificateless signature," in Proc. IEEE Conf. Commun. Netw. Security (CNS), 2017, pp. 1–9.

[32] K. Xue, X. Zhang, Q. Xia, D. S. Wei, H. Yue, and F. Wu, "SEAF: A secure, efficient and accountable access control framework for information centric networking," in Proc. IEEE Conf. Comput. Commun. (INFOCOM), 2018, pp. 2213–2221.

[33] I. O. Nunes and G. Tsudik, "KRB-CCN: Lightweight authentication and access control for private content-centric networks," in Proc. 16th Int. Conf. Appl. Cryptogr. Netw. Security (ACNS), vol. 10892, 2018, pp. 598–615.

[34] T. Mick, R. Tourani, and S. Misra, "LASeR: Lightweight authentication and secured routing for NDN IoT in smart cities," IEEE Internet Things J., vol. 5, no. 2, pp. 755–764, Apr. 2018.

Integrated IoT-Driven System with Fuzzy Logic and V2X Communication for Real-Time Speed Monitoring and Accident Prevention in Urban Traffic

Khadiza Tul Kubra[1], Tajim Md. Niamat Ullah Akhund[2]*, Waleed M. Al-Nuwaiser[3]*,
Md Assaduzzaman[4], Md. Suhag Ali[5], M. Mesbahuddin Sarker[6]
Institute of Information Technology, Jahangirnagar University, Dhaka, Bangladesh[1,6]
Department of Computer Science and Engineering, Daffodil International University, Dhaka 1216, Bangladesh[2,4]
Graduate School of Science and Engineering, Saga University, Saga, 8408502, Japan[2]
Computer Science Department, College of Computer and Information Sciences,
Imam Mohammad Ibn Saud Islamic University (IMSIU), Riyadh, Saudi Arabia[3]
Department of Software Engineering, Daffodil International University, Dhaka 1216, Bangladesh[5]

***Abstract*—Road safety is a critical concern globally, with speeding being a leading cause of traffic accidents. Leveraging advanced technologies can significantly enhance the ability to monitor and control vehicle speeds in real time. Traditional methods of speed monitoring are often limited in their ability to provide real-time, adaptive interventions. Existing systems do not adequately integrate sensor data and decision-making processes to prevent speeding-related accidents effectively. This paper aims to address these limitations by proposing a novel system that utilizes Internet of Things (IoT) technology combined with fuzzy logic to monitor vehicle speeds and prevent accidents in real time. The proposed system integrates IoT sensors for continuous vehicle speed monitoring and employs a Fuzzy Inference System (FIS) to make decisions based on variables such as speed, alcohol presence, and driver fitness. The system also facilitates interaction between drivers and law enforcement through Vehicle-to-Everything (V2X) communication. The FIS implementation demonstrated effective speed control capabilities, accurately assessing and responding to various risk levels, thereby reducing the likelihood of speeding-related accidents. This research contributes to the advancement of road safety systems by integrating IoT and fuzzy logic technologies, offering a more adaptive and responsive approach to traffic management and accident prevention. Future enhancements will focus on incorporating machine learning techniques to dynamically adjust FIS rules based on real-time data and improve sensor network reliability to ensure more accurate and comprehensive monitoring.**

***Keywords*—Internet of Things; high-speed monitoring; alcohol detection; Matlab simulation; write fuzzy inference system**

I. Introduction

The Internet of Things (IoT) represents a collection of smart technologies and connected devices that are highly intelligent. IoT enables data exchange over a network without requiring human-to-human or human-to-computer interactions. This technology has made our daily tasks easier and smarter. IoT has been applied in various fields, such as electricity, gas, water, and transportation. In Bangladesh, the majority of the population relies heavily on road transportation. Currently, traffic accidents are a significant problem, with high-speed driving being the primary cause.

The alarming rate of traffic accidents in Bangladesh, primarily due to high-speed driving, underscores the need for an effective solution. Implementing a system to monitor and control vehicle speeds can significantly reduce the incidence of road accidents. Our motivation stems from the desire to enhance road safety and protect lives by leveraging IoT technology.

The implementation of a vehicle speed monitoring and control system can have profound social impacts. By reducing the number of traffic accidents, we can save lives and prevent injuries, thereby improving the overall well-being of society. Moreover, safer roads contribute to a more efficient transportation system, which can have positive economic implications.

The primary objectives of this study are:

- To develop an IoT-based system for monitoring and controlling vehicle speeds.
- To reduce the rate of traffic accidents caused by high-speed driving.
- To enhance the communication between vehicles and control authorities for prompt intervention.

The statistics of fatalities in road accidents in Bangladesh are shown in Table I. To address the high rate of traffic accidents, a fixed vehicle speed should be enforced on the highways of Bangladesh. By controlling vehicle speeds, it is possible to reduce the accident rate. In this study, we have developed an IoT-based system to monitor and control vehicle speeds effectively. Often, vehicles exceed speed limits at the start of their journeys, which is a primary cause of road accidents. Our device monitors vehicle speeds and sends alerts to both the nearest police control room and the driver when overspeeding is detected. The mortality rate of road accidents in the last 8 years in Bangladesh (2015-2022) is shown in Fig. 1.

Each vehicle is equipped with a smart card/board that has a unique identification number. The device sends a message to the nearest police control room with the smart card/board

*Corresponding authors.

TABLE I. STATISTICS OF FATALITIES IN ROAD ACCIDENTS (SOURCE: DAILY NEWSPAPERS OF BANGLADESH)

Sl. No.	Year	Number of Fatalities	Reason	Type of Vehicle
1	2015	6823	Excessive speed	Bus and Truck
2	2016	3412	Hazardous overtaking	Motorcycle and Van
3	2017	4284	Careless driving	Microbus and Rickshaw
4	2018	7221	Poor road construction	Motorbike and Jeep
5	2019	7855	Unfit vehicles	Truck and Covered Van
6	2020	5431	Pedestrian carelessness	Truck and Pickup Van
7	2021	6284	Drunk driving	Minibus and Microbus
8	2022	4587	Lack of footpaths	Pickup and Van

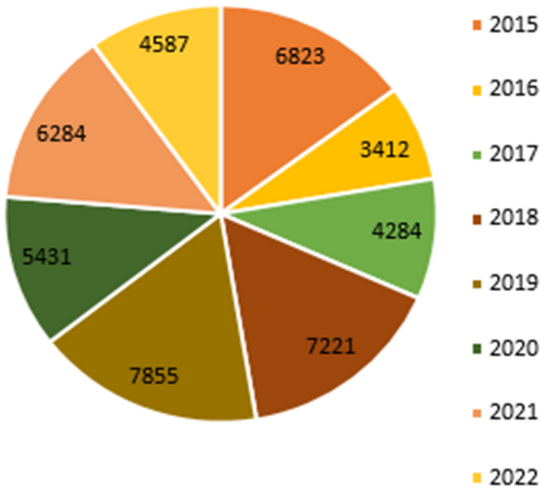


Fig. 1. Mortality rate of road accidents in the last 8 years in Bangladesh (2015-2022).

number when a vehicle is speeding. Simultaneously, it alerts the driver about the overspeeding. These devices are internet-enabled, allowing for monitoring and control.

The prevalence of road accidents globally, especially in developing countries like Bangladesh, is a critical concern. Bangladesh ranks among the highest in terms of accident rates. The primary cause of fatalities in accidents is the lack of emergency assistance. Timely intervention can save lives, making response time crucial. The vision of the Internet of Things (IoT) has rapidly expanded into new computing domains. IoT impacts human life more effectively and enhances functionality [1], [2].

The rest of the paper is organized as Section II discusses the previous works, Section III discusses the methodology, Section IV shows the implementation, Section V shows the results and discussion and finally Section VI shows the concluding remarks.

II. LITERATURE REVIEW

A. Related Works

Kishorkumar C. S. et al. proposed an intelligent car speed control system using Arduino and speed sensors. G. Kirankumar et al. developed an intelligent vehicle system for speed cameras using RuBee. Ravi Kishore Kodali et al. designed a system that enables traffic authorities to monitor all vehicles from the control room itself [3]. Sumit Deshpande et al. created a system that records the speed of vehicles and notifies relevant authorities of violations without human intervention [4]. P. Saichaitanya et al. proposed a car velocity tracking device using an accelerometer. Akriti et al. discussed the trade-offs in accidental control devices, including high costs and non-portability [5]. Chatrapathi et al. developed an efficient routing algorithm for ambulances [6]. Raut and Sachdev proposed a notification device using the XBee WiFi module, XBee Shield, and GPS module [7]. Ali and Alwan presented a device to detect low and high-speed vehicle crashes [8]. Aishwarya S. R. developed an approach that considered driver inconvenience [9]. Koneti S. et al. provided a solution for accidents caused by drunk driving [10]. Pratiksha R. et al. designed a device that detects accidents and monitors the car engine condition [11]. Kishwer K. et al. proposed an incident detection approach using sensors and hardware [12]. Namrata H. et al. implemented a push-button switch for boundary detection and microcontroller triggering [13]. Yadav et al. identified accidents and reported the reasons to registered numbers [14]. Reddy and Rao developed a system to detect vehicle fires and other failures [15]. Kavya and Geetha proposed a method to reduce delays caused by ambulances [16]. P. A. Targe and M. P. Satone developed a real-time algorithm using VANET communication [17]. Poorani K. et al. suggested using image processing techniques to monitor drivers [18]. Kim and Jeong developed an algorithm to detect crashes using crash probability data [19]. Patel K. H. et al. created software to detect accidents using accelerometer sensors [20]. Sonali N. and Maheshwari R. addressed obstacle detection in motorcycles with an intelligent approach. Chris T. and White J. designed a smartphone-based incident detection and notification system [21]. Prabha and Sunitha developed an automatic detection and messaging device for traffic accidents using a GSM modem and GPS [22]. Elie Nasr and Elie Kfoury proposed an IoT method for detecting, reporting, and navigating street accidents, suggesting a rescue system by reporting the location [23]. Vijay Savania and Hardik Agravata created a system for preventing accidents using vehicular alcohol detection sensors [24]. Syedul Amin and Jubayer Jalil proposed an incident detection and reporting device using GPS, GPRS, and GSM technology [25]. Suvarnanandyal et al. suggested a smart parking framework using an IoT module [26]. This review also integrates several recent papers that further emphasize the use of IoT in various applications, including health screening [27], robotic arm control [28], e-health databases [29], highway monitoring [30], Parkinson's prediction [31], and the impact of COVID-19 on education [32]. Additional contributions include low-cost robots for disabled individuals [33], smart farming [34], medical robotics [35], human activity recognition [36] and renewable energy generation [37]. The reviewed literature demonstrates various innovative approaches to vehicle speed control and accident detection using IoT and related

technologies. Each study contributes uniquely to enhancing road safety, reducing accident rates, and improving emergency response times. By incorporating these technologies, significant advancements can be achieved in the monitoring and management of traffic, ultimately leading to safer roads and reduced fatalities. The collective insights from these studies highlight the versatility and potential of IoT in transforming various sectors and improving overall quality of life.

B. Research Gap

In reviewing the existing literature, it is evident that while numerous studies have explored the use of IoT and fuzzy logic in traffic management, these approaches often fall short of addressing the complexity and variability of real-world driving conditions. For instance, several works focus on isolated aspects such as monitoring vehicle speed or detecting specific events like accidents or traffic violations through sensor data and GPS/GSM modules. However, these systems tend to operate within predefined parameters, lacking the flexibility to adapt to the dynamic nature of road environments. Furthermore, many studies fail to integrate a comprehensive decision-making framework that considers multiple risk factors simultaneously, such as speed, driver condition, and environmental context. This narrow focus limits their applicability to broader traffic safety scenarios. Additionally, most existing systems do not facilitate real-time interaction between drivers and law enforcement, which is crucial for timely interventions. This research gap highlights the need for a more holistic approach that combines continuous monitoring, adaptive decision-making, and seamless communication to enhance road safety comprehensively. The proposed study seeks to fill this gap by developing an integrated IoT and fuzzy logic-based system that not only monitors vehicle speed in real-time but also incorporates additional risk factors and enables proactive communication between drivers and authorities, thus providing a more robust solution to prevent speeding-related accidents.

C. Novelty of the Proposed System and Comparison with Existing Systems

The proposed system offers several novel features compared to existing systems:

- Integration of fuzzy logic for real-time speed monitoring and decision-making.
- The ability of drivers and police to interact with the system, enhances cooperative safety measures.
- Use of V2X communication for comprehensive traffic management.

Existing systems often focus solely on vehicle speed or accident detection using conventional methods [38], [39], [40], [41], [42], [43]. Our approach enhances these systems by incorporating fuzzy logic and enabling real-time interaction between drivers and law enforcement. Table II shows a Comparison of the Proposed System with Existing Systems.

TABLE II. COMPARISON OF PROPOSED SYSTEM WITH EXISTING SYSTEMS

System	Key Features	Limitations
[38]	Accident detection using speed sensors	Lacks real-time interaction
[39]	GPS and GSM-based monitoring	Limited to vehicle speed
[40]	Traffic management system	No fuzzy logic integration
[41]	Intelligent transport systems	Limited V2X communication
[42]	V2V communication for safety	No driver-police interaction
[43]	Road condition monitoring	Not focused on speed control
Proposed System	Fuzzy logic, V2X, driver-police interaction	Needs further integration

III. METHODOLOGY

The significance of defining a research problem lies in addressing a gap in the literature. The parameters considered in previous work include monitoring gas leaks through sensors, using GPS and GSM modules for communicating vehicle speed, and recording vehicle positions via sensors. However, for accident detection, only the speed monitored by the sensor was considered [38]. We believe this algorithm is more useful for specific intersections than for general traffic accidents. Therefore, the existing work needs to be modified to support road traffic crashes more broadly [39]. In our research, both drivers and police can use this system. They can sign into the system; drivers can monitor and control high speed, while police can detect overspeeding, track current locations, and notify the police control room. Both can receive message notifications from the system.

A. Proposed System Architecture

In this system, the driver signs in and monitors the car's speed. If high speed is detected, a message notification is sent automatically, prompting the driver to control the speed. The driver then signs out of the system. Similarly, police can sign in, detect high speed, receive SMS alerts for high speed, track the current location, and inform the police control room before signing out. The proposed system architecture is shown in Fig. 2.

The flow chart of research work stages and the proposed model for high speed is shown in Fig. 3 and Fig. 4. The following steps outline the flow chart of the research work stages and the proposed system model for high speed.

B. Fuzzy Logic Description

Fuzzy logic is a method of computing based on multivalued logic rather than the standard Boolean "true or false" (1 or 0) logic on which modern computing is based. Fuzzy logic relies on human perception and is used in situations where available information is partially true, making decision-making complex. Fuzzy Logic Controller (FLC) is used here for making decisions on critical events based on sensor data and Vehicle-to-Everything (V2X) communication.

There are some basic units of Fuzzy Inference System (FIS):

- Fuzzification Unit
- Knowledge-Based Rules

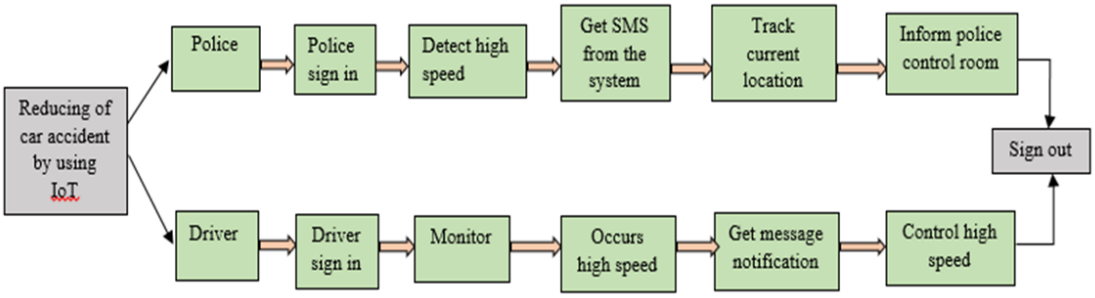


Fig. 2. Proposed system architecture.

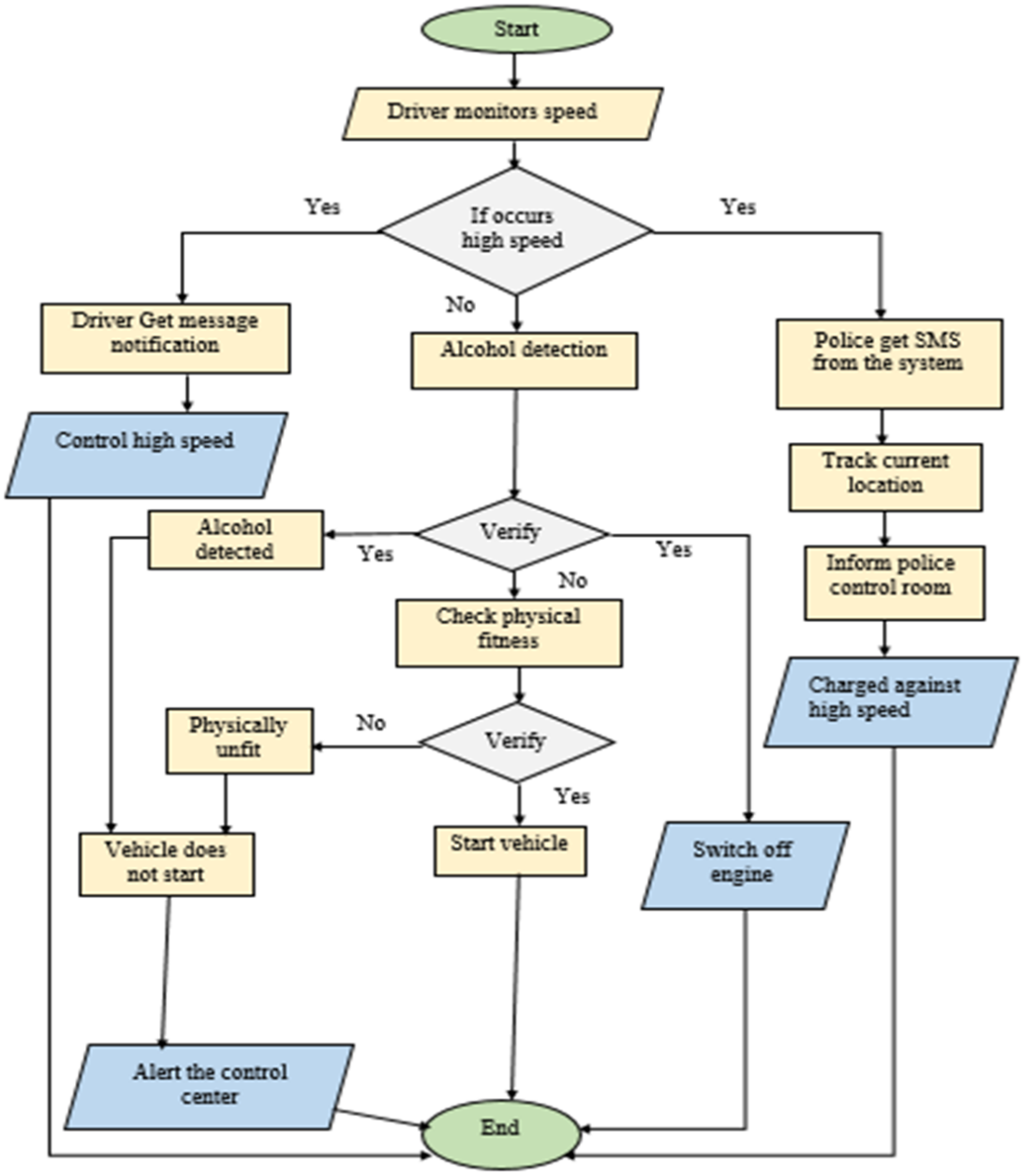


Fig. 3. Flow chart for work stages.

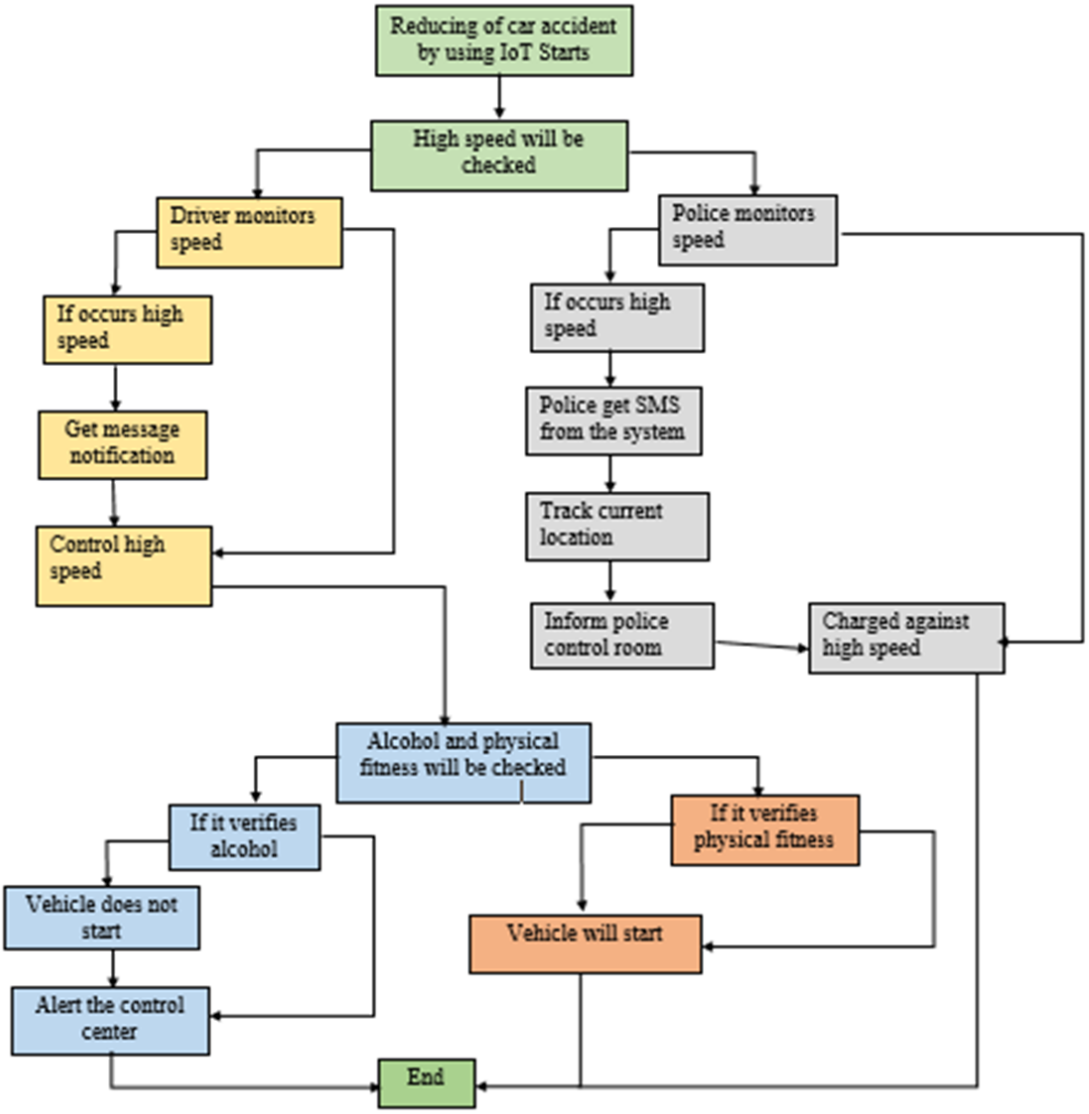


Fig. 4. High speed model.

- Decision or Controller Unit
- Defuzzification Unit

The block diagram of the FIS, shown in Fig. 5, illustrates these four basic units.

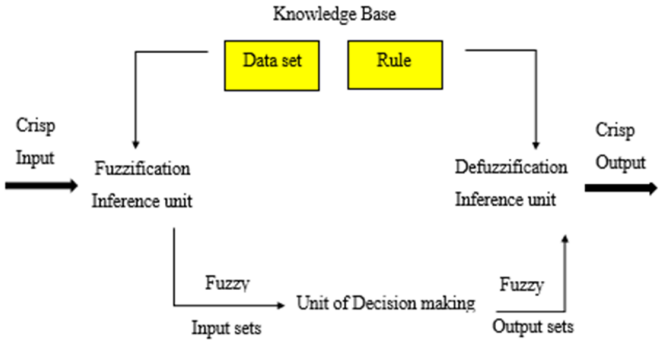


Fig. 5. Block diagram of FIS.

C. Mathematical Formulation

Let's consider the speed of the vehicle v, monitored in real-time. The speed threshold for high speed is denoted as v_{th}. If $v > v_{th}$, an alert is triggered. The alert system can be mathematically represented as:

$$A(v) = \begin{cases} 1 & \text{if } v > v_{th} \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

The position of the vehicle, $P(t)$, at time t is tracked using GPS coordinates $(x(t), y(t))$. The rate of change of position, which indicates the speed, is given by:

$$v(t) = \sqrt{\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2} \tag{2}$$

If $v(t) > v_{th}$, a notification N is sent to both the driver and the police:

$$N = \begin{cases} 1 & \text{if } v(t) > v_{th} \\ 0 & \text{otherwise} \end{cases} \tag{3}$$

Using fuzzy logic, the decision-making process involves fuzzification, applying knowledge-based rules, and defuzzification. The speed $v(t)$ is fuzzified into linguistic variables like "low", "medium", and "high". The rule-based system then decides the action based on these variables. For instance:

- If speed is "high" and location is "urban area", then issue a severe alert.
- If speed is "medium" and location is "highway", then issue a moderate alert.

Defuzzification converts these fuzzy outputs into a crisp value, which determines the exact nature of the alert.

These mathematical formulations and fuzzy logic rules ensure that the system is capable of making nuanced decisions based on real-time data, improving both road safety and the efficiency of emergency responses.

IV. Implementation

This section provides a detailed explanation of the configurations used to complete this research and ensure its functionality. It gives an overview of the software implementation and describes the process of reducing car accidents using MATLAB.

A. Software Implementation

The system is implemented using MATLAB, specifically employing fuzzy logic. A Fuzzy Inference System (FIS) is utilized to achieve the desired functionality. The software components used in the system are:

- MATLAB
- Fuzzy Inference System (FIS)

B. Reducing Car Accidents with High Speed Control in MATLAB

The system is implemented using a Fuzzy Inference System (FIS) in MATLAB. A FIS can be loaded from a .fis file using the `readfis` function. To save a FIS to a file, the `writeFIS` function is used. The `evalfis` function evaluates the FIS for given input values and returns the resulting output values. It evaluates the FIS using predefined evaluation options. The algorithm used in MATLAB is shown in Algorithm 1.

Algorithm 1 Fuzzy Traffic Control

```
f ← readfis('VehicleProtection.fis')
a ← input('Vehicle speed (km/h): ')
g ← evalfis([a, 0, 1], f)
if g < 1 then
    disp('Message to driver is sent')
    disp('Message to police with current location')
    disp('Control room informed')
    disp('Driver punished for high speed')
    b ← input('Is alcohol present (0:no, 1:yes): ')
    c ← input('Is the driver physically fit? (0:no, 1:yes): ')
    g ← evalfis([0, b, c], f)
    if g < 1 then
        disp('Other factors need to be checked')
    else
        disp('Other factors are okay')
    end if
else
    disp('The vehicle is good to go')
end if
```

C. Emergency Speed Control Warning

In this system, the user can observe different speed values. If the speed values are below 65 km/h, the system will indicate that the vehicle is good to go. If the speed is exactly 65 km/h, the system will provide different messages based on various parameters. If the speed exceeds 65 km/h, the system will also provide messages based on the parameters. Table III presents the analysis of speed, alcohol presence, and physical fitness.

TABLE III. Speed, Alcohol, and Physical Fitness Analysis

Speed (km/h)	Alcoholic	Physical Fitness	Message
30	×	×	The vehicle is good to go
50	×	×	The vehicle is good to go
65	0	1	Other factors are okay
65	1	1	Other factors need to be checked
65	0	0	Other factors need to be checked
90	1	0	Other factors need to be checked

V. Results and Discussion

This section discusses the results obtained from our Fuzzy Inference System (FIS) implementation and the performance analysis. The FIS uses fuzzy variables containing sets, where each set has an associated membership function. The membership function defines the form of the sentence and is used to "fuzzify" the values of the variable by associating a degree of membership (DOM) with a value. Sharpe's triangular membership functions are employed in this system. Fig. 6 shows the Mamdani FIS.

A. Obtained Features

The key features obtained from the FIS include:

- Three input variables for speed, alcohol presence, and physical fitness.
- Input and output mapping functions to process and evaluate the data.

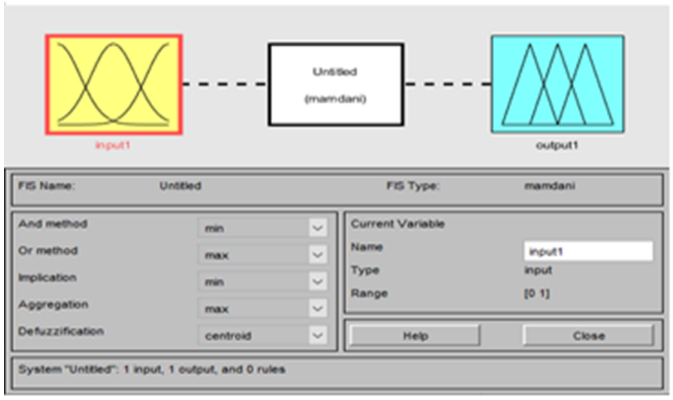


Fig. 6. Mamdani FIS.

- A rule base to determine the actions based on input conditions.
- Surface plots to visualize the relationship between inputs and outputs.

B. Performance Analysis of FIS

1) Three input variables and I/O mapping functions for FIS: The FIS accepts various input variables. Fig. 7 shows the three input variables for this system. The input mapping function receives blocks of data, and the output mapping function returns intermediate results. An input reduction function reads the intermediate results and produces a final result. Thus, mapping calculations are typically split into two related parts, with functions reduced separately. The input and output mapping features of the FIS can be seen in Fig. 8.

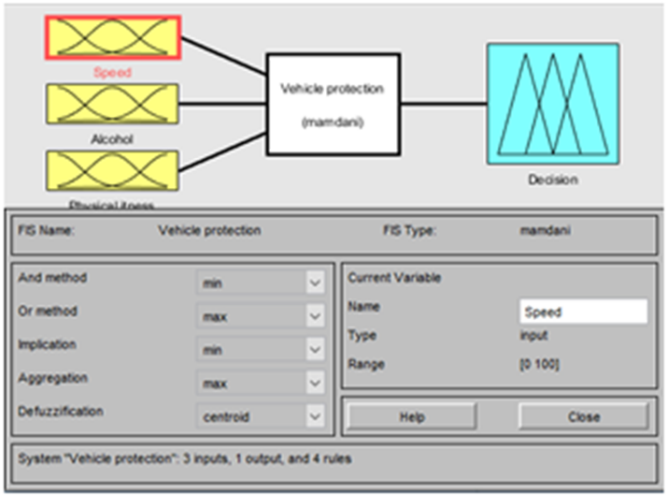


Fig. 7. Three input variables for FIS.

2) Rule base and rule view for FIS: Vehicles often travel at high speeds on roads. For emergency speed control, if speed v is measured in km/h, the rules are:

1) Low: $v < 30$
2) Medium: $20 \leq v \leq 65$
3) High: $v \geq 60$

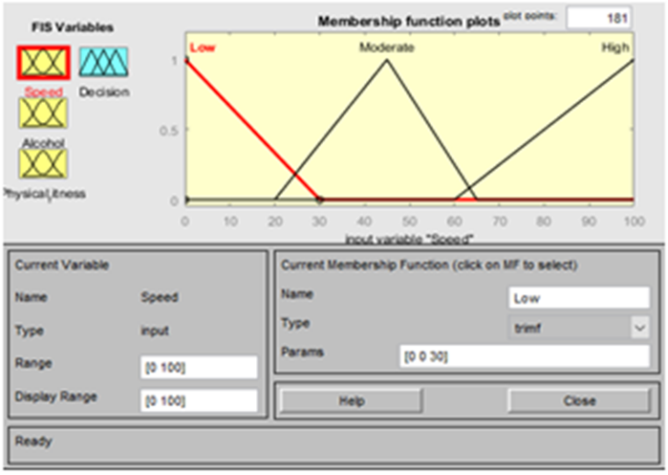


Fig. 8. Input/Output mapping functions for FIS.

Table IV shows the rule base, and Table V shows the decision-making process for emergency speed control to reduce car accidents on highways.

TABLE IV. RULE BASE FOR SPEED CONTROL

SL NO.	Rule Base
1	speed is low: the decision is the vehicle is clear
2	speed is moderate: the decision is the vehicle is clear
3	speed is high: the vehicle stops and necessary steps are taken by the police

TABLE V. DECISION TABLE

Speed	Decision
Low	Vehicle is clear
Moderate	Vehicle is clear
High	Vehicle stops and necessary steps are taken by the police

The rule view for the determination of decisions according to input is shown in Fig. 9.

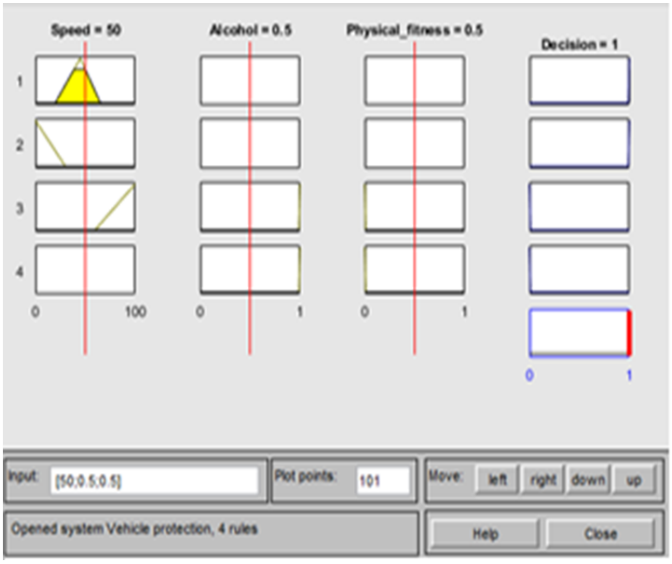


Fig. 9. Rule view for FIS.

3) Surface for FIS: The surface plot from the FIS shows the relationship between input and output variables. Fig. 10 depicts the surface according to the input.

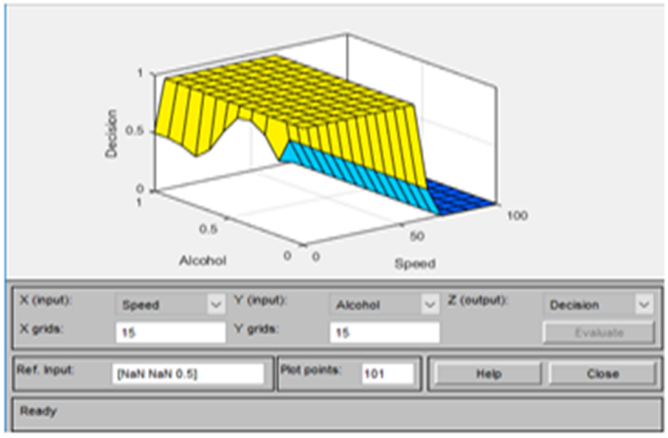


Fig. 10. Surface for FIS.

C. Limitations

While the FIS model provides a robust framework for speed control and accident reduction, it has some limitations:

- The system relies heavily on accurate sensor data, which may not always be available.
- The model's effectiveness is constrained by the predefined rules and membership functions, which might not cover all real-world scenarios.
- Integration with other traffic management systems is not addressed in this implementation.

D. Future Works

Future improvements and extensions to this research could include integrating the system with other traffic management systems for a more comprehensive solution, expanding the FIS rules to cover a broader range of driving conditions and scenarios, incorporating machine learning techniques to adaptively modify the FIS rules based on real-time data, and enhancing the sensor network to improve data accuracy and reliability.

VI. Conclusion

This study introduces an innovative approach to enhance road safety through the integration of IoT and fuzzy logic technologies for real-time vehicle speed monitoring and accident reduction. Our system leverages IoT sensors to continuously monitor vehicle speeds and employs fuzzy logic for decision-making based on variables like speed, alcohol presence, and driver fitness. This approach enables timely interventions by notifying both drivers and law enforcement agencies of potential speeding violations, thereby mitigating the risk of accidents. The system's capability to facilitate interaction between drivers and police, coupled with Vehicle-to-Everything (V2X) communication, enhances cooperative safety measures and improves traffic management. Although our Fuzzy Inference System (FIS) demonstrated effective speed control capabilities in our analysis, its performance is contingent on accurate sensor data and predefined rules, suggesting avenues for future enhancements. By integrating machine learning for adaptive rule adjustments and refining sensor networks for enhanced reliability, future research aims to further optimize our system's efficacy in ensuring safer and more efficient transportation systems globally.

References

[1] A. Zanella, N. Bui, A. Castellani, L. Vangelista, and M. Zorzi, "Internet of things for smart cities," *IEEE Internet of Things journal*, vol. 1, no. 1, pp. 22–32, 2014.

[2] K. T. Kubra, B. Barua, M. M. Sarker, and M. S. Kaiser, "An iot-based framework for mitigating car accidents and enhancing road safety by controlling vehicle speed," in *2023 7th International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud)(I-SMAC)*. IEEE, 2023, pp. 46–52.

[3] R. K. Kodali and M. Sairam, "Over speed monitoring system," in *2016 2nd international conference on contemporary computing and informatics (ic3i)*. IEEE, 2016, pp. 752–757.

[4] S. Deshpande, V. Bhole, P. Dudhade, N. Gourka, and S. Darade, "Implementing a system to detect over speeding & inform authorities in case of any violations," *International Research Journal of Engineering and Technology*, vol. 4, pp. 2445–2449, 2017.

[5] A. Singhal, R. Tomar *et al.*, "Intelligent accident management system using iot and cloud computing," in *2016 2nd international conference on next generation computing technologies (NGCT)*. IEEE, 2016, pp. 89–92.

[6] C. Chatrapathi, M. N. Rajkumar, and V. Venkatesakumar, "Vanet based integrated framework for smart accident management system," in *2015 International Conference on Soft-Computing and Networks Security (ICSNS)*. IEEE, 2015, pp. 1–7.

[7] P. Raut and V. Sachdev, "Car accident notification system based on internet of things," *International Journal of Computer Applications*, vol. 107, no. 17, 2014.

[8] H. M. Ali and Z. S. Alwan, *Car accident detection and notification system using smartphone*. Lap Lambert Academic Publishing Saarbrucken, 2017.

[9] S. Aishwarya, A. Rai, M. Prasanth, S. Savitha *et al.*, "An iot based accident prevention & tracking system for night drivers," *International Journal of Innovative Research in Computer and Communication Engineering*, vol. 3, no. 4, pp. 3493–3499, 2015.

[10] K. Sandeep, P. Ravikumar, and S. Ranjith, "Novel drunken driving detection and prevention models using internet of things," in *2017 International Conference on Recent Trends in Electrical, Electronics and Computing Technologies (ICRTEECT)*. IEEE, 2017, pp. 145–149.

[11] P. R. Shetgaonkar, V. NaikPawar, and R. Gauns, "Proposed model for the smart accident detection system for smart vehicles using arduino board, smart sensors, gps and gsm," *Int. J. Emerg. Trends Technol. Comput. Sci*, vol. 4, no. 4, 2015.

[12] K. A. Khaliq, A. Qayyum, and J. Pannek, "Prototype of automatic accident detection and management in vehicular environment using vanet and iot," in *2017 11th International Conference on Software, Knowledge, Information Management and Applications (SKIMA)*. IEEE, 2017, pp. 1–7.

[13] N. H. Sane, D. S. Patil, S. D. Thakare, and A. V. Rokade, "Real time vehicle accident detection and tracking using gps and gsm," *International Journal on Recent and Innovation Trends in Computing and Communication*, vol. 4, no. 4, pp. 479–482, 2016.

[14] U. Yadav and K. Kannan, "Smart vehicle monitoring system using iot," *International Journal for Development of Computer Science and Technology*, vol. 5, no. 3, pp. SW–31, 2017.

[15] M. S. Reddy and K. R. Rao, "Fire accident detection and prevention monitoring system using wireless sensor network enabled android application," *Indian Journal of Science and Technology*, vol. 9, no. 17, pp. 1–5, 2016.

[16] K. Kavya and C. Geetha, "Accident detection and ambulance rescue using raspberry pi," *International Journal of Engineering and Techniques*, vol. 2, no. 3, 2016.

[17] P. A. Targe and M. Satone, "Vanet based real-time intelligent transportation system," *International Journal of Computer Applications*, vol. 145, no. 4, pp. 34–38, 2016.

[18] K. Poorani, A. Sharmila, and G. Sujithara, "Iot based live streaming of vehicle, position accident prevention and detection system," *International Journal of Recent Trends in Engineering and Research*, vol. 3, pp. 52–55, 2017.

[19] T. Kim and H.-Y. Jeong, "A novel algorithm for crash detection under general road scenes using crash probabilities and an interactive multiple model particle filter," *IEEE Transactions on Intelligent Transportation Systems*, vol. 15, no. 6, pp. 2480–2490, 2014.

[20] K. Patel, "Utilizing the emergence of android smartphones for public welfare by providing advance accident detection and remedy by 108 ambulances," *International Journal of Engineering Research & Technology (IJERT)*, vol. 2, no. 9, pp. 1340–1342, 2013.

[21] J. White, C. Thompson, H. Turner, B. Dougherty, and D. C. Schmidt, "Wreckwatch: Automatic traffic accident detection and notification with smartphones," *Mobile Networks and Applications*, vol. 16, pp. 285–303, 2011.

[22] C. Prabha, R. Sunitha, R. Anitha *et al.*, "Automatic vehicle accident detection and messaging system using gsm and gps modem," *International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering*, vol. 3, no. 7, pp. 10 723–10 727, 2014.

[23] E. Nasr, E. Kfoury, and D. Khoury, "An iot approach to vehicle accident detection, reporting, and navigation," in *2016 IEEE international multidisciplinary conference on engineering technology (IMCET)*. IEEE, 2016, pp. 231–236.

[24] V. Savania, H. Agravata, and D. Patela, "Alcohol detection and accident prevention of vehicle," *International Journal of Innovative and Emerging Research in Engineering*, vol. 2, no. 3, pp. 55–59, 2015.

[25] M. S. Amin, J. Jalil, and M. B. I. Reaz, "Accident detection and reporting system using gps, gprs and gsm technology," in *2012 International Conference on Informatics, Electronics & Vision (ICIEV)*. IEEE, 2012, pp. 640–643.

[26] S. Nandyal, S. Sultana, and S. Anjum, "Smart car parking system using arduino uno," *International Journal of Computer Applications*, vol. 169, no. 1, pp. 13–18, 2017.

[27] T. Akhund, N. Newaz, and M. M. Sarker, "Internet of things based low-cost health screening and mask recognition system," *International Journal of Computing and Digital Systems*, vol. 15, no. 1, pp. 259–269, 2024.

[28] T. M. N. U. Akhund, Z. A. Shaikh, I. De La Torre Díez, M. Gafar, D. H. Ajabani, O. Alfarraj, A. Tolba, H. Fabian-Gongora, and L. A. D. López, "Iost-enabled robotic arm control and abnormality prediction using minimal flex sensors and gaussian mixture models," *IEEE Access*, vol. 12, pp. 45 265–45 278, 2024.

[29] A. Tabassum, T. Islam, and T. M. N. U. Akhund, "Data-medi: A web database system for e-health," in *Intelligent Sustainable Systems: Selected Papers of WorldS4 2022, Volume 2*. Springer, 2023, pp. 619–628.

[30] M. Rahman, M. F. I. Suny, J. Tasnim, M. S. Zulfiker, M. J. Alam, and T. M. N. U. Akhund, "Iot and ml based approach for highway monitoring and streetlamp controlling," in *International Conference on Machine Intelligence and Emerging Technologies*. Springer, 2022, pp. 376–385.

[31] S. Afroz, T. M. N. U. Akhund, T. Khan, M. U. Hasan, R. Jesmin, and M. M. Sarker, "Internet of sensing things-based machine learning approach to predict parkinson," in *International Congress on Information and Communication Technology*. Springer, 2023, pp. 651–660.

[32] T. M. N. U. Akhund, "Covid-19 effects on private tuition in bangladesh and internet of things based support system," *International Journal of Computing and Digital Systems*, vol. 13, no. 1, pp. 1153–1163, 2023.

[33] T. M. N. U. Akhund, M. Hossain, K. Kubra, Nurjahan, A. Barros, and M. Whaiduzzaman, "Iot based low-cost posture and bluetooth controlled robot for disabled and virus affected people," *International Journal of Advanced Computer Science and Applications*, vol. 13, no. 8, 2022. [Online]. Available: http://dx.doi.org/10.14569/IJACSA.2022.0130879

[34] T. M. Akhund, N. Ullah, N. T. Newaz, Z. Zaman, A. Sultana, A. Barros, and M. Whaiduzzaman, "Iot-based low-cost automated irrigation system for smart farming," in *Intelligent Sustainable Systems*. Springer, 2022, pp. 83–91.

[35] A. H. Himel, F. A. Boby, S. Saba, T. M. Akhund, N. Ullah, and K. Ali, "Contribution of robotics in medical applications a literary survey," in *Intelligent Sustainable Systems*. Springer, 2022, pp. 247–255.

[36] T. M. N. U. Akhund and W. M. Al-Nuwaiser, "Human iot interaction approach for modeling human walking patterns using two-dimensional levy walk distribution," *International Journal of Advanced Computer Science and Applications*, vol. 15, no. 6, 2024. [Online]. Available: http://dx.doi.org/10.14569/IJACSA.2024.01506151

[37] M. R. Rashel, M. Islam, S. Sultana, M. Ahmed, T. M. Akhund, N. Ullah, J. N. Sikta *et al.*, "Internet of things platform for advantageous renewable energy generation," in *Proceedings of International Conference on Advanced Computing Applications*. Springer, 2022, pp. 107–117.

[38] A. E. S. Leni *et al.*, "Instance vehicle monitoring and tracking with internet of things using arduino," *International Journal on Smart Sensing and Intelligent Systems*, vol. 10, no. 5, pp. 123–135, 2017.

[39] P. Wang, S. Fang, L. Zhang, and J. Wang, "A vehicle collision detection algorithm at t-shaped intersections based on location-based service," in *New Frontiers in Road and Airport Engineering*, 2015, pp. 308–317.

[40] J. Smith *et al.*, "Traffic management system: A review of technologies," *IEEE Transactions on Intelligent Transportation Systems*, vol. 20, no. 8, pp. 3050–3060, 2019.

[41] A. Johnson *et al.*, "Intelligent transport systems: Design and deployment," *Transportation Research Part C: Emerging Technologies*, vol. 95, pp. 414–432, 2018.

[42] H. Lee *et al.*, "V2v communication for safety and traffic efficiency," *IEEE Vehicular Technology Magazine*, vol. 15, no. 3, pp. 140–147, 2020.

[43] M. Chen *et al.*, "Road condition monitoring using sensor networks," *Sensors*, vol. 19, no. 24, p. 5360, 2019.

TGMoE: A Text Guided Mixture-of-Experts Model for Multimodal Sentiment Analysis

Xueliang Zhao[1], Mingyang Wang[2]*, Yingchun Tan[3], Xianjie Wang[4]
College of Computer and Control Engineering, Northeast Forestry University, Harbin, China[1,2,3]
Harbin Institute of Technology, Harbin, China[4]

***Abstract*—Multimodal sentiment analysis seeks to determine the sentiment polarity of targets by integrating diverse data types, including text, visual, and audio modalities. However, during the process of multimodal data fusion, existing methods often fail to adequately analyze the sentimental relationships between different modalities and overlook the varying contributions of different modalities to sentiment analysis results. To address this issue, we propose a Text Guided Mixture-of-Experts (TGMoE) Model for Multimodal Sentiment Analysis. Based on the varying contributions of different modalities to sentiment analysis, this model introduces a text guided cross-modal attention mechanism that fuses text separately with visual and audio modalities, leveraging attention to capture interactions between these modalities and effectively enrich the text modality with supplementary information from the visual and audio data. Additionally, by employing a sparsely gated mixture of expert layers, the TGMoE model constructs multiple expert networks to simultaneously learn sentiment information, enhancing the nonlinear representation capability of multimodal features. This approach makes multimodal features more distinguishable concerning sentiment, thereby improving the accuracy of sentiment polarity judgments. The experimental results on the publicly available multimodal sentiment analysis datasets CMU-MOSI and CMU-MOSEI show that the TGMoE model outperforms most existing multimodal sentiment analysis models and can effectively improve the performance of sentiment analysis.**

***Keywords*—Multimodal fusion; sentiment analysis; cross modal; mixture of experts**

I. Introduction

With the rapid growth of text data such as social media and online comments, sentiment analysis has become an increasingly important research field. The goal of sentiment analysis tasks is to classify the sentiment information contained in raw data into different sentiment polarities such as positive, negative, or neutral. However, in many practical scenarios, sentiment data often not only contains textual information but also includes multimodal data such as images, videos, audio, etc. Compared to unimodal data lacking diversity, these multimodal data can provide more information for sentiment analysis, and the complementarity of this information can enhance the accuracy of sentiment analysis.

Existing multimodal sentiment analysis methods include Tensor-based fusion [1], which directly connects feature tensors from different modalities for analysis. However, this method generates very large feature tensors, requiring a lot of storage space and computational resources, and does not consider the interaction of information between different modalities. To address these issues, researchers have developed other deep learning-based fusion methods. Huddar et al. [2] proposed multi-level feature optimization, extracting feature tensors from multiple modalities and using LSTM to extract contextual information between adjacent utterances at multiple levels. However, this method did not examine the correlation of different modal information with the sentiment analysis results, making it unable to fully understand the target sentiment comprehensively and accurately. Tsai et al. [3] proposed a multimodal routing method that dynamically adjusts the relative weights between input samples and output representations by exploring the correlation between modalities and identifying the relative importance of single-modal and cross-modal features.

However, although previous approaches have made progress in multimodal fusion, they often fail to adequately account for the varying contributions of different modalities to sentiment information, overlooking the importance of sentimental information from different modalities. In the field of sentiment analysis, while audio and visual modalities indeed contain crucial sentimental information, the distribution of sentiment information across modalities is unevenly distributed. Neglecting the differences in contributions of different modalities to sentiment analysis may result in multimodal fusion representations lacking crucial sentiment information from key modalities, thus reducing the accuracy of sentiment analysis [4].

To address the above issues, this paper proposes a text guided mixture-of-experts model for multimodal sentiment analysis. The model aims to better capture the differences in sentiment information between different modalities, obtaining more targeted sentiment features. TGMoE leverages pre-trained models for feature extraction from three modalities. It integrates visual and audio modality information into the text modality through a text guided cross-modal fusion mechanism to obtain multimodal fusion features. Subsequently, for the sentiment prediction task, multiple highly specialized experts are simultaneously trained by a trainable gating network to selectively handle sentiment features. This approach delves deeper into uncovering potential connections among modal data, thereby enhancing the accuracy of sentiment prediction in the model.The contributions of TGMoE model can be summarized as follows:

- Proposing a text guided cross-modal Transformer network that integrates sentiment information from visual and audio modalities into the text modality through a text guided attention mechanism.
- TGMoE uses a sparsely gated mixture-of-experts mechanism to selectively process multi-modal fusion

features, enhancing the model's ability to learn and represent complex emotional information.

- Extensive experimental results on two benchmark datasets demonstrate that the proposed TGMoE outperforms several existing methods in multimodal sentiment analysis tasks.

II. Related Work

With the advent of the information age, we have access to a large amount of multimodal data (videos, audio, and text), providing a more abundant source of features for sentiment analysis tasks. Accurate and rapid analysis of human emotions can offer better services for daily work and life. Multimodal sentiment analysis aims to understand the sentiment of the target in video data (including text, audio, and visual modalities), uncovering deep sentimental information in each modality to reduce bias in single-modal sentimental information. Learning how to capture interaction information within and between modalities and effectively integrate multimodal information is a key challenge faced by multimodal sentiment analysis tasks [5].

To address this issue, researchers have proposed various multimodal fusion methods for modeling. With the advancement of deep learning, model-based fusion has received more attention. Zadeh et al. [1] introduced the Tensor Fusion Network, which computes the Cartesian product of three modalities and concatenates the resulting tensor along a certain dimension. The concatenated tensor is then fed into a deep neural network for sentiment classification. Building on this, Liu et al. [6] decomposed the weights of the fusion tensor into a set of low-rank factors to improve efficiency, making the computational complexity linearly related to the number of modalities. Hou et al. [7] recursively integrated local correlations into global correlations through multilinear fusion. Mai et al. [8] adopted a divide-and-conquer approach by partitioning multimodal features into blocks, applying tensor fusion to each block to capture local interaction information, and then combining local information to obtain global multimodal interaction information.

Model-based fusion methods can effectively preserve sentimental information within modalities but struggle to consider contextual relationships between modalities. Graph neural networks, due to their excellent structural learning capabilities, are widely applied in multimodal sentiment analysis tasks. Yang et al. [9] proposed a modal-temporal attention graph for unaligned multimodal data, where each sub-feature of each modality in the sequential data is treated as a node. They construct modality-type edges between different modalities and temporal-type edges within the same modality. By applying a pruning algorithm, they fuse and align asynchronous distributed multimodal sequential data. Hu et al. [10] constructed a fully connected heterogeneous graph for conversational data, considering each modality of each utterance as a node. They connect each node to nodes representing the same utterance but from different modalities and to nodes representing the same modality from the same conversation. Different aggregation mechanisms for various types of edges are designed to learn multimodal dynamics in the graph network.

To further explore sentimental information within and across modalities, researchers have started integrating attention mechanisms into multimodal sentiment analysis. For example, Wang et al. [11] proposed a recurrent attentional variable embedding network that combines attention mechanisms to investigate facial and speech features. They learn displacement information generated during the vocabulary representation process for multimodal representation. Building upon this, Rahman et al. [12] incorporated multimodal representations into Transformer models, using visual and audio features to learn representation shifts and apply them to text modality for sentiment analysis. While these methods have partially addressed the issue of insufficient fusion between modalities, they often treat each modality equally, overlooking the varying contributions of different modalities to the final sentiment analysis results.

Therefore, this paper addresses the issue of disparate contributions between different modalities by enhancing the role of the text modality in sentiment analysis. It utilizes a mixture of experts to further extract sentiment information from multimodal features, enhancing nonlinear representation capabilities and obtaining more abstract fusion features of sentiment information.

III. Method

A. Overall Model Architecture

Multimodal sentiment analysis employs three modalities - audio (X_a), visual (X_v), and text (X_t) - from the same video segment to determine the sentiment polarity of the target. The proposed TGMoE model also aims to effectively integrate information from the three modalities to enhance the effectiveness of sentiment analysis. Fig. 1 illustrates the overall architecture of the TGMoE model. The model framework consists of three parts: the feature extraction module, the text guided cross-modal feature fusion module, and the sparsely gated mixture of experts module.

Feature Extraction Module: For each modality, appropriate pre-trained models are used to gain incipient modality features.

Text Guided Cross-Modal Feature Fusion Module: Utilizing cross-modal attention to capture the interaction information between audio, visual, and text features, the module adds this information to the text features. This encourages the text features to incorporate information from other modalities, providing high-quality fused features for sentiment prediction.

Sparsely Gated Mixture-of-Experts Module: Training multiple experts to handle multimodal features with different sentimental biases, to deeply explore the potential connections between data and improve the accuracy of sentiment prediction.

B. Feature Extraction Module

To better extract features of single modality data, different feature extraction methods are adopted for different modalities. For the text modality, the language pre-trained model SimCSE [13] is utilized as the feature extractor for text discourse. The hidden state output from the last layer is taken as the feature vector for the text modality.

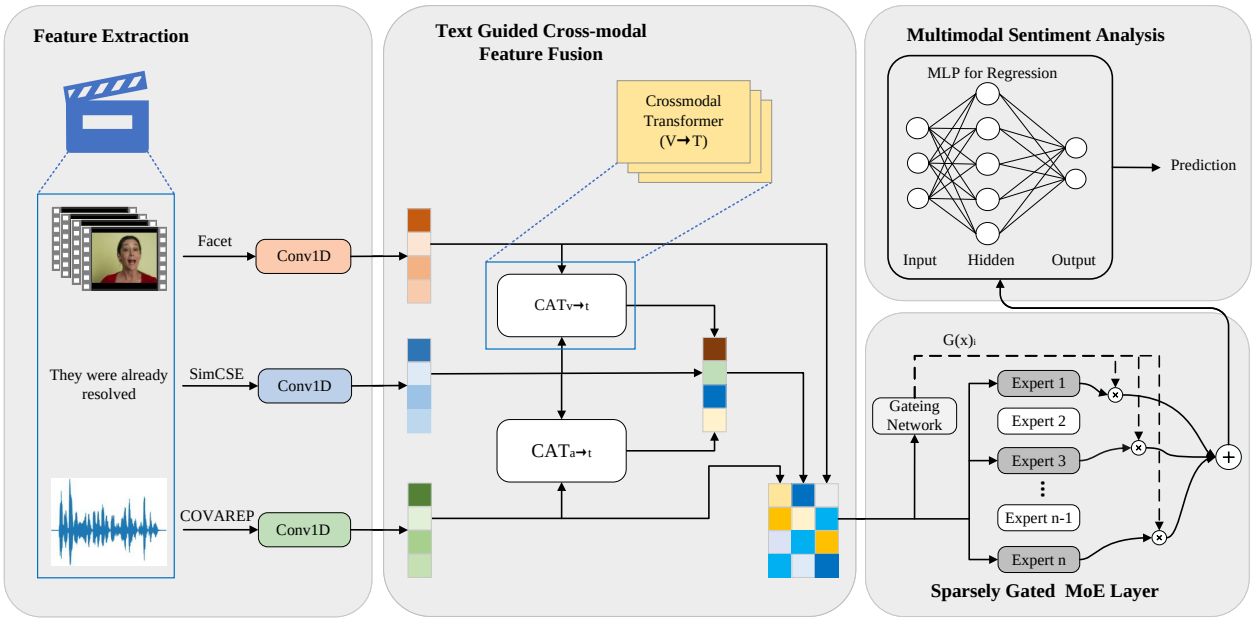


Fig. 1. The overall structure of the TGMoE model.

$$X_t' = \text{SimCSE}(X_t; \theta_t) \in \mathbb{R}^{s_t \times d_t} \tag{1}$$

Where X_t' represents the result of text modality feature extraction and θ_t is the SimCSE model's parameter. s_t means the sequence length and d_t is the feature dimension of text modalities.

For the audio modality, audio features are extracted using the COVEREP [14] acoustic framework. These features include pitch, volume, Mel-Frequency Cepstral Coefficients, and more, denoted as X_a. For the visual modality, visual features are extracted using Facet. These features include facial action units, facial landmarks, head pose, and other features, denoted as X_v. The features for audio and visual modalities can be obtained through the CMU-Multimodal SDK.

To achieve better fusion in the upcoming work, 1D temporal convolution is used to unify the feature dimensions of the three modalities while ensuring that each element of the input sequence has sufficient awareness of its neighboring elements. The features of the three modalities are fed into a 1D temporal convolutional layer:

$$f_m = \text{Conv1D}(X_m', k_m) \in \mathbb{R}^{s_m \times d_m}, m \in \{t, v, a\} \tag{2}$$

where k_m is the size of the convolutional kernel. f_m is output of 1D temporal convolutional layer.

C. Text Guided Cross-Modal Feature Fusion Module

In the traditional Transformer model, changing the positions of the input sequence does not alter the final output. To enable the model to capture the sequential information of the input sequence, positional embeddings (PE) are added to the representation of each modality based on the practice outlined in Transformer [15]:

$$H_m = f_m + \text{PE}_m \in \mathbb{R}^{s_m \times d_m}, m \in \{t, v, a\} \tag{3}$$

where PE_m means the PE of each modal, H_m represents the feature vector after each modal adds PE.

The text modality is the most basic and intuitive form of reflecting the speaker's sentiment, containing more sentiment-related information compared to video and audio modalities. Therefore, based on the idea of MulT [16], this paper presents a text guided cross-modal fusion module, which utilizes cross-modal attention mechanisms to calculate the attention weights between the text modality and the audio-visual modalities. This promotes the reception of information from the other two modalities by the text modality, potentially integrating emotion-related features from the audio and visual modalities into the text features for better encoding of emotional information across all three modalities. Additionally, considering the characteristics of the sentiment analysis task, the dominant role of the text modality in the feature fusion process is reinforced to incorporate emotional information from different modalities. The text-guided cross-modal feature fusion is illustrated in Fig. 2. Each cross-modal Transformer consists of L layers of cross-modal attention.

$$\text{CAT}_{v \to t} = \text{Softmax}(\frac{W_Q H_t \times W_K^T H_v}{\sqrt{d_k}}) W_V H_v \tag{4}$$

$$\text{CAT}_{v \to t} = \text{LN}(\text{CAT}_{v \to t}) \tag{5}$$

Where $\text{CAT}_{v \to t}$ indicates the visual modality transmission of information to the text modality. Softmax(·)represents a normalized exponential function, it can compress a K-dimensional vector z containing arbitrary real numbers into another K-dimensional vector $\sigma(z)$ such that the range of

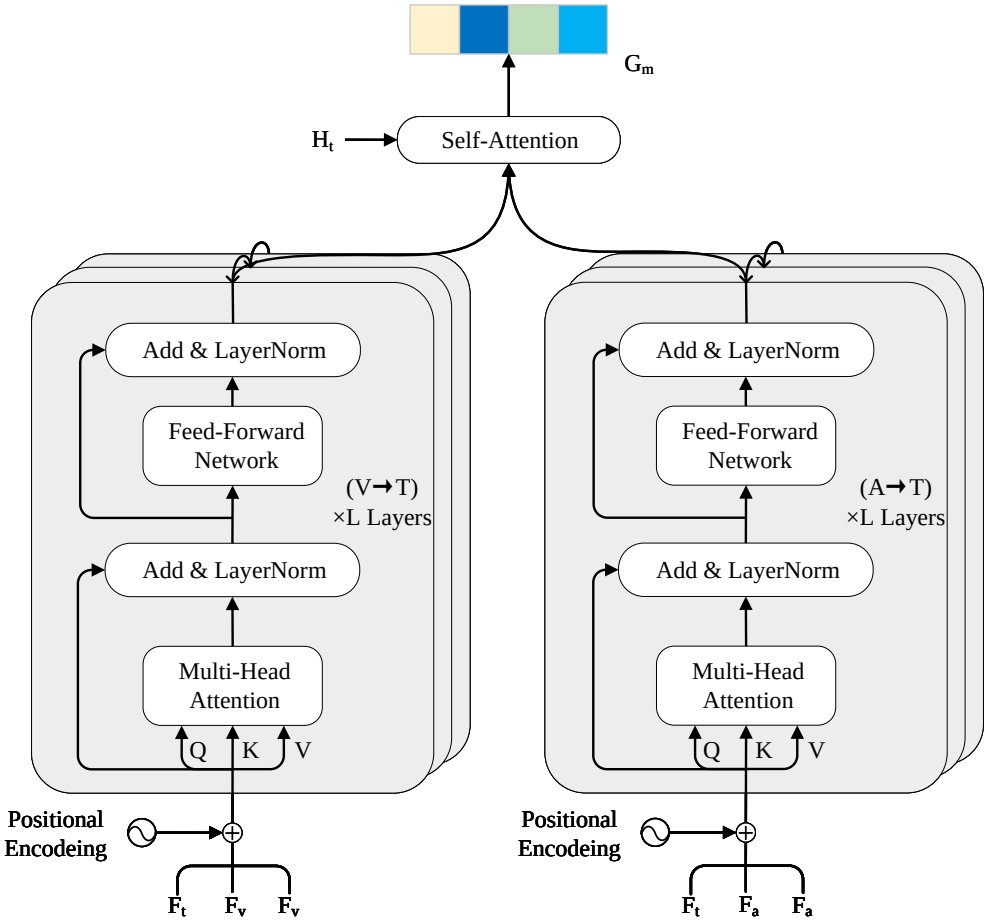


Fig. 2. Text guided cross-modal feature fusion module.

each element is between (0,1), and the sum of all elements is 1. The attention mechanism uses numbers between 0 and 1 to represent the importance of different data blocks. Where $Q = W_Q H_t, K = W_K^T H_v, V = W_V H_v$.

As each layer of the network can capture different abstract levels of input data, we stack L layers of cross-modal Transformers to learn hierarchical sentimental information within the features. Subsequently, cross-modal attention is further computed between the audio modality and the text modality, as exemplified by $(A \rightarrow T)$ in Fig. 2. The two fusion results are then concatenated with the text features and passed through self-attention blocks to capture the interactive information between the text features and the fusion features. Finally, the fusion features are combined with the features from the three modalities through convolutional layers to generate the ultimate fused feature representation. Adapting from the low-level features is beneficial for the model to retain the original information of each modality.

$$Z_m = H_t + \text{CAT}_{v \rightarrow t} + \text{CAT}_{a \rightarrow t} \tag{6}$$

$$G_m = \text{Self-Attention}(Z_m) \tag{7}$$

$$F_m = [G_m, H_a, H_v, H_t] \tag{8}$$

During the text guided cross-modal feature fusion process, the text modes constantly update their sequences through external information from multi-head cross-modal attention. By taking advantage of the fact that the text contains more sentiment-related information [17], the sentimental information of the text features is strengthened, and the sentimental information of the vision and audio modes is fully integrated into the text modes to obtain multi-modal features containing more sentimental information.

D. Sparsely Gated Mixture-of-Experts Module

During the process of text guided cross-modal feature fusion, sentiment may be expressed differently across different modalities. For example, in the sentence "You look beautiful today," the sentiment conveyed in the text modality is positive, but if accompanied by a pouting expression, the sentiment

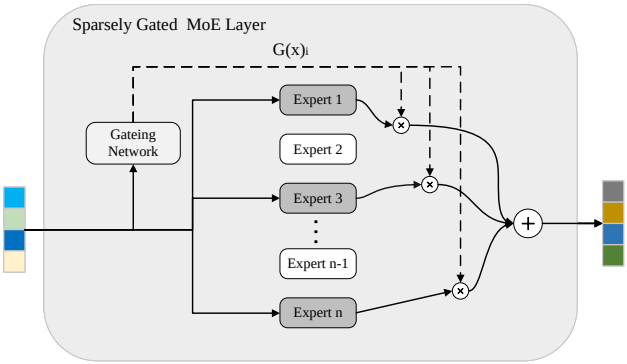


Fig. 3. Sparsely gated mixture-of-experts layer.

in the visual modality becomes negative. There are certain differences in the distribution of sentiment across different samples of data, so sentimental information in multi-modal features can be enhanced through a feed-forward network (FFN). However, for traditional deep learning models, activating the entire model for each sample when the training data is large can lead to significant spatial and time costs. To address this issue, the TGMoE model introduces a neural network component: the mixture of experts (MoE). By using gating mechanisms, the number of experts involved in the work can be effectively controlled, thus compressing the model's computational costs.

MoE is a special type of feed-forward network. In this structure, each model unit is referred to as an expert, and there is a gating network to select a combination of experts, combining the weight of each model as the final output. The difference from training data individually in traditional FFNs is that the mixture of experts' networks can enhance the non-linear representation capability of multi-modal features by allowing multiple experts to learn simultaneously. This enhances the distinctiveness of multi-modal features in terms of sentiments, thereby improving the classification accuracy of data samples.

The MoE layer in the TGMoE model consists of several experts and a trainable gating network. Each expert is an independent FFN that learns similar or different features from each other. The gating network learns parameters to select a sparse combination of numerous experts to process each input. The output of the gating network is a sparse n-dimensional vector, which is used to weigh the selected combination of experts. Each expert has the same architecture but distinct parameters. The structure of the MoE layer is illustrated in Fig. 3.

For a given input x, we define $G(x)$ as the output of the gating network; $E_i(x)$ is the output of the i-th expert network. Therefore, the output of the Sparsely Gated Mixture-of-Experts module is:

$$y = \sum_{i=1}^{n} G(x)_i E_i(x) \tag{9}$$

When $G(x)_i = 0$, the model does not need to compute $E_i(x)$. Therefore, although the model includes numerous neural networks, only a small number of neural networks will be utilized for each sample, significantly reducing computational complexity and time cost.

Sparsely Gated Network: The input x is multiplied by a trainable weight matrix W_g, then the initial architecture of the gating network is completed by applying the Softmax function.

$$G(x) = \text{Softmax}(x \cdot W_g) \tag{10}$$

x represents the input of sparsely gated network, and W_g represents a randomly generated matrix. Constructed in this way, the gating network outputs a non-sparse vector. Therefore, to ensure the sparsity of the gated output, the top k values of the gated output are retained. In addition, adjustable Gaussian noise is introduced to ensure that each neural network receives roughly the same amount of training data. The amount of noise for each gate is controlled by another adjustable parameter matrix W_{noise}.

$$\text{KeepTopK}(v_i, k) = \begin{cases} v_i, & v_i \in \text{TopK}(k) \\ -\infty, & v_i \notin \text{TopK}(k) \end{cases} \tag{11}$$

$$H(x) = x \cdot W_g + \text{LN}(x \cdot W_{noise}) \tag{12}$$

$$G(x) = \text{Softmax}(\text{KeepTopK}(H(x), k)) \tag{13}$$

where $\text{LN}(\cdot)$ represents data standardization, $v_i \in \text{TopK}(k)$ means that v_i belongs to the top K elements.

The MoE layer is placed after the text guided cross-modal feature fusion module. After passing through the text guided cross-modal fusion module layer, each multimodal fusion feature will invoke the MoE once, thereby selecting different combinations of experts to enhance the sentimental information in the multimodal representation. For the multi-modal feature F'_m, multiple expert networks are simultaneously trained through the MoE layer to deeply explore the potential connections between data, as described below:

$$\begin{aligned} z = \sum_{i=1}^{n} \text{Softmax}(\text{KeepTopK}((F_m \cdot W_g \\ + \text{LN}(F_m \cdot W_{noise})), k)) E_i(F_m) \end{aligned} \tag{14}$$

E. Model Prediction and Loss Function

The vector z output from the MoE layer is fed into a Multilayer Perceptron (MLP) for sentiment prediction. The MLP consists of three linear layers. The TGMoE model uses MAE as the basis for computing the loss function for the entire task.

$$Y'_m = \text{MLP}(Z_m; \theta_m), m \in \{t, v, a\} \tag{15}$$

$$\text{Loss}_m = \frac{1}{N} \sum_{i=1}^{N} (|\text{pred}^i - y^i|) \tag{16}$$

where N represents the number of training samples, and y represents the true labels of the multimodal data.

IV. Experimental Results

A. Datasets and Evaluation Metrics

1) Datasets: This paper evaluates the performance of the TGMoE model using two publicly available datasets.

The CMU-MOSI dataset [18] is a commonly used dataset in the field of multimodal sentiment analysis. This dataset consists of 93 videos from YouTube where the reviews of movies are discussed. The dataset is divided into 2,199 subjective discourse-level video segments. Each segment has a real-valued sentiment score in the range of [-3, +3] to express the intensity of the sentiment polarity of the characters.

The CMU-MOSEI dataset [19] is an extension of the CMU-MOSI dataset, with a larger number of utterances, more diverse samples, speakers, and topics. This dataset contains 23,453 video segments with sentiment-labeled tags from 5,000 videos. Each video segment in the MOSI and MOSEI datasets contains a sentiment score in the range of [-3, 3], where a higher value indicates a stronger positive sentiment polarity.

2) Evaluation metrics: When considering sentiment analysis on the CMU-MOSI and CMU-MOSEI datasets as a regression task, the predictive performance of the models is measured using Mean Absolute Error (MAE) and Pearson correlation (Corr). When viewed as a classification task, evaluation methods include seven-class accuracy (Acc-7), binary accuracy (Acc-2), and F1 score. Here, Acc-7 represents the accuracy of predicting values falling into seven intervals within [-3, +3], while Acc-2 and F1 represent the accuracy and F1 score of the binary classification task, respectively.

B. Experimental Settings

We developed the TGMoE model in the PyTorch framework, using Mean Absolute Error (MAE) as the loss function, Adam as the optimizer, and PyCharm as the integrated development environment. All experiments in this study were conducted on an RTX 4090 GPU, and multiple validations and analyses were performed to obtain the best set of hyperparameters.

We set the batch size to 32, epochs to 50, learning rate to 1e-3, and sequence length to 50 for all three modalities. The text feature dimension is 768, the visual feature dimension is 35, the audio feature dimension is 74, the dropout rate is 0.1, and the number of heads in multi-head attention is set to 5. The number of stacked layers in the attention layer is set to 5.

C. Baselines

To assess the relative performance of the TGMoE model, we will compare it with the following baseline models.

TFN [1]: Tensor fusion network decomposes unimodal vectors into tensors through the Cartesian product, then fuses the outer product of tensors to learn interactions within and between modes.

LMF [6]: Efficient low-rank multimodal fusion decomposes stacked high-order tensors into many low-rank factors, then efficiently fuses them based on these low-rank factors to improve efficiency.

MFM [20]: Multimodal representation learning factors link a multi-modal discriminative network with a generative network possessing intermediate modality-specific factors to facilitate the reconstruction of the fusion process and optimize the discriminative loss.

MulT [16]: Multimodal Transformer constructs a Transformer between modalities through attention mechanisms, integrating multimodal information and optimizing the fusion process.

MAG-BERT [12]: The multimodal adaptive gate designs an alignment gate for integrating visual and audio information and integrates it into a standard BERT model.

MISA [21]: The modality invariance and specificity of multimodal sentiment analysis project features into two separate independent spaces with specific constraints, taking into account the invariance and specificity of modalities, and then complete fusion on the features of both spaces.

BIMHA [22]: Enhancing bimodal information for arbitrary pairs of modalities using a multi-head attention mechanism, utilizing tensor fusion to capture interactions between modalities, effectively integrating information carried by different modalities, and improving sentiment prediction.

SELF-MM [23]: Self-supervised multi-task learning assigns a single-modal training task with self-generated labels to each modality, aiming to learn the consistency between modalities and the specificity within each modality.

CubeMLP [24]: By mixing relevant modality features on three axes using three independent MLP units and flattening the mixed multimodal features for task prediction.

TETFN[4]: By utilizing visual features extracted by the Vision Transformer, combined with audio features, learning text-oriented cross-modal mappings in pairs, in order to obtain efficient multimodal representations for emotion prediction.

D. Results

1) Comparison experiment with the baseline model: Tables I and II provide the experimental results of various models on the CMU-MOSI and CMU-MOSEI datasets. For Acc-2 and F1 values, there are two sets of evaluation results: on the left, positive and neutral sentiment samples are considered positive examples, and negative samples are considered negative examples to calculate accuracy. On the right, positive sentiment samples are considered positive examples, and negative samples are considered negative examples to calculate accuracy. Similarly, F1 values are calculated accordingly to obtain the corresponding F1 scores.

It can be seen that the proposed TGMoE model achieved the best performance on both datasets. On the CMU-MOSEI dataset, the Acc-2 and F1 scores were improved by 1.11%/0.33% and 1.4%/0.59%, respectively compared to previous methods. On the CMU-MOSI dataset, Acc-7, Acc-2 (left), and F1 scores (left) were improved by 0.1%, 1.32%, and 1.4% compared to previous methods. This indicates that the TGMoE model can adequately integrate different modalities of sentimental information, enhance non-textual modality

sentimental information to contribute more to textual modality sentiment representation in sentiment analysis, and effectively balance the semantic gap between different modalities.

TABLE I. EXPERIMENTAL RESULTS OF THE TGMOE MODEL AND BASELINE MODELS ON THE CMU-MOSI DATASET

Model	MAE	Corr	Acc-7	Acc-2	F1
TFN	0.901	0.698	34.90	-/80.80	-/80.70
LMF	0.917	0.695	33.20	-/82.50	-/82.40
MFM	0.877	0.706	35.40	-/81.70	-/81.60
MulT	0.861	0.711	-	81.50/84.10	80.60/83.90
MAG-BERT	0.727	0.781	43.62	82.37/84.43	82.50/84.61
MISA	0.804	0.764	-	80.79/82.10	80.77/82.03
BIMHA	0.925	0.671	36.44	78.57/80.30	78.57/80.30
SELF-MM	0.712	0.795	45.79	82.54/84.77	82.68/84.91
CubeMLP	0.770	0.767	45.50	-/85.60	-/85.50
TETFN	0.717	0.800	-	84.05/86.10	83.83/86.07
TGMoE	**0.760**	**0.767**	**45.89**	**85.37/85.64**	**85.23/85.71**

TABLE II. EXPERIMENTAL RESULT OF THE TGMOE MODEL AND BASELINE MODELS ON THE CMU-MOSEI DATASET

Model	MAE	Corr	Acc-7	Acc-2	F1
TFN	0.593	0.700	50.20	-/82.50	-/82.10
LMF	0.623	0.677	48.00	-/82.00	-/82.10
MFM	0.568	0.717	51.30	-/84.40	-/84.30
MulT	0.580	0.703	-	-/82.50	/82.30
MAG-BERT	0.543	0.755	52.67	82.51/84.82	82.77/84.71
MISA	0.568	0.724	-	82.59/84.23	82.67/83.97
BIMHA	0.559	0.731	52.11	84.07/83.96	83.35/83.50
SELF-MM	0.529	0.767	53.46	82.68/84.96	82.95/84.93
CubeMLP	0.529	0.760	54.90	-/85.10	-/84.50
TETFN	0.551	0.748	-	84.25/85.18	84.18/85.27
TGMoE	**0.535**	**0.757**	**53.70**	**85.36/85.51**	**85.58/85.86**

2) Ablation studies: In the field of sentiment analysis, compared to audio and visual modalities, the text modality contains more sentimental semantic information. Therefore, the model proposed in this paper is text-centric, where the sentimental information from audio and visual modalities is extensively extracted and integrated into the text modality. Subsequently, a sparsely gated MoE network is utilized to select different expert combinations to analyze and process the sentimental information.

To validate the rationality of the proposed fusion approach, we further explored the impact of different modalities and sparsely gate MoE on sentiment analysis results of the CMU-MOSEI dataset. Experimental results are shown in Tables III and IV, where T, V, and A, respectively represent text, visual, and audio modalities. TGMoE-NoMoE indicates the TGMoE model without the hybrid expert module, where the extracted multimodal features are directly connected to a fully connected layer for sentiment prediction. This leads to the model being unable to learn sufficient nonlinear representations from multimodal features. TGMoE-FFN represents the TGMoE model replacing the hybrid expert module with an FFN layer, where the multimodal features output by the text guided cross-modal feature fusion module are input into the FFN layer to strengthen the multimodal features. This results in the model only being able to learn limited sentimental information, directly impacting the accuracy of sentiment analysis results.

TABLE III. EXPERIMENTAL RESULTS OF ABLATION EXPERIMENTS INVOLVING DIFFERENT MODALITIES IN FUSION ON THE CMU-MOSEI DATASET

Num	Model	MAE	Corr	Acc-7	Acc-2	F1
1	A	0.839	0.012	41.36	71.02/62.85	83.06/77.19
2	V	0.810	0.217	41.92	75.74/70.77	67.85/60.84
3	T	0.589	0.713	50.16	80.92/82.77	80.82/83.13
4	A+V	0.810	0.229	41.39	64.99/63.43	65.41/65.15
5	T+A	0.575	0.720	51.86	79.87/84.31	79.29/84.35
6	T+V	0.584	0.703	50.91	82.08/83.41	82.17/83.89
7	A+V+T	0.535	0.757	53.70	85.36/85.51	85.58/85.86

TABLE IV. EXPERIMENTAL RESULT OF THE EFFECTIVENESS EXPERIMENT OF MOE MODULE ON THE CMU-MOSEI DATASET

Num	Model	MAE	Corr	Acc-7	Acc-2	F1
1	TGMoE-NoMoE	0.587	0.714	50.88	81.28/83.73	82.13/83.65
2	TGMoE-FFN	0.564	0.729	52.09	83.62/84.36	83.65/84.58
3	TGMoE	0.535	0.757	53.70	85.36/85.51	85.58/85.86

E. Discussion

From the perspective of the number of modalities, information from each modality can complement each other, and the addition of more modal information will lead to better sentiment analysis results. For the single modality baseline, the performance of the text modality is superior to the audio and visual modalities, indicating that the text modality contributes more to multimodal sentiment analysis than the audio and visual modalities. In real life, people also prefer to use language to express their direct sentiments.

From the perspective of model modules, when the MoE module is removed, all metrics of the model are lower compared to the complete model. Substituting the MoE module with an FFN (Feedforward Neural Network) layer leads to an improvement in model performance compared to TGMoE-NoMoE, indicating that FFN can strengthen the sentimental information in multimodal features, but with limited performance. The complete TGMoE model outperforms all metrics. Experimental results suggest that MoE plays a crucial role in multimodal fusion, and removing or simply replacing MoE results in the model's inability to learn sufficient nonlinear representations from multimodal features, thereby failing to guarantee that the final fused features contain abstract sentimental information, consequently affecting sentiment prediction performance.

In conclusion, the experiments above validate the efficacy of the text guided mixture of experts model TGMoE. This model adeptly harnesses sentimental data across various modalities, facilitating efficient fusion, mitigating the impact of sentimental information discrepancies across modalities on the final sentiment, and bolstering the efficacy of multimodal sentiment analysis.

V. CONCLUSION

To address the issue that the obtained multimodal fusion representation may be defective in capturing sentimental information due to ignoring the different contributions of various modalities to sentiment analysis, this paper proposes a

text guided mixture-of-experts model TGMoE for multimodal sentiment analysis. The TGMoE model is structured around three key modules. Firstly, features are extracted for each of the three modalities respectively to capture the inherent consistency within each modality. Secondly, a text guided cross-modal fusion mechanism is proposed: cross-modal attention mechanisms are used for text-visual and text-audio modalities, respectively, to capture the interactive information of visual and audio modalities with the text modality, supplementing the text modality with the sentimental information from the visual and audio modalities. Finally, a sparsely gated mixture of expert networks is employed to fortify the nonlinear representational capacity within multimodal features, engender more abstract fusion features, and elevate the precision of sentiment polarity classification. Comparative evaluations against existing multimodal sentiment analysis frameworks demonstrate a pronounced performance boost, underscoring the efficacy of the proposed text guided cross-modal interactive approach and the utility of employing mixture of expert networks for sentiment analysis enhancement.

In real-world scenarios of multimodal sentiment analysis, users may not provide information from all modalities simultaneously. For example, they might only provide text while missing audio or visual data. Therefore, our future research will focus on effectively handling cases of missing modalities, developing more robust sentiment analysis models, and enhancing the practical application value and user experience of these models.

ACKNOWLEDGMENT

This work was supported the National Natural Science Foundation of China (Grant No. 71473034), and the Heilongjiang Provincial Natural Science Foundation of China (Grant No. LH2019G001).

REFERENCES

[1] A. Zadeh, M. Chen, S. Poria, E. Cambria, and L.-P. Morency, "Tensor fusion network for multimodal sentiment analysis," in *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, 2017, pp. 1103–1114.

[2] M. G. Huddar, S. S. Sannakki, and V. S. Rajpurohit, "Multi-level feature optimization and multimodal contextual fusion for sentiment analysis and emotion classification," *Computational Intelligence*, vol. 36, no. 2, pp. 861–881, 2020.

[3] Y.-H. H. Tsai, M. Q. Ma, M. Yang, R. Salakhutdinov, and L.-P. Morency, "Multimodal routing: Improving local and global interpretability of multimodal language analysis," in *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing*, 2020, p. 1823.

[4] D. Wang, X. Guo, Y. Tian, J. Liu, L. He, and X. Luo, "Tetfn: A text enhanced transformer fusion network for multimodal sentiment analysis," *Pattern Recognition*, vol. 136, p. 109259, 2023.

[5] D. Gkoumas, Q. Li, C. Lioma, Y. Yu, and D. Song, "What makes the difference? an empirical comparison of fusion strategies for multimodal language analysis," *Information Fusion*, vol. 66, pp. 184–197, 2021.

[6] Z. Liu, Y. Shen, V. B. Lakshminarasimhan, P. P. Liang, A. B. Zadeh, and L.-P. Morency, "Efficient low-rank multimodal fusion with modality-specific factors," in *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics*, 2018, pp. 2247–2256.

[7] M. Hou, J. Tang, J. Zhang, W. Kong, and Q. Zhao, "Deep multimodal multilinear fusion with high-order polynomial pooling," *Advances in Neural Information Processing Systems*, vol. 32, 2019.

[8] S. Mai, H. Hu, and S. Xing, "Divide, conquer and combine: Hierarchical feature fusion network with local and global perspectives for multimodal affective computing," in *Proceedings of the 57th annual meeting of the association for computational linguistics*, 2019, pp. 481–492.

[9] J. Yang, Y. Wang, R. Yi, Y. Zhu, A. Rehman, A. Zadeh, S. Poria, and L.-P. Morency, "Mtag: Modal-temporal attention graph for unaligned human multimodal language sequences," in *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 2021, pp. 1009–1021.

[10] J. Hu, Y. Liu, J. Zhao, and Q. Jin, "Mmgcn: Multimodal fusion via deep graph convolution network for emotion recognition in conversation," in *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, 2021, pp. 5666–5675.

[11] Y. Wang, Y. Shen, Z. Liu, P. P. Liang, A. Zadeh, and L.-P. Morency, "Words can shift: Dynamically adjusting word representations using nonverbal behaviors," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 33, no. 01, 2019, pp. 7216–7223.

[12] W. Rahman, M. K. Hasan, S. Lee, A. Zadeh, C. Mao, L.-P. Morency, and E. Hoque, "Integrating multimodal information in large pretrained transformers," in *Proceedings of the conference. Association for Computational Linguistics. Meeting*, vol. 2020, 2020, p. 2359.

[13] T. Gao, X. Yao, and D. Chen, "Simcse: Simple contrastive learning of sentence embeddings," in *EMNLP 2021-2021 Conference on Empirical Methods in Natural Language Processing, Proceedings*, 2021.

[14] G. Degottex, J. Kane, T. Drugman, T. Raitio, and S. Scherer, "Covarep—a collaborative voice analysis repository for speech technologies," in *2014 ieee international conference on acoustics, speech and signal processing*, 2014, pp. 960–964.

[15] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," *Advances in neural information processing systems*, vol. 30, 2017.

[16] Y.-H. H. Tsai, S. Bai, P. P. Liang, J. Z. Kolter, L.-P. Morency, and R. Salakhutdinov, "Multimodal transformer for unaligned multimodal language sequences," in *Proceedings of the conference. Association for Computational Linguistics. Meeting*, vol. 2019, 2019, p. 6558.

[17] Z. Sun, P. Sarma, W. Sethares, and Y. Liang, "Learning relationships between text, audio, and video via deep canonical correlation for multimodal language analysis," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 34, no. 05, 2020, pp. 8992–8999.

[18] A. Zadeh, R. Zellers, E. Pincus, and L.-P. Morency, "Multimodal sentiment intensity analysis in videos: Facial gestures and verbal messages," *IEEE Intelligent Systems*, vol. 31, no. 6, pp. 82–88, 2016.

[19] A. B. Zadeh, P. P. Liang, S. Poria, E. Cambria, and L.-P. Morency, "Multimodal language analysis in the wild: Cmu-mosei dataset and interpretable dynamic fusion graph," in *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 2018, pp. 2236–2246.

[20] Y.-H. H. Tsai, P. P. Liang, A. Zadeh, L.-P. Morency, and R. Salakhutdinov, "Learning factorized multimodal representations," in *International Conference on Representation Learning*, 2019.

[21] D. Hazarika, R. Zimmermann, and S. Poria, "Misa: Modality-invariant and-specific representations for multimodal sentiment analysis," in *Proceedings of the 28th ACM international conference on multimedia*, 2020, pp. 1122–1131.

[22] T. Wu, J. Peng, W. Zhang, H. Zhang, S. Tan, F. Yi, C. Ma, and Y. Huang, "Video sentiment analysis with bimodal information-augmented multi-head attention," *Knowledge-Based Systems*, vol. 235, p. 107676, 2022.

[23] W. Yu, H. Xu, Z. Yuan, and J. Wu, "Learning modality-specific representations with self-supervised multi-task learning for multimodal sentiment analysis," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 35, no. 12, 2021, pp. 10 790–10 797.

[24] H. Sun, H. Wang, J. Liu, Y.-W. Chen, and L. Lin, "Cubemlp: An mlp-based model for multimodal sentiment analysis and depression estimation," in *Proceedings of the 30th ACM international conference on multimedia*, 2022, pp. 3722–3729.

A Simple and Efficient Approach for Extracting Object Hierarchy in Image Data

Saravit Soeng[1], Vungsovanreach Kong[2], Munirot Thon[3], Wan-Sup Cho[4], Tae-Kyung Kim[5]*

Department of Big Data, Chungbuk National University, Cheongju, South Korea[1,2,3,4,5]
Department of Management Information Systems, Chungbuk National University, Cheongju, South Korea[5]

***Abstract*—An object hierarchy in images refers to the structured relationship between objects, where parent objects have one or more child objects. This hierarchical structure is useful in various computer vision applications, such as detecting motorcycle riders without helmets or identifying individuals carrying illegal items in restricted areas. However, extracting object hierarchies from images is challenging without advanced techniques like machine learning or deep learning. In this paper, a simple and efficient method is proposed for extracting object hierarchies in images based on object detection results. This method is implemented in a standalone package compatible with both Python and C++ programming languages. The package generates object hierarchies from detection results by using bounding box overlap to identify parent-child relationships. Experimental results show that the proposed method accurately extracts object hierarchies from images, providing a practical tool to enhance object detection capabilities. The source code for this approach is available at https://github.com/saravit-soeng/HiExtract.**

***Keywords*—Object hierarchy; object relationship; object detection; computer vision**

I. Introduction

Computer vision, particularly object detection and hierarchy extraction in digital images, plays a crucial role in enabling computers to perceive and understand digital images similarly to humans. These technologies facilitate a range of tasks, such as image classification, object detection, and instance segmentation, by allowing the classification of images, identification of objects within images, and segmentation of objects from images [1], [2]. These tasks have been successfully accomplished using advanced deep learning techniques, notably Convolutional Neural Networks (CNNs). By leveraging big data and powerful computational resources, CNNs have significantly enhanced prediction performance and accuracy [3], elevating the field of computer vision. In the current digital era, computer vision tasks have uncovered applications across a broad spectrum of research areas [2], [4].

Object detection is a computer vision task that deals with identifying instances of objects such as humans, cars, motorcycles, or animals in digital images [5], [6], [7]. The results of object detection can be applied in various contexts and fields, including facial recognition, medical image analysis, surveillance systems, robotics, and autonomous driving. Consequently, object detection is one of the most ubiquitous tasks in these diverse applications [6], [7], emphasizing its critical importance in real-world scenarios.

However, obtaining an object hierarchy from images or videos is a challenging task without the assistance of advanced techniques like machine learning or deep learning. Existing object detection methods primarily focus on identifying individual objects within an image and do not establish hierarchical relationships. For example, algorithms such as YOLO (You Only Look Once), SSD (Single Shot MultiBox Detector), Fast R-CNN, and Faster R-CNN offer state-of-the-art approaches for identifying various objects in digital images or videos but do not provide a hierarchical structure that elucidates relationships between objects. This gap in existing methodologies highlights the need for a novel approach to extract hierarchical relationships among detected objects.

This study aims to extract object hierarchies by leveraging the unique combination of object detection results and hierarchical structuring techniques. The task of object hierarchy extraction focuses on establishing parent-child relationships between identified objects within digital images. To construct the object hierarchy, a simple method based on the criteria of overlapping bounding boxes obtained from object detection results is employed. Objects with overlapping bounding boxes are considered to have a parent-child relationship. Thus, object detection serves as a fundamental component of the proposed approach. The study implements the proposed approach as a standalone package compatible with both Python and C++ programming languages, facilitating easy integration into various applications.

This research significantly contributes to the field of computer vision by providing a novel and efficient method for extracting object hierarchies using object detection results. The approach is practical due to its implementation as a standalone package compatible with Python and C++. The versatility of this approach allows it to be integrated into a wide range of applications. The significance of this research lies in its potential to offer new possibilities for understanding and processing digital images in a more structured and relational manner. This work not only extends object detection technology but also opens up possibilities for more complex and nuanced computer vision tasks, marking an important advancement in the field of computer vision.

This paper is organized as follows: Section II reviews related work in object detection and visual relationship extraction. Section III details our proposed method for hierarchical structuring of detected objects. Section IV presents

* Corresponding author.

This work was supported by the research grant of the Chungbuk National University in 2024.

experimental results demonstrating the effectiveness of our approach across various scenarios. Section V explores practical applications and use cases for the extracted object hierarchies. Section VI provides insights, discusses limitations, and offers future research directions. Finally, Section VII concludes the paper.

II. Related Work

Object detection and visual relationship extraction have been active areas of research in computer vision. This section provides an overview of recent advancements in these fields and identifies the gap that our research aims to address.

A. Object Detection

Recent years have seen significant progress in object detection algorithms, with several state-of-the-art methods emerging:

1) YOLO (You Only Look Once): Introduced by Redmon et al. [8], YOLO revolutionized object detection by treating it as a regression problem, enabling real-time detection with high accuracy.

2) SSD (Single Shot MultiBox Detector): Liu et al. [9] proposed SSD, which improved upon YOLO by using multi-scale feature maps for detection, enhancing accuracy for objects of various sizes.

3) Fast R-CNN and Faster R-CNN: Girshick [10] and Ren et al. [11] developed these region-based convolutional network methods, which significantly improved detection speed and accuracy.

While these algorithms excel at identifying individual objects, they do not establish hierarchical relationships between detected objects, which is the focus of our research.

B. Visual Relationship Detection

To bridge this gap, several novel frameworks have been proposed to detect visual relationships between objects. Dai et al. [12] introduced a deep relational network that leverages statistical dependencies between objects to detect visual relationships, demonstrating superior performance on two large datasets compared to other state-of-the-art methods. Kolesnikov et al. [13] proposed a model utilizing a Box Attention mechanism to detect visual relationships such as "person riding motorcycle" or "bottle on table," enabling the modeling of pairwise interactions between objects using standard object detection pipelines.

Lu et al. [14] incorporated language priors from semantic word embedding to guide model predictions, training separate models for objects and predicates independently before combining them to predict multiple relationships per image. Zhu et al. [15] introduced deep structured learning for visual relationship detection, employing both feature-level and label-level predictions to learn relationships, thereby capturing dependencies between objects and predicates and enhancing the understanding of visual relationships.

Additional research [16], [17], [18], [19], [20], has contributed various approaches for extracting visual relationships between objects in digital images and videos, achieving notable results. These studies typically formulate visual relationships as <subject-predicate-object> structures.

C. Research Gap and Our Contribution

While existing research has made significant strides in object detection and visual relationship extraction, there remains a gap in efficiently extracting hierarchical relationships among objects in images. Most current methods focus on identifying individual objects or pairwise relationships, but do not provide a comprehensive hierarchical structure.

Our research addresses this gap by proposing a simple yet effective method for extracting object hierarchies from images based on object detection results. Unlike previous work that focuses on complex visual relationships, our approach simplifies the task by establishing a parent-child hierarchy among objects. This method provides a novel perspective on object detection and structuring in computer vision, with potential applications in scene understanding, image captioning, and visual question answering.

By leveraging the output of existing object detection algorithms and introducing a hierarchical structuring technique, our approach offers a practical solution that can be easily integrated into various computer vision applications. This research thus bridges the gap between individual object detection and complex visual relationship extraction, providing a middle ground that captures essential object hierarchies in a computationally efficient manner.

III. Hierarchical Structuring Method of Detected Objects

The proposed research provides a novel method for efficiently creating a hierarchy of detected objects in a digital image. This method simplifies the process of understanding object relationships and interactions, improving accuracy and reducing computational resources. It utilizes a unique combination of object detection algorithms and a hierarchical structuring technique that distinguishes it from existing methods. The proposed method leverages the output of an object detection algorithm, which identifies and locates objects within an image. The object detection results provide bounding boxes and confidence scores for each detected object. These bounding boxes represent the spatial locations of the objects, while the confidence scores indicate the likelihood that the detected objects are actually present in the scene [8].

The proposed method for extracting object hierarchies from object detection results has been implemented as a versatile, cross-platform package, facilitating seamless integration into diverse applications across multiple domains. The package is available for both Python and C++ programming languages, ensuring wide accessibility and compatibility with the preferred development environments of a vast range of developers. To leverage the capabilities of this package, users simply need to provide the object detection results as input, typically comprising predicted object classes and their corresponding bounding box coordinates. The package then processes this input data, employing sophisticated algorithms to analyze the spatial relationships between the detected objects and construct

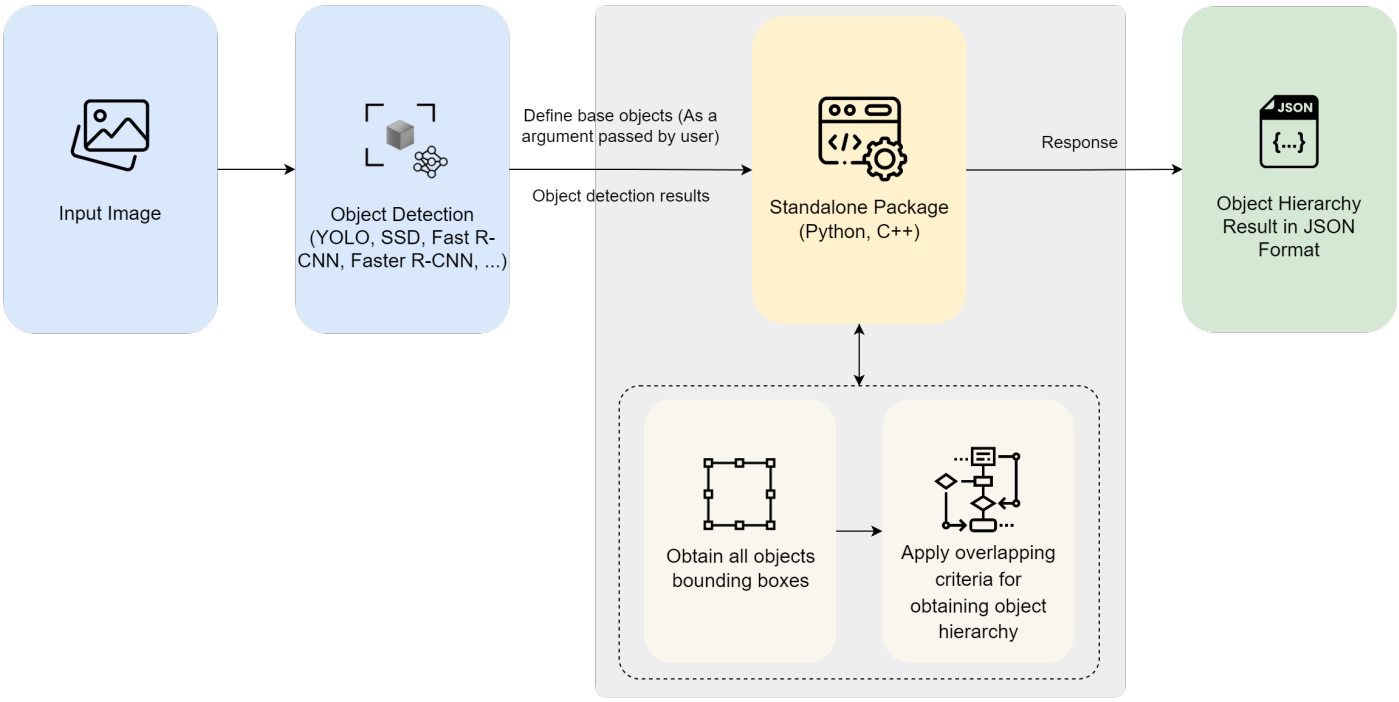


Fig. 1. A flowchart illustrating the process of object detection and hierarchy establishment.

Algorithm 1 Extract the object hierarchy from digital image based on object detection

```
function EXTRACT_OBJECT_HIERARCHY(base_objects, result)
    get predicted classes and bounding boxes from detection result
    combine predicted classes and bounding boxes: object_with_boxes
    get all class names
    initial empty result list: result_list []
    for each object in base_objects do
        get base object class index from original classes
        if object_with_boxes do not empty then
            for each object in object_with_boxes do
                if class index of current object == class index of base object then
                    create children object: children []
                    for each object in object_with_boxes do
                        if not base object then
                            if current object overlaps with base object then
                                add child object to children list
                            end if
                        end if
                    end for
                    if children object is not empty then
                        create a parent object with child objects
                        add parent object to result_list
                    end if
                end if
            end for
        end if
    end for
    create results object with result_list data
    return results object
end function
```

a hierarchical representation of the scene. The object hierarchy extraction process is illustrated in Fig. 1.

The given pseudo-code in Algorithm 1 describes a function called extract_object_hierarchy that takes two inputs: base_objects and result. The core algorithm accepts only two primary inputs:

1) Base_objects: This parameter specifies the parent objects from which we aim to extract the hierarchy. It allows flexibility in defining the root nodes of our hierarchical structure.

2) Result: This parameter encompasses the complete object detection results, typically including bounding box coordinates and class predictions for all detected objects in the image.

The purpose of this function is to analyze the output of an object detection algorithm and organize the detected objects into a hierarchical structure based on their spatial relationships (overlapping) and their predicted classes.

Initially, the package combines the predicted object classes and their associated bounding boxes into a unified data structure. The package then iterates over each detected object, referred to as the "base object," retrieving its corresponding class index from the original set of classes. If the detection results contain overlapping objects, the package performs an additional loop to identify objects whose class indices match that of the current base object. For each matching object, the package creates a new list to store objects that spatially overlap with the base object. It iterates over the detection results once more, comparing the bounding boxes of the objects against the base object's bounding box. If an overlap is detected, the current object is appended to the 'children' list.

Upon completing the overlap detection process, if the 'children' list is not empty, a new "parent object" is constructed, encapsulating the base object and its associated children objects. This parent object is then added to the 'result_list', effectively building the hierarchical structure.

After processing all base objects and their corresponding overlapping objects, the package consolidates the 'result_list' data into a final results object, which is then returned to the user. This results object represents the complete hierarchical structure of the detected objects, enabling users to seamlessly integrate and leverage this information within their applications.

IV. Experimental Results

The experiments were conducted on various images in different scenarios to evaluate the proposed approach's effectiveness in extracting object hierarchies. The proposed method demonstrated impressive performance in extracting object hierarchies from images. However, the accuracy of the object hierarchy extraction process heavily relies on the object detection results.

The first experiment involved an image of a person holding a cell phone and a cup. Based on the proposed method, the person was defined as the single base or parent object, while the cell phone and cup were detected as child objects. This experiment utilized the YOLOv8s [21] model for object detection from images, as illustrated in Fig. 2.



Fig. 2. An experimental result on single base object.

Another experiment was conducted on an image containing multiple common base objects. In this case, a custom YOLOv5 model was employed to detect objects from the image. The custom model identified three distinct objects: rider, helmet, and non-helmet. The rider was defined as the base object, and through the extraction process, multiple parent-child relationships were established between the rider object and the helmet or non-helmet objects. Fig. 3 illustrates the results obtained in this experiment.

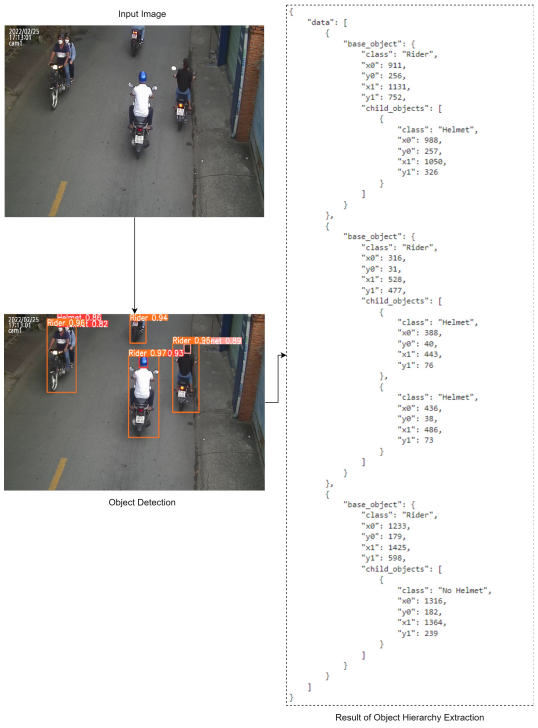


Fig. 3. An experimental result on multiple common base objects.

In addition to defining a common base object, the proposed approach was tested on multiple base objects. Fig. 4 shows the results obtained from an image featuring a person, a dog, and a tennis ball. In this test, both the person and the dog were defined as base objects.

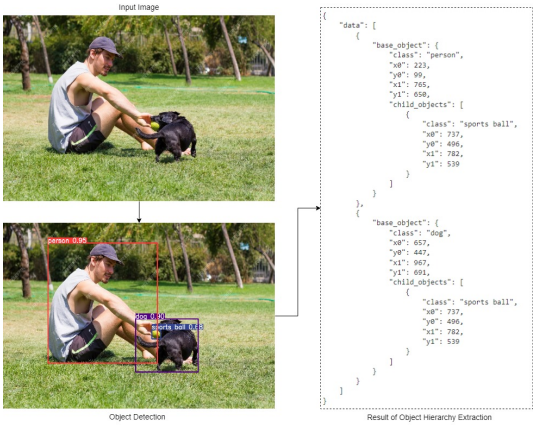


Fig. 4. An experimental result on different base objects.

The proposed approach demonstrated its versatility by handling various object hierarchies, ranging from a single base object to multiple base objects coexisting within an image. However, it is crucial to acknowledge that the accuracy of the extracted object hierarchy heavily depends on the performance of the underlying object detection model. Improvements in object detection algorithms could further enhance the accuracy and robustness of the proposed object hierarchy extraction method.

V. Application Use Cases

The extracted object hierarchy results can be leveraged in a variety of research and practical applications. In this research, we explored two distinct use cases to demonstrate the utility of the proposed object hierarchy extraction method.

The first use case involved the detection of motorcycle riders without helmets, a critical safety concern. We employed a custom YOLOv5-based model to detect three distinct objects: rider, helmet, and non-helmet. Leveraging the proposed method, a standalone package was developed to generate object hierarchy results from the object detection outputs. These object hierarchy results enabled the identification of illegal actions by motorcycle riders, specifically those not wearing helmets. Fig. 5 illustrates an example of detecting motorcycle riders without helmets using the generated object hierarchy results.

In this application, the object hierarchy played a crucial role in distinguishing between riders wearing helmets and those without helmets. By establishing the parent-child relationships between the rider object and the helmet or non-helmet objects, the system could effectively identify instances where a rider was associated with a non-helmet object, indicating a potential safety violation.

The second use case explored the application of object hierarchy results in a different domain or scenario, leveraging the versatility of the proposed approach. In this use case, the object hierarchy results enable the determination of whether specific items belong to a particular person within a defined area. By establishing the parent-child relationships between the person object and the item objects (e.g., handbag, backpack, or any other prohibited item), the system can effectively identify instances where an individual is carrying a restricted item into a prohibited area.

Fig. 6 demonstrates an example of detecting a person carrying a handbag into a banned area based on the proposed method. The object hierarchy results allow the system to associate the handbag object with the person object, indicating that the individual is in possession of the item while entering the restricted area.

The versatility of the proposed object hierarchy extraction method allows for its application in diverse domains, showcasing its potential for enhancing situational awareness, decision-making processes, and automated analysis of complex visual data.

VI. Discussion

The proposed method for extracting object hierarchies from images based on object detection results offers a novel approach to understanding spatial relationships between objects. This section discusses the implications of our findings, limitations of the current study, and potential directions for future research.

A. Implications of the Findings

Our approach demonstrates the feasibility of extracting meaningful hierarchical relationships between objects in images using a straightforward method based on bounding box overlap. This finding carries several important implications:

1) Simplified scene understanding: By organizing detected objects into a hierarchical structure, our method provides a more intuitive representation of the scene, potentially simplifying downstream tasks such as image captioning or visual question answering.

2) Computational efficiency: Compared to more complex visual relationship detection methods, our approach is computationally efficient, making it suitable for real-time applications.

3) Versatility: As demonstrated in our use cases, the extracted hierarchies can be applied to various domains, from traffic safety to security applications.

B. Limitations of the Study

While our method demonstrates potential, it is important to acknowledge its limitations:

1) Dependence on object detection accuracy: The quality of the extracted hierarchy is heavily dependent on the accuracy of the underlying object detection algorithm. Errors in object detection can propagate to the hierarchy extraction process.

2) Simplistic relationship model: Our method primarily relies on spatial overlap to determine relationships, which may not capture more complex or abstract relationships between objects.

3) Lack of semantic understanding: The method does not incorporate semantic knowledge about object classes, which could potentially improve the accuracy of the extracted hierarchies.

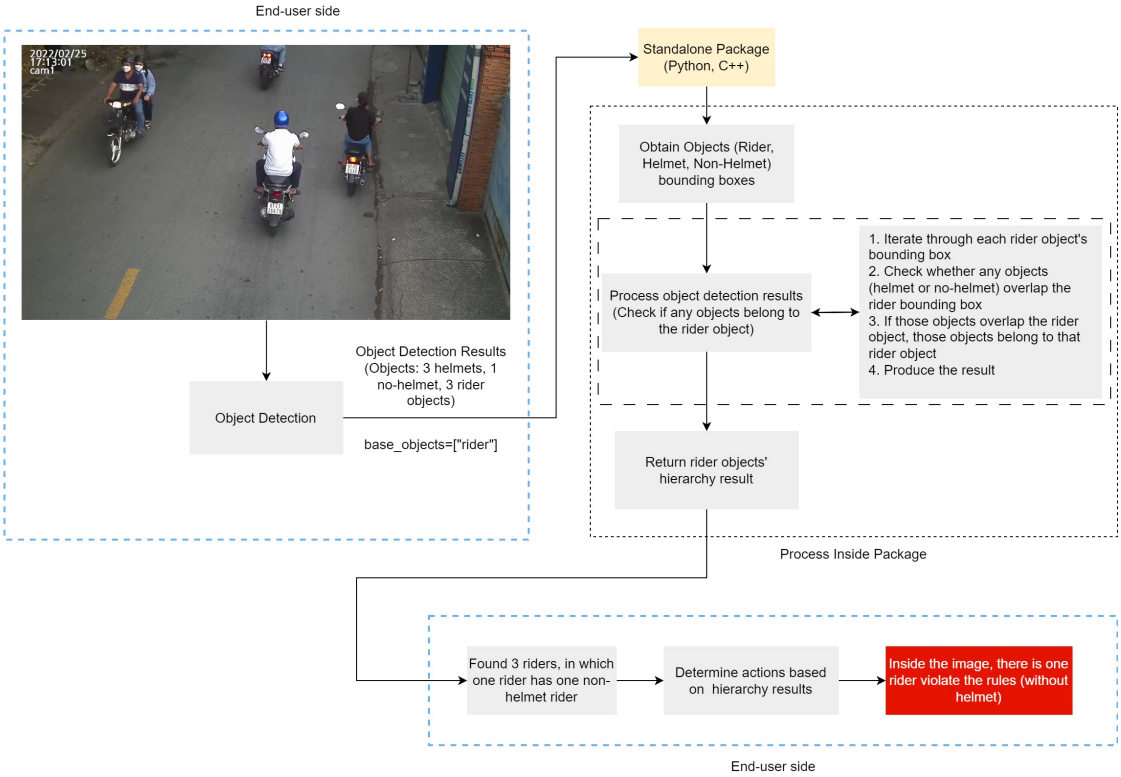


Fig. 5. An example of detecting motorcycle riders without wearing helmets using the object hierarchy results.

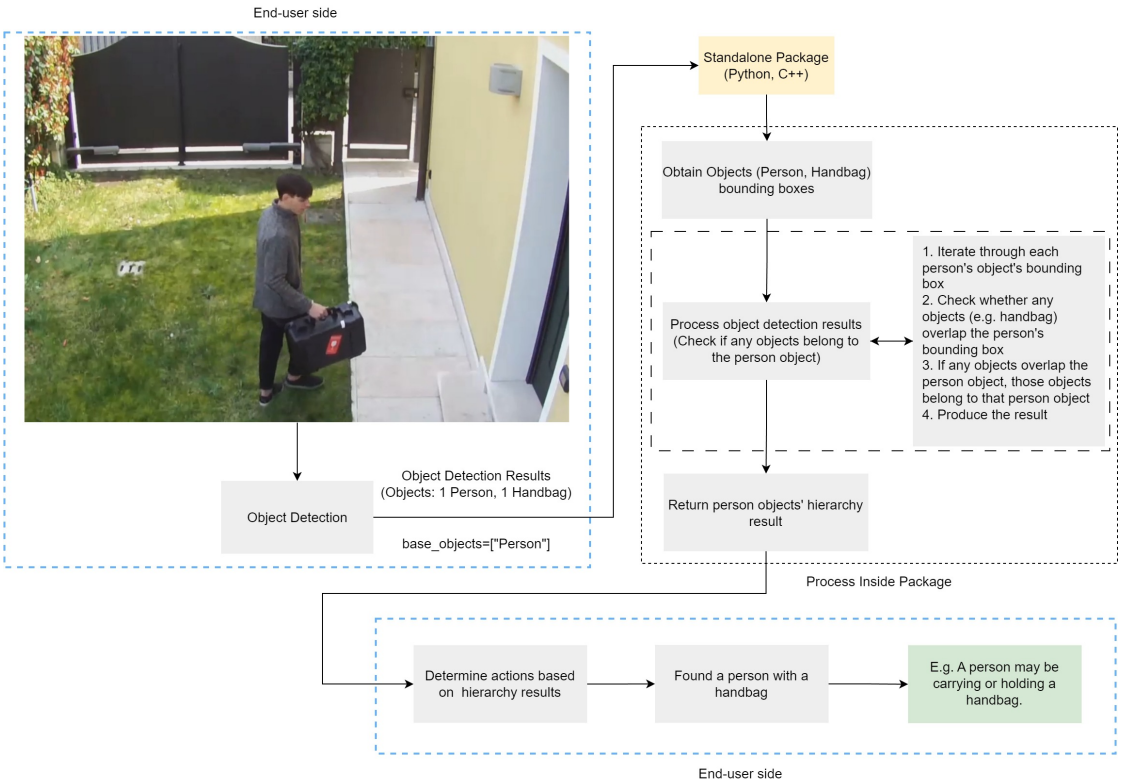


Fig. 6. An example of detecting a person carrying a handbag into a banned area based on the proposed method.

C. Future Research Directions

Based on the findings and limitations of this study, several avenues for future research emerge:

1) Integration with semantic knowledge: Incorporating class-specific semantic information could enhance the accuracy of the hierarchy extraction process.

2) Temporal hierarchies: Extending the approach to video data to capture temporal relationships between objects over time.

3) Machine learning enhancement: Developing a machine learning model to predict hierarchical relationships based on both spatial and semantic features could improve the method's accuracy and robustness.

4) Expanding object detection compatibility: Future versions of the standalone package should support a wider range of object detection algorithms beyond YOLO.

While our proposed method offers a simple and efficient approach to extracting object hierarchies, there is significant potential for further refinement and expansion of this technique. Future research should focus on addressing the current limitations and exploring more sophisticated approaches to understanding object relationships in visual data.

VII. Conclusion

In conclusion, this research has significantly contributed to the field of computer vision by introducing a novel approach for extracting object hierarchies from digital images. Utilizing the results of object detection, the proposed research presents a unique method that identifies parent-child relationships between objects based on overlapping bounding boxes criteria. This approach is also practical, as it has been implemented as a standalone package compatible with both Python and C++ programming languages. The versatility of this approach allows it to be integrated into a wide range of applications. The implications of this research are solid, offering new possibilities for understanding and processing digital images in a more structured and relational manner. This work not only expands object detection techniques but also opens up possibilities for more complex and subtle computer vision tasks, marking a significant step forward in the field. Based on the experiments, the results have demonstrated the effectiveness of the proposed approach in extracting the object hierarchy from the images. However, in terms of standalone package implementation, object detection is limited to the YOLO algorithms. In the next release version, we intend to expand to other object detection algorithms and improve the approach to work on more complexities of hierarchy extraction.

AUTHORS' CONTRIBUTION

Saravit Soeng: Conceptualization; methodology; coding; writing original draft. Vungsovanreach Kong: Conceptualization; validation; writing review and editing. Munirot Thon: Conceptualization; writing review and editing. Wan-Sup Cho: Investigation; resources. Tae-Kyung Kim: Conceptualization; methodology; supervision; project administration.

Acknowledgments

This research was funded by PentaGate, a company based in Republic of Korea.

Conflict of Interest

The authors declare no potential conflict of interests.

Data Availability Statement

The images used in this study were obtained from publicly available sources. No new images were generated as part of this research.

References

[1] A. Esteva, K. Chou, S. Yeung, N. Naik, A. Madani, A. Mottaghi, Y. Liu, E. Topol, J. Dean, and R. Socher, "Deep learning-enabled medical computer vision," *NPJ digital medicine*, vol. 4, no. 1, p. 5, 2021.

[2] D. Bhatt, C. Patel, H. Talsania, J. Patel, R. Vaghela, S. Pandya, K. Modi, and H. Ghayvat, "Cnn variants for computer vision: History, architecture, application, challenges and future scope," *Electronics*, vol. 10, no. 20, p. 2470, 2021.

[3] N. O'Mahony, S. Campbell, A. Carvalho, S. Harapanahalli, G. V. Hernandez, L. Krpalkova, D. Riordan, and J. Walsh, "Deep learning vs. traditional computer vision," in *Advances in Computer Vision: Proceedings of the 2019 Computer Vision Conference (CVC), Volume 1 1*. Springer, 2020, pp. 128–144.

[4] S. Paneru and I. Jeelani, "Computer vision applications in construction: Current state, opportunities & challenges," *Automation in Construction*, vol. 132, p. 103940, 2021.

[5] Z. Zou, K. Chen, Z. Shi, Y. Guo, and J. Ye, "Object detection in 20 years: A survey," *Proceedings of the IEEE*, vol. 111, no. 3, pp. 257–276, 2023.

[6] Y. Amit, P. Felzenszwalb, and R. Girshick, *Object Detection*. Cham: Springer International Publishing, 2021, pp. 875–883. [Online]. Available: $https://doi.org/10.1007/978-3-030-63416-2_660$

[7] Z.-Q. Zhao, P. Zheng, S.-t. Xu, and X. Wu, "Object detection with deep learning: A review," *IEEE transactions on neural networks and learning systems*, vol. 30, no. 11, pp. 3212–3232, 2019.

[8] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real−time object detection," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 779–788.

[9] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, "Ssd: Single shot multibox detector," in *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part I 14*. Springer, 2016, pp. 21–37.

[10] R. Girshick, "Fast r-cnn," in *Proceedings of the IEEE international conference on computer vision*, 2015, pp. 1440–1448.

[11] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," *IEEE transactions on pattern analysis and machine intelligence*, vol. 39, no. 6, pp. 1137–1149, 2016.

[12] B. Dai, Y. Zhang, and D. Lin, "Detecting visual relationships with deep relational networks," in *Proceedings of the IEEE conference on computer vision and Pattern recognition*, 2017, pp. 3076–3086.

[13] A. Kolesnikov, A. Kuznetsova, C. Lampert, and V. Ferrari, "Detecting visual relationships using box attention," in *Proceedings of the IEEE/CVF international conference on computer vision workshops*, 2019, pp. 0–0.

[14] C. Lu, R. Krishna, M. Bernstein, and L. Fei-Fei, "Visual relationship detection with language priors," in *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part I 14*. Springer, 2016, pp. 852–869.

[15] Y. Zhu and S. Jiang, "Deep structured learning for visual relationship detection," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 32, no. 1, 2018.

[16] K. Liang, Y. Guo, H. Chang, and X. Chen, "Visual relationship detection with deep structural ranking," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 32, no. 1, 2018.

[17] X. Shang, T. Ren, J. Guo, H. Zhang, and T.-S. Chua, "Video visual relation detection," in *Proceedings of the 25th ACM international conference on Multimedia*, 2017, pp. 1300–1308.

[18] Y. Zhan, J. Yu, T. Yu, and D. Tao, "On exploring undetermined relationships for visual relationship detection," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 5128–5137.

[19] J. Zhang, Y. Kalantidis, M. Rohrbach, M. Paluri, A. Elgammal, and M. Elhoseiny, "Large-scale visual relationship understanding," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 33, no. 01, 2019, pp. 9185–9194.

[20] B. Zhuang, L. Liu, C. Shen, and I. Reid, "Towards context-aware interaction recognition for visual relationship detection," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 589–598.

[21] G. Jocher, A. Chaurasia, and J. Qiu, "Ultralytics YOLO," Jan. 2023. [Online]. Available: https://github.com/ultralytics/ultralytics

Priority-Based Service Provision Using Blockchain, Caching, Reputation and Duplication in Edge-Cloud Environments

Tarik CHANYOUR[1], Seddiq EL KASMI ALAOUI[2], Mohamed EL GHMARY[3]
CSS Lab-Science Faculty of Ain-Chock, Hassan II University, Maarif Casablanca, Morocco[1,2]
FSDM, Sidi Mohamed Ben Abdellah University, Atlas-Fez, Morocco[3]

***Abstract*—The integration of Multi-access Edge Computing (MEC) and Dense Small Cell (DSC) infrastructures within 5G and beyond networks marks a substantial leap forward in communication technologies. This convergence is critical for meeting the stringent low latency demands of services delivered to Smart Devices (SDs) through lightweight containers. This paper introduces a novel split-duplicate-cache technique seamlessly embedded within a secure blockchain-based edge-cloud architecture. Our primary objective is to significantly shorten the service initiation durations in high density conditions of SDs and ENs. This is executed by meticulously gathering, verifying, and combining the most optimal chunk candidates. Concurrently, we ensure that resource allocation for services within targeted ENs is meticulously evaluated for every service request. The system challenges and decisions are modeled then represented as a mixed-integer nonlinear optimization problem. To tackle this intricate problem, three solutions are developed and evaluated: the Brute-Force Search Algorithm (BFS-CDCA) for small-scale environments, the Simulated Annealing-Based Heuristic (SA-CDCA) and the Markov Approximation-Based Solution (MA-CDCA) for complex, high-dimensional environments. A comparative analysis of these methods is conducted in terms of solution quality, computational efficiency, and scalability to assess their performance and identify the most suitable approach for different problem instances.**

***Keywords*—Multi-access Edge-cloud Computing; container base image chunks; replication; fragmentation; service provision; blockchain; Markov approximation**

I. Introduction

A. Preliminary

Multi-access Edge Computing (MEC) [1], [2] represents a distributed computing framework, fostering a distributed computing environment at the network edge. By meticulously placing computational resources at access points, edge routers, gateways, base stations or dedicated edge servers, MEC enables the efficient deployment of high-speed, real-time systems and solutions for end-users or Smart Devices (SDs). This approach is optimized for latency-critical, high-capacity data processing, making it ideal for emerging mobile apps and IoT devices.

MEC presents a transformative range of applications, notably within the Internet of Things (IoT) domain, alongside augmented and virtual reality, autonomous vehicles, and smart factories [3]. It empowers localized data processing and analysis, reducing reliance on centralized cloud data centers and effectively addressing latency and bandwidth challenges. Moreover, MEC architectures seamlessly integrate with a wide spectrum of wireless technologies, including next-generation cellular networks, legacy cellular and wireless local area networks. This integration ensures that MEC solutions are versatile and adaptable across different network environments, whether public or private.

Dense Small Cell (DSC) networks, characterized by a high density of small cell base stations within a limited geographic area, have emerged as a critical component of modern cellular architectures. These networks are designed to deliver high-capacity, high-speed connectivity to users. However, the surge in user demands necessitates rapid service provisioning and virtualization [4] to maintain seamless and superior user experiences. Virtualization offers substantial benefits, including enhanced service delivery, scalability, simplified management, and improved performance. This is especially advantageous in the constrained physical environments and dynamic traffic patterns of Dense Small Cell (DSC) networks. By optimizing resource utilization and enabling on-demand provisioning, virtualization helps DSC networks effectively meet the growing demands of users [5]. In addition, containerized applications enable service virtualization, which provides a more sophisticated method for effective service deployment and management in resource-constrained environments. Containerization enables to create lightweight, self-contained service instances that improve operational flexibility and streamline management procedures. This method not only makes resource allocation easier but also supports advanced interference mitigation techniques like network slicing, which guarantee optimal performance and scalability in dynamic network conditions. Nevertheless, Delivering swift service provisioning in DSC environments necessitates a combination of high-speed network connectivity for rapid data transfer, advanced service orchestration to optimize network performance, and low-latency data processing to support real-time applications. These essential components work together to ensure timely and reliable service accessibility for end-users.

Furthermore, by integrating blockchain technology with caching mechanisms [6], [7], [8], containers data can be fragmented and dispersed within multiple ENs. This distribution not only enhances data availability but also reinforces data integrity through distributed verification. Thus, initiating container-based MEC services using splitting/duplicating/caching entails distributing container image fragments across multiple secure nodes using the blockchain technology [9]. Subsequent verification, orchestration, and

service delivery to the user ensure privacy, confidentiality, reliability, and fault tolerance. This decentralized fragmentation approach also alleviates network congestion and improves performance by caching and replicating frequently requested service data.

B. Motivation

This paper aims to optimize service delivery by minimizing container-based service data collection time, bandwidth consumption, and network hops between the user and the target Edge Node (EN). The service data comprises base image chunks of service containers, which require optimal caching, replication, and transfer to the designated target EN. To address this, we formulate an optimization problem considering chunks transfer paths and the availability of diverse resources at each Edge Node. Three solution approaches are proposed: a brute-force search algorithm (BFS-CDCA) for small-scale environments, a simulated annealing-based heuristic (SA-CDCA), and a Markov approximation-based solution for larger, more complex scenarios.

C. Contributions

The distinctive features of this paper can be enumerated as follows:

- A method based on reputation and blockchain for secure service provision, emphasizing the strategic optimized collection of containers' base image from multiple edge nodes, was introduced within a multi-user MEC network.
- An adaptive approach for Containers' Base Image Chunks (CBIC) collection is proposed, ensuring efficient service provision by strategically considering CBIC with the possibility of duplication across Edge Nodes (ENs). The focus is placed on security and efficiency within a dense small cell network environment.
- An optimization problem was formulated to minimize a derived cost function, considering constraints imposed by network bandwidth, cached chunk availability, and strict service initiation deadlines. Notably, the availability of critical resources was enhanced through the implementation of service penalization based on priority.
- Given the NP-hard nature of the formulated optimization problem, a time-efficient heuristic scheme was proposed, incorporating a simulated annealing-based algorithm and a Markov approximation-based solution, with a thorough evaluation of their respective performances.

D. Paper Organization

The remaining sections of this paper are detailed following the structure : Section II presents relevant works related to the study. Section III elaborates on the framework under examination. Section V outlines the formulation of the optimization problem. Following that, Section VII provides an overview of resolution methods, and Section VIII offers an overview of the main evaluation results. Finally, Section IX serves as the conclusion, offering insights into future directions.

II. Related Works

Recent studies, such as [10], have shown a rising interest in using edge computing for the Internet of Things (IoT). Many, like the authors in [11], have investigated how it can enhance performance, primarily by optimizing resources. The authors of [10] focused on reducing costs related to task completion, and employed specialized techniques to lower energy use during tasks. Meanwhile, in [11], wireless charging and task offloading were combined to save energy. Liu et al. developed a new strategy for data storage in edge computing, aiming to boost service providers' profits [12]. Yet, studies such as [13], [14] took a broader view, improving system performance through task offloading and data storage techniques. However, these studies highlight the limitations of individual Edge Nodes (EN) in providing services.

Addressing these shortcomings, the authors of [15]-[16] advocated for EN collaboration to enhance resource utilization and equitably distribute workloads. Specifically, [15] and [17] centered on diminishing task completion durations. While [15] explored collaborative methodologies, [17] added resource allocation to its examination. Feng et al. in [18] concentrated on minimizing user delays and conserving energy. On a different note, [16] embarked on ascertaining the optimal count of collaborating ENs. However, a shared assumption across these studies is that providers freely extend their services, leading to potential reservations about their spontaneous participation absent of incentives.

The subsequent works have primarily centered on enhancing data caching and data sharing by mobile devices. In this context, a plethora of contemporary studies have delved into issues related to edge caching for content sharing. For example, Huang et al. [19] tackled the specific issue of ensuring fairness in data sharing for caching within MEC environments. In a related vein, Asheralieva et al. [20] probed the data caching dilemma using D2D-based method, allowing users to transparently reveal their content sharing costs, diminish average network expenses, and stabilize the queuing framework. Yin et al. [21] put forth a hierarchical strategy for data sharing applications within the same environments, focusing on the challenge of efficient sharing management in MEC networks to enhance user device mobility and heterogeneity.

Progressing further, the authors of [22], [23] amalgamated machine learning methodologies with secure computing techniques to redress the limitations observed in antecedent models. Specifically, in [22], devices were harnessed to execute federated calculations, with an accent on fine-tuning task offloading choices. Concurrently, Cui et al., in their study [23], synergistically integrated secure data storage mechanisms with a blockchain framework, solidifying data integrity. Notwithstanding these innovative strides, it's worth noting that conventional mathematical approaches might grapple when faced with intricate network configurations. To elevate system performance to its zenith, striking a harmonious equilibrium between competitive and collaborative paradigms is paramount. In recent work by [24], a model was proposed concentrating on a singular container-based service. However, this study overlooked the significance of considering the blockchain block size as a crucial parameter, limiting its scope. In contrast, our investigation delves into the integration of multiple container-based services, addressing their specificities

and intricacies. Additionally, we explore the impact of the blockchain operational delay as a critical parameter, considering its implications on the overall system dynamics and performance.

III. The Multi-tier Framework Under Study

In this section, the foundational framework upon which our study is built is outlined. A comprehensive overview of the framework's architecture, components, and key methodologies is provided. The framework aims to provide Container-based Services (CSs) to smart devices (SDs) by offering Container Base Image (CBI) caching with intelligent fragmentation and duplication. It employs a blockchain network to secure the exchange of the resulting Container Base Image Chunks (CBICs) among the ENs constituting a DSC network situated within a two-dimensional geographical area.

Afterwards, the main components of the proposed framework, as illustrated in Fig. 1, are presented in this section.

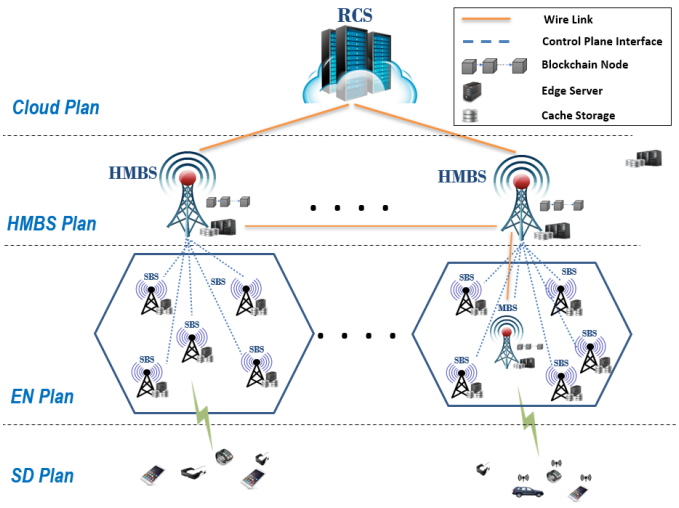


Fig. 1. Overview of the framework's main components.

A. Cloud Server and Edge Nodes

The framework handles a Remote Cloud Server (RCS) and several Small/Macro Base Stations. The RCS is expected to have extensive resources, including virtually unlimited capacities for disk storage, memory, CPU, GPU, etc. The base stations, acting as Edge Nodes (EN), provide wireless network access for smart devices (SDs) within their coverage areas. Additionally, interconnections between ENs are established through wired connections, utilizing high-bandwidth Ethernet cables or fiber-optic connections. Furthermore, each EN supports many users/subscribers and is provisioned with a dedicated edge server that has finite resources, thereby enhancing their capacity to perform localized data processing and service provisioning.

The set of all ENs is organized into multiple regions, each orchestrated by a designated Head Macro Base Station (HMBS) serving as the region head. The clustering of ENs into distinct regions, overseen by HMBSs, enables streamlined network operation and facilitates the delivery of services to end-users, resulting in improved reliability and performance. Additionally, HMBSs take on the role of caching nodes, overseeing image and transaction management to ensure the reliability and security of CBICs while disseminating them across the network. They orchestrate the geographical distribution of published containers' CBICs and handle service initiation requests by refining collection procedures to reduce service initiation time.

B. Container-based Service and Reputation

Each EN can provide only a limited set of independent CSs, due to limited capacities, where each running service uses a container instance and serves one SD only. Advancing further, the base image of the i-th CS, is partitioned into multiple chunks according to the specified chunk size, and subsequently, each CBIC is replicated D_i times.

Furthermore, a reputation management model is integrated, playing a crucial role in fostering trust among entities and enabling secure and informed caching decisions. This model dynamically allocates a reputation score to each CS i within each region r based on its historical usage and interactions. This reputation reflects the size of the demand for the service within that specific region. As a result, the degree of duplication of a CS within a region is closely tied to its reputation. Essentially, the more reputable a CS is perceived to be, the higher the likelihood of it being duplicated within that region. This approach is driven by the principle that services with a stronger reputation are in higher demand and therefore benefit from having more duplicates available to meet user requests promptly and efficiently.

Practically, the reputation scores are computed based on all pertinent trust information and feedback provided by the network entities. These scores serve to establish trust relationships among caching entities, including service providers and edge nodes.

C. Blockchain Network

The blockchain network operations in the proposed secure caching framework comprises the following four phases:

1) System initialization: During this phase, the Trusted Authority (TA) executes a Setup procedure to compute public parameters. TA generates cyclic groups with a prime order, selects random exponents and hash functions, computes essential parameters to release the public parameters to all involved entities of the blockchain network.

2) Entity registration: Upon joining the blockchain network, both the Service Provider (SP) and the ENs undergo a registration phase, during which the Trusted Authority (TA) authenticates their identities. TA executes a key generation procedure to generate random values and computes secret keys for encryption, signing, and verification. TA assigns the signing key to SP and the decryption key to edge nodes (EN) through a secure communications channel.

3) Block creation and validation: Registered SPs encrypt message chunks with a randomly generated symmetric secret key and define access structures to control access to encrypted messages by target edge nodes (ENs). SPs then signcrypt the secret key under the access structure and send the resulting ciphertext to the blockchain network for validation. This record

includes the SP's public key, pseudo identity, block hash, and signcrypted ciphertext along with the SP's signature. All registered SPs are considered authority candidates for validation. A genesis block is created at the initiation of the permissioned chain, and time intervals are allocated for affixing single blocks to the chain. In the case of multiple authorities, one leader per interval collects and validates records before passing them to other candidate authorities. Validated records are included in new blocks along with blockheaders containing relevant metadata.

4) Chunks distribution: Upon receiving a new block, the network may opt to forward the blockchain header to the edge nodes (ENs), allowing them to decide whether to request the signed ciphertext (ST) and signature (π) from specific blocks via a pull request. Upon receiving ST and π from the blockchain, ENs decrypt to recover the symmetric key (keysym), then decrypt the chunk message and verify its integrity and signature. After obtaining keysym, ENs use it to decrypt the message and calculate verification parameters, comparing them to π to confirm that the message has not been altered.

IV. System Model

This section offers a comprehensive overview of the modelization of the primary components essential to our study. Here, to simplify our notation, we will assign the variables i, j, k, n, p, r and m to refer container-based services, chunks, chunks' duplicates, edge nodes, paths, regions and resources, respectively.

A. Nodes and Resources

The CSs are denoted by the set $\mathcal{S} = \{S_1, S_2, ..., i, ..., S_{\sigma_s}\}$, distributed among the Edge Nodes (ENs) represented by the set $\mathcal{N} = \{N_1, N_2, ..., n, ..., N_{\sigma_n}\}$. Each EN is equipped with a edge server which in turn offers a diverse array of resources across multiple categories, encompassing CPU, GPU, TPU, FPGA, memory, storage, network bandwidth, etc. Here, the set of possible σ_z resources is denoted $\mathcal{Z} = \{r_1, r_2, ..., r_{\sigma_z}\}$. Accordingly, every node n is characterized by its capacity set in terms of resources which we denote $\mathcal{Z}_n^{cap} = \{Z_{n,1}^c, Z_{n,2}^c, ..., Z_{n,\sigma_z}^c\}$. Here, $Z_{n,m}^c$ denotes the maximum quantity of resource r_m that node n can provide. Furthermore, its current resource utilization is represented by: $\mathcal{Z}_n^{use} = \{Z_{n,1}^u, Z_{n,2}^u, ..., Z_{n,\sigma_z}^u\}$, where $Z_{n,m}^u$ indicates the amount of resource r_m utilized in node n.

B. Container-based Service and Chunks

For convenience, we will interchangeably use the term Container-based Service (CS) or its user, and denote CS S_i as i. Then, to summarize the operational parameters related to CS i, we use the notation Ω_i, defined as :

$$\Omega_i \triangleq \langle R_i, \pi_i, \pi_i^{min}, \pi_i^{max}, \rho_i, N_i^t, \mathcal{M}_i, B_i^{ser}, \mathcal{Z}_i^{dem}, D_i, t_i^{trans}, A_i \rangle.$$

As shown in Table I, the operational parameters for CS i encompass essential details regarding its initiation and execution.

TABLE I. Summary of Operational Parameters for CS i

Parameter	Description
Ω_i	The operating parameters of CS i
R_i	Region receiving the service initiation request for CS i.
$\mathcal{C}_i$	The set of chunks related to CS i with the size σ_i
D_i	Number of duplicates of the CS i.
A_i	Total data amount of CS i.
π_i	Priority score of CS i.
π_i^{min}, π_i^{max}	Priority score bounds of CS i.
ρ_i	Reputation score of CS i.
N_i^t	EN receiving the service initiation request for CS i.
$\mathcal{M}_i$	Localization matrix of all available duplicates of chunks associated with CS i.
B_i^{ser}	Minimum permissible data rate for the available bandwidth between the node hosting CS i and N_i^t.
$\mathcal{Z}_i^{dem}$	Resource demand of CS i, quantified by the number of standardized virtual resource units.
t_i^{trans}	Maximum permissible deployment delay for CS i.

Specifically, $\mathcal{Z}_i^{dem} = \{Z_{i,1}^d, Z_{i,2}^d, ..., Z_{i,\sigma_z}^d\}$ denotes the resource demand of CS i in terms of resources, where the resources are quantified by the number of standardized virtual resource units. Here, σ_z denotes the number of resource types, and $z_{i,m}^d$ indicates the required quantity of resource r_m.

Afterwards, the set of chunks of CS i is denoted $C_i = \{C_{i,1}, C_{i,2}, ..., C_{i,\sigma_i}\}$ where σ_i is the chunks count of CS i. For ease of use the j chunk of C_i is denoted as $C_{i,j}$ ($j \in \mathcal{C}_i = [\![1; \sigma_i]\!]$) and defined by the following key parameters: $\Omega_{i,j} \triangleq \langle I_{i,j}, D_{i,j} \rangle$. Here, $I_{i,j}$ is the identifier used to determine the order of the chunk within the CS i, and $D_{i,j}$ denotes the total data size of the chunk, measured in bytes. Moreover, in region R_i the σ_i duplicates of all chunks belonging to CS i are distributed across its ENs. The location information is provided by a set of σ_i matrices that are continually updated by the HMBS. This set specifies, for each CS i, the identifiers of the ENs caching all its chunk duplicates. The matrix $\mathcal{M}_i$ linked to CS i is of dimensions $(\sigma_i \times D_i)$. For convenience, the k-th duplicate ($k \in \mathcal{K} = \{1, 2, ..., D_i\}$) of the j-th chunk of CS i is denoted as $C_{i,j,k}$ and located in EN $\mathcal{M}_{i,j,k}$.

C. Service Priority and Reputation

Penalization of services in dense small cell networks is an unavoidable challenge, often stemming from users diversity and resource constraints such as limited bandwidth, computational capabilities, or storage/memory availability. Capacity constraints within a system may lead to the prioritization of certain users over others, achieved through a deliberate reduction in the pace at which their service requests are fulfilled. This selective prioritization strategy is implemented to effectively manage resource limitations, ensuring that critical users or tasks receive timely attention while acknowledging and potentially delaying less urgent requests. Such prioritization mechanisms play a crucial role in optimizing resource utilization and maintaining system stability under high demand scenarios. In response to these challenges, we adopt a prioritization model wherein users with higher priorities are subjected to lesser penalties, if necessary. Accordingly, a priority π_i is attributed to CS i such that:

$$\pi_i \in \left\{\pi_i^{min}, ..., \pi_i^{max}\right\} \cup \{M, M + \rho_i\} \quad ; i \in \mathcal{S} \qquad (1)$$

Here, π_i^{min} and π_i^{max} respectively represent the minimum and maximum allowable priority scores for CS i. M is a very

big number such that $\pi_i^{max} \ll M$, $\forall i \in \mathcal{S}$. The priority factors are used to favour the transfer of CS. In particular, CS with priority at least $\pi_i = M$ are high priority containers that are solicited for urgent transfer. Other priorities are determined according to the service continuity requirement in terms of downtime or the service level agreement.

Simultaneously, users who have been penalized in one round may be perceived as potential priority users in subsequent rounds. This dynamic approach recognizes that users experiencing penalties in a time slot might require special attention or resource allocation in subsequent slots to ensure fair access and equitable long term treatment. By adapting priorities based on historical user interactions, the system aims to efficiently handle specific needs and requirements of individual users over time. As a result, within our model, the penalization score undergoes adjustments based on the following criteria:

- If user i holds high-priority score $M + \rho_i$, its priority score remains unchanged. Here, ρ_i, referred the reputation score which is introduced to establish a hierarchical order among high-priority users to ensuring their precedence.
- Alternatively, if user i does not possess high-priority status:
  - If its request is fulfilled during the current time slot, its priority π_i is reset to its initial value π_i^{min}.
  - If its request remains unsatisfied during the current slot, its priority in the subsequent time slot is either incremented by 1 or set to M: if π_i reaches the maximum allowable value π_i^{max}, π_i is set to M, otherwise it is elevated to $\pi_i + 1$.

D. Network Model

The network model is given by $\mathcal{P} = \{\mathcal{P}_{n\to n'} \,/\, n \in \mathcal{N}; n' \in \mathcal{N} \setminus \{n\}\}$ composed of all sufficient paths connecting all pairs of distinct nodes (n, n'). Also, we use $\mathcal{P}_{i,j,k}^n$ to denote the set $\mathcal{P}^{\mathcal{M}_{i,j,k}\to n}$ of paths connecting EN $\mathcal{M}_{i,j,k}$ housing chunk $C_{i,j,k}$ to node n. Without loss of generality, we assume that the set $\mathcal{P}_{i,j,k}^n$ is precalculated and given at the decision time slot.

Moreover, each available path $p \in \mathcal{P}_{n\to n'}$ that is used for multi-hop data transmission is characterized by [5], [25]:

- its hop count $\mathcal{H}_{n\to n'}^p$
- its allocatable bandwidth $\mathcal{B}_{n\to n'}^p$
- traversing delay $\mathcal{D}_{n\to n'}^p$

Thus, if EN n is the transfer destination of the CS i's chunks, the backhaul bandwidth between n and the target node N_i^t, serving the user of CS i's, is given by:

$$\mathcal{B}_i^n = \begin{cases} \infty & ;n = N_i^t \\ \max\limits_{p\in\mathcal{P}_{n\to N_i^t}} \left\{\mathcal{B}_{n\to N_i^t}^p\right\} & ;n \neq N_i^t \end{cases} \quad ;i \in \mathcal{S}; n \in \mathcal{N} \qquad (2)$$

E. Delays

The cumulative delay while transferring a data byte of chunk $C_{i,j,k}$ using path p in the set $\mathcal{P}_{i,j,k}^n$ is denoted $\mathcal{D}_{i,j,k}^{n,p}$.

In this study, the service collection process involves transferring the base image of its container from the most suitable caching ENs to the nearest node possible to node N_i^t to cater to the user's request. According to the specific service being requested, the provider's CBI might be exclusively stored in the RCS, completely located on an EN, or distributed as multiple chunks across regional ENs. Upon receiving a new service request, the associated node acquires necessary instructions from the HMBS. The MBS selects the most appropriate procedure from three potential options based on specific circumstances. Notably, the focus of this study is the third option, which encompasses the characteristics of the other two scenarios. Consequently, the transfer latency experienced by a specific chunk, denoted as $C_{i,j,k}$, to EN n is determined by the cumulative transfer delays of all selected chunks from their respective caching nodes. These delay are influenced by the underlying blockchain network architecture and prevailing transfer conditions. As such, they can be decomposed into the following components:

1) Blockchain-related operational delays: [26], [27] this delay encompasses the time it takes for a node to respond and send a requested block back to the requester. This delay is intricately tied to factors like node processing power, data volume, network bandwidth, and the blockchain protocol employed. In our approach, we extend this modeling by incorporating the block size as an additional parameter, recognizing its influence on delay. Specifically, we characterize this delay as a linear function of both the chunk size and block size, introducing two parameters associated with the hosting EN and the HMBS. This refinement allows for a more comprehensive representation of the operational delay in the blockchain network.

2) Network-related transfer delays: which refers to the time it takes for the transfer of the CBIC to propagate through the network from its holding node to its final decided node. This duration is influenced by several factors, including network topology, congestion, and the number of hops.

The first delay, denoted ${}^{BC}T_{i,j,k}$, is given in the next formula where $a_{i,j,k}$, $b_{i,j,k}$ and $c_{i,j,k}$ are the delay-related parameters associated with the hosting EN $\mathcal{M}_{i,j,k}$:

$${}^{BC}T_{i,j,k} = a_{i,j,k} * D_{i,j} + b_{i,j,k} * L^{Bloc} + c_{i,j,k} \qquad (3)$$

Here, $i \in \mathcal{S}$, $j \in \mathcal{C}_i$, $k \in \mathcal{K}$, $D_{i,j}$ is the total data amount of chunk $C_{i,j}$, the terme $b_{i,j,k} * L^{Bloc}$ represents the delay overhead related to the adopted block-chain block size L^{Bloc}.

Achieving precise estimates for the parameters $a_{i,j,k}$, $b_{i,j,k}$ and $c_{i,j,k}$ within the blockchain edge-cloud network involves a comprehensive strategy. Initial empirical experiments provide a foundational dataset for response time with varying chunk sizes. Employing regression analysis offers initial parameter estimates. To enhance precision, integrate machine learning (ML) techniques, utilizing supervised learning algorithms and considering features like hosting EN attributes, blockchain protocol, and chunk characteristics. Advanced ML methods, including neural networks, contribute to capturing intricate

relationships. Continuous monitoring and adaptation based on real-world performance data ensure ongoing accuracy. This hybrid empirical-ML approach establishes a robust framework for dynamic and precise estimation of the delay-related parameters in the blockchain edge-cloud network.

The second delay depends on the decided target node n and it is given by:

$$^{NET}T^{n,p}_{i,j,k}=\begin{cases} 0 & ;n=M_{i,j,k} \\ D_{i,j}\times\mathcal{D}^{n,p}_{i,j,k} & ;n\neq M_{i,j,k} \end{cases} \quad (4)$$

where $i\in\mathcal{S}; j\in\mathcal{C}_i; k\in\mathcal{K}; n\in\mathcal{N}; p\in\mathcal{P}^n_{i,j,k}$.

Deploying a CS in this work refers to the transfer of all its CBICs from the distributed hosting nodes to the decided MEC server in order to make it ready to start serving the requester user. Thus, a new incoming service request from a user trigger the service deployment from their hosting nodes to the best available nearby EN.

Hence, the transfer process delay of duplicate $C_{i,j,k}$ to EN n is composed of the blockchain operational delay as well as the transfer delay from the storing nodes to the decided hosting node n. Using Eq. 3 and 4 we can formulate the overall transfer delay $T^{n,p}_{i,j,k}$ related to the duplicate $C_{i,j,k}$ as follows:

$$T^{n,p}_{i,j,k}={}^{BC}T_{i,j,k}+{}^{NET}T^{n,p}_{i,j,k} \quad (5)$$

Lastly, Table II provides an inventory of the primary notations employed in this paper.

TABLE II. MAIN NOTATIONS

Notation	Definition
$\mathcal{N}$	The set of edge-cloud nodes
$\mathcal{S}$	The set of container-based services
σ_n, σ_c, σ_r	The the total number of nodes, CS and resources
$\mathcal{P}_{n\to n'}$	The set of available paths connecting nodes N_n and $N_{n'}$
$\mathcal{B}^p_{n\to n'}$	The bandwidth of path $p\in\mathcal{P}_{n\to n'}$
$\mathcal{H}^p_{n\to n'}$	The hop count of path $p\in\mathcal{P}_{n\to n'}$
$D_{i,j}$	The total data amount of the container, chunk $C_{i,j}$
$\mathcal{P}^n_{i,j,k}$	The set of available paths connecting nodes N_n and $N_{n'}$
$\mathcal{D}^n_{i,j,k}$	The transfer cumulative delay using path $p\in\mathcal{P}^n_{i,j,k}$
$Z^d_{i,m}$	The CS i's demand in terms of resource r_m
$Z^u_{n,m}$	The available resource of node N_n in terms of resource r_m

V. THE OPTIMIZATION PROBLEM

In this section, the fundamental problem that serves as the focal point of our study is articulated. The variables, objectives, and constraints are clearly defined to establish a comprehensive formulation and delineate the overarching objective that requires optimization.

A. Decision Variables

To accurately represent the operations within our system, the decision variables employed in our model are introduced as follows:

The transfer binary decision variable of CS i to node n is denoted α^n_i where $\alpha^n_i=1$ refers to the decision to deploy C_i to node n, otherwise, $\alpha^n_i=0$.

$$\alpha^n_i\in\{0;1\} \qquad ;i\in\mathcal{S}; n\in\mathcal{N} \quad (6)$$

The duplicate choice binary decision variable of chunk duplicate $C_{i,j,k}$ is denoted $\beta_{i,j,k}$ where $\beta_{i,j,k}=1$ refers to the decision to transfer $C_{i,j,k}$, otherwise, $\beta_{i,j,k}=0$.

$$\beta_{i,j,k}\in\{0;1\} \qquad ;i\in\mathcal{S}; j\in\mathcal{C}_i; k\in\mathcal{K} \quad (7)$$

Additionally, when transferring chunk $C_{i,j,k}$ from its caching node $M_{i,j,k}$ to node n the decision variable to select the migration path p among the possible paths set $\mathcal{P}^n_{i,j,k}=\mathcal{P}_{M_{i,j,k}\to n}$ is the binary variable $\gamma^{n,p}_{i,j,k}$ where $\gamma^{n,p}_{i,j,k}=1$ refers to the decision to use the p-th path in $\mathcal{P}^n_{i,j,k}$ to migrate $C_{i,j,k}$ from node $M_{i,j,k}$ to n, otherwise $\gamma^{n,p}_{i,j,k}=0$.

$$\gamma^{n,p}_{i,j,k}\in\{0;1\} \qquad ;i\in\mathcal{S}; j\in\mathcal{C}_i; k\in\mathcal{K}; n\in\mathcal{N}; p\in\mathcal{P}^n_{i,j,k} \quad (8)$$

In situations where the available system resources are insufficient to support the initiation of a CS, our proposed model offers the option to defer the initiation of this CS. To represent this penalization decision for C_i, we utilize the binary variable δ_i, where $\delta_i=0$ indicates penalization, while $\delta_i=1$ denotes no penalization.

$$\delta_i\in\{0;1\} \qquad ;i\in\mathcal{S} \quad (9)$$

B. The Cost Model

In this section, we present the proposed Cost Model, a fundamental component of our system design aimed at comprehensively assessing and optimizing the framework decisions.

1) Delay cost: In our model, when deploying CSs to the designated hosting nodes, the delay cost incurred during the transmission of data transfer flows is contingent on the congestion levels of the network's links. Consequently, with a placement decision vector α, a duplicates selection vector β, and a path selection vector γ as well as using Eq. 5 we can formulate the transfer delay related to the CS i s as follows:

$$T_i(\alpha,\beta,\gamma)=\sum_{j\in\mathcal{C}_i}\sum_{k\in\mathcal{K}}\sum_{n\in\mathcal{N}}\sum_{p\in\mathcal{P}^n_{i,j,k}}\alpha^n_i\beta_{i,j,k}\gamma^{n,p}_{i,j,k}\left(T^{n,p}_{i,j,k}\right) \quad (10)$$

As a result, the comprehensive delay-related cost can be calculated as follows:

$$\Psi^{delay}(\alpha,\beta,\gamma,\delta)=\sum_{i\in\mathcal{S}}\delta_i T_i(\alpha,\beta,\gamma) \quad (11)$$

Additionally, for normalization purpose we use the total maximal allowable transfer delay for all CS given by:

$$\Psi^{delay}_{max}=\sum_{i\in\mathcal{S}}t^{trans}_i \quad (12)$$

2) Backhaul cost: Once all the required container chunks related to S_i have been transferred to destination EN n, the backhaul bandwidth between the connected EN N_i^t and n plays a crucial role in determining the service-related user's perceived delay. Ideally, the user's experience is best when $n = N_i^t$, and we aim to prioritize such decisions. To achieve this, our model takes into account two main costs related to EN n: the available bandwidth between EN n and N_i^t as well as the hop count between them. Utilizing a placement decision vector α and employing Eq. 2, the serving bandwidth for the user linked to CS S_i can be expressed as:

$$\mathcal{B}_i(\alpha) = \sum_{n \in \mathcal{N}} \alpha_i^n \mathcal{B}_i^n \qquad ; i \in \mathcal{S} \tag{13}$$

The resulting backhaul cost function, denoted as $\Psi_{i,n}^{back}$, is defined in such a way that it equals 0 when $n = N_i^t$. However, in all other cases, it takes on a value as follows:

$$\Psi_{i,n}^{back} = \min_{p \in \mathcal{P}_{n \to N_i^t}} \left\{ W_r \frac{\min\limits_{p' \in \mathcal{P}_{n \to N_i^t}} \mathcal{B}^{p'}_{n \to N_i^t}}{\mathcal{B}^{p}_{n \to N_i^t}} + W_h \frac{\mathcal{H}^{p}_{n \to N_i^t}}{\max\limits_{p' \in \mathcal{P}_{n \to N_i^t}} \mathcal{H}^{p'}_{n \to N_i^t}} \right\} \tag{14}$$

Here, $i \in \mathcal{S}$, $n \in \mathcal{N}$ and the weights W_r and W_h are two adjustable weights to fine-tune the optimization process for different scenarios, where W_r represents the weight associated with available bandwidth, and W_h represents the weight associated with hop count costs, with the constraint that $W_r + W_h = 1$. Moreover, the cost function $\Psi_{i,n}^{back}$ falls within the range [0,1], and the use of the max and min expressions for fractions serves the purpose of normalization.

Therefore, with the decision vector α, the overall user backhaul cost can be obtained as:

$$\Psi^{back}(\alpha, \delta) = \sum_{i \in \mathcal{S}} \left\{ \delta_i \sum_{n \in \mathcal{N}} \alpha_i^n \Psi_{i,n}^{back} \right\} \tag{15}$$

3) Load balancing cost: The load balancing procedure targets an equitable distribution of the incoming service workloads, aiming to minimize deviations from the average value. Within our model, the associated cost must consider the processing workloads across all ENs. This includes the current workload of running services on each EN, along with anticipating the additional load expected from services that will be selected to run on them.

Initially, following the CS deployment, the anticipated resource utilization $Z_{n,m}^u$ of processing resource r_m on node n can be determined using the following equation:

$$Z_{n,m}^a(\alpha, \delta) = Z_{n,m}^u + \textstyle\sum_{i \in \mathcal{S}} \alpha_i^n \delta_i Z_{i,m}^d \qquad ; n \in \mathcal{N}; m \in \mathcal{Z}^p. \tag{16}$$

Next, we establish the processing load ratios $\theta_{n,m}$ associated with resource r_m in EN n, along with their mean value $\bar{\theta}_m$, defined as follows:

$$\theta_{n,m}(\alpha, \delta) = \frac{Z_{n,m}^a(\alpha, \delta)}{Z_{n,m}^c} \in [0,1] \qquad ; n \in \mathcal{N}; m \in \mathcal{Z}^p \tag{17}$$

$$\overline{\theta_m}(\alpha, \delta) = \sum_{n \in \mathcal{N}} \frac{\theta_{n,m}(\alpha, \delta)}{\sigma_n} \in [0,1] \qquad ; m \in \mathcal{Z}^p \tag{18}$$

Subsequently, the processing load concerning EN n across all processing resource types in $\mathcal{Z}^p$ is defined as:

$$\Psi_n^{load}(\alpha, \delta) = \sum_{m \in \mathcal{Z}^p} \frac{|\theta_{n,m}(\alpha, \delta) - \overline{\theta_m}(\alpha, \delta)|}{\sigma_z} \qquad ; n \in \mathcal{N} \tag{19}$$

Finally, the resulting overall processing load is as follows:

$$\Psi^{load}(\alpha, \delta) = \sum_{n \in \mathcal{N}} \sum_{m \in \mathcal{Z}^p} \frac{|\theta_{n,m}(\alpha, \delta) - \overline{\theta_m}(\alpha, \delta)|}{\sigma_n \sigma_z} \in [0,1] \tag{20}$$

C. The Multi-Objective Function

The adopted multi-objective function, to be elaborated upon, consists of two components: the overall cost and the degree of penalization for users. We begin by introducing the overall cost model below. It is constructed using a multi-objective function, which combines the detailed cost metrics into a weighted sum using the weight aggregation approach:

$$\Psi(\alpha,\beta,\gamma,\delta) = W_d \frac{\Psi^{delay}(\alpha,\beta,\gamma,\delta)}{\Psi_{max}^{delay}} + W_b \frac{\Psi^{back}(\alpha,\delta)}{\sigma_c} + W_l \Psi^{load}(\alpha,\delta) \tag{21}$$

Here, W_d, W_b, and W_l serve as regulatory weight constants that determine the priority attributed to each cost. Their values range between 0 and 1, satisfying the condition $W_d + W_b + W_l = 1$. Furthermore, the denominators in this function act as normalization factors for each cost function. The metrics normalizing procedure involves transforming the values of the three studied metrics into dimensionless costs between 0 and 1. This ensures the ability to addition and comparison within the cost function by standardizing the metrics' scales.

Next, a penalty function is introduced in order to minimize the number of penalized users. It is proposed such that high-priority users could be penalized, if necessary, only if all non priority users are penalized. Accordingly, we adopt the following penalty function given by:

$$\Pi(\delta) = \sum_{i \in \mathcal{S}} (1 - \delta_i) \pi_i \tag{22}$$

The formulation of this function establishes a penalization hierarchy among the users, where the decision to penalize any CS guarantees that high-priority containers are penalized last, if required.

At this point, we can finally state the following equation, which defines the adopted overall multi-objective function, referring to the objective function targeted for minimization:

$$\Theta(\alpha, \beta, \gamma, \delta) = \Psi(\alpha, \beta, \gamma, \delta) + \Pi(\delta) \tag{23}$$

D. The Constraints

In our model, the case when CS i is not transferred (penalized by setting $\delta_i = 0$) is represented by setting $\alpha_i^n = 0$ for $n \in \mathcal{N}$ and if transferred, only one target node is selected. Accordingly, the transfer decision of CS i has to meet the following constraint:

$$\sum_{n\in\mathcal{N}} \alpha_i^n = \delta_i \qquad ; i \in \mathcal{S} \tag{24}$$

Additionally, if i is not penalized, only one duplicate of chunk $C_{i,j,k}$ must be selected, resulting in the following constraint:

$$\sum_{k\in\mathcal{K}} \beta_{i,j,k} = \delta_i \qquad ; i \in \mathcal{S}; j \in \mathcal{C}_i \tag{25}$$

Also, all the σ_i chunks of CS i, if not penalized, must be transferred to the selected node which lead to the following constraint:

$$\sum_{j\in\mathcal{C}_i}\sum_{k\in\mathcal{K}} \beta_{i,j,k} = \delta_i\sigma_i \qquad ; i \in \mathcal{S} \tag{26}$$

By selecting only one path p in $\mathcal{P}^n_{i,j,k}$ to serve the transfer flow of chunk $C_{i,j,k}$, if i is not penalized, from its node $\mathcal{M}_{i,j,k}$ to target node n, the next constraint has to be satisfied:

$$\sum_{p\in\mathcal{P}^n_{i,j,k}} \gamma^{n,p}_{i,j,k} = \delta_i \qquad ; i \in \mathcal{S}; j \in \mathcal{C}_i; k \in \mathcal{K}; n \in \mathcal{N} \tag{27}$$

Therefore, the anticipated demand for each resource r_m on each node n must fulfill every resource type requirement. This can be formally formulated as:

$$Z^a_{n,m}(\alpha,\delta) \leq Z^c_{n,m} \qquad ; n \in \mathcal{N}; m \in \mathcal{Z}. \tag{28}$$

The subsequent constraint ensures that the total delay incurred during the collection of all chunks associated with CS i does not surpass its tolerated transfer delay t_i^{trans}.

$$\delta_i T_i(\alpha,\beta,\gamma) \leqslant t_i^{trans} \qquad ; i \in \mathcal{S} \tag{29}$$

Finally, the bandwidth constraint associated with CS i, utilizing the maximum available bandwidth specified in Eq. (13), is formalized as follows:

$$\mathcal{B}_i(\alpha) \geqslant \delta_i B_i^{ser} \qquad ; i \in \mathcal{S} \tag{30}$$

E. The Problem Formulation

Based on the elucidated problem context, the following optimization problem, designated as $\mathscr{P}1$, aims to optimize the aforementioned dual objectives. The solution aims to maximize user benefits while respecting resource constraints, meeting transfer delay and bandwidth requirements, and minimizing penalties associated with priority.

$$\mathscr{P}1 : \underset{\{\alpha,\beta,\gamma,\delta\}}{\text{minimize}}\ \Theta(\alpha,\beta,\gamma,\delta)$$
$$\text{s.t. } (6),(7),(8),(9),(24),(25),(26),(27),(28),(29),(30)$$

VI. Problem Decomposition

The general problem $\mathscr{P}1$ involves making decisions regarding CSs placement (α decision variables), selecting CBICs along with path determination (β and γ decision variables) as well as the final penalization decision (δ decision variables). Next, $\mathscr{P}1$ is decomposed into sub-problem as well as a general problem. The sub-problem (CSPDP) deals exclusively with the aspect concerning the selection of CBICs along with path determination and penalization decisions. It is assumed that the decisions regarding container placement have been made. Subsequently, the general problem (GCP) iterates over the placement possibilities and employs the solution from the CSPDP to derive the overall solution.

A. CSPDP Sub-Problem

The formulation of this problem can be presented as follows:

$$\mathcal{CSPDP} : \underset{\{\beta,\gamma,\delta^\alpha\}}{\text{minimize}}\ \Theta^\alpha(\beta,\gamma,\delta^\alpha)$$
$$\text{s.t. } (7),(8),(9),(25),(26),(27),(29)$$

In this formulation, with fixed decisions α, the objective function value is obtained using Eq. (23). Additionally, the penalty vector δ^α pertains only to containers for which resource and bandwidth constraints ((28) and (30) respectively) are satisfied based on decisions α. That is to say, if a certain container i is penalized with decisions α ($\delta_i = 0$), then δ_i^α always equals 0 and is not considered a variable for the CSPDP problem. The remaining penalty decisions of δ^α are related to transfer time constraints that depend on selecting CBICs along with path determination (β and γ decision variables).

Subsequently, a proposition related to the set of feasible solutions of the problem will be demonstrated.

Proposition 1. The set $\mathcal{S}_f^\alpha$ of feasible solutions of $\mathcal{CSPDP}$ is non-empty.

Proof: By utilizing binary decisions, constraints $(7),(8),(9)$ are fulfilled. Additionally, $\forall\alpha$, with a penalty vector $\delta^+ = \mathbf{0}$, resource constraints ((28) and (30)) will be satisfied. Let the decision vectors β^+ and γ^+ be constructed such that $\beta^+ = \mathbf{0}$ and $\gamma^+ = \mathbf{0}$. Since $\delta_i^+ = 0\ \forall i$, all remaining constraints $(25),(26),(27),(29)$ are satisfied. Therefore, $(\beta^+,\gamma^+,\delta^+) \in \mathcal{S}_f^\alpha \Rightarrow \mathcal{S}_f^\alpha \neq \emptyset$. ■

B. GCP Global Problem

The general problem $\mathscr{P}1$ involves traversing all possible decisions given by the containers placement α and solving the sub-problem CSPDP at each iteration. This means that for each potential configuration α, problem CSPDP is solved to obtain a corresponding optimal solution. This iterative process is repeated until all possibilities of α are explored, thereby determining the best global solution for problem $\mathscr{P}1$.

Thus, by using the next constraint, this challenging placement problem, denoted $\mathcal{GCP}$, can be further formulated:

$$(\beta,\gamma,\delta^\alpha) \in \mathcal{S}_f^\alpha \tag{31}$$

and the GCP Global Problem is as follows:

$$\mathcal{GCP} : \underset{\{\alpha\}}{\text{minimize}}\ \Theta^{\alpha}\left(\beta, \gamma, \delta^{\alpha}\right)$$
$$\text{s.t. } (6), (24), (28), (30), (31)$$

However, this problem's placement decisions lies in optimally placing σ_s containers within σ_n edge nodes. This placement must consider resource constraints, including available resources and bandwidth capacity, which contribute to the inherent complexity of the problem. Consequently, the next proposition 2 is derived.

Proposition 2. The optimization problem $\mathcal{GCP}$ is NP-hard.

Proof: Given σ_s services to place within σ_n ENs. To find the best placement solution for problem $\mathcal{GCP}$, even when all resources are sufficient, the problem remains one of determining the optimal placement of σ_s services among σ_n nodes, the computation complexity of $\mathcal{GCP}$ can reach $O(\sigma_n^{\sigma_c})$. Therefore, problem $\mathcal{GCP}$ is NP-hard and cannot be well solved in polynomial time. ■

Subsequently, an efficient resolution of $\mathscr{P}1$ (or its equivalent form $\mathcal{GCP}$) will be presented in the next section.

VII. Problems Resolution

Now, the resolution of the obtained optimization problem $\mathscr{P}1$ is explored. The identified optimization-related challenges will be tackled through a detailed approach, and the derived solutions will be presented. The end of this section will provide insights into our efforts to effectively address the problem and demonstrate our commitment to achieving optimal results.

A. GCP Problem Resolution

Initially, two conditions are established regarding the ENs that can potentially serve as candidates for the placement of service i. The first condition (32) ensures that each EN must satisfy the resource constraints (28) and the service bandwidth constraints (30).

$$\begin{cases} Z_{i,z}^{d} + \sum\limits_{i' \in \mathcal{C} \setminus i} \alpha_{i'}^{n} Z_{i',z}^{d} \leq Z_{n,z}^{c} \quad z \in \mathcal{Z} \\ \max\limits_{p \in \mathcal{P}_{n \to N_i^t}} \left\{ \mathcal{B}_{n \to N_i^t}^{p} \right\} \geqslant B_i^{ser} \end{cases} \tag{32}$$

Based on this first condition (32), we can define the next set $E_i(\alpha)$ associated with CS i as the collection of candidate ENs with their minimum hop count to target node N_i^t .

$$E_i(\alpha) = \left\{ \left(n, \min_{p \in \mathcal{P}_{n \to N_i^t}} \left\{ \mathcal{H}_{n \to N_i^t}^{p} \right\} \right) ; n \in \mathcal{N} \text{ and (32) is satisfied} \right\} \tag{33}$$

A tuple (n, h) is included in this set if EN n is reachable from target node N_i^t with a hop count h and satisfies the conditions in (32).

Then, using $E_i(\alpha)$, we define a second condition to identify a refined set $\mathcal{E}_i(\alpha, \Delta_h)$ of candidate placement ENs relevant to CS i. This condition use a threshold Δ_h to ensure that the minimum number of hops between each potential node in $E_i(\alpha)$ and the target node N_i^t is less than Δ_h. If no nodes satisfy this condition, the set will instead comprise the node from the set $E_i(\alpha)$ that is closest to N_i^t, regardless of this condition, provided such node exist. Since an empty subset signifies an unavoidable penalty for the associated service, this condition guarantees that the set $\mathcal{E}_i(\alpha, \Delta_h)$ is empty only when $E_i(\alpha)$ is also empty. Moreover, the use of the threshold Δ_h is primarily driven by the need to restrict the solution search space, thus excluding trivially non-feasible combinations. Additionally, it serves the purpose of maintaining control over the execution time, particularly in scenarios involving non-feasible configurations.

Accordingly, and in relation to all services, we define a general vector $\mathcal{E}(\alpha, \Delta_h)$ containing all subsets $\mathcal{E}_i(\alpha, \Delta_h)$ as follows:

$$\mathcal{E}(\alpha, \Delta_h) = \{ \mathcal{E}_i(\alpha, \Delta_h) \ / \ i \in \mathcal{S} \} \tag{34}$$

Subsequently, the total number $\chi(\alpha, \Delta_h)$ of placement possibilities for "not yet penalized" services can be derived as:

$$\chi(\alpha, \Delta_h) \leftarrow \prod_{\{x \in \mathcal{E}(\alpha, \Delta_h) \text{ and } x \neq \emptyset\}} (|x|) \tag{35}$$

Now that we've established these key concepts, let's delve into the solution implementation. Here's the Algorithm 1, named GSPA, that outlines the steps involved:

Algorithm 1 : Global Service Placement Algorithm (GSPA)

Require: $\mathcal{S}$, $\mathcal{N}$, $\mathcal{C}$, $\mathcal{K}$, $\mathcal{P}$, $\mathcal{M}$, Ω, $\mathcal{D}$ and Δ_h.
Ensure: Global solution α^*, β^*, γ^*, δ^* with cost $Cost^*$

1: $Cost^* \leftarrow \infty$;
2: build $\mathbf{E}^0 = \mathcal{E}(\mathbf{0}, \Delta_h)$ using (34);
3: $N \leftarrow \chi(\mathbf{0}, \Delta_h)$ using (35);
4: **for** $l = 0$ à $N - 1$ **do**
5: **for** each service i in $\mathcal{S}$ **do**
6: $\alpha_i \leftarrow \mathbf{0}$;
7: **if** $|\mathbf{E}_i^0| = 0$ **then**
8: $\delta_i \leftarrow 0$; //service i penalization
9: **else**
10: $\delta_i \leftarrow 1$; //no penalization of service i
11: n_i is the node within $\mathbf{E}_i^0$ at index $\left(l \ mod \ |\mathbf{E}_i^0|\right)$;
12: $\alpha_i^{n_i} \leftarrow 1$; // service i placement at node n_i
13: $l \leftarrow \ l \ div \ |\mathbf{E}_i^0|$;
14: **end if**
15: **end for**
16: $(\beta, \gamma, \delta, X) \leftarrow subSolution(\alpha, \delta)$;
17: **if** $X < Cost^*$ **then**
18: $(\alpha^*, \beta^*, \gamma^*, \delta^*, Cost^*) \leftarrow (\alpha, \beta, \gamma, \delta, X)$
19: **end if**
20: **end for**
21: **return** $(\alpha^*, \beta^*, \gamma^*, \delta^*, Cost^*)$

In the outer for loop (line 4), all feasible placements are iterated over, with each iteration involving the construction of the placement vector α and the initial penalization decisions δ (lines 5 to 15). Subsequently, the CSPDP sub-problem is resolved, and the current solution along with its cost are updated (lines 16 to 19).

B. CSPDP Sub-Problem Resolution

In this subsection, the process of solving this sub-problem is demonstrated. Three solutions are proposed and detailed: The first involves an exact solution based on a Brute Force Search (BFS) method, while the remaining two utilize approximate solutions based on Simulated Annealing (SA) and Markov Approximation (MA) methods, respectively.

1) Brute-force-search-based scheme: To determine the optimal chunks duplicate Transfer decisions provided by the placement decisions α, we conduct an exhaustive search across all potential solutions using a Brute Force Search for Chunks Duplicates Collection, denoted as BFS-CDCA. Algorithm 2 presents this solution. An exhaustive search over all possible combinations of chunk duplicates and paths is conducted with respect to the placement decision α. This search aims to identify the optimal solution $(\beta^*, \gamma^*, \delta^*)$ with the minimum cost S^*. The cost function is computed based on the decisions made for α, β, γ, and δ, considering the constraints of problem $\mathcal{CSPDP}$. The algorithm iterates, evaluates and updates the optimal solution whenever a lower cost is encountered. The function $minPenalisation$ facilitates the determination of the penalty vector that minimizes the overall objective function Θ, considering that the decisions α, β, and γ have already been established. Finally, the algorithm returns the intermediate optimal solution $(\beta^*, \gamma^*, \delta^*)$ along with the corresponding cost S^*.

Algorithm 2 Brute Force Search based CBIC Distributed Collection Algorithm (BFS-CDCA)

Require: $\mathcal{S}$, $\mathcal{N}$, $\mathcal{C}$, $\mathcal{K}$, $\mathcal{P}$, $\mathcal{M}$, Ω, $\mathcal{D}$ and α.
Ensure: Intermediate solution $(\beta^*, \gamma^*, \delta^*)$ with cost S^*;
1: $S^* \leftarrow \infty$
2: $N \leftarrow$ combinations count;
3: **for** $l = 0$ à $N - 1$ **do**
4: construct decisions vectors β, γ from l;
5: $\delta \leftarrow minPenalisation(\alpha, \beta, \gamma)$
6: **if** constraints of $\mathcal{CSPDP}$ are satisfied **then**
7: $S \leftarrow \Theta(\alpha, \beta, \gamma, \delta)$ according to (23);
8: **if** $S < S^*$ **then**
9: $(\beta^*, \gamma^*, \delta^*, S^*) \leftarrow (\beta, \gamma, \delta, S)$
10: **end if**
11: **end if**
12: **end for**
13: **return** $(\beta^*, \gamma^*, \delta^*, S^*)$

2) Simulated-annealing-based scheme: In Algorithm 3, the proposed heuristic solution based on simulated annealing is introduced. Simulated annealing, a widely utilized optimization technique, is renowned for its simplicity, general applicability, and efficiency compared to alternative methods.

Algorithm 3 uses a probabilistic approach that allows for potential degradation in cost to prevent being trapped in local minima. It utilizes a cost function as an analogy to the energy of a thermodynamic system. Throughout the iteration in the solution space, the acceptance of the current state depends on whether the new state has lower energy. If the new state has higher energy, acceptance is probabilistic, determined by a temperature parameter and the Boltzmann distribution. As the temperature decreases, the system becomes less likely to accept higher energy states. This adjustment in the probability of accepting a penalizing transition seeks a balance between exploring new solutions and exploiting known solutions in the space.

Algorithm 3 : Simulated Annealing based CBIC Distributed Collection Algorithm (SA-CDCA)

Require: $\mathcal{S}$, $\mathcal{N}$, $\mathcal{C}$, $\mathcal{K}$, $\mathcal{P}$, $\mathcal{M}$, Ω, $\mathcal{D}$, Δ, L^{max}, T_0 and α.
Ensure: Intermediate solution $(\beta^*,\gamma^*,\delta^*)$ with cost S^*;
1: Generate initial decisions (β,γ)
2: $\delta \leftarrow minPenalisation(\alpha,\beta,\gamma)$;
3: $S^* \leftarrow \Theta(\alpha,\beta,\gamma,\delta)$ according to (23);
4: **for** l=1 to L^{max} **do**
5: $T \leftarrow T_0 e^{-0.5l\frac{1}{2\Delta}}$;
6: $\beta' \leftarrow rand_neighbour(\beta)$;
7: $(\gamma',\delta') \leftarrow bestTransfer(\alpha,\beta')$;
8: **if** constraints of $\mathcal{CSPDP}$ are satisfied **then**
9: Calculate $S'=\Theta(\alpha,\beta',\gamma',\delta')$ using 23;
10: $\Delta_S \leftarrow S' - S$
11: **if** $\Delta_S<0$ or $e^{\frac{-|\Delta_S|}{T}} \geqslant random(0,1)$ **then**
12: $(\beta,\gamma,\delta,S) \leftarrow \left(\beta',\gamma',\delta',S'\right)$
13: **if** $S<S^*$ **then**
14: $(\beta^*,\gamma^*,\delta^*,S^*) \leftarrow (\beta,\gamma,\delta,S)$
15: **end if**
16: **end if**
17: **end if**
18: **end for**
19: **return** $X^*=(\beta^*,\gamma^*,\delta^*,S^*)$

The algorithm enters a simulated annealing loop (line 4), where it iterates over a specified number of iterations L^{max}. In each iteration, the temperature T is updated based on the current iteration l, and a random neighbor solution β' is generated. The corresponding best γ' is then determined using β'. The penalty vector δ' for the new solution is obtained by minimizing penalization. If the constraints of problem $\mathcal{CSPDP}$ are satisfied, the cost S' is calculated, and a decision is made based on the Metropolis criterion to accept or reject the new solution. If accepted, the current solution is updated, and if it improves the global solution, it is recorded. Finally, the algorithm returns the global intermediate solution $X^* = (\beta^*, \gamma^*, \delta^*)$ along with the corresponding cost S^*.

3) Markov approximation-based scheme: Markov Approximation is a technique used in optimization to approximate complex problems by simplifying them into a Markov chain model. It is particularly useful for problems with large solution spaces where exact methods become computationally infeasible. In this method, the problem is represented as a Markov decision process, where each state corresponds to a possible solution, and transitions between states are determined by a stochastic process based on the problem's constraints and objectives. By iteratively updating the probabilities of transitioning between states, the algorithm converges towards an optimal or near-optimal solution.

a) Log-sum-exp approximation: Now, we delve into a Log-sum-exp method aimed at approximating the mathematical expression of the obtained problem. Let the configuration $c = \{\alpha; \beta; \gamma; \delta\} \in \mathcal{S}_f^\alpha$ be a solution, where $\mathcal{S}_f^\alpha$ is the Feasible Solution Set (FSS) of problem $\mathcal{CSPDP}$ with objective

function Θ^{α}. Sure, problem $\mathcal{CSPDP}$ is equivalent to:

$$\min_{c\in\mathcal{S}_f^{\alpha}} \Theta^{\alpha}(c) \tag{36}$$

According to Appendix A of [28], by associating a probability p_c with the adoption of a configuration c, the optimal solution of the problem $\max_{c\in\mathcal{S}_f^{\alpha}} \Theta^{\alpha}(c)$ is the same as that of the problem $\max_{p\geq 0} \sum_{c\in\mathcal{S}_f^{\alpha}} p_c\Theta^{\alpha}(c)$ where $\sum_{c\in\mathcal{S}_f^{\alpha}} p_c = 1$. It follows that the optimal solution of problem $\mathcal{CSPDP}$ is the same as that of the following problem:

$$\min_{p\geq 0} \sum_{c\in\mathcal{S}_f^{\alpha}} p_c\Theta^{\alpha}(c) \text{ subject to } \sum_{c\in\mathcal{S}_f^{\alpha}} p_c = 1. \tag{37}$$

Here $p = (p_c)_{c\in\mathcal{S}_f^{\alpha}}$ is a probability distribution associated with the possibility universe $\mathcal{S}_f^{\alpha}$.

Subsequently, the formulation of the log-sum-exp approximation can be derived such that [29]:

- Firstly, for any strictly positive constant τ, we have:

$$0 \leq \min_{c\in\mathcal{S}_f^{\alpha}} \Theta^{\alpha}(c) + \tfrac{1}{\tau}\ln\left(\sum_{c\in\mathcal{S}_f^{\alpha}} e^{-\tau\Theta^{\alpha}(c)}\right) \leq \tfrac{\ln\left|\mathcal{S}_f^{\alpha}\right|}{\tau} \tag{38}$$

- Secondly, let g_{τ} be the log-sum-exp function defined on $\mathbb{R}^m$ by:

$$g_{\tau}(x_1;...;x_m) = \frac{1}{\tau} ln(\sum_{1\leq i\leq m} e^{\tau x_i}) \tag{39}$$

  where $m = |\mathcal{S}_f^{\alpha}|$. As $\tau \to \infty$ and according to Eq. (38), the approximation gap $\frac{\ln|\mathcal{S}_f^{\alpha}|}{\tau} \to 0$, and thus the proposed approximation in Eq. 38 becomes exact.

- Thirdly, by using the conjugate function of g_{τ} and according to [30]-p.93, we can find that:

$$g_{\tau}(x) = \max_{\left\{p\geq 0 \text{ s.t. } \sum_{1\leq i\leq m} p_i=1\right\}} \sum_{1\leq i\leq m} p_i x_i - \tfrac{1}{\tau} p_i ln(p_i) \tag{40}$$

- Finally, the approximation defined by Eq. 40 is also the solution to the following optimization problem:

$$\min_{\left\{p\geq 0 \text{ s.t. } \sum_{c\in\mathcal{S}_f^{\alpha}} p_c=1\right\}} \sum_{c\in\mathcal{S}_f^{\alpha}} p_c\Theta^{\alpha}(c) + \frac{1}{\tau}\sum_{c\in\mathcal{S}_f^{\alpha}} p_c ln(p_c) \tag{41}$$

b) The Markov chaine: : In this subsection, the Metropolis-Hastings algorithm will be employed to construct an irreducible Markov chain $(X_n)_{n\geq 0}$ with state space S_f. This chain will have p^* as its reversible probability distribution, thereby ensuring its stationarity.

Firstly, the Lagrangian of problem (41) is given by:

$$L(p,\lambda) = \sum_{c\in\mathcal{S}_f^{\alpha}} p_c\Theta^{\alpha}(c) + \frac{1}{\tau}\sum_{c\in\mathcal{S}_f^{\alpha}} p_c\ln(p_c) + \lambda\left(\sum_{c\in\mathcal{S}_f^{\alpha}} p_c - 1\right) \tag{42}$$

where λ is the Lagrange multiplier [29].

Secondly, solving the Karush-Kuhn-Tucker conditions yields the following two equations $\sum_{c\in\mathcal{S}_f^{\alpha}} p_c^* = 1$ and $\Theta^{\alpha}(c) + \frac{1}{\tau}(\ln({p_c}^*) + 1) + \lambda^* = 0 \left(\forall c \in \mathcal{S}_f^{\alpha}\right)$, where $({p_c}^*)_{c\in S_{\alpha}}$ is the optimal solution of the primal problem and λ^* is the optimal solution of the dual problem. Thus, from the second equation we can get:

$$p_c^* = e^{-\tau(\Theta^{\alpha}(c)+\lambda^*)-1}; \forall c \in \mathcal{S}_f^{\alpha} \tag{43}$$

This result and the condition $\sum_{c\in\mathcal{S}_f^{\alpha}} p_c^* = 1$, give the result:

$$\lambda^* = \frac{1}{\tau}\left(\ln\left(\sum_{c\in\mathcal{S}_f^{\alpha}} e^{-\tau\Theta^{\alpha}(c)}\right) - 1\right) \tag{44}$$

Finally, combining Eq. (43) and (44) leads to the following probability distribution:

$$p_c^* = \frac{e^{-\tau\Theta^{\alpha}(c)}}{\sum_{u\in\mathcal{S}_f^{\alpha}} e^{-\tau\Theta^{\alpha}(u)}}; \forall c \in \mathcal{S}_f^{\alpha} \tag{45}$$

here p_c^* represents the optimal solution of problem (Eq. 41) and thus a quasi-optimal solution of problem (Eq. 36).

c) MA-CDCA algorithm: Our problem-solving process starts with constructing an initial configuration, denoted as c_0. Once c_0 is established, we employ the Metropolis-Hastings algorithm to create an irreducible Markov chain, represented by $(X_n)_{n\geq 0}$. This Markov chain has the desirable property of converging to a specific stationary probability distribution, denoted by p^*. The core of the algorithm lies in a probabilistic approach based on the Metropolis-Hastings principle, which is described in detail next as follows :

if at time t, we have $X_t = c \in \mathcal{S}_f^{\alpha}$, a state c' is randomly drawn from $\mathcal{S}_f^{\alpha}$ according to the distribution $q_{c,c'}$ given by:

$$q_{c,c'} = \frac{p_c^* + p_{c'}^*}{|\mathcal{S}_f^{\alpha}| - 1} \tag{46}$$

then we calculate the acceptance probability $A_{c,c'} = min\left\{\frac{p_{c'}^* q_{c',c}}{p_c^* q_{c,c'}}; 1\right\}$ simplified as:

$$A_{c,c'} = min\left\{e^{-\tau(\Theta^{\alpha}(c')-\Theta^{\alpha}(c))}; 1\right\} \tag{47}$$

Hence, we accept the transition $X_{t+1} = c'$ with probability $A_{c,c'}$ and reject it with probability $1 - A_{c,c'}$.

The pseudo-code of the solution is presented in Algorithm 4.

Algorithm 4 : Markov Approximation based CBIC Distributed Collection Algorithm (MA-CDCA)

Require: $\mathcal{N}$, $\mathcal{C}$, $\mathcal{K}$, $\mathcal{P}$, $\mathcal{M}$, Ω, $\mathcal{D}$, T^{max} and α.
Ensure: Intermediate solution $X^*=(\beta^*,\gamma^*,\delta^*)$;
1: Generate an initial chunks selection (β_0) based on α;
2: $(\gamma_0,\delta_0)\leftarrow bestTransfer(\alpha,\beta_0)$;
3: Build the initial solution $X_0=(\alpha,\beta_0,\gamma_0,\delta_0)$;
4: **for** t=0 to T^{max} **do**
5: Select CS i such that $CS_Gain(X_t,i)>0$;
6: Generate a chunks selection $\beta_i(t+1)$ related to CS i such that $Chunk_Gain(X_t,i,\beta_i(t+1))>0$;
7: Update $\beta^{'}$ according to shift $\beta_i(t)\rightarrow\beta_i(t+1)$;
8: $(\gamma^{'},\delta^{'})\leftarrow bestTransfer(\alpha,\beta^{'})$;
9: Build the new configuration $c^{'}\leftarrow(\alpha,\beta^{'},\gamma^{'},\delta^{'})$
10: Calculate acceptation probability $A_{X_t,c'}$ using 47;
11: **if** $A_{X_t,c'}>$random(0,1) **then**
12: Accept the new transition $X_t\rightarrow c'$ and set $X_{t+1}=c'$
13: **else**
14: Reject the new transition $X_t\rightarrow c'$ and set $X_{t+1}=X_t$
15: **end if**
16: **end for**
17: **return** $X^*=X_{T^{max}+1}$

VIII. Evaluation

To evaluate the proposed solutions, we conducted a series of experiments using a range of metrics. These experiments were meticulously designed to test each solution under various conditions, ensuring a comprehensive assessment.

A. Simulation Setup

All the experiments were conducted on a personal computer running Windows 10, equipped with a 2.4GHz Intel Core i5 processor and 16GB of RAM. This setup ensured a consistent and controlled environment for testing, minimizing the potential influence of hardware variability on the performance metrics.

The parameters of the simulation experiments are detailed in Table III. These parameters were meticulously chosen to reflect realistic and challenging conditions for the proposed solutions.

TABLE III. Simulation Parameters

Parameter	Values
σ_s;σ_n;σ_i; D_i	variable
σ_r	3 resource types: CPU, RAM, Storage
$\lvert\mathcal{P}_{n\rightarrow n'}\rvert$	$\in [\![3;5]\!]$
τ	10^{10}
W_r; W_h	0.5; 0.5
W_d; W_b; W_l	0.5; 0.5; 0.5
L^{Bloc}	10.0 MB
$a_{i,j}$	$\in [5,30]\times 10^{-3}$ s/MB
$b_{i,j}$	$\in [1,5]\times 10^{-3}$ s/MB
$c_{i,j}$	$\in [1,10]\times 10^{-4}$ s

B. Cost and Execution Time Analysis

This section presents a comparative analysis of the three solutions: BFS-CDCA, MA-CDCA, and SA-CDCA focusing on their normalized cost and execution time under varying conditions. In this analysis, we consider a scenario with 8 services (users) and a variable count of edge nodes σ_n ranging from 2 to 9. For all services i, a duplication factor $D_i = 2$ is used and two settings for chunks fragmentation are considered, $\sigma_i \in \{3;5\}$, and these values are used as suffixes in the names of each solution to denote their specific configurations. For example, MA-CDCA-3 and MA-CDCA-5 refer to the Markov Approximation-based solution with a chunk fragmentation settings of $\sigma_i = 3$ and $\sigma_i = 5$, respectively.

Fig. 2 shows how the average total cost changes with the number of edge nodes.

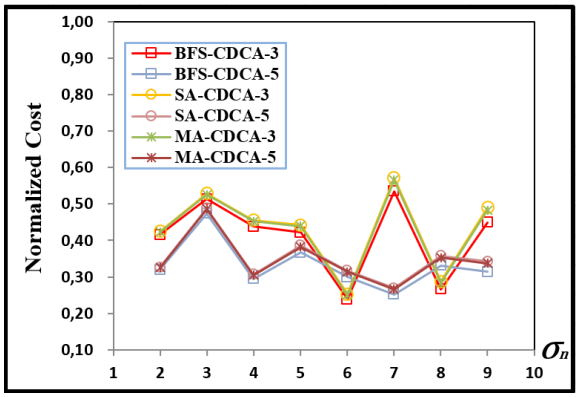


Fig. 2. Normalized Cost; $\sigma_s = 8$, $D_i = 2$, $\sigma_i \in \{3;5\}$.

The results indicate a significant similarity between the three solutions' performance, particularly for $\sigma_n \in \{2;3;4\}$ across both settings of σ_i. Furthermore, for $\sigma_c \in \{5;6;7;8;9\}$, the experiment demonstrates that both MA-CDCA and SA-CDCA solutions achieve costs within 1.1% and 1.3% margin of difference compared to the exact BFS-CDCA solution, respectively.

Likewise, Fig. 3 shows how average execution time changes with the number of edge nodes σ_n.

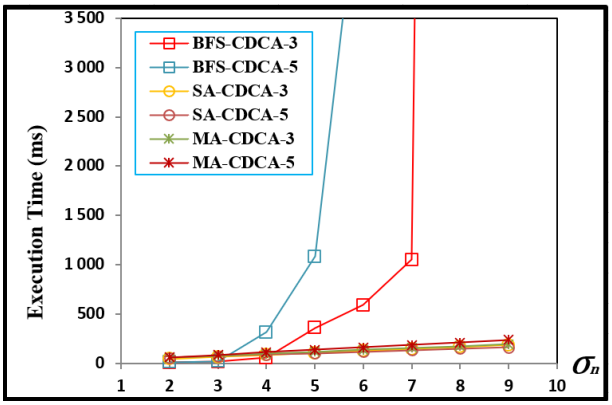


Fig. 3. Execution Time; $\sigma_s = 8$, $D_i = 2$, $\sigma_i \in \{3;5\}$.

The trend in Fig. 3 is confirmed by the data. Execution time increases with the number of edge nodes σ_n, particularly for the BFS-CDCA solution, which exhibits exponential growth. In contrast, the MA-CDCA and SA-CDCA solutions demonstrate stable execution times. For example, with a fragmenta-

tion parameter σ_i of 3 and 9 edge nodes, MA-CDCA-3 and SA-CDCA-3 require only $210.0ms$ and $229.0ms$, respectively, while the exact BFS-CDCA solution reaches impractically high execution times of $6452677.1ms$. Similarly, with a fragmentation parameter of 5 and 9 edge nodes, MA-CDCA-5 and SA-CDCA-5 require only $320.0ms$ and $389.0ms$, respectively, whereas the exact BFS-CDCA solution reaches impractically high execution times of $10877142.4ms$.

While the number of edge nodes σ_n has the most significant impact on execution time, the results also reveal the influence of the fragmentation parameter σ_i. As σ_i increases, all solutions show a trend of requiring more execution time. This suggests that careful consideration must be given to the fragmentation parameter σ_i to avoid additional processing overhead for the adopted solutions.

C. Service Provision Analysis

Now, this section examines the satisfaction rate for service provision requests, focusing on three main factors: fragmentation, duplication, and service priority. The influence of these factors is analyzed to determine their effects on the overall performance of both the system and the proposed solutions.

1) Chunks count and duplication factors: In this experiment, the impact of fragmentation on request satisfaction rates is investigated. This is achieved by controlling two factors related to each service: the per-service duplication factor D_i and The chunk count factor σ_i. In essence, the experiment examines how dividing services into smaller chunks (fragmentation) affects service requests satisfaction. For each service i, the D_i factor is set in $\{1; 3\}$) while the chunk count was varied from $\sigma_i = 1$ (where only a single piece of service data is considered) to $\sigma_i = 8$. Also, the number of services is set to either $\sigma_s = 8$ or $\sigma_s = 16$. These services are deployed across 10 edge nodes ($\sigma_n = 10$) and are assigned equal priority. Moreover, considering the significant impact of data size on the studied timing metric, the CSs' data sizes A_i are generated within the range of [1, 8] MB. For the configuration where $\sigma_s = 8$, the obtained average data size A_i is 4.6 MB, whereas for the $\sigma_s = 16$ configuration, the average is 5.7 MB.

Fig. 4 shows how the Average Collection Time of the selected chunks changes with the number of chunks.

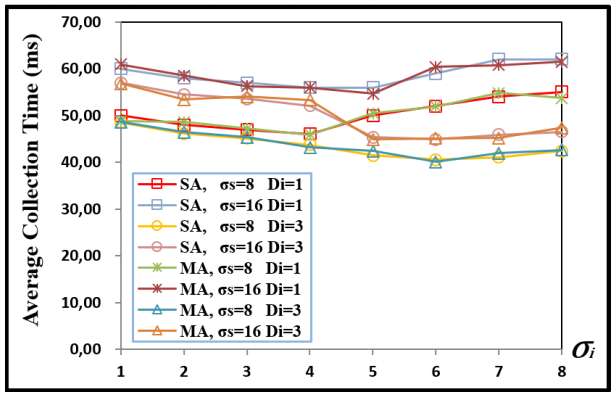


Fig. 4. Average Chunks Collection Time; $\sigma_s \in \{8; 16\}$, $\sigma_n = 10$, $D_i \in \{1; 3\}$.

A preliminary analysis of the results reveals a high degree of similarity between the SA-CDCA and MA-CDCA solutions across most configurations examined. This suggests that both approaches achieve comparable efficiency. For fragmentation levels (σ_i) ranging from 1 to 4, the average collection time consistently decreases across all combinations of service count (σ_s) and duplication factor (D_i). This suggests that a moderate level of fragmentation may improve efficiency in collecting data. In fact, regardless of the service count (σ_s), increasing the duplication factor (D_i) from 1 to 3 consistently reduces the average collection time. This implies that data redundancy introduced by duplication might be beneficial for faster collection. Interestingly, the behavior changes beyond a fragmentation level of 4. Indeed, when $D_i = 1$ (no duplication), the collection time tends to increase regardless of the service count (σ_s). This suggests that excessive fragmentation without redundancy becomes detrimental for collection efficiency. Conversely, with three duplicates of each chunk ($D_i = 3$), the collection time continues to decrease until a fragmentation level of 6, after which it slightly increases. This implies that a higher level of redundancy can tolerate a wider range of fragmentation levels before seeing a performance drop. These observations highlight the importance of considering both fragmentation and duplication when optimizing data collection strategies.

2) Service priority factor: The next experiment investigates the influence of service priority under critical resource limitations, where the available resources can meet the requirements of only two CS. Twenty services, all requiring the same resources, were divided into two groups of ten. The first group (GU) received a uniform priority ($\pi_i = 1$). The second group (GG) received graded individual priorities assigned from 10 (highest) to 1 (lowest). The experiment was conducted over five rounds, with served services removed from the pool for subsequent rounds. This setup allowed the study of the global penalization function, which evaluates the overall penalty incurred due to unmet service requests based on priority levels. Additionally, the obtained average normalized cost was analyzed to understand how the changing priority of the variable service affects the efficiency of service delivery.

Fig. 5 illustrates how the Average Normalized Cost and the overall penalization evolve as the rounds progress.

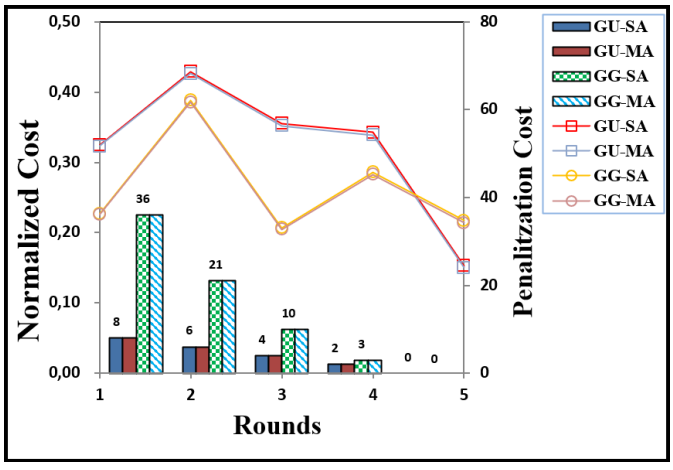


Fig. 5. Normalized cost (line chart), penalization cost (bar chart); $\sigma_s = 20$, $\sigma_n = 12$, $D_i = 2$, $\sigma_i = 2$.

In terms of Normalized Cost, the two algorithms, SA-CDCA and MA-CDCA, yield nearly identical costs for the cost related objective function. However, the MA-CDCA algorithm has a minor edge in terms of cost making it slightly more effective. Furthermore, the penalization cost for all priority groups (GU, GG) and algorithms (SA, MA) consistently decreases across the rounds. This indicates a general improvement in performance over execution rounds in terms of penalized CS. Indeed, in the GU group, where all equipment possesses a uniform priority ($\pi_i = 1$), the consistent decrease in penalization cost by 2 units per round implies that two CS are being served consistently each round. Similarly, in the GG group, the decreasing rate in penalization cost implies that the highest priority equipment is being served consistently, leading to a reduction in penalties. In other words, the adopted prioritization strategy effectively prioritizes high-priority CS.

D. Blockchain-based Caching Analysis

The third experiment investigates how replicating services influences time overhead. To understand the effect of varying fragmentation levels, four different chunks' count scenarios ($\sigma_i = 1$ to $\sigma_i = 4$) are considered, each with 1 and 3 duplicate scenarios for each of the σ_s services ($\sigma_s \in \{8, 16\}$). This experiment aims to analyze how different levels of fragmentation and duplication impact the overall time overhead. Throughout the experiment, all services are assigned the same priority, and a fixed number of edge nodes ($\sigma_n = 10$) is maintained. For each service, the Blockchain-related (BC) Time Overhead is calculated as the maximum operational time across all the involved blockchain nodes, as given by equation 3. The maximum is taken because BC operations are performed in parallel. Then, the total time overhead for the system is obtained by summing these maximum times for all σ_s services.

Fig. 6 shows how the Overall Blockchain-related Time Overhead changes as the fragmentation progresses.

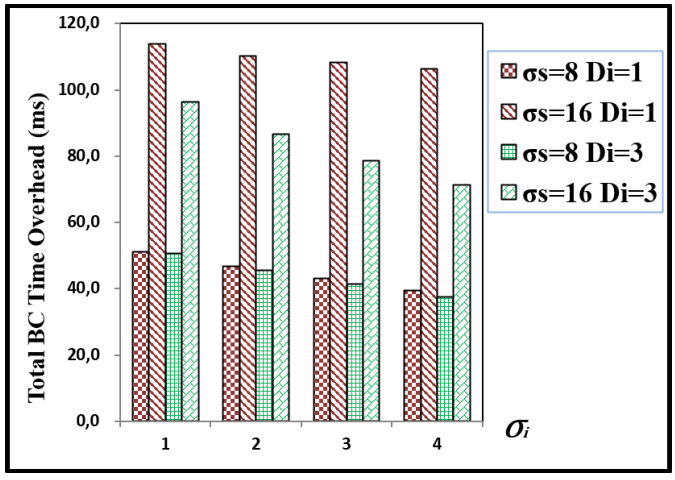


Fig. 6. Overall blockchain-related time overhead; $\sigma_s \in \{8; 16\}$, $\sigma_i \in \{1; 2; 3; 4\}$, $\sigma_n = 10$, $D_i \in \{1; 3\}$.

As the services are divided into an increasing number of chunks (σ_i), the time overhead associated with the blockchain operations decrease. Indeed, with lower fragmentation levels, the time overhead is relatively high due to fewer parallel operations being performed. However, as the fragmentation level increases, more chunks are processed in parallel across the blockchain nodes, resulting in a reduction in processing time. This trend is observed consistently, regardless of the number of services or duplicates involved. Additionally, the total time overhead clearly increases with a higher number of services, as more cache blocs are processed. Conversely, the total time overhead decreases with an increase in the number of duplicates, as the redundant data allows for more flexible decisions regarding the timing characteristics of the caching blockchain nodes. Hence, the results clearly highlight the positive impact of service fragmentation and replication on reducing the time overhead caused by blockchain-related operations. These insights are essential for the development of efficient cache management mechanisms within the investigated blockchain-based systems.

IX. Conclusions and Perspectives

This paper investigates the fusion of Multi-access Edge Computing (MEC) capabilities with blockchain technology to address the low-latency and safety requirements of Smart Devices (SDs). The emphasis was placed on examining the role of a split-duplicate-cache method within a secure blockchain-enhanced MEC system, with the objective of enhancing service provision. An optimization problem was derived and solved with the objective to reduce the backhaul bandwidth and the number of hops between the serving node and the selected deployment node. The experimentation assessed the performance of the three proposed solutions: MA-CDCA, SA-CDCA, and BFS-CDCA. The analysis revealed that BFS-CDCA performed best in smaller-scale settings, while MA-CDCA and SA-CDCA showed efficient execution times, especially in configurations with a high number of CBICs, replicas, ENs, and CSs. Additionally, it was demonstrated that the fragmentation-replication methodology positively impacts the Blockchain operational time overhead.

Future research could focus on evaluating the scalability of MA-CDCA and SA-CDCA in larger, more complex scenarios and assessing their real-world deployment in smart cities and IoT networks. Additionally, exploring the security implications, energy efficiency, and impact on end-user experience could provide valuable insights into the practical benefits and limitations of these solutions.

References

[1] Q.-V. Pham, F. Fang, V. N. Ha, M. J. Piran, M. Le, L. B. Le, W.-J. Hwang, and Z. Ding, "A survey of multi-access edge computing in 5g and beyond: Fundamentals, technology integration, and state-of-the-art," *IEEE access*, vol. 8, pp. 116 974–117 017, 2020.

[2] Y. Hmimz, T. Chanyour, M. El Ghmary, and M. O. Cherkaoui Malik, "Energy efficient and devices priority aware computation offloading to a mobile edge computing server," in *2019 5th International Conference on Optimization and Applications (ICOA)*, 2019, pp. 1–6.

[3] T. Chanyour, Y. Hmimz, M. El Ghmary, and M. O. Cherkaoui Malki, "Delay-aware and user-adaptive offloading of computation-intensive applications with per-task delay in mobile edge computing networks," *International Journal of Advanced Computer Science and Applications*, vol. 11, no. 1, 2020. [Online]. Available: http://dx.doi.org/10.14569/IJACSA.2020.0110190

[4] V. K. Kaliappan, S. Yu, R. Soundararajan, S. Jeon, D. Min, and E. Choi, "High-secured data communication for cloud enabled secure docker image sharing technique using blockchain-based homomorphic encryption," *Energies*, vol. 15, no. 15, p. 5544, 2022.

[5] T. Chanyour and M. O. Cherkaoui Malki, "Deployment and migration of virtualized services with joint optimization of backhaul bandwidth and load balancing in mobile edge-cloud environments," *International Journal of Advanced Computer Science and Applications*, vol. 12, no. 3, 2021. [Online]. Available: http://dx.doi.org/10.14569/IJACSA.2021.0120368

[6] G. Wang, C. Li, Y. Huang, X. Wang, and Y. Luo, "Smart contract-based caching and data transaction optimization in mobile edge computing," *Knowledge-Based Systems*, vol. 252, p. 109344, 2022.

[7] J. Guo, C. Li, and Y. Luo, "Blockchain-assisted caching optimization and data storage methods in edge environment," *The Journal of Supercomputing*, vol. 78, no. 16, pp. 18 225–18 257, 2022.

[8] R. Aghazadeh, A. Shahidinejad, and M. Ghobaei-Arani, "Proactive content caching in edge computing environment: A review," *Software: Practice and Experience*, vol. 53, no. 3, pp. 811–855, 2023.

[9] H. Chai, S. Leng, M. Zeng, and H. Liang, "A hierarchical blockchain aided proactive caching scheme for internet of vehicles," in *ICC 2019-2019 IEEE International Conference on Communications (ICC)*. IEEE, 2019, pp. 1–6.

[10] S. Chen, Y. Zheng, W. Lu, V. Varadarajan, and K. Wang, "Energy-optimal dynamic computation offloading for industrial iot in fog computing," *IEEE Transactions on Green Communications and Networking*, vol. 4, no. 2, pp. 566–576, 2019.

[11] R. Malik and M. Vu, "On-request wireless charging and partial computation offloading in multi-access edge computing systems," *IEEE Transactions on Wireless Communications*, vol. 20, no. 10, pp. 6665–6679, 2021.

[12] Y. Liu, Q. He, D. Zheng, X. Xia, F. Chen, and B. Zhang, "Data caching optimization in the edge computing environment," *IEEE Transactions on Services Computing*, vol. 15, no. 4, pp. 2074–2085, 2020.

[13] Z. Chen and Z. Zhou, "Dynamic task caching and computation offloading for mobile edge computing," in *GLOBECOM 2020-2020 IEEE Global Communications Conference*. IEEE, 2020, pp. 1–6.

[14] G. Zhang, S. Zhang, W. Zhang, Z. Shen, and L. Wang, "Joint service caching, computation offloading and resource allocation in mobile edge computing systems," *IEEE Transactions on Wireless Communications*, vol. 20, no. 8, pp. 5288–5300, 2021.

[15] X. Ma, A. Zhou, S. Zhang, and S. Wang, "Cooperative service caching and workload scheduling in mobile edge computing," in *IEEE INFOCOM 2020-IEEE Conference on Computer Communications*. IEEE, 2020, pp. 2076–2085.

[16] P. Yuan, S. Shao, L. Geng, and X. Zhao, "Caching hit ratio maximization in mobile edge computing with node cooperation," *Computer Networks*, vol. 200, p. 108507, 2021.

[17] S. Zhong, S. Guo, H. Yu, and Q. Wang, "Cooperative service caching and computation offloading in multi-access edge computing," *Computer Networks*, vol. 189, p. 107916, 2021.

[18] H. Feng, S. Guo, L. Yang, and Y. Yang, "Collaborative data caching and computation offloading for multi-service mobile edge computing," *IEEE Transactions on Vehicular Technology*, vol. 70, no. 9, pp. 9408–9422, 2021.

[19] Y. Huang, X. Song, F. Ye, Y. Yang, and X. Li, "Fair and efficient caching algorithms and strategies for peer data sharing in pervasive edge computing environments," *IEEE Transactions on Mobile Computing*, vol. 19, no. 4, pp. 852–864, 2019.

[20] A. Asheralieva and D. Niyato, "Combining contract theory and lyapunov optimization for content sharing with edge caching and device-to-device communications," *IEEE/ACM Transactions on Networking*, vol. 28, no. 3, pp. 1213–1226, 2020.

[21] J. Yin, M. Zhan, Z. Zhang, L. Wang, D. Zhang, and X. Xiao, "Research on the content sharing system for mobile edge caching networks: a hierarchical architecture," in *2022 15th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI)*. IEEE, 2022, pp. 1–6.

[22] S. Zarandi and H. Tabassum, "Federated double deep q-learning for joint delay and energy minimization in iot networks," in *2021 IEEE International Conference on Communications Workshops (ICC Workshops)*. IEEE, 2021, pp. 1–6.

[23] L. Cui, X. Su, Z. Ming, Z. Chen, S. Yang, Y. Zhou, and W. Xiao, "Creat: Blockchain-assisted compression algorithm of federated learning for content caching in edge computing," *IEEE Internet of Things Journal*, vol. 9, no. 16, pp. 14 151–14 161, 2020.

[24] T. Chanyour and A. Kaddari, "Blockchain-based distributed caching with replication for efficient service provision in edge-cloud environments," in *Proceedings of the 6th International Conference on Networking, Intelligent Systems & Security*, ser. NISS '23. Larache, Morocco: Association for Computing Machinery, 2023. [Online]. Available: https://doi.org/10.1145/3607720.3607768

[25] Z. Ma, S. Shao, S. Guo, Z. Wang, F. Qi, and A. Xiong, "Container migration mechanism for load balancing in edge network under power internet of things," *IEEE Access*, vol. 8, pp. 118 405–118 416, 2020.

[26] F. Wilhelmi, S. Barrachina-Muñoz, and P. Dini, "End-to-end latency analysis and optimal block size of proof-of-work blockchain applications," *IEEE Communications Letters*, vol. 26, no. 10, pp. 2332–2335, 2022.

[27] T. Pflanzner, H. Baniata, and A. Kertesz, "Latency analysis of blockchain-based ssi applications," *Future Internet*, vol. 14, no. 10, p. 282, 2022.

[28] W. Pu, X. Li, J. Yuan, and X. Yang, "Resource allocation for millimeter wave self-backhaul network using markov approximation," *IEEE Access*, vol. 7, pp. 61 283–61 295, 2019.

[29] X. Li and C. Zhang, "Semi-dynamic markov approximation-based base station sleep with user association for heterogeneous networks," *IET Communications*, vol. 17, no. 6, pp. 704–711, 2023.

[30] S. P. Boyd and L. Vandenberghe, *Convex optimization*. Cambridge university press, 2004.

A Data Augmentation Approach to Sentiment Analysis of MOOC Reviews

Guangmin Li[1], Long Zhou[2], Qiang Tong[3], Yi Ding[4], Xiaolin Qi[5], Hang Liu[6]
School of Computer and Information Engineering, Hubei Normal University, Huangshi, China[1,2,3,4]
Academic Affairs Office, Wuhan Technology and Business University, Wuhan, China[5]
College of Physical Science and Technology, Central China Normal University, Wuhan, China[6]

***Abstract*—To address the lack of Chinese online course review corpora for aspect-based sentiment analysis, we propose Semantic Token Augmentation and Replacement (STAR), a semantic-relative distance-based data augmentation method. STAR leverages natural language processing techniques such as word embedding and semantic similarity to extract high-frequency words near aspect terms, learns their word vectors to obtain synonyms and replaces these words to enhance sentence diversity while maintaining semantic consistency. Experiments on a Chinese MOOC dataset show STAR improves Macro-F1 scores by 3.39%-8.18% for LCFS-BERT and 1.66%-8.37% for LCF-BERT compared to baselines. These results demonstrate STAR's effectiveness in improving the generalization ability of deep learning models for Chinese MOOC sentiment analysis.**

***Keywords*—Data augmentation; sentiment analysis; MOOC; natural language processing; deep learning**

I. Introduction

The rise of information technology has facilitated the sharing of experiences via online platforms, leading to significant growth in User Generated Content (UGC). Researchers have utilized Natural Language Processing (NLP) and machine learning to extract valuable insights from UGC on topics such as product attribute extraction [1], [2], [3], consumer preference patterns [4], [5], [6], and public sentiment monitoring [7], [8]. These studies enhance text mining applications and support data-driven consumer behavior analysis, product improvement, and market strategies.

Massive Open Online Courses (MOOCs) have generated extensive online course reviews, providing insights into learner preferences and teaching effectiveness. Analyzing these reviews aids educators in improving courses. Automated sentiment analysis, using machine learning algorithms, efficiently processes large volumes of review data, offering insights for student feedback [9], course evaluation [10], and teaching quality assessment [11]. These techniques facilitate data-driven decision-making in education.

Despite the growing interest in sentiment analysis for Chinese MOOC review data, challenges still exist including a scarcity of annotated corpora, leading to overfitting and class imbalance issues. Data augmentation (DA) techniques can expand training datasets while preserving labels. Although DA techniques have been applied in various NLP tasks, including natural language inference [12], [13] and sentiment analysis [14], [15], [16], there remains a research gap in DA strategies for Chinese educational review data.

This paper proposes a novel approach, Semantic Token Augmentation and Replacement (STAR), to address Chinese text augmentation challenges in the educational domain. STAR uses semantic relative distance calculation to augment training datasets and balance sentiment polarity in Chinese MOOC reviews.

The main contributions of this work are as follows:

1) We propose STAR, a novel augmentation method that utilizes external knowledge bases and semantic relative distance calculation to rapidly augment small sample datasets. STAR maintains the semantic accuracy of original review sentences while balancing sentiment distribution in augmented samples. We demonstrate its effectiveness across three BERT model variants (LCFS-BERT, LCF-BERT, and BERT-SPC) for Chinese MOOC sentiment analysis.
2) We conduct a comparative analysis of Word2Vec and BERT Encoding for word vector generation in STAR, employing uniform sampling for synonym replacement. This comparison provides insights into the optimal word embedding approach for data augmentation in aspect-based sentiment analysis of Chinese MOOC reviews.
3) To the best of our knowledge, we first present the synonym replacement-based data augmentation to Chinese review datasets, demonstrating its effectiveness in improving aspect-based sentiment analysis performance in this specific domain.

The rest of this paper is organized as follows: Section II presents related work. Section III and Section IV describe the proposed method in detail. The experimental results and analysis are presented in Section V and Section VI. Section VII concludes our study and points out the research work in the future.

II. Related Work

Data augmentation aims to increase the diversity of training samples by synthesizing new data from existing datasets [17], [18]. It has been widely applied in various NLP tasks, including named entity recognition[19], [20], natural language inference [21], and sentiment analysis [22], [23]. Wei et al. [24] introduced an Easy Data Augmentation (EDA) that employs methods like synonym replacement, random insertion, deletion, and swapping. Their experiments improved significantly performance on text classification tasks, especially for smaller datasets. To address the limitations of random insertion

and swapping approaches, Karimi et al. [25] introduced an Easier Data Augmentation (AEDA), which randomly inserts punctuations into the original text while maintaining word order. This method offers easier implementation and preserves all input information, leading to enhanced generalization performance. For sentence-level named entity extraction, Dai et al. [18] improved experimental results through label-oriented token and synonym replacement, demonstrating their approach using Transformer models on biomedical and materials science datasets.

Synonym Replacement (SR) stands out as a simple yet effective data augmentation method. It replaces selected words with synonyms from WordNet or similar words from word embeddings, maintaining the original sentence semantics [26], [27]. Claude et al. [28]implemented synonym replacement and spell checking via Cloud API, achieving 4.3%-21.6% accuracy improvements in deep networks like LSTM and BiLSTM. For Aspect-Based Sentiment Analysis (ABSA), Zhang Rong et al. [29] developed Multi-Level Data Augmentation (MLDA), improving Accuracy and Macro-F1 by 1.2% 3% on the Rest dataset compared to LSA-BERT.

However, most studies have focused on English text, with limited exploration of Chinese text augmentation, particularly for educational review data. To address this gap, we propose Semantic Token Augmentation and Replacement (STAR), a novel approach leveraging semantic relative distance calculation for Chinese MOOC course reviews. STAR utilizes external corpus knowledge to augment training data and balance sentiment polarity distribution. We evaluate STAR using three BERT variants (LCFS-BERT, LCF-BERT, and BERT-SPC), demonstrating significant improvements in model performance and generalization capability for educational data mining tasks.

III. Methodology

Let $S = \{t_1, t_2, \ldots, t_n\}$ denote a sentence of n tokens, where t_n represents the nth token. We define the aspect term set $A = \{t_{m+1}, \ldots, t_{m+k}\}$ and the opinion term set $O = \{t_{n-k+1}, \ldots, t_n\}$, where $A, O \subseteq S$. The sentiment polarity set is given by $P = \{\text{positive, negative}\}$. The sentiment classification model computes the sentiment polarity of various aspect terms within a review. It computes a function $f : A \times O \to P$, which maps aspect terms and their associated opinion terms to sentiment polarities based on local context analysis.

For instance, in the sentence, as shown in Fig. 1, *The course is really good, but the picture is blurred.* we have aspect terms A = ”course”, ”picture” and opinion terms O = ”good”, ”blurred”. The model would classify f(”course”, ”good”) as positive and f(”picture”, ”blurred”) as negative, demonstrating its ability to discern different sentiment polarities for various aspects within a review.

The lack of annotated corpus resources presents a significant challenge in training deep learning models, especially in specialized domains requiring expert evaluation of annotation quality. To address this limitation and enhance the generalization capability of neural network models, researchers have employed various data augmentation techniques, including word transformation, sentence order alteration, and back-translation

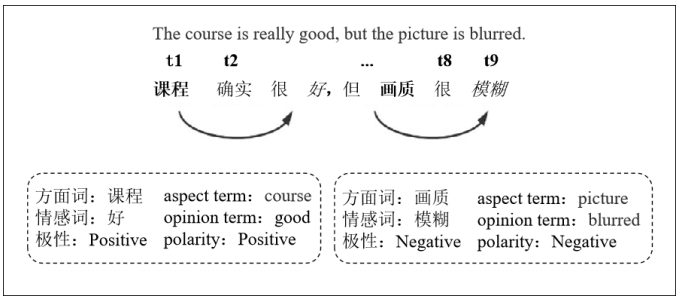


Fig. 1. Course aspect sentiment analysis example.

[16], [24], [30], [31]. These techniques aim to generate diverse training sentences while preserving semantic integrity.

To address these challenges, we propose three data augmentation approaches tailored specifically to Chinese MOOC course review data. These approaches aim to expand the training sample size while preserving the original labels, thereby reducing the loss function value during the deep learning model training process, as shown in Eq. 1. This approach seeks to enhance the extraction and identification of course-related content and its associated sentiment polarities.

$$\arg\min_{\theta} \frac{1}{N} \sum_{i=1}^{N} L(y_i, f(x_i, \theta)) \tag{1}$$

Where $L(,)$ denotes the cross-entropy loss function, θ represents the optimization parameter set, $f(x_i, \theta)$ is the decision function, and N is the number of training samples.

Due to the imbalance in the distribution of sentiment polarity in the training data, we employ Eq. 2 to keep the number of positive and negative samples the same size, thereby addressing the overfitting issue during the training process.

$$N' = (1 + N) \times rate \tag{2}$$

Here, N represents the original dataset sample size, $rate$ denotes the augmentation rate, and N' indicates the sample size in the new dataset after augmentation, accounting for the augmentation rate and sentiment polarity distribution ratio.

IV. Proposed Data Augmentation Methods

We propose and implement three augmentation methods, namely, Semantic Token Augmentation and Replacement (STAR), Aspect Replacement (AR), and Token Replacement TF-IDF (TRT). We first focus on the STAR method, followed by AR and TRT. The STAR method is the core augmentation approach in this study, leveraging semantic relative distance calculations to enhance the training dataset. The AR method focuses on aspect word replacement, while the TRT method utilizes TF-IDF values to identify replacement candidates. These methods aim to expand the training dataset while preserving original labels, thereby improving model performance in sentiment analysis tasks.

A. Semantic Token Augmentation and Replacement (STAR)

This approach employs token replacement based on semantic relative distance, utilizing Word2Vec to identify similar words and generate a new training dataset. To preserve semantic integrity, the augmentation process avoids replacing words near core words, as these typically better express sentiment or opinion. Our analysis revealed that 75% of sample sentences exceed 18 words in length. Consequently, we set the Semantic-Relative Distance (SRD) threshold to 5. When a non-stop word's semantic relative distance from the core word surpasses this threshold, it is replaced with a similar word, generating a new sentence as shown in Fig. 2.

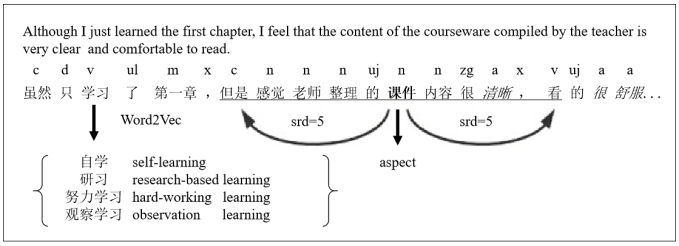


Fig. 2. STAR augmentation approach.

To maintain the semantic relationships of sentiment polarity words in the original sentence, adjective and adverb tokens outside the aspect word area remain unchanged during augmentation.

The SRD calculation process involves the following steps:

1) Initialize the SRD threshold to 5 and set the augmented quantity *augNumI* to 0. While *augNumI* is less than the specified augmentation quantity *augNum*, iterate through the original data and calculate the quantity to be augmented.
2) Designate aspect words from the specified area as stop words. Use the Jieba word segmentation tool to annotate parts of speech and add adjectives to the stop word set. The remaining words become the *randomWords* for augmentation. During candidate word traversal, replace words using similar words from the augmentation dictionary via uniform sampling.
3) Regulate the sample augmentation quantity to ensure the training set's augmentation meets requirements. Save results in *augData* and return the augmented dataset.

Algorithm **??** presents the STAR algorithm's pseudo-code, with input parameters including *index* (augmentation dictionary), *dataframe* (data to be augmented), and *augNum* (augmentation quantity). The algorithm outputs the augmented dataset.

B. Aspect Replacement (AR)

This approach leverages BERT for aspect word semantic similarity matching. It randomly selects aspect words from a synonym dictionary using uniform sampling, replaces original aspect words in the sentence with these selections, and updates the corresponding aspect words in the training data.

Algorithm 1 Semantic Token Augmentation and Replacement (STAR) algorithm pseudo-code

```
Input: index, dataframe, augNum;
Output: Augment dataset augData
srd = 5
augNumI = 0
while augNumI ¡ augNum do
    for each row_i in dataframe do
        augProportion = maxNum of augment per sentence
        left = aspect_index - srd
        right = aspect_index + srd
        srdWords = words[left : right + 1]
        adjWords = Filter tags starting with A by jieba
        stopWords = stopWords ∪ srdWords ∪ adjWords
        randomWords = Wordlist ∉ stopWords
        rowAug = 0
        for each word_i in randomWords do
            if word not in index or similar words is NULL then
                continue
            end if
            synonymsList = synonym for word_i in index
            for each synonym_s in synonymsList do
                sentence_n = Replace word in sentence with synonym_s
                augData = sentence ∪ augData
                augNumI = augNumI + 1
                rowAug = rowAug + 1
                if augNumI = augNum then
                    return augData
                end if
                if rowAug ¿ augProportion then
                    break
                end if
            end for
        end for
        if rowAug ¿ augProportion then
            break
        end if
    end for
end while
```

C. Token Replacement TF-IDF (TRT)

Token Replacement TF-IDF (TRT) follows these steps:

1) Calculate the TF-IDF value for each token.
2) Identify replacement candidates: non-stop words, non-aspect words, with a distance greater than 1 from aspect words.
3) For each candidate, compute its weight as $w_i = \text{tfidf}_{\text{max}} - \text{tfidf}$.
4) Determine each candidate's position in the sentence.
5) Using the synonym dictionary in the augmentation index, uniformly sample one token from 10 candidate synonyms for each replacement candidate. If no synonyms are available, retain the original token.
6) Save the augmented dataset and return the results.

These data augmentation methods expand the training dataset while preserving original labels, thereby enhancing model performance in downstream tasks.

V. Experimental Settings

We conducted experiments on a real-world dataset of Chinese MOOC course comments to evaluate the effectiveness of our proposed data augmentation methods. Our experiments focused on aspect-level sentiment classification, comparing the performance of LCFS-BERT, LCF-BERT, and BERT-SPC models across different training set sizes. We analyzed the impact of our data augmentation methods on model performance and evaluated the effectiveness of these methods in enhancing model generalization capability.

A. Data Source

The data used in this study is sourced from course reviews on the China University MOOC website. After data cleaning and annotation, we obtained 1,971 valid data points, which were classified into positive and negative categories based on sentiment polarity. Among them, there are 1,550 positive reviews and 421 negative reviews, with a positive-to-negative sample ratio of approximately 3.7:1. Due to this imbalance in the number of positive and negative samples, there is a certain degree of imbalance in the performance of the training samples across these categories.

We analyzed the distribution of positive and negative polarities across various course aspects, as detailed in Table I. For sentences containing multiple aspect words, we split them into multiple samples to ensure that each sample contains only one aspect word.

B. Experimental Procedure

In the experiment, we first randomly shuffle the original data and then split it into training and testing sets at an 8:2 ratio. To ensure that the testing set and training set are mutually exclusive, meaning that testing samples do not appear during training, we implemented strict data partitioning measures.

Subsequently, we extract 50, 150, 300, and 500 samples from the training set to form new training subsets, ensuring that the sentiment polarity ratio in the new training subsets remains consistent with the original training set. For each new training subset, we use three data augmentation methods to expand it, ensuring that the augmented training data remains balanced in terms of sentiment polarity.

To compare the effectiveness of different data augmentation methods, we kept the sentiment polarity ratio unchanged in the testing set. For example, STAR_150_3 indicates that 150 samples are extracted according to the sentiment polarity ratio in the original training set. Using the STAR augmentation method, we expand it to 600 samples, including the original samples. After replacing aspect terms with *T*, these data are input into the deep learning model for training.

C. Parameter Configuration

To balance the training efficiency and generalization ability of each model, ensuring good performance on MOOC data, training stops when the loss function value on the validation set is minimal for the LCFS-BERT, LCF-BERT, and BERT-SPC models.

In terms of model training, the Chinese BERT [32] is used as the pre-trained language model, with the Adam optimizer updating the model weights. The learning rate is set to 0.00002 to control the rate at which the model weights are updated. The training dataset is iterated for 10 epochs, with a batch size of 16 for each iteration. To reduce the risk of overfitting and improve the model's generalization ability, a dropout rate of 0.5 is used. Additionally, to control sequence length, reduce the model's computational complexity, and avoid negative impacts on the loss function from overly long sequences, the maximum sequence length is set to 80.

D. Experimental Results

In the experiment, aspect-based sentiment classification models based on BERT were selected: LCFS-BERT [33], LCF-BERT [34], and BERT-SPC [35]. We compared the performance of our three proposed augmentation methods against the baseline results from training on the original dataset. Furthermore, we analyzed the performance of these models across different training set sizes, using Macro-F1 as the evaluation metric for the model's effectiveness.

As shown in Table II, we can see the performance of the three augmentation methods across different sample sizes and models. Bold underlined values indicate the highest score in the current training batch.

With the increase in training data, different augmentation methods showed the varying performance improvement on each model. For instance, after applying the STAR method, the Macro-F1 score of the LCFS-BERT model improved from 71.9% to 92.27%. Models trained with augmentation methods outperformed those trained with original data alone. For the LCFS-BERT model, after using STAR, AR, and TRT methods, the Macro-F1 scores were 92.27%, 78.92%, and 78.2%, respectively. However, for the BERT-SPC model, the original dataset's Macro-F1 score was better than the results with data augmentation methods at 150 training samples. It was mainly due to the introduction of excessive noise text in the augmentation experiment, weakening the semantic correlation between aspect words and sentiment words.

To comprehensively evaluate and compare the effectiveness of the STAR augmentation method across different models and dataset sizes, the model performance curve shown in Fig. 3 visually demonstrates the improvement effect of the STAR augmentation method on model performance.

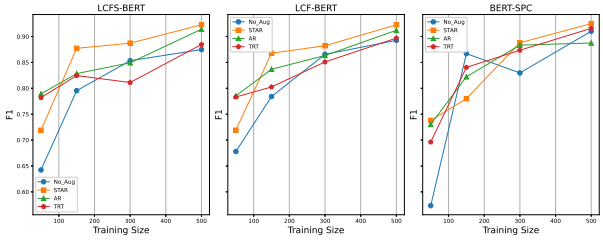


Fig. 3. Model performance comparison chart.

As shown in Fig. 3, the STAR method consistently outperforms the other two augmentation techniques, particularly in enhancing model generalization on small datasets. Compared

TABLE I. STATISTICS FOR SENTIMENT POLARITY OF COURSE ASPECTS

Sentiment Polarity	**Teaching Evaluation**	**Teaching Content**	**Teaching Cost**	**Platform Health**	**Teaching Interaction**	**Teacher**	**Course Structure**	**Course Video**
Positive (count)	62	571	16	11	639	76	117	58
Negative (count)	74	133	3	38	65	23	33	52
Total	136	704	19	49	704	99	150	110
Positive Ratio	45.59%	81.11%	84.21%	22.45%	90.77%	76.77%	78.00%	52.73%

TABLE II. COMPARISON OF MODELS' PERFORMANCE(%)

Model	Method	50	150	300	500
LCFS-BERT	**STAR**	71.90	**87.72**	**88.72**	**92.27**
	AR	**78.92**	82.84	84.93	91.41
	TRT	78.20	82.46	81.12	88.46
	No_Aug	64.23	79.54	85.33	87.48
LCF-BERT	**STAR**	71.90	**86.78**	**88.22**	**92.26**
	AR	**78.54**	83.68	86.34	91.20
	TRT	78.32	80.26	85.09	89.69
	No_Aug	67.79	78.41	86.56	89.26
BERT-SPC	**STAR**	**73.80**	78.00	**88.80**	**92.49**
	AR	73.03	82.22	88.34	88.75
	TRT	69.62	84.04	87.35	91.58
	No_Aug	57.34	**86.70**	82.97	91.02

to training results without augmentation, all three augmentation methods generally improved Macro-F1 scores, with STAR showing outstanding comprehensive performance across all three model types.

VI. DISCUSSION AND ANALYSIS

This paper explores the application of augmentation methods in sentiment analysis tasks for text data. The analysis of experimental results indicates that augmentation methods positively impact model performance. The following analysis is conducted from two aspects: the implementation ideas of the augmentation methods and the performance of the models.

- From the perspective of the implementation ideas of the augmentation methods:
  - The AR augmentation method differs from the other two augmentation methods. It replaces aspect words in sentences and uses BERT Encoding for synonym generation, which can reduce the impact of replacement words on the context. According to experimental results, AR improves the model's classification performance compared to the baseline model, but the improvement is not as significant as that of STAR. The main limitation of AR is that it only replaces aspect terms without altering other tokens in the sentence. Consequently, AR generates sentences with a uniform structure, offering less potential for improving the model's generalization capability compared to the other augmentation methods.
  - For the TRT augmentation method, replacing words within the local context of a term can introduce textual noise. This can impact the model's ability to extract syntactic features, thereby affecting its classification performance.
- From the perspective of model performance:
  - The overall enhancement effect of the LCF-BERT model is slightly inferior to that of the LCFS-BERT model. Considering that the LCF-BERT model calculates semantic relative distance based on word distance, both AR and TRT methods perform word replacements close to aspect words, so the actual word distance does not change. However, for the LCFS-BERT model, which calculates semantic relative distance based on syntax trees, the model can better understand the local context related to aspect words.
  - Both AR and TRT methods perform word replacements within the local context of aspect words, introducing some noise in the model's recognition of local context. Therefore, their effectiveness is weaker than that of the STAR method.

According to experimental results, the STAR augmentation approach shows better enhancement effects on small sample training sets compared to AR and TRT methods and can also quickly augment according to the specified multiplier. Currently, the experimental results only provide a statistical analysis of the classification performance of each model when the original dataset is augmented by a single multiplier. However, for high-multiplier augmentation methods, further experimental verification is needed to compare the performance of

the three augmentation methods.

VII. Conclusion and Future Work

In this paper, we propose a novel semantic relative distance-based augmentation method (STAR) to address the lack of labeled data corpus for Chinese course reviews. This method combines augmentation techniques with the BERT pre-trained model to mine the sentiment tendencies in the text. To improve the generalization ability of the deep learning model and discover more sentiment tendencies related to course teaching in the text, the STAR method integrates external knowledge and semantic relative distance calculation and then conducts experiments.

The experimental results show that after using the STAR method, the Macro-F1 values of the LCFS-BERT model and the LCF-BERT model increased by 3.39% to 8.18% and 1.66% to 8.37%, respectively, indicating the effectiveness of this augmentation method. Our experiments are based on aspect-level sentiment classification models using a local context attention mechanism.

In the future, it is necessary to further explore and validate the effectiveness of high-ratio data augmentation and the applicability of the STAR augmentation method under other language models. Additionally, it is essential to deeply explore and optimize data augmentation techniques to adapt to the characteristics of Chinese education review data, thus providing more accurate and effective support for data analysis and decision-making in the education field.

Acknowledgment

This research was supported by Postgraduate Education and Teaching Reform Research Project of Hubei Province (2023392), Funds For Philosophy and Social Science of Hubei Province (21Y145), Huangshi Innovation and Development Joint Fund Project (2024AFD002) and Special Fund for Scientific Planning of Private Higher Education (GA202003).

References

[1] A. Brinkmann, R. Shraga, and C. Bizer, "Product attribute value extraction using large language models," *arXiv preprint arXiv:2310.12537*, 2023.

[2] K. Roy, P. Goyal, and M. Pandey, "Attribute value generation from product title using language models," in *Proceedings of The 4th Workshop on e-Commerce and NLP*, 2021, pp. 13–17.

[3] K. Kumar and A. Saladi, "Pave: Lazy-mdp based ensemble to improve recall of product attribute extraction models," in *Proceedings of the 31st ACM International Conference on Information & Knowledge Management*, 2022, pp. 3233–3242.

[4] S. M. Babu, P. P. Kumar, S. Devi, K. P. Reddy, M. Satish *et al.*, "Predicting consumer behaviour with artificial intelligence," in *2023 IEEE 5th International Conference on Cybernetics, Cognition and Machine Learning Applications (ICCCMLA)*. IEEE, 2023, pp. 698–703.

[5] R. Sleiman, K.-P. Tran, and S. Thomassey, "Natural language processing for fashion trends detection," in *2022 International Conference on Electrical, Computer and Energy Technologies (ICECET)*. IEEE, 2022, pp. 1–6.

[6] L. Ren, B. Zhu, and Z. Xu, "Robust consumer preference analysis with a social network," *Information Sciences*, vol. 566, pp. 379–400, 2021.

[7] H. Liang, U. Ganeshbabu, and T. Thorne, "A dynamic bayesian network approach for analysing topic-sentiment evolution," *IEEE Access*, vol. 8, pp. 54 164–54 174, 2020.

[8] R. Koonchanok, Y. Pan, and H. Jang, "Tracking public attitudes toward chatgpt on twitter using sentiment analysis and topic modeling," *arXiv preprint arXiv:2306.12951*, 2023.

[9] X. Chen, F. L. Wang, G. Cheng, M.-K. Chow, and H. Xie, "Understanding learners' perception of moocs based on review data analysis using deep learning and sentiment analysis," *Future Internet*, vol. 14, no. 8, p. 218, 2022.

[10] B. Du, "Research on the factors influencing the learner satisfaction of moocs," *Education and Information Technologies*, vol. 28, no. 2, pp. 1935–1955, 2023.

[11] K. F. Hew, X. Hu, C. Qiao, and Y. Tang, "What predicts student satisfaction with moocs: A gradient boosting trees supervised machine learning and sentiment analysis approach," *Computers & Education*, vol. 145, p. 103724, 2020.

[12] M. Sadat and C. Caragea, "Learning to infer from unlabeled data: A semi-supervised learning approach for robust natural language inference," *arXiv preprint arXiv:2211.02971*, 2022.

[13] J. Li and Y. Ning, "Anti-asian hate speech detection via data augmented semantic relation inference," in *Proceedings of the International AAAI Conference on Web and Social Media*, vol. 16, 2022, pp. 607–617.

[14] Z. Feng, H. Zhou, Z. Zhu, and K. Mao, "Tailored text augmentation for sentiment analysis," *Expert Systems with Applications*, vol. 205, p. 117605, 2022.

[15] B. Wang, L. Ding, Q. Zhong, X. Li, and D. Tao, "A contrastive cross-channel data augmentation framework for aspect-based sentiment analysis," *arXiv preprint arXiv:2204.07832*, 2022.

[16] G. Li, H. Wang, Y. Ding, K. Zhou, and X. Yan, "Data augmentation for aspect-based sentiment analysis," *International Journal of Machine Learning and Cybernetics*, vol. 14, no. 1, pp. 125–133, 2023.

[17] S. Y. Feng, V. Gangal, J. Wei, S. Chandar, S. Vosoughi, T. Mitamura, and E. Hovy, "A survey of data augmentation approaches for nlp," *arXiv preprint arXiv:2105.03075*, 2021.

[18] X. Dai and H. Adel, "An analysis of simple data augmentation for named entity recognition," *arXiv preprint arXiv:2010.11683*, 2020.

[19] S. Chen, G. Aguilar, L. Neves, and T. Solorio, "Data augmentation for cross-domain named entity recognition," *arXiv preprint arXiv:2109.01758*, 2021.

[20] T. Kang, A. Perotte, Y. Tang, C. Ta, and C. Weng, "Umls-based data augmentation for natural language processing of clinical research literature," *Journal of the American Medical Informatics Association*, vol. 28, no. 4, pp. 812–823, 2021.

[21] J. Singh, B. McCann, N. S. Keskar, C. Xiong, and R. Socher, "Xlda: Cross-lingual data augmentation for natural language inference and question answering," *arXiv preprint arXiv:1905.11471*, 2019.

[22] G. Li, H. Wang, Y. Ding, K. Zhou, and X. Yan, "Data augmentation for aspect-based sentiment analysis," *International Journal of Machine Learning and Cybernetics*, vol. 14, no. 1, pp. 125–133, 2023.

[23] H. Q. Abonizio, E. C. Paraiso, and S. Barbon, "Toward text data augmentation for sentiment analysis," *IEEE Transactions on Artificial Intelligence*, vol. 3, no. 5, pp. 657–668, 2021.

[24] J. Wei and K. Zou, "Eda: Easy data augmentation techniques for boosting performance on text classification tasks," *arXiv preprint arXiv:1901.11196*, 2019.

[25] A. Karimi, L. Rossi, and A. Prati, "Aeda: an easier data augmentation technique for text classification," *arXiv preprint arXiv:2108.13230*, 2021.

[26] X. Zhang, J. Zhao, and Y. LeCun, "Character-level convolutional networks for text classification," *Advances in neural information processing systems*, vol. 28, 2015.

[27] S. Y. Feng, V. Gangal, D. Kang, T. Mitamura, and E. Hovy, "Genaug: Data augmentation for finetuning text generators," *arXiv preprint arXiv:2010.01794*, 2020.

[28] C. Coulombe, "Text data augmentation made simple by leveraging nlp cloud apis," *arXiv preprint arXiv:1812.04718*, 2018.

[29] L. Y. ZHANG Rong, "Multi-level data augmentation method for aspect-based sentiment analysis," *Frontiers of Data and Computing*, vol. 5, pp. 140–153, 2023.

[30] H. Shi, K. Livescu, and K. Gimpel, "Substructure substitution: Structured data augmentation for nlp," *arXiv preprint arXiv:2101.00411*, 2021.

[31] C. Shorten, T. M. Khoshgoftaar, and B. Furht, "Text data augmentation for deep learning," *Journal of big Data*, vol. 8, no. 1, p. 101, 2021.

[32] Y. Cui, W. Che, T. Liu, B. Qin, and Z. Yang, "Pre-training with whole word masking for chinese bert," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 29, pp. 3504–3514, 2021.

[33] M. H. Phan and P. O. Ogunbona, "Modelling context and syntactical features for aspect-based sentiment analysis," in *Proceedings of the 58th annual meeting of the association for computational linguistics*, 2020, pp. 3211–3220.

[34] B. Zeng, H. Yang, R. Xu, W. Zhou, and X. Han, "Lcf: A local context focus mechanism for aspect-based sentiment classification," *Applied Sciences*, vol. 9, no. 16, p. 3389, 2019.

[35] Y. Song, J. Wang, T. Jiang, Z. Liu, and Y. Rao, "Attentional encoder network for targeted sentiment classification," *arXiv preprint arXiv:1902.09314*, 2019.

Design and Implementation of Style-Transfer Operations in a Game Engine

Haechan Park[1], Nakhoon Baek[2]
School of Computer Science and Engineering, Kyungpook National University, Daehak-ro 80, Daegu 41566, Korea[1,2]
Graduate School of Data Science, Kyungpook National University, Daehak-ro 80, Daegu 41566, Korea[2]
Data-Driven Intelligent Mobility ICT Research Center, Kyungpook National University, Daehak-ro 80, Daegu 41566, Korea[2]

***Abstract*—The image style transfer operations are a kind of high-level image processing techniques, in which a target image is transformed to show a given style. These kind of operations are typically acquired with modern neural network models. In this paper, we aim to achieve the image style-transfer operations in real time, with the underlying computer games. We can apply the style-transfer operations to the all or part of rendering textures in the existing games, to change the overall feeling and appearance of those games. For a computer game or its underlying game engine, the style-transfer neural network models should be executed so fast to maintain the real-time execution of the original game. Efficient data management is also required to achieve deep learning operations while maintaining overall performance of the game as much as possible. This paper compares several aspects of style-transfer neural network models, and its executions in the game engines. We propose a design and implementation way for the real-time style-transfer operations. The experimental result shows a set of technical points to be considered, while applying neural network models to a game engine. We finally shows that we achieved real-time style-transfer operations, with the Barracuda module in the Unity game engine.**

***Keywords*—Style transfer; neural network models; game engine; rendering textures; real-time operations**

I. Introduction

Game engines are now general development tools, which help users develop games conveniently through providing the functionalities needed to develop any kind of computer games. A game engine typically includes a *rendering module* that draws objects on the screen, a *physics simulation module* for adding physically-simulating effects, such as collisions and/or gravity effects, a *sound module* for background music and/or sound effects, and an *event processing system* for user input and system events [1].

It can also support network communication features, external database connections, and on-line storage connections for storing or retrieving information about users or data for the game. A game engine is now typically a complex and heavy program, since it provides various sets of functions, as show here.

Recently, game engines are emerged to support machine learning features, typically as *add-on* modules. Actually, in artificial intelligence and machine learning fields, various researches and developments have drawn much attentions from general public persons. For example, *AlphaGo*[2] which is artificial intelligence in the game of *Go*, *AlphaStar*[3] which is artificial intelligence in the *StarCraft 2*, and *Vizoom* [4] which is a study that trains neural networks to play the first-person shooter game *Doom* through visual information. These artificial intelligence works are implemented based on the *deep reinforcement learning* [5]. Due to the remarkable achievements of deep reinforcement learning, many studies of artificial intelligence used in the computer games are mainly focused on reinforcement learning.

In fact, the artificial intelligence methods can be used in a variety of ways, even in game-specific applications. Recently, artificial intelligence has also been used in the areas of creation, such as painting pictures as sophisticated as photographs [6], writing [7], or music composing [8]. Several major gaming companies carry out research on the generation of images, animations, music, etc. to be used in games across various neural networks. These game resources are usually created outside of the game engine, and the generated resources are provided in the general file formats. The game engines read those resource files, and then use them in the games.

We expect that it would be innovative to generate resources such as images, videos, and sounds through neural networks inside the game engine and apply them to the game in real-time. However, most generative neural networks are too slow for real-time applications. Thus, applications of neural network models to real-time games are limited to relatively simple neural network models to quickly generate game resources.

In this work, we focused on the convergence development of real-time games and artificial intelligence models, through applying generative neural networks in real-time. More precisely, we focused on the *image style transfers* [9] to generate the texture images in real-time, as game resources, as shown in Fig. 1.

The texture images are used for various purposes in the game, and thus, our work can also be extended to the various applications. In contrast, the style transfer is not possible to be implemented in other ways except machine learning methods. Thus, it is a good case study for game engines, which support machine learning features. The image style transfer is even more practical, since it can be easily adopted for the special effects in games.

Applying style transfer methods to the real-time games shows some technical issues. Firstly, the resulting images of the neural network models should be generated quickly so that it can be applied in real-time. Thus, efficient architectural models for running the game itself and also the neural network models simultaneously are needed.

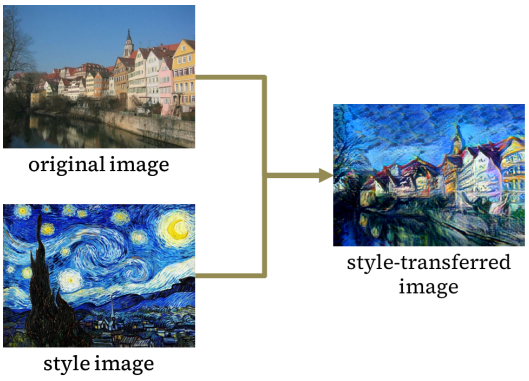


Fig. 1. An example of the style-transfer operation.

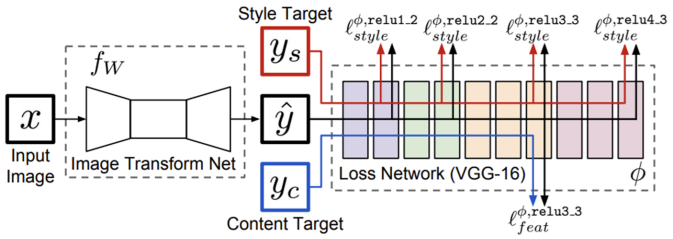


Fig. 2. The overall structure of the feed-forward style-transfer neural network model [11].

In this paper, we will solve those technical issues. From the viewpoint of execution speeds, we aim to achieve real-time performance even with applying neural networks that perform style transfer. We will use a commercial game engine, as the case study, which provides machine learning model drivers to analyze and optimize the inter-working method to check performance and limitations.

II. Related Works

The *style transfer method* can be achieved by neural network models which create the resulting images by transferring the image styles of the given style images to the content images. The work performed by the neural network model can be checked through the resulting image created with the content image and style image, as shown in Fig. 1.

The key concept in the style transfer method is that it is possible to separate content and style from an image, by extracting features of the image through a set of trained convolution layers [10]. The initial style transfer method [9] operated by gradually transforming the input image through gradient descent, and it took a long time to obtain the resulting image, making it difficult to process video streams or real-time applications.

Later, the *Feed-Forward style transfer method* [11] was proposed, which solved the problem of slow conversion speed. This method generates the transformed image at once through the *encoder-decoder* networks. As shown in Fig. 2, this method creates a transformed image by receiving the content image from the *Image Transform Net*, which is actually an encoder-decoder network. The input image is separated into the content image and the style image. The trained *VGG* (Visual Geometry Group) neural network [12] calculates content loss (L_c) and style loss (L_s), respectively, and then trains the *Image Transform Net* through the back-propagation. Since the entire neural network is trained on a single style image, it can be suitable for a single style transfer.

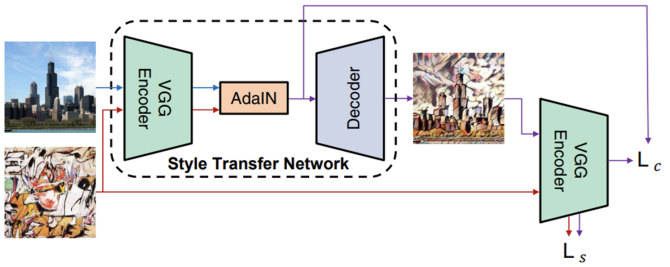


Fig. 3. The overall structure of the AdaIN style-transfer neural network model [13].

Style transfer methods with *Adaptive Instance Normalization* (AdaIN) [13] can be used for arbitrary number of style images, even in real-time or at least in pseudo real-time. The AdaIN layer calculates the *mean* and *variance* for each channel of the features extracted from the content images and the style images. Then, it normalizes the means and variances of the content images, with respect to the means and variances of the style images.

The *AdaIN* operations can be summarized as the follow equation:

$$\mathrm{AdaIN}(x, y) = \sigma(y)\frac{x - \mu(x)}{\sigma x} + \mu(y), \qquad (1)$$

where μ is the mean and σ is the variance function, from statistics. The parameter x is the feature values extracted from the content image through the VGG encoder, while the parameter y is the feature values from the style image through the VGG encoder.

As shown in Fig. 3, VGG encoder is trained by classification tasks and network weights are fixed. Since AdaIN has no parameters used for learning, training is performed only on the VGG decoder part. There are two loss functions used for training: *content loss* (L_c) and *style loss* (L_s) functions. The differences between the results of applying VGG encoder to the images that has been constructed through VGG decoder, and the results obtained when the content images go through VGG encoder and the AdaIN layer, are defined as the content losses. Similarly, the differences between the results of applying VGG encoder to the images that has been constructed through VGG decoder and the style images are defined as the style losses. The final objective function is the sum of content losses and style losses, and the VGG encoder is trained to minimize it.

For a computer game or its underlying game engine, the style-transfer neural network models should be executed so fast to maintain the real-time execution of the original game. Unfortunately, the previous works are insufficient to get real-time results, and we aims to get the style-transfer operations in real-time or even in pseudo real-time. In the next sections, we will show our experimental details and the results.

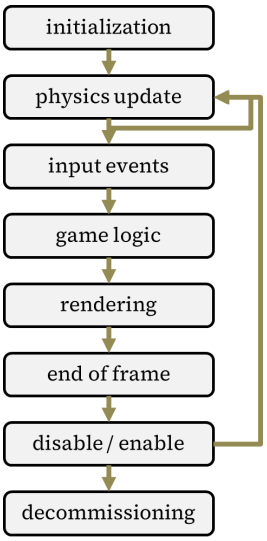


Fig. 4. A typical execution order of Unity event functions.

III. Experimental Environment

A. System Configurations

Our system configuration for the experiments are as follows:

- CPU : Intel Core i7, 2.30 GHz
- GPU : NVIDIA GeForce RTX 2060
- Main Memory : 16GB
- Operating System : Windows 10 Pro 64bit Edition

As shown here, we used typical commercial PCs with mid-tire computing powers, rather than high-tire ones, for more practical uses and more real-world experimental results.

B. Game Engine Integration

Our work is carried out on the *Unity* game engine. It manages various functions necessary for game development in module units. Each module must be operated efficiently at an appropriate time, according to the characteristics of the game to be made, to obtain a positive response from game service users. Each module is controlled through *event functions*. Event functions perform actions that need to be handled at a specific point in time or situation in the game.

The Unity game engine provides many event functions to control modules in various situations, so it is inadequate to mention all of them in this paper. Taking it into consideration, the execution orders of the event functions are iterated in groups, according to their purpose, as shown in Fig. 4.

At the start of the event processing loop, event functions for *initializing* the game are called, and then event functions related to *physical effects* are called. Physical effects are updated separately from the main thread at a specified time through a reliable timer system. By applying this method, if the physical effects are not updated within a fixed time interval, it can be applied inaccurately unlike in reality, but it is possible to prevent unexpected situations, in which screen rendering is delayed until all the physical effects are applied.

TABLE I. Comparison of *Unity* and *Self-Engine* Game Engines, According to the Machine Learning Features

game engine	Unity	Self-Engine
machine learning module	ML-Agents	Neuro
programming language	Python, C#	C++
training	Pytorch, Tensorflow	cuDNN
inferencing	Barracuda	cuDNN
additional features	support reinforcement learning	lightweight

Then, the *input* event functions are called to handle user inputs to interact with the game. Next, *game logic* event functions that perform calculations for various decisions to be performed in the game are called, and event functions that draw the screen with the data updated so far are called. Next, the event function that defines the action to be applied to the *end of the frame* is called, and the event function specifies the action to be performed when it is decided to make the object be *disabled or enabled* in the game scenario is called. Finally, when the game ends, event functions for resource *decommissioning* are called.

C. Machine Learning Modules for Game Engines

Before starting our full-scale experiments, we investigated several game engines, for their features which support machine learning models, and also the overall environment for the inference executions. At that time, the Unity engine is one of the best-suitable commercial game engines, and it officially supports machine learning features. We also compared those features with the *Self-Engine* [14], which is an open-source, lightweight game engine which supports machine learning models, as shown in Table I.

Unity supports the *Unity Machine Learning-Agents Toolkit* (or shortly, *ML-Agents*) [15], which is an add-on module to apply machine learning features to the games. In ML-Agents, the *C# scripting language* [16] is used to train neural networks in the game engine environment. *Python terminals* [17] work in conjunction with it, for neural network training.

For the general cases, the machine learning modules for neural network training in the Python environment can be selected from *PyTorch* [18] and *TensorFlow* [19], which are already familiar to AI researchers. In contrast, since we use ML-Agents for functional inferences, we had better to use *Barracuda* [20], which is a specialized module that executes pre-trained neural network models with C# scripts. Additionally, ML-Agents support other various functions to efficiently perform reinforcement learning.

Self-Engine uses *Neuro* [21] for its machine learning module. Both the game engine itself and the machine learning module are implemented in the same C++ programming language and thus, the machine learning module is naturally integrated into the game engine. With the integrated machine learning module, neural network training and inferences are performed in the similar manner. In the Self-Engine, the machine learning

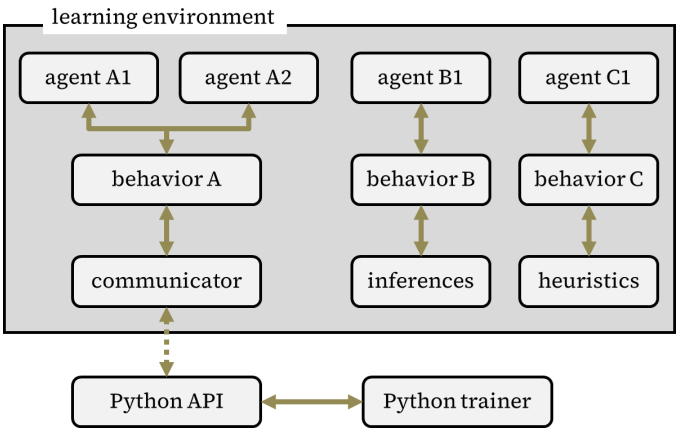


Fig. 5. The internal structure of the ML-Agent module.

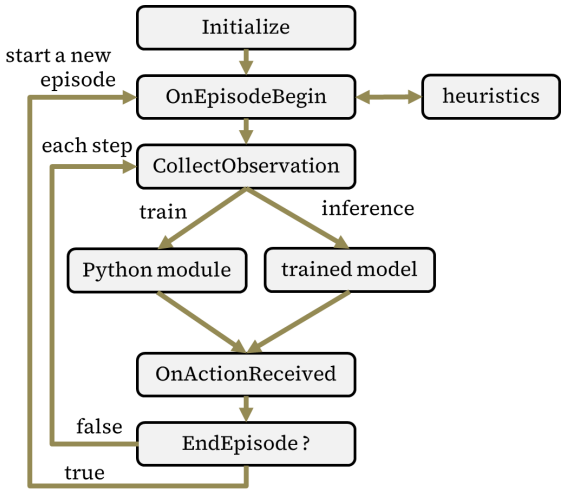


Fig. 6. The internal structure and function calling orders in the agent class.

module of Neuro internally uses *cuDNN* [22] as its underlying deep neural network operation tools.

Actually, Self-Engine is a lightweight game engine implemented as an open-source, and its internal structure is less complex than that of commercial game engines. However, at least at this time, most of artificial intelligence researchers prefer PyTorch and TensorFlow, and Self-Engine has fewer convenient features compared to Unity, which officially supports machine learning modules. Conclusively, in our experiments, we use the Unity game engine, as our major target system.

IV. Implementation Details

A. ML-Agents Features

The ML-Agents performs training between the *Unity editor* and the *Python terminal* via *Inter-Process Communication* (IPC), as shown in Fig. 5. The two processes conduct training by exchanging data with each other in socket communication. For this purpose, ML-Agents must be installed in both Unity editor and Python terminal respectively. ML-Agents installed on the Unity editor side is implemented in the C# programming language, which is part of the Learning Environment. ML-Agents installed on the Python side is the Python API part, as shown in Fig. 5.

After each installation, the Unity editor needs to set up a project to use ML-Agents. An object to be used as an agent must be placed in the scene of the Unity editor, and Behavior Parameters and Decision Requester script must be added to the object as components. In the Behavior Parameters script, the agent specifies the settings for the information the agent wants to observe in the environment, and when not used in the learning mode, the neural network model saved as an *ONNX file* [23] to be used for inference.

After that, we implemented the actions to be performed by the agent by inheriting the Agent Class, and also the observation data to be sent to the Python module during training and the processing to be performed when the data is received from the Python module. After implementing the internal functions by inheriting the Agent Class, write a Python script using the ML-Agent Python Low-Level API to access the Unity process from the Python terminal. Fig. 6 shows the overall structure of the agent class and its internal functions with their calling orders.

The *Initialize* function inside the *Agent Class* performs necessary works when the game starts. The *OnEpisodeBegin* function implements the initialization work to be performed when an episode starts at each time. After implementing the inside of the *CollectObservation* function that obtains observation information about the environment to be used for the training or the inference, it is divided into the training mode and the inference mode, and the next task to be performed is determined. Mode selection can be set in *Behavior Type* of the *Behavior Parameters* script.

The result from the neural network models being trained in Python terminals or the pre-trained neural network models used for the inference can be applied to the agent through the *OnActionReceived* function. When the episode ends by applying the action output from the neural network model, a new episode is started and training proceeds again. In other cases, the process of receiving observation data in the next step is repeated. The *heuristic* function is used when a user controls an agent directly, without using a neural network model or with the user-provided artificial intelligence.

After implementing the internal functions for inheriting the Agent Class, write a Python script using the ML-Agent Python low-level API to access the Unity process from the Python terminals. When a script is written using the Python low-level API of ML-Agents, the approximate structure of the script code is shown in Fig. 7.

The neural network training method in ML-Agents starts with selecting Unity environments to control with Python APIs. If the build completed executable is selected as an environment to use, the path to the executable must be specified, and if the path is left blank, it will automatically connect to the Unity editor. The successful connection to the Python terminals will result in the reset operation of the Unity environment at the first start. Then Unity conducts simulation by its agents and the Python terminal starts training either PyTorch or TensorFlow neural network models with the received observation data, as the *get observations* step in Fig. 7.

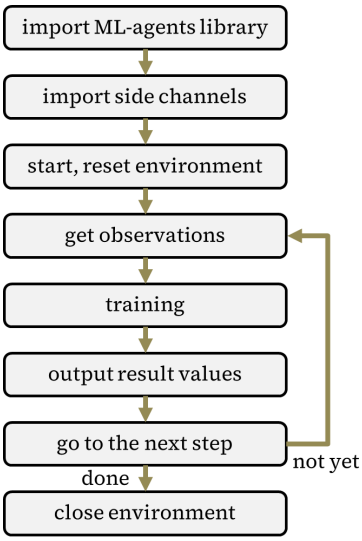


Fig. 7. The ML-Agent python API script structure.

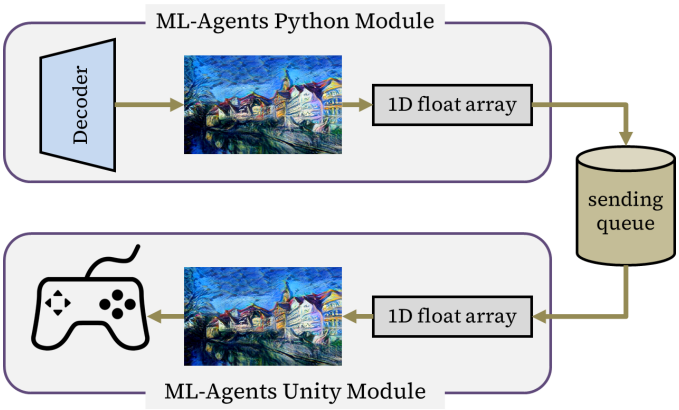


Fig. 8. Our data transmission model, with the side channel.

After that, the training proceeds with repeating the number of steps set in the Python script. At each step during training, information about the agent that needs a decision on the next action to be performed and the agent whose episode has ended collects the output of the neural network results, as the *output result values* stage, and applies it to the next step for agents that need a decision, as the *go to next step* stage. During training, user input does not work in the Unity editors or the built environment, since the Python terminal controls input events in Unity.

Additional work is required to perform the *style-transfer* operations with the ML-Agents. ML-Agents can transmit the information observed through the camera sensors of agents in the Unity environment to the Python terminal. This feature is provided for training deep reinforcement learning networks with environmental observation data obtained from the camera sensors. Performing style transfer on this image is applying style transfer to the screen that the agent sees. However, the Python module of ML-Agents does not support sending images to the Unity process. Currently, ML-Agents only considered transmitting the behavior of the agent to be performed in the next step. To solve this problem, we inherited the Side channel class and used it for sending user-defined data.

The *side channel* is a supported function to transmit user-defined data or inform the user about a specific state during the neural network training process. To use this function, we need to implement the internal functions directly from inheriting the Side Channel class. The side channel implemented in this way can be used in the *import side channels* step, as shown in Fig. 7.

Similar to the structure of ML-Agents, *Side Channel Class* must be implemented in Unity and Python, respectively. On the Unity side, the *Side Channel C# class* is inherited, and on the Python side, the *Side Channel python class* is inherited. The process of data transmission through the side channel is shown in Fig. 8. In Python terminals, after converting the style-transferred image into a one-dimensional array of floating-point values, it is transmitted to the Unity side through the side channel. In the Unity game engine, the received floating-point one-dimensional array is converted into a texture image, which will be rendered on the screen. However, this method transfers the image pixel-by-pixel through the send/receive queue and copies the received float array pixels one-by-one to the texture, which requires a large amount of CPU computing power.

Our style-transfer neural network model is based on the *AdaIN* layer. The model was implemented in PyTorch and used only for inference, when the training was completed. In the case of VGG encoder, the number of filters in the convolutional layer increased from 256 to 512 except for the end of the VGG19 neural network model.

In the VGG encoder, the feature map is extracted from the image by repeatedly stacking the Convolution (Conv2D) layers and the Rectified Linear Unit (ReLU) [24] activation function layers. The subsequent MaxPool2D operation reduces the horizontal and vertical resolution of the image by half. The ReflectionPad2D operation [25] adds spaces to the edges of the output tensors of the current layers, as if reflected in a mirror so that the inverted image appears repeatedly [26].

The VGG decoder also has a convolution layer and a ReLU activation function layer like the VGG encoder [27]. The difference is that the number of filters in the convolution layer is reduced by half, and the resolution of the output tensor is doubled through the added Upsample layer. If the mode of the Upsample layer is set to nearest, the extended part is filled with the same value as the element value of the nearest feature map.

B. Using the Barracuda Module

Barracuda is an inference-specific module built into the Unity game engine. Barracuda executes neural network models stored in ONNX file formats, through C# scripts. Barracuda internally uses the compute shaders [28] to handle the operations required for the neural network inference. These operations are quickly processed in parallel, using the multiple cores of the *Graphics Processing Unit* (GPU). Since Barracuda can run the neural network itself, there is no need to use PyTorch or TensorFlow, there is no cost required for inter-working between the two processes.

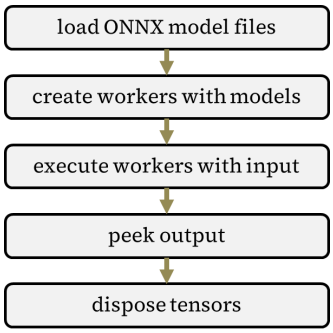


Fig. 9. Overall operations in the barracuda module.

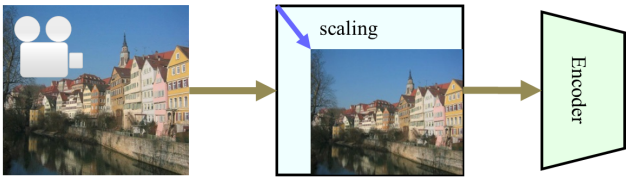


Fig. 10. Input data preprocessing process.

A neural network model executed with Barracuda can use Texture2D data types, which works as an input tensor in the Unity game engine, and vice versa. In the case of ML-Agents, the CPU reads pixel values from the resulting tensor, and injects them into the Texture2D object, actually in a step-by-step manner. In contrast, Barracuda runs faster by simultaneously processing multiple pixels at once, with the GPU. Since Barracuda was developed specifically only for the inference, it does not support any training of the neural network models.

The operating sequences of the Barracuda module are as shown in Fig. 9. A target neural network model will be stored in the ONNX file format, and it will be loaded through the C# script, at the *load ONNX model files* stage. The module creates a worker object corresponding to the loaded neural network model, at the *create worker with models* stage. A Texture2D object is passed as the input tensor to the neural network model, and also as an argument to the worker object, at the *execute workers with input* stage. The resulting tensor will be obtained with the *PeekOutput* function. The final result will be returned as a Tensor type, and can be converted to other types including Texture2D and floating-point number arrays.

To achieve the style-transfer operations with the Barracuda module, we provide the trained style-transfer neural network models as the ONNX files. Those files are imported into the Unity game engine. In our experiments, we use the ONNX files for the neural network models trained with a single style image, in a feed-forward manner [11].

To clearly check the performance of the style-transfer operations, we applied the style-transfer operation to the real-time texture images, from a webcam-based video stream. The video stream provides 60 frames per second, in the 1920×1080 resolution. Since the video resolution is too large

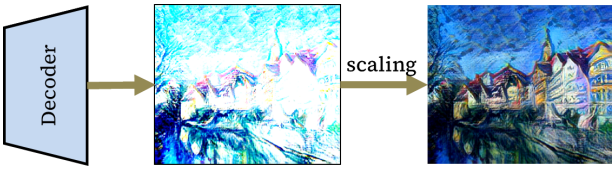


Fig. 11. Post-processing at the decoder.

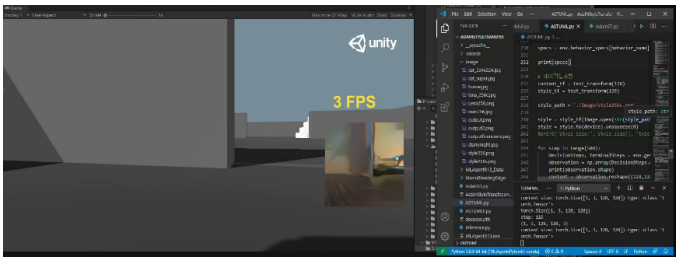


Fig. 12. A brute-force implementation of the style-transfer operation, with the python terminal in the unity engine.

to achieve real-time style-transfer, the video streams are scaled down to the internal render-texture area, with lower resolution. We achieved the highly efficient texture transfers, by using Shader functions of Unity to modify texture images at high speed through GPU's parallel processing features, as shown in Fig. 10.

In addition, appropriate post-processing is needed to the resulting tensor, to be properly used in the games. For example, the final pixel values from the Barracuda module may be normalized floating-point numbers between 0.0 and 1.0, while the rendering module needs 8bit unsigned integer values, as the corresponding color values. We also integrated these kinds of post-processing operations into the Shader programs, for more efficient and faster operations, as shown in Fig. 11.

V. Experiment Result

A. Python-based Implementation

For comparison purposes, we implemented the style transfer operations in a brute-force way, to directly link the Unity editor to the Python terminals, as shown in Fig. 12. As the starting point, the underlying game shows 170 frames per second, without connecting to the Python terminals. Since it was expected that this direct connection would be inefficient, we use low-resolutions of 128×128, for the input video streams. The style-transfer neural network model is executed on the Python side, as an AdaIN network. The resulting style-transferred image is in 256×256 resolution, and read directly from the Python terminal.

Even the style-transfer operations can be achieved efficiently in Python environment, we found that the bottleneck is the data transmission between the Python terminal and the Unity game engine. To check the transmission speed, we once used a dummy Python kernel, which just re-transmit the input data as the result. Since they use network sockets for the data transmission, it shows very slow result of even 3 frames per second, as shown in Fig. 12.



Fig. 13. An example of feed-forward style transfer.



Fig. 14. Applied style image.

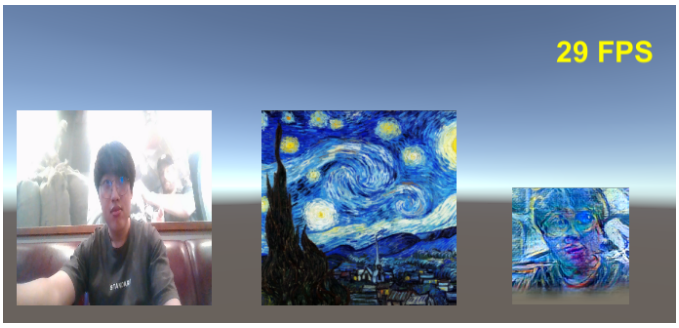


Fig. 15. Experimental results from a sample AdaIN style-transfer implementation.

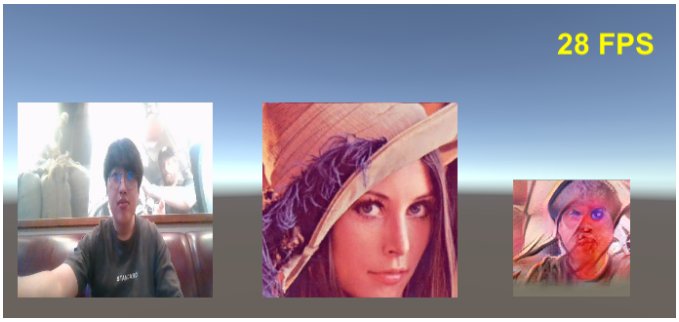


Fig. 16. Experimental results when applying the different style image.

Even worse, the received data need additional post-processing on the CPU side. For example, the color values should be converted from floating-point values to 8bit unsigned integer values for the graphics rendering. Although using shared memory seems more appropriate for the images, this simple use of Python terminals for the style-transfer purpose should consider the optimal connect of two different programming languages: Python and the C# programming language used in the Unity engine. Conclusively, we need another fully efficient way of providing style-transfer operations to the Unity engine.

B. Our Barracuda-based Implementation

In this work, we used a special device driver module, named Barracuda, in the Unity game engine. The Barracuda module is built into the Unity engine, and drives a trained neural network for inference. We first extracted input data of the real-time video streams, from our small size Web Cameras (or webcams). The input data stream is actually texture images extracted from the webcam video streams. It is then resized to be processed by the neural networks.

The style transfer neural networks are used in two aspects: a neural network trained on a single style image in a feed-forward manner, and another neural network with AdaIN, which can apply style images to the target texture, in real-time. The input images extracted from the video camera are converted into the 256×256 resolution render textures, which are used as inputs to the neural networks. The final resulting images are then used as surface textures for cube objects, which are actually a physically-simulated moving object, in our scenario, as shown in Fig. 13.

Fig. 14 shows our sample style image used for this experiment. Since there is little data transmission/reception cost, the final frame rate of the game engine is affected by the resulting texture image generation speed of the neural network models. In our experiment, we confirmed that the game engine works fast enough to be used in real-time.

VI. Analysis of Experimental Results

In our experiments, we found that the *Barracuda* module is specialized in inference, and suitable for the game engines to use the neural network features. This module also shows some limitations: it currently works well with limited sets of neural network models. Actually, we used two style-transfer models: the *feed-forward model* and the *AdaIN model*. The feed-forward model shows fast conversion speeds, and also limitations of being best suitable for a single style image training. In contrast, the AdaIN model can be used for several style images even in real-time, as shown in Fig. 15 and 16.

However, the AdaIN model was suitable for small-size images, due to the real-time requirement. Fig. 17 shows the speed of the style transfer neural networks for input data of various sizes. In this case of the AdaIN models, we found that the size of the image used as the style image was best suitable with 256×256 resolutions. The style transfer performance was greatly decreased with the image sizes larger than this.

For the game engine, it can achieve about 170 frames per second, without any style-transfer operations. Applying the style-transfer operations, the maximum texture image size was 300×300, with our experiment environment settings. Since we used mid-powered PCs, rather than the best-performance

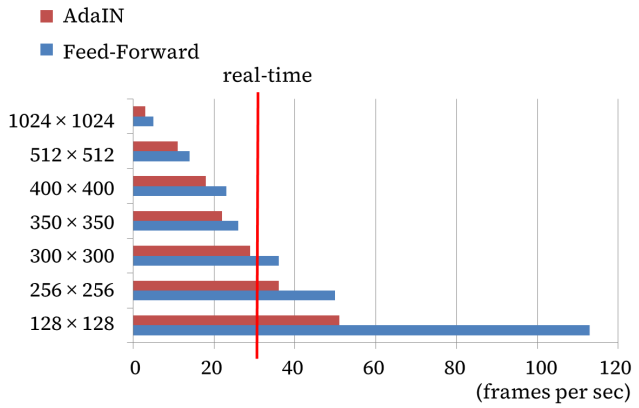


Fig. 17. Real-time style transfer performance based on input image sizes.

computing environment, we expect that the performance and also the size of the input texture would be improved with enhanced computing power machines.

VII. Conclusions and Future Works

Through this simulation, we examined the real-time style transfer performance in the game engine and efficient data processing methods. When interlocking a neural network that generates game resources in real-time with a game engine, the data transmission and reception method between the neural network and the game engine must be efficient. In addition, input and output data from neural networks must be fast and efficient when converted to game resources. In the paper, the Shader function of the game engine was used to efficiently convert and copy data through GPU parallel processing.

This paper shows a simulation of applying a deep neural network to textures on the game in real-time. Since the texture is used in various ways in the game, it will be possible to conduct various studies through future applications. It will also be necessary to study the performance optimization problems that may arise during game development using deep neural networks.

Additionally, research will be needed on performance optimization issues that may arise while developing games using neural network models. Just as the optimization techniques applied are different depending on the genre or characteristics of the game itself, optimizations based on the characteristics of the used neural network models should be investigated. Considerations for the mobile environment are also needed.

Acknowledgment

This research was supported by the MSIT (Ministry of Science and ICT), Korea, under the ITRC (Information Technology Research Center) support program (IITP-2024-RS-2024-00437756) supervised by the IITP (Institute for Information & Communications Technology Planning & Evaluation).

This study was supported by the BK21 FOUR project (AI-driven Convergence Software Education Research Program) funded by the Ministry of Education, School of Computer Science and Engineering, Kyungpook National University, Korea (4199990214394).

References

[1] J. Gregory, *Game Engine Architecture, Third Edition*. CRC Press, 2018.

[2] D. Silver, A. Huang, C. J. Maddison, A. Guez, L. Sifre, G. van den Driessche, J. Schrittwieser, I. Antonoglou, V. Panneershelvam, M. Lanctot, S. Dieleman, D. Grewe, J. Nham, N. Kalchbrenner, I. Sutskever, T. Lillicrap, M. Leach, K. Kavukcuoglu, T. Graepel, and D. Hassabis, "Mastering the game of go with deep neural networks and tree search," *Nature*, vol. 529, pp. 484–503, 2016.

[3] O. Vinyals, I. Babuschkin, W. M. Czarnecki, M. Mathieu, A. Dudzik, J. Chung, D. H. Choi, R. Powell, T. Ewalds, P. Georgiev, J. Oh, D. Horgan, M. Kroiss, I. Danihelka, A. Huang, L. Sifre, T. Cai, J. P. Agapiou, M. Jaderberg, A. S. Vezhnevets, R. Leblond, T. Pohlen, V. Dalibard, D. Budden, Y. Sulsky, J. Molloy, T. L. Paine, C. Gulcehre, Z. Wang, T. Pfaff, Y. Wu, R. Ring, D. Yogatama, D. Wünsch, K. McKinney, O. Smith, T. Schaul, T. Lillicrap, K. Kavukcuoglu, D. Hassabis, C. Apps, and D. Silver, "Grandmaster level in StarCraft II using multi-agent reinforcement learning," *Nature*, vol. 575, no. 7782, pp. 350–354, 2019.

[4] M. Wydmuch, M. Kempka, and W. Jaśkowski, "Vizdoom competitions: Playing doom from pixels," *IEEE Transactions on Games*, vol. 11, no. 3, pp. 248–259, 2019.

[5] V. Mnih, K. Kavukcuoglu, D. Silver, A. A. Rusu, J. Veness, M. G. Bellemare, A. Graves, M. A. Riedmiller, A. Fidjeland, G. Ostrovski, S. Petersen, C. Beattie, A. Sadik, I. Antonoglou, H. King, D. Kumaran, D. Wierstra, S. Legg, and D. Hassabis, "Human-level control through deep reinforcement learning," *Nature*, vol. 518, no. 7540, pp. 529–533, 2015.

[6] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Proc. of the 27th International Conference on Neural Information Processing Systems - Volume 2*. Cambridge, MA, USA: MIT Press, 2014, pp. 2672–2680.

[7] T. Brown, B. Mann, N. Ryder, M. Subbiah, J. D. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell, S. Agarwal, A. Herbert-Voss, G. Krueger, T. Henighan, R. Child, A. Ramesh, D. Ziegler, J. Wu, C. Winter, C. Hesse, M. Chen, E. Sigler, M. Litwin, S. Gray, B. Chess, J. Clark, C. Berner, S. McCandlish, A. Radford, I. Sutskever, and D. Amodei, "Language models are few-shot learners," in *Advances in Neural Information Processing Systems*, H. Larochelle, M. Ranzato, R. Hadsell, M. F. Balcan, and H. Lin, Eds., vol. 33. Curran Associates, Inc., 2020, pp. 1877–1901.

[8] C.-Z. A. Huang, A. Vaswani, J. Uszkoreit, N. Shazeer, I. Simon, C. Hawthorne, A. M. Dai, M. D. Hoffman, M. Dinculescu, and D. Eck, "Music transformer," 2018. [Online]. Available: http://arxiv.org/abs/1809.04281

[9] L. A. Gatys, A. S. Ecker, and M. Bethge, "Image style transfer using convolutional neural networks," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 2414–2423.

[10] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Commun. ACM*, vol. 60, no. 6, pp. 84–90, 2017.

[11] J. Johnson, A. Alahi, and L. Fei-Fei, "Perceptual losses for real-time style transfer and super-resolution," in *Computer Vision – ECCV 2016*, B. Leibe, J. Matas, N. Sebe, and M. Welling, Eds. Cham: Springer International Publishing, 2016, pp. 694–711.

[12] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*, Y. Bengio and Y. LeCun, Eds., 2015.

[13] X. Huang and S. Belongie, "Arbitrary style transfer in real-time with adaptive instance normalization," in *2017 IEEE International Conference on Computer Vision (ICCV)*, 2017, pp. 1510–1519.

[14] H. Park and N. Baek, "Developing an open-source lightweight game engine with dnn support," *Electronics*, vol. 9, no. 9, 2020.

[15] A. Juliani, V. Berges, E. Vckay, Y. Gao, H. Henry, M. Mattar, and D. Lange, "Unity: A general platform for intelligent agents," *CoRR*, vol. abs/1809.02627, 2018.

[16] A. Hejlsberg, M. Torgersen, S. Wiltamuth, and P. Golde, *The C# Programming Language*, 3rd ed. Addison-Wesley Professional, 2008.

[17] G. V. Rossum and F. L. Drake, *Python 3 Reference Manual*. Scotts Valley, CA: CreateSpace, 2009.

[18] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Köpf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai, and S. Chintala, *PyTorch: an imperative style, high-performance deep learning library*. Red Hook, NY, USA: Curran Associates Inc., 2019, no. 721.

[19] M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G. Irving, M. Isard, M. Kudlur, J. Levenberg, R. Monga, S. Moore, D. G. Murray, B. Steiner, P. Tucker, V. Vasudevan, P. Warden, M. Wicke, Y. Yu, and X. Zheng, "Tensorflow: a system for large-scale machine learning," in *Proc. of the 12th USENIX Conference on Operating Systems Design and Implementation*. USA: USENIX Association, 2016, pp. 265–283.

[20] Unity, *Unity Barracuda: a lightweight and cross-platform Neural Net inference library for Unity*, retrieved in July 2024. [Online]. Available: https://docs.unity3d.com/Packages/com.unity.barracuda0.3/-manual/index.html

[21] Cr33zz, "Neuro_: C++ implementation of neural networks library with keras-like API," retrieved in July 2024. [Online]. Available: https://github.com/Cr33zz/Neuro_

[22] S. Chetlur, C. Woolley, P. Vandermersch, J. Cohen, J. Tran, B. Catanzaro, and E. Shelhamer, "cudnn: Efficient primitives for deep learning," *CoRR*, vol. abs/1410.0759, 2014.

[23] ONNX, "Open Neural Network Exchange," retrieved in July 2024. [Online]. Available: https://onnx.ai/

[24] V. Nair and G. E. Hinton, "Rectified linear units improve restricted boltzmann machines," in *Proc. of the 27th International Conference on International Conference on Machine Learning*. Madison, WI, USA: Omnipress, 2010, p. 807–814.

[25] G. Liu, A. Dundar, K. J. Shih, T.-C. Wang, F. A. Reda, K. Sapra, Z. Yu, X. Yang, A. Tao, and B. Catanzaro, "Partial convolution for padding, inpainting, and image synthesis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 45, no. 5, pp. 6096–6110, 2023.

[26] P. Baldi, "Autoencoders, unsupervised learning and deep architectures," in *Proc. of the 2011 International Conference on Unsupervised and Transfer Learning Workshop - Volume 27*. JMLR.org, 2011, p. 37–50.

[27] P. Vincent, H. Larochelle, I. Lajoie, Y. Bengio, and P.-A. Manzagol, "Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion," *J. Mach. Learn. Res.*, vol. 11, pp. 3371–3408, Dec. 2010.

[28] V. S. Gordon and J. Clevenger, *Computer Graphics Programming in OpenGL with C++, Third Edition*. De Gruyter, 2024.

Under Sampling Techniques for Handling Unbalanced Data with Various Imbalance Rates: A Comparative Study

Esraa Abu Elsoud[1], Mohamad Hassan[2], Omar Alidmat[3], Esraa Al Henawi[4],
Nawaf Alshdaifat[5], Mosab Igtait[6], Ayman Ghaben[7], Anwar Katrawi[8], Mohmmad Dmour[9]
Department of Computer Science-Faculty of Information Technology, Zarqa University, Zarqa, Jordan[1,2,3,4]
Faculty of Information Technology, Applied Science Private University, Amman, Jordan[5]
Department of Data Science and Artificial Intelligence, Zarqa University, Zarqa, Jordan[6,8]
Department of Cyber Security-Faculty of Information Technology, Zarqa University, Zarqa, Jordan[7]
Department of Computer Science-Faculty of Information Technology, Zarqa University, Zarqa, Jordan[9]

***Abstract*—Unbalanced data sets represent data sets that contain an unequal number of examples for different classes. This dataset represents a problem faced by machine learning tools; as in datasets with high imbalance ratios, false negative rate percentages will be increased because most classifiers will be affected by the major class. Choosing specific evaluation metrics that are most informative and sampling techniques represent a common way to handle this problem. In this paper, a comparative analysis between four of the most common under-sampling techniques is conducted over datasets with various imbalance rates (IR) range from low to medium to high IR. Decision Tree classifier and twelve imbalanced data sets with various IR are used for evaluating the effects of each technique depending on Recall, F1-measure, gmean, recall for minor class, and F1-measure for minor class evaluation metrics. Results demonstrate that Clusters Centroid outperformed Neighborhood Cleaning Rule (NCL) based on recall for all low IR datasets. For both medium, and high IR datasets NCL, and Random Under Sampling (RUS) outperformed the rest techniques, while Tomek Link has the worst effect.**

***Keywords*—Clusters centroid; decision tree; neighborhood cleaning rule; random under sampling; Tomek Link under sampling; unbalanced datasets**

I. Introduction

Machine learning and statistics have been used for classification in many fields such as security [1]–[8], medical [9]–[15], text classification [16]–[18], and others [19], [20]. Classifications are defined as building a training model based on previous experiences or examples. Recently, there has been a rapid growth in data that has been collected from different environments but, unfortunately, there is a lack of quality data.

The quality of the data means a balanced distribution for all classes, the range of values is normalized, and no missing values, and so on. The point is, that several traditional machine learning techniques assumed that the target classes have a balanced distribution in the data [21]–[23].

Multi-class classifiers such as Support Vector Machine (SVM), Decision Tree (DT), Logistic Regression (LR), and others are sensitive for imbalanced class distribution problems [24] while one class classifiers such as Isolation forest, local outlier factor (LOF), OC-SVM and other were not.

Due to the fact that, it is rare to find balanced datasets, that contain equal or nearly equal numbers of instances for each class in real-life classification problems; Building classification models under highly imbalanced datasets is an issue in machine learning algorithms.

In unbalanced datasets the most important class (class of interest) has fewer examples than other classes like rare disease datasets [25], therefore, the classification performance of the classifier will be affected by skewing to the majority class instances, which is usually not class of interest [26].

There are two main approaches for handling unbalanced dataset problems: "algorithm-driven approach", and "data-driven approach". The first approach concerns adjusting the classifier to improve its learning from the minority class samples [27].

On the other hand, the second approach concentrates on changing the data distribution in two ways either by adding new minor class examples (over-sampling) or removing some major class instances (under-sampling). Each way has its own advantages and disadvantages, where over-sampling is considered more overhead than under-sampling and can lead to overfitting problems while under-sampling causes the loss of important information [27].

In the literature, most researchers either propose some under-sampling techniques and commonly use these techniques in research, or compare these techniques using a specific dataset.

To our knowledge, there does not exist any research in the literature comparing the effects of specific under sampling techniques in the classification performance over datasets with low, medium, and high IR. This notice represents the rationale behind this paper where the main goal from this paper is to compare the influence of four common under sampling techniques called Tomek Link, NCL, Clusters Centroid, and Random Under Sampling (RUS) in the classification results for different datasets with various IR ranges from low to high IR datasets.

To accomplish this goal twelve datasets from the Keel collection with different IR variate between low, medium, and

high IR have been used for comparing the performance of the decision tree classifier using each one of the previous four under-sampling techniques based on average recall, average F1 measure, gmean, minor class recall, and minor class F1 measure.

The rest of this paper is organized as follows: Section II displays the most recent studies about handling imbalance datasets in the literature. Section III-A describes Tomek Link, NCL, Clusters Centroid, and RUS under-sampling techniques that are compared. methodology has been demonstrated in Section III. In Section III-B the datasets are displayed. In Section IV results are presented. Finally, Section V contains the conclusion.

II. Related Work

A large number of domains with significant environmental, vital, or commercial importance encounter the class imbalance problem. The class imbalance problem means that there is a majority of one or more class spreads in the datasets [28], [29]. Moreover, it has been shown in some instances to significantly impede the performance achievable by conventional learning techniques that assume a balanced distribution of the classes and produce biased classifiers. Also, it degrades the performance of machine learning classifiers [30].

Many proposals have been presented in the literature to solve the imbalanced dataset. One of the most well-known techniques is the cluster-based under-sampling approach. It has been widely used to solve the imbalance of class distribution. In [31]–[35] a cluster-based under-sampling approach has been used to select the representative data as training data. Thus, the classification accuracy for minority classes will be improved. The experimental results show that the proposed approach outperforms other under-sampling techniques in the previous studies.

Random Under Sampling (RUS) is considered an under-sampling technique that is used for class imbalance problems. Many proposals used RUS to maintain a balanced class distribution [36]–[38]. In [39] a combination of T-Link and, Synthetic Minority Technique (SMOTE) and another sampling method such as RUS, and ROS in order to produce balance data.

Additionally, RUS has been used with different ratios to detect the performance of some of the machine learning classifiers as [40] eight random undersampling (RUS) ratios which are no sampling, 999:1, 99:1, 95:5, 9:1, 3:1, 65:35, and 1:1 have been used. Moreover, to show the performance of these ratios seven different classifiers are employed which are LightGBM (LGB), Decision Tree (DT), Random Forest (RF), Naive Bayes (NB), Logistic Regression (LR) CatBoost (CB), and XGBoost (XGB).

In [41] an ensemble feature selection has been proposed to classify the attack using the BoT-IoT dataset. The proposed approach is centered on the building of predictive models that are based on different classifiers. RUS has been used to solve the imbalance BoT-IoT dataset. The results show that the best RUS ratio was 1:1 or 1:3.

In [42] a new hybrid under sampling-based ensemble approach (HUSBoost) has been proposed. The main objective of HUSBoost is to handle imbalanced data using three main steps which are data cleaning, data balancing, and classification. At first, we remove the noisy data using Tomek-Links. RUS has been applied to create several balanced subsets.

The neighborhood cleaning rule (NCL) method has been used in many proposals in literature to deal with imbalance data [43]–[45] while other studies used hybrid approaches instead of NCL alone [46]. In [47] a combination of under-sampling and oversampling methods have been used to solve imbalance cases. Their proposal used is NCL under-sampling method and Adaptive Semi unsupervised Weighted Oversampling (A-SUWO) for the oversampling method.

Tomek link Tomek link technique is used in many studies to overcome the challenges of data imbalances that affect the performance of supervised learning-based [48]. In [49] Synthetic minority oversampling technique (SMOTE) and T-link have been used for imbalanced data. In addition, a Naïve Bayes classifier, support vector machine, and k-nearest neighbors together have been used for performance evaluation.

In [50] Cluster Based, Tomek Link, and Condensed Nearest neighbours have been used to handle the class imbalance problem by equalizing the number of instances. This is done by under-sampling the majority class based on some particular criteria [51]–[54]. The performance evaluation was done based on applied different machine learning classifiers such as K-Nearest Neighbor, Decision Tree, and Naive Bayes. The results showed that Decision Tree outperformed other machine learning techniques using the proposed technique.

Up to our best knowledge and based on an extensive literature review search, we noticed that most of the previous works exist in the literature compare the performance of specific under sampling techniques versus a hybrid version of these techniques over specific datasets with specific IR. However, in this paper, four of the most common under sampling techniques were applied over three categories of datasets that were categorized based on IR into three categories (low IR, medium IR, and high IR), and the effects of each technique on the classifier performance were compared.

The main purpose of this paper is to conclude a standard relationship between the compared under sampling techniques and dataset IR value in order to guide researchers in choosing the most suitable under sampling technique from the compared ones based on the dataset IR. In this paper, twelve imbalanced data sets with various IR have been used for evaluation in addition to DT for classification.

III. Methodology

For each dataset Work starts by dividing the original file into two parts: 70% for learning the classifier (training data set) and 30% for evaluating the effectiveness of the constructed model (testing set) based on average recall, average F1 measure, gmean, minor class recall, and minor class F1 measure [55].

For the training dataset, four under sampling techniques have been applied in order to make it balanced for learning decision tree classifier then using the generated model for classifying the test set and evaluating its performance, as shown in Fig. 1.

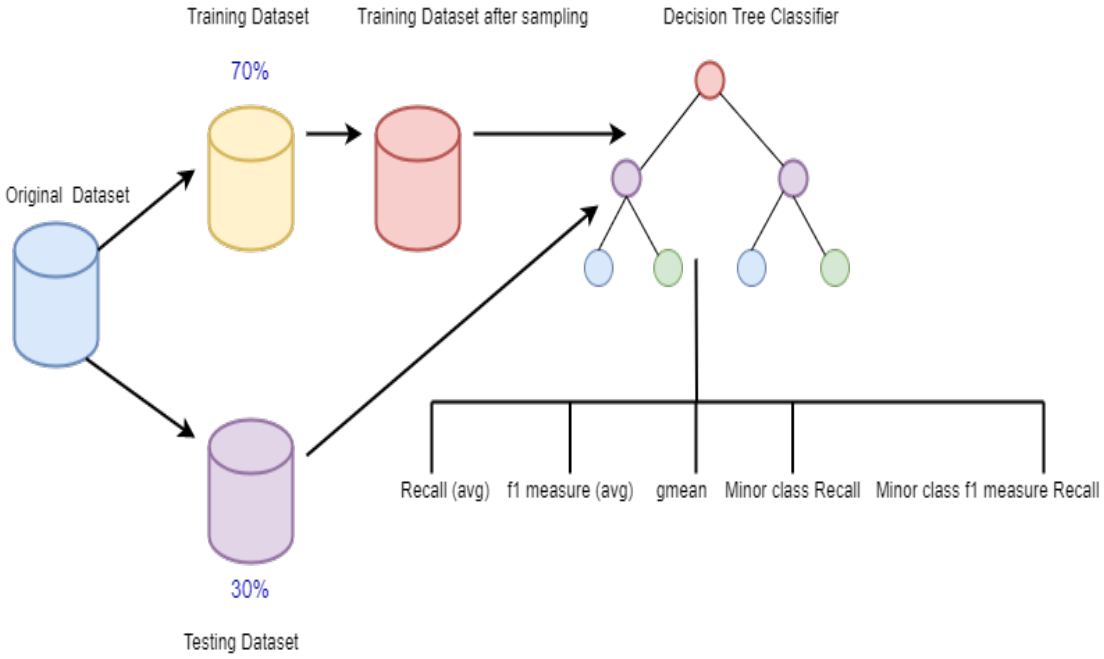


Fig. 1. Workflow for the used methodology

For each dataset, the best results that have been generated by DT classifier with each under sampling technique after adjusting its parameters are documented and compared.

In this paper, python is used for conducting all experiments through the PyCharm integrated development environment (IDE). Python libraries sci-kit-learn and imbalanced-learn are deployed for generating the results [56].

A. Under Sampling Techniques

1) Tomek link Tomek link refers to a pair of nearest neighbors where each one belongs to different classes. Under-sampling is done by removing all majority class samples from all Tomek Links [57].
2) Neighborhood Cleaning Rule (NCL) It was proposed by Laurikkala in 2001, where in this method for each sample in the training set three nearest neighbors for it must be defined then if it belongs to the major class while all of its selected neighbours belong to the minor class then it will be removed as a noise sample but if this sample belongs to the minor class and its three nearest neighbours belongs to the major class then these neighbours must be removed now. This method needs numerous computations with large-size datasets [58].
3) Clusters Centroid This method undersamples the majority class by replacing a cluster of majority samples as it finds the clusters of the majority class with the K-mean algorithm then it keeps the Cluster Centroids of the N clusters as the new majority samples [5].
4) Random Under Sampling (RUS) RUS works by removing some of the majority class samples randomly to change the distribution of data in the imbalanced dataset in order to convert it to a more balanced one for improving the classifier learning process in machine learning but this method sometimes means losing important information which considered as one drawback according to using this technique [59].

B. Datasets

Twelve datasets from Keel [8] collection with different IR are used in this paper. Table I summarizes these datasets properties.

Datasets are divided into three groups based on their IR, where the first five datasets with IR smaller than 9 represent a low IR group while the medium IR group contains datasets with IR greater than 9 and smaller than 50. Finally, datasets with IR greater than 50 are members of the high IR group in this paper.

C. Evaluation Metrics

This section is devoted to displaying the evaluation metrics that are used for evaluating the effects of TL, RUS, NCL, and CC under sampling techniques in the classifier performance for different datasets from various IR.

- Recall or TPR: It measures how often the classifier correctly detects the positive instances from all positive instances [60], [61], as shown in Eq. (1)

$$Recall = \frac{TP}{(TP + FN)} \quad (1)$$

- F1-Score: It combines the effects of precision and recall together [62], as shown in Eq. (2).

TABLE I. PROPERTIES OF TWELVE IMBALANCED DATASETS WITH DIFFERENT IR FROM KEEL REPOSITORY

Imbalance Category	Dataset Num.	Dataset Name	features	examples	Minor class	Major class	imbalance rate
Low	D1	glass1	9	214	76	138	1.82
	D2	ecoli_0_vs_1	7	220	77	139	1.86
	D3	vehicle0	18	846	199	647	3.25
	D4	ecoli3	7	336	35	301	8.6
	D5	page-blocks0	10	5472	559	4913	8.79
Medium	D6	glass4	9	214	13	201	15.47
	D7	car-good	6	1728	69	1659	24.04
	D8	kr-vs-k-one_vs_fifteen	6	2244	78	2166	27.77
High	D9	kr-vs-k-zero_vs_eight	6	1460	27	1433	53.07
	D10	Winquality	11	691	10	681	68.1
	D11	kr-vs-k-one_vs_fifteen	6	2193	27	2166	80.22
	D12	abalone19	8	4174	32	4142	129.44

$$F1 - Score = \frac{(2TP)}{(2TP + FP + FN)} \quad (2)$$

- gmean: It is a combination of TPR, and TNR metrics, as shown in Eq. (3).

$$gmean = \sqrt{(TPR * TNR)} \quad (3)$$

All metrics depend on four main parameters explained below:

- True Positive (TP): represents a number of actually positive instances classified as “positive”.
- True Negative (TN): represents a number of actually negative instances classified as “negative”.
- False Positive (FP): represents a number of actually negative instances classified as “positive”.
- False Negative (FN): represents a number of actually positive instances classified as “negative”.

IV. EXPERIMENTAL RESULTS AND ANALYSIS

This section discusses the results that are generated from comparing the effects of Tomek Link, NCL, Clusters Centroid, and RUS in the performance of Decision Tree (DT) classifier for all groups of low, medium, and high IR datasets based on average Recall in subsection IV-A, then based on average F1 measure in subsection IV-B. Later the results of gmean, recall of minor class, and F1 measure of minor class results were discussed and analyzed in subsections IV-C, IV-D, and IV-E, consequently.

A. Recall

From Table II, we can conclude the following results by comparing the effects of the selected under sampling techniques in decision tree classifier average recall value for low, medium and high IR groups of datasets. From Fig. 2 we can concludes the following points for all low IR dataset groups recall value

- NCL provides better performance based on recall than Tomek link for all datasets.
- Also we concluded that NCL provides better recall value than RUS for first three datasets.

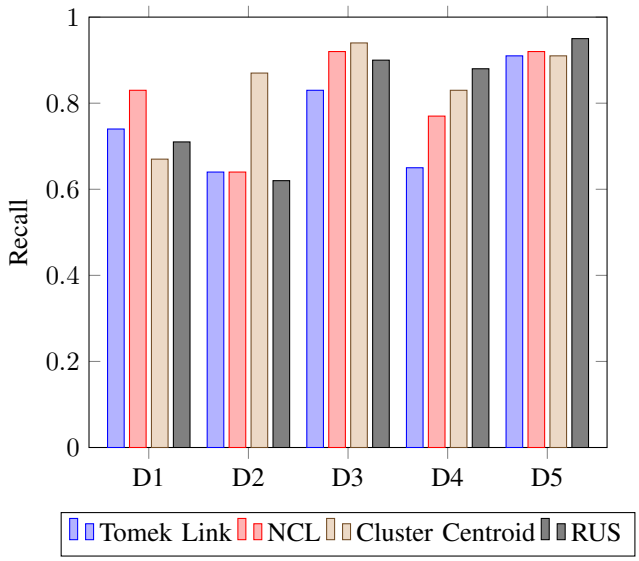


Fig. 2. Recall results for all low imbalance rate datasets.

- Cluster Centroid has better performance than Tomek link and NCL for all datasets except the first one with imbalance rate = 1.82

From Fig. 4 we can conclude the following points for all medium IR dataset groups recall values.

- From Fig. 5 we show that Tomek link provides the worst performance for all datasets.
- NCL provides better performance than Cluster Centroid for all datasets
- RUS outperformed all other techniques for all datasets

From Fig. 5 we can concludes the following points for all high IR dataset groups recall value

- NCL provides better performance based on recall than Tomek link for all datasets.
- RUS outperformed all other techniques for all datasets except the last one with imbalance rate = 129.44.
- Tomek link provides the worst performance for all datasets

TABLE II. RECALL RESULTS FOR ALL EXAMINED DATASETS USING THE SELECTED UNDERSAMPLING TECHNIQUES

Dataset	Tomek Link	NCL	Cluster Centroids	RUS
glass1	.74	.83	.67	.71
ecoli_0_vs_1	.64	.64	.87	.62
vehicle0	.83	.92	.94	.9
ecoli3	.65	.77	.83	.88
page-blocks0	.91	.92	.91	.95
glass4	.49	.98	.91	.99
car-good	.92	.97	.95	.99
kr-vs-k-one_vs_fifteen	1	1	.98	1
kr-vs-k-zero_vs_eight	.92	.96	.93	.99
Winquality	.49	.5	.76	.81
kr-vs-k-zero_vs_fifteen	1	1	1	1
abalone19	.56	.68	.77	.71

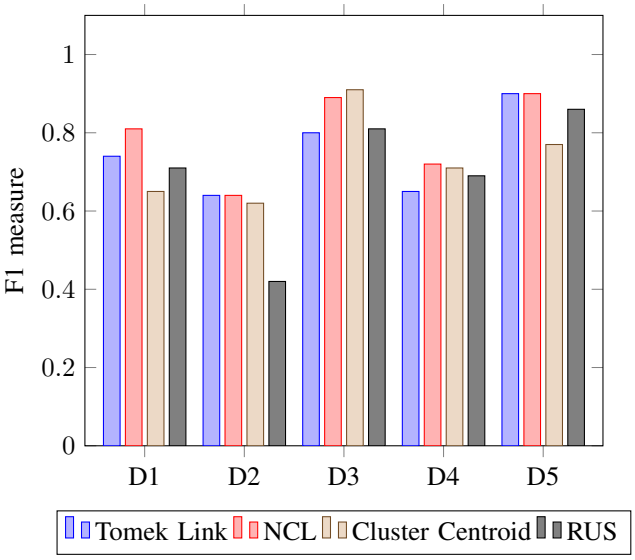


Fig. 3. F1 measure results for all low imbalance rate datasets.

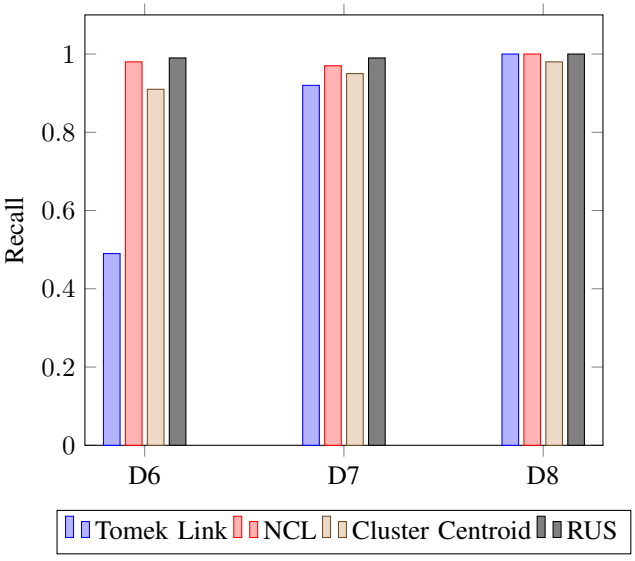


Fig. 4. Recall results for all medium imbalance rate datasets.

B. F1-Measure

From Table III, Fig. 3, 6 to 16 we can concluded the following points

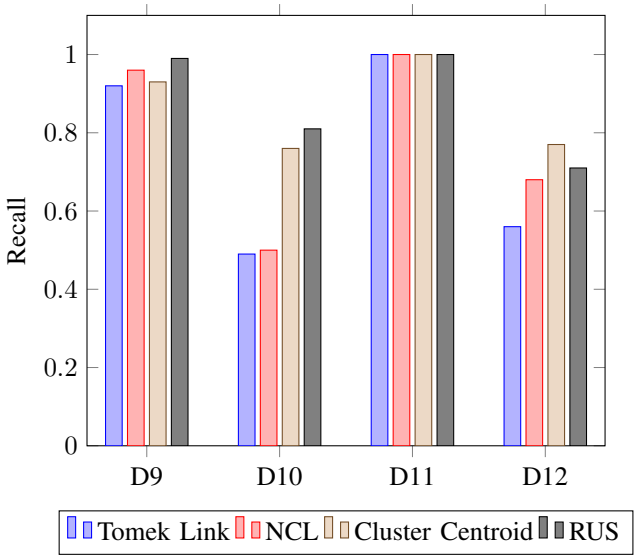


Fig. 5. Recall results for all high imbalance rate datasets.

- RUS provides less performance than NCL for all low imbalance rate datasets
- NCL provides better or equal performance than Tomek link for all medium imbalance rate group datasets
- Clusters Centroid provides better or equal performance than Tomek link for all medium imbalance rate group datasets
- RUS provides better or equal performance than Clusters Centroid for all medium imbalance rate group datasets
- Clusters Centroid provides the worst performance for all high imbalance rate group datasets except for last one with imbalance rate = 129.44

C. Gmean

From Table IV we can concluded the following points:

- Tomek link provides the worst performance for all low, medium, and high imbalance rate datasets
- NCL has better performance than Cluster Centroid for all datasets in the low imbalance rate group except the first one with imbalance rate = 1.82

TABLE III. F1-MEASURE RESULTS FOR ALL EXAMINED DATASETS USING THE SELECTED UNDERSAMPLING TECHNIQUES

Dataset	Tomek Link	NCL	Cluster Centroids	RUS
glass1	.74	.81	.65	.71
ecoli-0_vs_1	.64	.64	.62	.42
vehicle0	.8	.89	.91	.81
ecoli3	.65	.72	.71	.69
page-blocks0	.9	.9	.77	.86
glass4	.49	.74	.52	.83
car-good	.93	.96	.77	.92
kr-vs-k-one_vs_fifteen	1	1	.79	.95
kr-vs-k-zero_vs_eight	.95	.96	.6	.88
Winquality	.49	.5	.37	.4
kr-vs-k-zero_vs_fifteen	1	1	.89	.95
abalone19	.54	.44	.46	.41

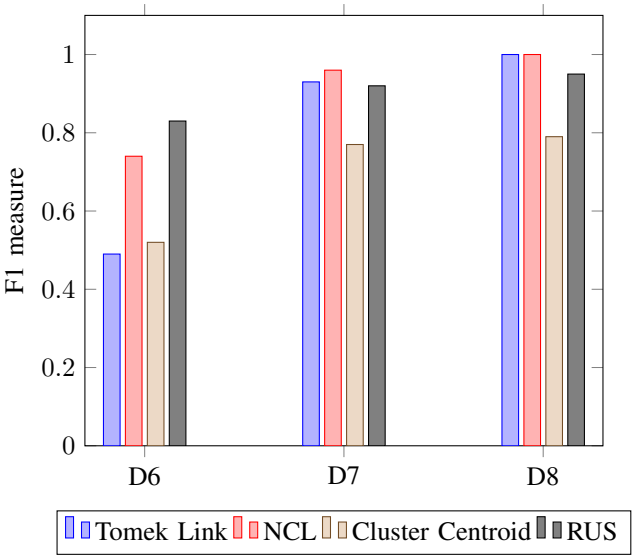


Fig. 6. F1 Measure results for all medium imbalance rate datasets.

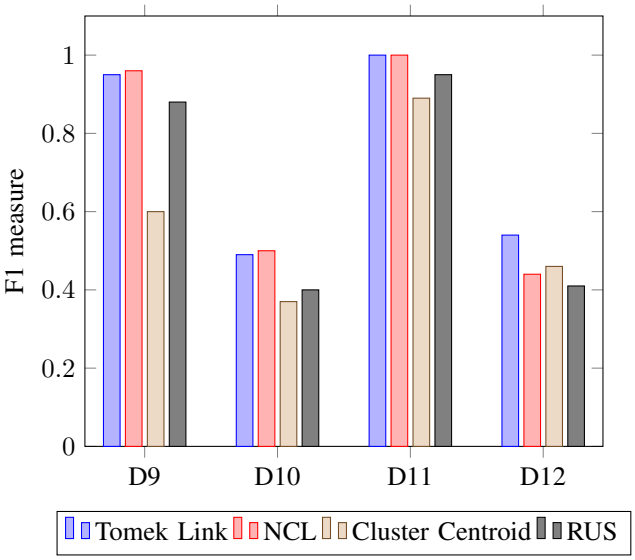


Fig. 7. F1 Measure results for all high imbalance rate datasets.

- NCL and RUS provide the same gmean value =1 for all datasets in the medium imbalance rate group and these sampling techniques outperformed Cluster

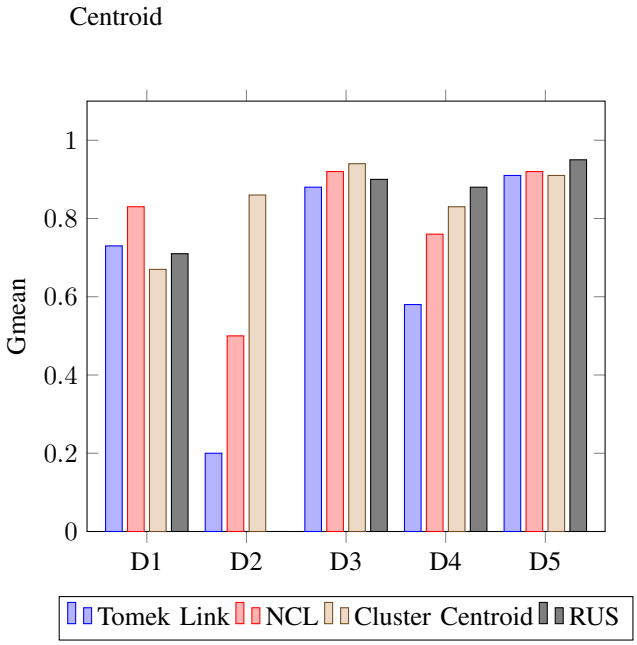


Fig. 8. Gmean results for all low imbalance rate datasets.

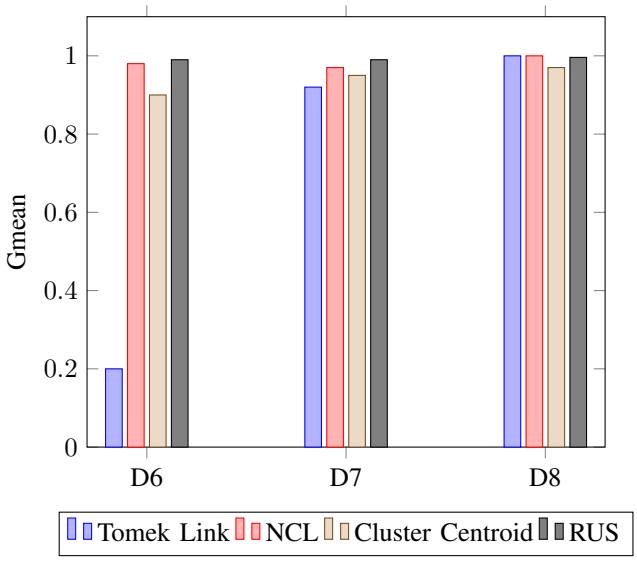


Fig. 9. Gmean results for all medium imbalance rate datasets.

TABLE IV. GMEAN RESULTS FOR ALL EXAMINED DATASETS USING THE SELECTED UNDERSAMPLING TECHNIQUES

Dataset	Tomek Link	NCL	Cluster Centroids	RUS
glass1	.73	.83	.67	.71
ecoli-0_vs_1	.2	.5	.86	0
vehicle0	.88	.92	.94	.9
ecoli3	.58	.76	.83	.88
page-blocks0	.91	.92	.91	.95
glass4	.2	.98	.9	.99
car-good	.92	.97	.95	.99
kr-vs-k-one_vs_fifteen	1	1	.97	.996
kr-vs-k-zero_vs_eight	.91	.96	.92	.99
Winquality	0	0	.73	.78
kr-vs-k-zero_vs_fifteen	1	1	.996	.998
abalone19	.35	.68	.77	.71

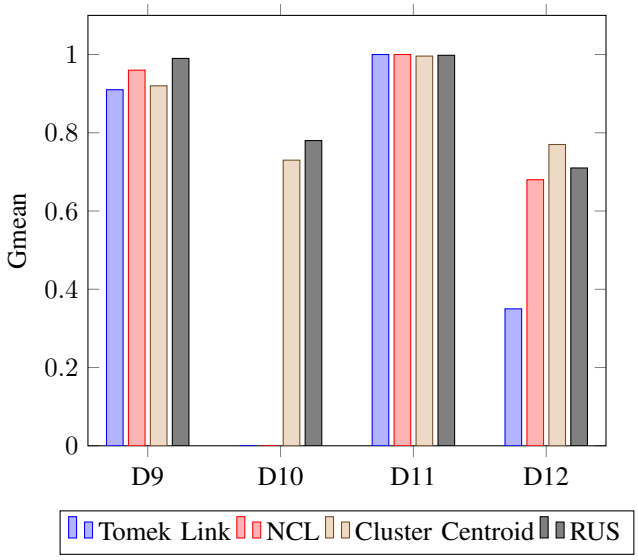


Fig. 10. Gmean results for all high imbalance rate datasets.

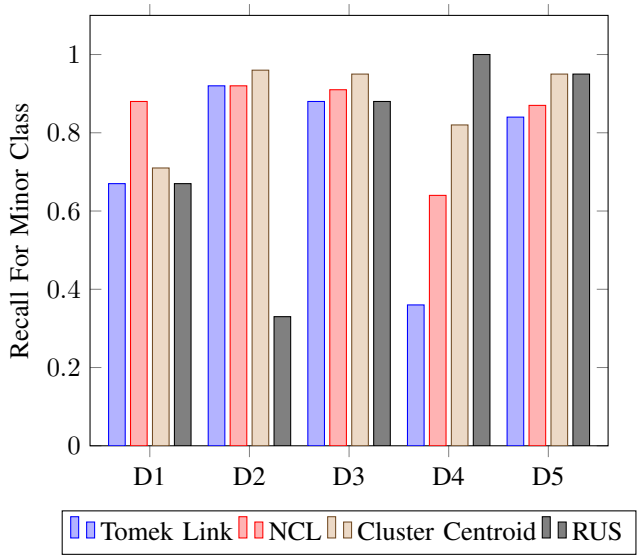


Fig. 11. Recall for minor class results for all low imbalance rate datasets.

D. Recall for Minor Class

From Table V we can concluded the following points:

- NCL outperformed Cluster Centroid for all datasets in the Low imbalance rate group except the first one with an imbalance rate = 1.82.
- NCL provides better or equal performance than Tomek link for all Low imbalance rate datasets.
- Clusters Centroid provides better performance than Tomek link for all Low imbalance rate datasets.
- Clusters Centroid and RUS provide recall value = 1 for the minor class for all datasets in the medium imbalance rate group.
- Tomek link provides the worst performance for all medium and high imbalance rate datasets.
- Clusters Centroid and RUS provide the same recall value for the minor class for all datasets in the high imbalance rate group.

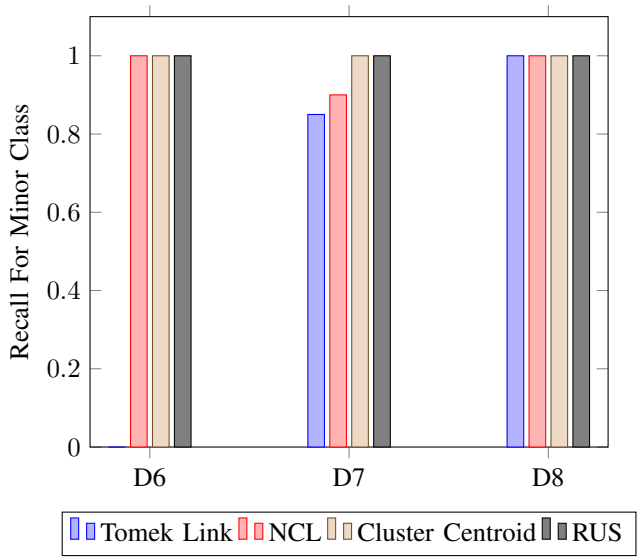


Fig. 12. Recall for minor class results for all medium imbalance rate datasets.

E. F1 Measure for Minor Class

From Table VI we can conclude the following points:

- NCL provides better or equal performance than Tomek link and Cluster Centroid for all low imbalance rate

TABLE V. RECALL FOR MINOR CLASS

Dataset	Tomek Link	NCL	Cluster Centroids	RUS
glass1	.67	.88	.71	.67
ecoli-0_vs_1	.92	.92	.96	.33
vehicle0	.88	.91	.95	.88
ecoli3	.36	.64	.82	1
page-blocks0	.84	.87	.95	.95
glass4	0	1	1	1
car-good	.85	.9	1	1
kr-vs-k-one_vs_fifteen	1	1	1	1
kr-vs-k-zero_vs_eight	.83	.92	1	1
Winquality	0	0	1	1
kr-vs-k-zero_vs_fifteen	1	1	1	1
abalone19	.12	.62	.75	.75

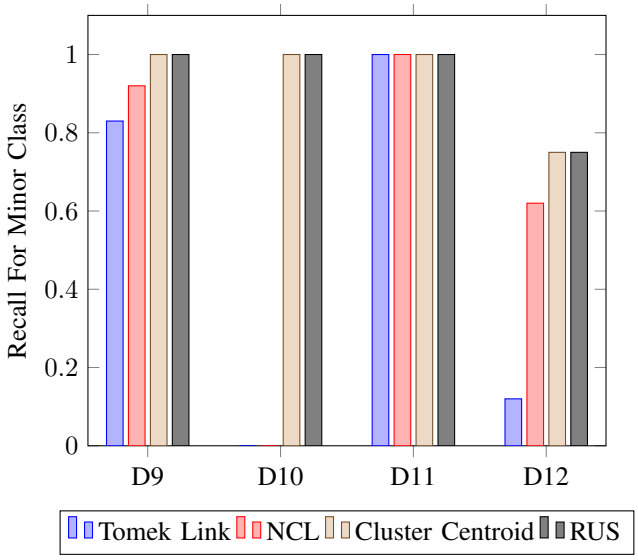


Fig. 13. Recall for minor class results for all high imbalance rate datasets.

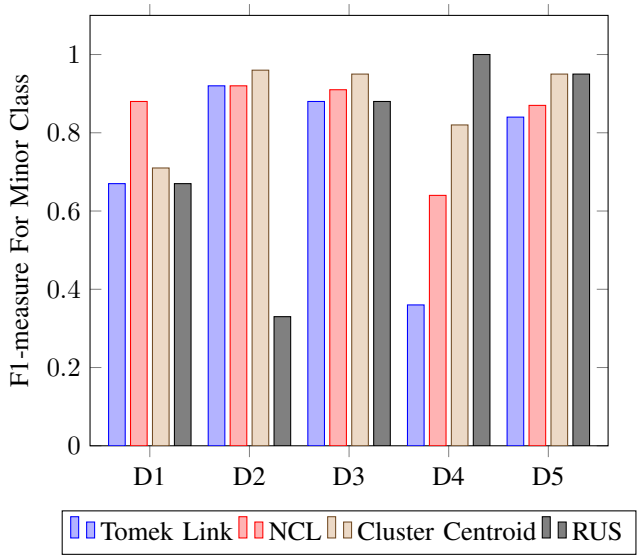


Fig. 14. F1-measure for minor class results for all low imbalance rate datasets.

datasets except vehicle0 dataset with imbalance rate = 3.25

- RUS provides less performance than NCL for all medium imbalance rate datasets except the first one with imbalance rate = 15.47
- Cluster Centroid provides less performance than NCL and RUS for all medium imbalance rate datasets
- RUS provides better performance than Clusters Centroid for all high imbalance rate datasets except for the abalone19 dataset with imbalance rate = 129.44

V. CONCLUSION AND FUTURE WORK

Sampling techniques are one of the most effective ways of handling imbalanced data set problems in machine learning.

This paper is concerned on comparing the effects of four common under-shambling techniques including Tomek Link, NCL, RUS, and Clusters Centroid in handling imbalance datasets problems for various Imbalance ratios ranges from low, medium, and high IR. Twelve imbalanced data sets with various IR have been used for comparison. DT has been used for classification.

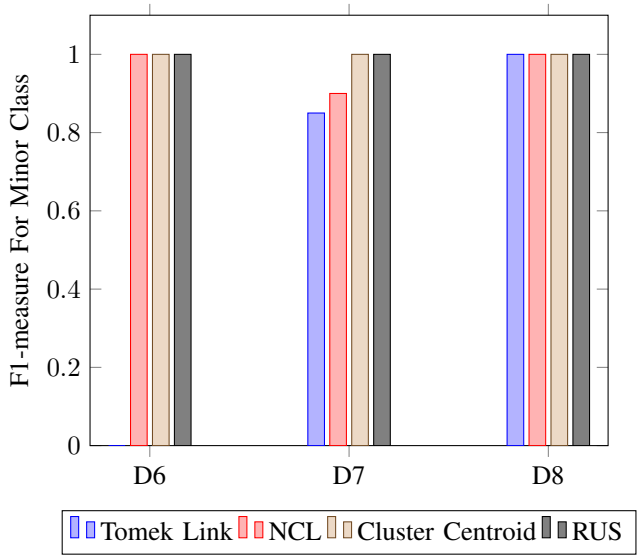


Fig. 15. F1-measure for minor class results for all medium imbalance rate datasets.

Results from all low IR datasets clearly show that NCL

TABLE VI. F1-MEASURE FOR MINOR CLASS

Dataset	Tomek Link	NCL	Cluster Centroids	RUS
glass1	.67	.78	.61	.64
ecoli-0_vs_1	.96	.96	.96	.74
vehicle0	.87	.83	.87	.87
ecoli3	.38	.52	.51	.51
page-blocks0	.82	.82	.6	.76
glass4	0	0.5	.14	.67
car-good	.87	.92	.28	.77
kr-vs-k-one_vs_fifteen	1	1	.61	.91
kr-vs-k-zero_vs_eight	.91	.92	.28	.77
Winquality	0	0	.04	.05
kr-vs-k-zero_vs_fifteen	1	1	.78	0.9
abalone19	.08	.03	.04	.03

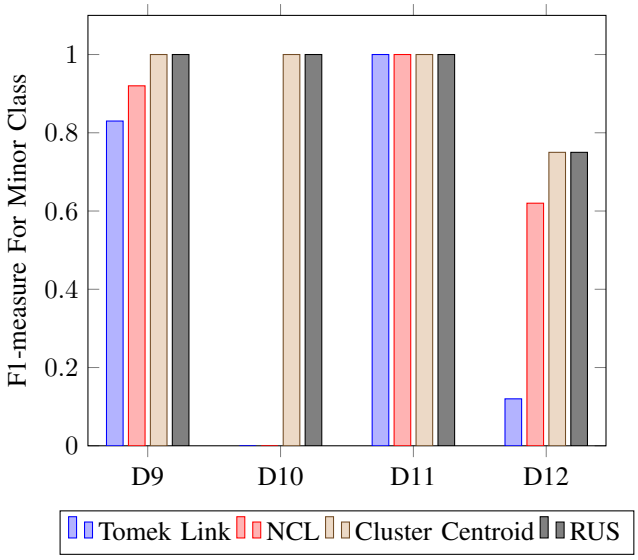


Fig. 16. F1-measure for minor class results for all high imbalance rate datasets.

outperformed both Tomek Link and RUS, while Clusters Centroid outperformed NCL based on recall. Based on minor class recall and gmean, NCL outperformed Clusters Centroid. RUS provides less performance than NCL for all low IR datasets.

For all medium IR datasets, Tomek Link provides the worst performance, while NCL and RUS outperformed other techniques in terms of recall, minor class recall, and gmean values. Based on the minor class F1 measure NCL outperformed RUS which outperformed Clusters Centroid based on the average F1 measure.

For all high IR datasets, Tomek Link provides the worst performance, while NCL and RUS outperformed other techniques based on recall, minor class recall, and gmean values. Based on the average F1 measure and minor class F1 measure, RUS provides better performance than Clusters Centroid.

Finally, the results presented in this paper were derived from the databases used here and according to the rates that were set to classify these databases for only four common under sampling techniques.

In the future, we need to compare the effect of these techniques by applying them to more databases in each IR category. Also, we can study more under-sampling techniques, and comparing them with other oversampling techniques.

REFERENCES

[1] N. S. Shikha Gupta, "Machine learning driven threat identification to enhance fanet security using genetic algorithm," *The International Arab Journal of Information Technology (IAJIT)*, vol. 21, no. 04, pp. 711 – 722, 1970.

[2] H. A. Owida, H. S. Migdadi, O. S. M. Hemied, N. F. F. Alshdaifat, S. F. A. Abuowaida, and R. S. Alkhawaldeh, "Deep learning algorithms to improve covid-19 classification based on ct images," *Bulletin of Electrical Engineering and Informatics*, vol. 11, no. 5, pp. 2876–2885, 2022.

[3] H. Owida, O. S. M. HEMIED, R. S. ALKHAWALDEH, N. F. F. ALSHDAIFAT, and S. F. A. ABUOWAIDA, "Improved deep learning approaches for covid-19 recognition in ct images," *Journal of Theoretical and Applied Information Technology*, vol. 100, no. 13, pp. 4925–4931, 2022.

[4] A. Y. Alhusenat, H. A. Owida, H. A. Rababah, J. I. Al-Nabulsi, and S. Abuowaida, "A secured multi-stages authentication protocol for iot devices." *Mathematical Modelling of Engineering Problems*, vol. 10, no. 4, 2023.

[5] S. ABUOWAIDA, E. ELSOUD, A. AL-MOMANI, M. ARABIAT, H. A. OWIDA, N. ALSHDAIFAT, and H. Y. CHAN, "Proposed enhanced feature extraction for multi-food detection method," *Journal of Theoretical and Applied Information Technology*, vol. 101, no. 24, 2023.

[6] H. Abu Owida, "Recent biomimetic approaches for articular cartilage tissue engineering and their clinical applications: narrative review of the literature," *Advances in Orthopedics*, vol. 2022, no. 1, p. 8670174, 2022.

[7] H. A. Owida, B. A.-h. Moh'd, and M. Al Takrouri, "Designing an integrated low-cost electrospinning device for nanofibrous scaffold fabrication," *HardwareX*, vol. 11, p. e00250, 2022.

[8] A. Al-Momani, M. N. Al-Refai, S. Abuowaida, M. Arabiat, N. Al-shdaifat, and M. N. A. Rahman, "The effect of technological context on smart home adoption in jordan," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 33, no. 2, p. 1186 – 1195, 2024.

[9] E. Alhenawi, R. Al-Sayyed, A. Hudaib, and S. Mirjalili, "Feature selection methods on gene expression microarray data for cancer classification: A systematic review," *Computers in Biology and Medicine*, vol. 140, p. 105051, 2022.

[10] Z. Salah and E. Abu Elsoud, "Enhancing network security: A machine learning-based approach for detecting and mitigating krack and kr00k attacks in ieee 802.11," *Future Internet*, vol. 15, no. 8, p. 269, 2023.

[11] H. Alazzam, A. Al-Adwan, O. Abualghanam, E. Alhenawi, and A. Alsmady, "An improved binary owl feature selection in the context of android malware detection," *Computers*, vol. 11, no. 12, p. 173, 2022.

[12] R. Al-Sayyed, E. Alhenawi, H. Alazzam, A. Wrikat, and D. Suleiman, "Mobile money fraud detection using data analysis and visualization techniques," *Multimedia Tools and Applications*, vol. 83, no. 6, pp. 17 093–17 108, 2024.

[13] S. Shukri, R. Al-Sayyed, H. Al-Bdour, E. Alhenawi, T. Almarabeh, and H. Mohammad, "Internet of things: Underwater routing based on user's health status for smart diving," *International Journal of Data and Network Science*, vol. 7, no. 4, pp. 1715–1728, 2023.

[14] M. Haj Qasem, M. Aljaidi, G. Samara, R. Alazaidah, A. Alsarhan, and M. Alshammari, "An intelligent decision support system based on multi agent systems for business classification problem," *Sustainability*, vol. 15, no. 14, p. 10977, 2023.

[15] ——, "An intelligent decision support system based on multi agent systems for business classification problem," *Sustainability*, vol. 15, no. 14, p. 10977, 2023.

[16] T. Sabbah, M. Ayyash, and M. Ashraf, "Hybrid support vector machine based feature selection method for text classification." *The International Arab Journal of Information Technology (IAJIT).*, vol. 15, no. 3A, pp. 599–609, 2018.

[17] E. Alhenawi, R. A. Khurma, P. A. Castillo, M. G. Arenas, and A. M. Al-Hinawi, "Effects of term weighting approach with and without stop words removing on arabic text classification," in *2023 9th International Conference on Optimization and Applications (ICOA)*, 2023, pp. 1–6.

[18] H. Alazzam, O. AbuAlghanam, A. Alsmady, and E. Alhenawi, "Arabic documents clustering using bond energy algorithm and genetic algorithm," in *2022 13th International Conference on Information and Communication Systems (ICICS)*. IEEE, 2022, pp. 4–8.

[19] O. K. A. Alidmat, K. Y. Umi, E. Alhenawi, H. J. Badarneh, R. Alazaidah, and L. Al-Rbabah, "Simulation of exit selection behavior evacuation based on an improved cellular automata model during fire disaster," in *2023 24th International Arab Conference on Information Technology (ACIT)*. IEEE, 2023, pp. 1–8.

[20] O. Alidmat, H. A. Owida, U. K. Yusof, A. Almaghthawi, A. Altalidi, R. S. Alkhawaldeh, S. Abuowaida, N. Alshdaifat, and J. AlShaqsi, "Simulation of crowd evacuation in asymmetrical exit layout based on improved dynamic parameters model," *IEEE Access*, 2024.

[21] R. Mohammed, J. Rawashdeh, and M. Abdullah, "Machine learning with oversampling and undersampling techniques: overview study and experimental results," in *2020 11th international conference on information and communication systems (ICICS)*. IEEE, 2020, pp. 243–248.

[22] R. Alazaidah, A. Al-Shaikh, M. Al-Mousa, H. Khafajah, G. Samara, M. Alzyoud, N. Al-Shanableh, and S. Almatarneh, "Website phishing detection using machine learning techniques," *Journal of Statistics Applications & Probability*, vol. 13, no. 1, pp. 119–129, 2024.

[23] R. Alazaidah, F. K. Ahmad, M. F. M. Mohsin, and W. A. AlZoubi, "Multi-label ranking method based on positive class correlations," *Jordanian Journal of Computers and Information Technology*, 2020.

[24] O. AbuAlghanam, H. Alazzam, E. Alhenawi, M. Qatawneh, and O. Adwan, "Fusion-based anomaly detection system using modified isolation forest for internet of things," *Journal of Ambient Intelligence and Humanized Computing*, pp. 1–15, 2022.

[25] V. Ganganwar, "An overview of classification algorithms for imbalanced datasets," *International Journal of Emerging Technology and Advanced Engineering*, vol. 2, no. 4, pp. 42–47, 2012.

[26] J. L. Leevy, T. M. Khoshgoftaar, R. A. Bauder, and N. Seliya, "A survey on addressing high-class imbalance in big data," *Journal of Big Data*, vol. 5, no. 1, pp. 1–30, 2018.

[27] B. Krawczyk, "Learning from imbalanced data: open challenges and future directions," *Progress in Artificial Intelligence*, vol. 5, no. 4, pp. 221–232, 2016.

[28] N. Japkowicz, "The class imbalance problem: Significance and strategies," in *Proc. of the Int'l Conf. on Artificial Intelligence*, vol. 56. Citeseer, 2000, pp. 111–117.

[29] X. Guo, Y. Yin, C. Dong, G. Yang, and G. Zhou, "On the class imbalance problem," in *2008 Fourth international conference on natural computation*, vol. 4. IEEE, 2008, pp. 192–201.

[30] N. Japkowicz and S. Stephen, "The class imbalance problem: A systematic study," *Intelligent data analysis*, vol. 6, no. 5, pp. 429–449, 2002.

[31] S.-J. Yen and Y.-S. Lee, "Under-sampling approaches for improving prediction of the minority class in an imbalanced dataset," in *Intelligent Control and Automation*. Springer, 2006, pp. 731–740.

[32] ——, "Cluster-based under-sampling approaches for imbalanced data distributions," *Expert Systems with Applications*, vol. 36, no. 3, pp. 5718–5727, 2009.

[33] Y.-P. Zhang, L.-N. Zhang, and Y.-C. Wang, "Cluster-based majority under-sampling approaches for class imbalance learning," in *2010 2nd IEEE International Conference on Information and Financial Engineering*. IEEE, 2010, pp. 400–404.

[34] M. M. Rahman and D. Davis, "Cluster based under-sampling for unbalanced cardiovascular data," in *Proceedings of the world congress on engineering*, vol. 3, 2013, pp. 3–5.

[35] H. A. Owida, N. Alshdaifat, A. Almaghthawi, S. Abuowaida, A. Aburomman, A. Al-Momani, M. Arabiat, and H. Y. Chan, "Improved deep learning architecture for skin cancer classification," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 36, no. 1, p. 501 – 508, 2024. [Online]. Available: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85200149026&doi=10.11591%2fijeecs.v36.i1.pp501-508&partnerID=40&md5=0b7a43e8a8aac2d4ae6e3688bd3f6f93

[36] T. Hasanin and T. Khoshgoftaar, "The effects of random undersampling with simulated class imbalance for big data," in *2018 IEEE international conference on information reuse and integration (IRI)*. IEEE, 2018, pp. 70–79.

[37] T. Hasanin, T. M. Khoshgoftaar, J. Leevy, and N. Seliya, "Investigating random undersampling and feature selection on bioinformatics big data," in *2019 IEEE Fifth International Conference on Big Data Computing Service and Applications (BigDataService)*. IEEE, 2019, pp. 346–356.

[38] M. Saripuddin, A. Suliman, S. Syarmila Sameon, and B. N. Jorgensen, "Random undersampling on imbalance time series data for anomaly detection," in *2021 The 4th International Conference on Machine Learning and Machine Intelligence*, 2021, pp. 151–156.

[39] T. Elhassan and M. Aljurf, "Classification of imbalance data using tomek link (t-link) combined with random under-sampling (rus) as a data reduction method," *Global J Technol Optim S*, vol. 1, 2016.

[40] R. Zuech, J. Hancock, and T. M. Khoshgoftaar, "Detecting web attacks using random undersampling and ensemble learners," *Journal of Big Data*, vol. 8, no. 1, pp. 1–20, 2021.

[41] J. L. Leevy, J. Hancock, T. M. Khoshgoftaar, and N. Seliya, "Iot reconnaissance attack classification with random undersampling and ensemble feature selection," in *2021 IEEE 7th International Conference on Collaboration and Internet Computing (CIC)*. IEEE, 2021, pp. 41–49.

[42] M. H. Popel, K. M. Hasib, S. A. Habib, and F. M. Shah, "A hybrid under-sampling method (husboost) to classify imbalanced data," in *2018 21st international conference of computer and information technology (ICCIT)*. IEEE, 2018, pp. 1–7.

[43] K. Agustianto and P. Destarianto, "Imbalance data handling using neighborhood cleaning rule (ncl) sampling method for precision student modeling," in *2019 International Conference on Computer Science, Information Technology, and Electrical Engineering (ICOMITEE)*. IEEE, 2019, pp. 86–89.

[44] Y. Wu, J. Yao, S. Chang, and B. Liu, "Limcr: Less-informative majorities cleaning rule based on naïve bayes for imbalance learning in software defect prediction," *Applied Sciences*, vol. 10, no. 23, p. 8324, 2020.

[45] Y. Zhang, H. Zhang, X. Zhang, and D. Qi, "Deep learning intrusion detection model based on optimized imbalanced network data," in *2018 IEEE 18th International Conference on Communication Technology (ICCT)*. IEEE, 2018, pp. 1128–1132.

[46] P. Gulati, "Hybrid resampling technique to tackle the imbalanced classification problem," *Applied Sciences*, 2020.

[47] S. Choirunnisa and J. Lianto, "Hybrid method of undersampling and oversampling for handling imbalanced data," in *2018 International Seminar on Research of Information Technology and Intelligent Systems (ISRITI)*. IEEE, 2018, pp. 276–280.

[48] Q. Ning, X. Zhao, and Z. Ma, "A novel method for identification of glutarylation sites combining borderline-smote with tomek links technique in imbalanced data," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 2021.

[49] E. F. Swana, W. Doorsamy, and P. Bokoro, "Tomek link and smote approaches for machine fault classification with an imbalanced dataset," *Sensors*, vol. 22, no. 9, p. 3246, 2022.

[50] A. Bansal and A. Jain, "Analysis of focussed under-sampling techniques with machine learning classifiers," in *2021 IEEE/ACIS 19th International Conference on Software Engineering Research, Management and Applications (SERA)*. IEEE, 2021, pp. 91–96.

[51] O. Tarawneh, Q. Saber, A. Almaghthawi, H. A. Owida, A. Issa, N. Alshdaifat, G. Jaradat, S. Abuowaida, and M. Arabiat, "The effect of pre-processing on a convolutional neural network model for dorsal hand vein recognition." *International Journal of Advanced Computer Science & Applications*, vol. 15, no. 3, 2024.

[52] H. A. Owida, M. R. Hassan, A. M. Ali, F. Alnaimat, A. Al Sharah, S. Abuowaida, and N. Alshdaifat, "The performance of artificial intelligence in prostate magnetic resonance imaging screening," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 14, no. 2, pp. 2234–2241, 2024.

[53] A. Al Sharah, H. A. Owida, F. Alnaimat, and S. Abuowaida, "Application of machine learning in chemical engineering: outlook and perspectives," *Int J Artif Intell*, vol. 13, no. 1, pp. 619–630, 2024.

[54] N. Alshdaifat, M. A. Osman, and A. Z. Talib, "An improved multi-object instance segmentation based on deep learning," *Kuwait Journal of Science*, vol. 49, no. 2, 2022.

[55] J. Wang and Y. Wang, "Fd technology for hss based on deep convolutional generative adversarial networks." *The International Arab Journal of Information Technology (IAJIT)*, vol. 21, no. 2, pp. 299–312, 2024.

[56] G. Lemaître, F. Nogueira, and C. K. Aridas, "Imbalanced-learn: A python toolbox to tackle the curse of imbalanced datasets in machine learning," *The Journal of Machine Learning Research*, vol. 18, no. 1, pp. 559–563, 2017.

[57] A. More, "Survey of resampling techniques for improving classification performance in unbalanced datasets," *arXiv preprint arXiv:1608.06048*, 2016.

[58] J. Laurikkala, "Improving identification of difficult small classes by balancing class distribution," in *Conference on artificial intelligence in medicine in Europe*. Springer, 2001, pp. 63–66.

[59] S. Kotsiantis, D. Kanellopoulos, P. Pintelas *et al.*, "Handling imbalanced datasets: A review," *GESTS international transactions on computer science and engineering*, vol. 30, no. 1, pp. 25–36, 2006.

[60] H. Rawashdeh, S. Awawdeh, F. Shannag, E. Henawi, H. Faris, N. Obeid, and J. Hyett, "Intelligent system based on data mining techniques for prediction of preterm birth for women with cervical cerclage," *Computational biology and chemistry*, vol. 85, p. 107233, 2020.

[61] H. Abu Owida, G. AlMahadin, J. I. Al-Nabulsi, N. Turab, S. Abuowaida, and N. Alshdaifat, "Automated classification of brain tumor-based magnetic resonance imaging using deep learning approach." *International Journal of Electrical & Computer Engineering (2088-8708)*, vol. 14, no. 3, 2024.

[62] A. K. Shukla, P. Singh, and M. Vardhan, "A hybrid gene selection method for microarray recognition," *Biocybernetics and Biomedical Engineering*, vol. 38, no. 4, pp. 975–991, 2018.

Multiclass Chest Disease Classification Using Deep CNNs with Bayesian Optimization

Maneet Kaur Bohmrah, Harjot Kaur
Department of Computer Science and Engineering
Guru Nanak Dev University
Amritsar, India, 143001

Abstract*—Ever since its outbreak, numerous research studies have been initiated worldwide as an attempt for an accurate and efficient diagnosis of COVID-19. In the recent past, patients suffering from various chronic lung diseases, either passed away due to COVID-19 or Pneumonia. Both of these pulmonary diseases are strongly correlated as they share a common set of symptoms and even for medical professionals, it has been difficult to perform discerned diagnosis for both of these diseases. The dire need of the current scenario is a chest-disease diagnosis framework for accurate, precise, real-time and automatic detection of COVID-19 because of its mass fatality rate. The review of various contemporary and previous research works show that the currently available computer-aided diagnosis systems are insufficient for real-time implementation of COVID-19 prediction due to their long training time, substantial memory requirements and excessive computations. This work proposes an optimized hybrid DNN-ML framework by combining Deep Neural Networks' (DNNs) models and optimized Machine Learning (ML) classifiers along with an efficacious image preprocessing approach. For feature extraction, Deep learning (DL) models namely GoogleNet, EfficientNetB0, and ResNet50 have been deployed and extracted features have been further fed to Bayesian optimized ML classifiers. The two major contributions of this study are, Edge based Region of Interest (ROI) extraction and use of Bayesian optimization approach for configuring optimal architectures of ML classifiers. With extensive experimentation, it has been observed that the proposed optimized hybrid DNN-ML model with encapsulated image preprocessing techniques performed much better as compared to various previously existing ML-DNN models. Based on the promising results obtained from this proposed light weight hybrid framework, it has been concluded that, this model can facilitate radiologists, while functioning as an accurate disease diagnosis and support system for early detection of COVID-19 and Pneumonia.

Keywords*—Deep neural networks; machine learning; Bayesian optimization; image preprocessing; COVID-19; pneumonia

I. Introduction

Recently, a new virus known as **COVID-19** emerged in China and began to spread globally as an respiratory illness. Since its outburst, COVID-19 has contributed significantly to the economic crisis of numerous nations and adversely affected human life. Due to its transmissible characteristics, it can spread vigorously with uncertain transmission methods and co-exist for a longer duration of time [1]. According to statistics received from World Health Organization (WHO) [2], approximately six million people worldwide have died till date because of COVID-19 and over forty million cases have been reported so far. People who are older or have chronic health conditions seem to be more susceptible to contacting COVID-19 infection. Various symptoms of COVID-19 include high fever, coughing, anxiety and breathlessness. This virus spreads quickly via respiratory droplets produced by an infected person's cough or sneeze [3].

Numerous medical professionals globally, have been developing vaccines and researching on treatments to combat this virus. Moreover, many medical techniques and therapies have been developed and are currently under development for treatment and recovery of the affected individuals. Unfortunately, despite several protective measures, the available medical systems have failed to combat and control the virus, because it has been continuously undergoing several mutations [4]. This study requires advanced diagnosis and treatment methodologies to control its menace. Presently, COVID-19 has been diagnosed using time consuming primitive methods like, administering RT-PCR (Reverse Transcription - Polymerase Chain Reaction) examination. Another alternative is *Computed Tomography (CT) scan and Chest X-ray (CXR)* images, that have emerged as an robust imagery techniques for diagnosis of the same [5]

While investigating CXR and CT scan images, notable clinical findings that can be inferred are, ground glass opacities (GGO), thickening of intertobular septa and air branchogram sign, with or without increased broncho-vascular markings that lead to diagnosis of disparate lung diseases. Among these two medical imaging tools, CT scan images have been considered more reliable due to their high contrast image properties. And, they are more effective medical imaging system for diagnosis of the chest diseases. As compared, CXR images, have been widely recommended by doctors, because they are more economical as compared to CT scan image and easily accessible for patients too. For effective screening and diagnosis of COVID-19 and/or other chest infections, CXR and CT scan images have to be manually examined and then clinically correlated with patient's symptoms. Nevertheless, this manual screening is a very time consuming process and might not be feasible in emergency cases [6]. Therefore, an accurate, precise, real-time and automated COVID-19 diagnostic system is the dire need of the current scenario.

With the advent of prominent AI tools and medical imaging, researchers have started proposing novel solutions to develop automatic tools for accurate detection of COVID-19. AI based deep learning (DL) models, i.e., *Convolutional Neural Networks (CNNs)* excel in domain of image classification and provide extraordinary performance in medical image analysis. Due to their complex architecture, CNNs can be used for both feature extraction and classification tasks, which makes

them distinguishable from rest of Machine Learning (ML) algorithms [7].

In medical imaging system, accurate and precise results, are always a major concern. For image classification task, a model should be well-versed with dominant features of an image. The ultimate motivating factor for the present study is that, all the image characteristics needed for feature extraction and classification should be noise free [8]. Henceforth, this study has focused more on image preprocessing techniques (segmentation, filtration and enhancement) in order to pertain the prominent features of the image. Afterwards, resultant segmented and enhanced image has been utilized by the pretrained CNN models for feature extraction. And finally, in order to classify the features, Bayesian optimized ML classifiers have been used in this work. Various innovative and novel contributions of the present study are as follows:

- Experimentation with the new enhanced model for the diagnosis of COVID-19 for two distinct types of COVID-19 images; the Chest X-ray and CT scan image datasets.
- Proposal of Edge based Adaptive Segmentation algorithms and their integration with image filtration and enhancement techniques for the preparation of enhanced dataset.
- Utilization of pretrained models and machine learning classifiers for feature extraction and classification process, respectively.
- Implementation of Bayesian optimization technique for disparate ML classifiers.
- Performance based comparative analysis of proposed DNN-ML models trained with and without image preprocessing techniques, thereby highlighting the advantages of latter.
- Comparative analysis based on classification accuracy of Bayesian optimized ML classifiers (BO-ML) with their non-optimized ML version, thereby investigating the significance of Bayesian optimization and deducing experimental insights on CXR and CT scan datasets.
- Comparison of the experimental findings with the other existing state-of-the-art models.

The remaining part of this article has been structured as follows, Section II highlights the state-of-the-art research studies conducted for automated diagnosis of COVID-19. Section III describes the datasets, the proposed framework, and methods used. Section IV presents various experiments performed along with the findings, discussions and comparative analysis. Finally, the conclusions and future work have been discussed in Section V.

II. Literature Review

In recent past, numerous research studies have been performed on automatic COVID-19 detection using disparate pretrained Deep Neural Networks (DNNs) and ML algorithms from lung images, in a stand-alone and hybrid mode. Kesav et al. [9] employed GoogleNet, a pretrained neural network for feature extraction and used different ML classifiers for performing the task of classification. On chest X-ray image dataset, Bayesian optimization has been utilized to optimize ML classifiers with accuracy score of 98.31% for binary and 98.60% for multi-class classification respectively.

Hamza et al. [10] proposed the use of Bayesian optimized neural networks and gradcam technique for visualization of chest x-rays to detect chest diseases. Arman et al. [11] designed modified Xception model based on Bayesian optimization and compared it with other DNN models such as VGG16, MobileNetV2 and InceptionV3. The proposed model achieved the highest value of classification accuracy, i.e., 99.4% to classify COVID-19, normal and pneumonia classes belonging to CXR image dataset. Canayaz et al. [12] presented a combination of ResNet50 and Bayesian optimized kNN to detect COVID-19 and this model accomplished an accuracy score of 96.42% when trained using CT scan dataset with 349 images for binary classification (COVID-19 and non COVID-19 categorization).

Awal et al. [13] experimented using various machine learning classifiers, i.e., Linear Discriminant Analysis (LDA), Quadratic Discriminant Analysis (QDA), Naive Bayes(NB), k-Nearest Neighbour (kNN), Decision Tree (DT), Random Forest (RF), extreme Gradient Boosting (XGBoost) for COVID-19 diagnosis. To optimize hyperparameters of aforementioned ML classifiers, Bayesian Optimization has been deployed and the results hence obtained showed that optimized XGBoost performed better as compared to other ML classifiers. Aslan et al. [14] utilized Artificial Neural Networks (ANN) based segmentation method for dataset enhancement and compared eight different pretrained DNNs. Further, three Bayesian optimized ML classifiers (i.e., SVM, kNN and NB) have been used in integration with these pretrained models. The results concluded that DenseNet201 and BO-SVM outperformed other competing models with 96.29% accuracy score. Nour et al. [15] proposed CNN for feature extraction and utilized Bayesian optimization based ML (SVM, kNN and DT) classifiers, achieving 98.97% accuracy.

Jaiswal et al. [16] deployed a pretrained DenseNet201 model to classify COVID-19 using CT scan images with an accuracy score of 96.25%. The combination of Bayes and SqueezeNet, has been utilized by Ucar et al [17] to detect COVID-19 using CXR images and the accuracy score accomplished was 98.30% . A Gravitational Search Algorithm (GSA) has been deployed to optimize hyperparameters of DenseNet121 (Ezzat et al.[18]) for classification of COVID-19 using CXR images and with the an accuracy value of 98.38%. Das et al. [19] presented a fully automated COVID-19 detection model. COVID-19 radiography dataset from Kaggle repository has been used that consisted of three classes: COVID-19 positive, pneumonia infection, and no infection. Two CNN models namely, VGG16 and ResNet50 have been implemented and compared in this work. With an accuracy score of 97.67%, VGG16 model provided the best performance in automated COVID-19 detection model.

According to Monshi et al. [20], the hyperparameter optimization of ML classifiers provided best results. The pretrained models have been implemented in this work to design a novel CovidXrayNet framework based on EfficientNet-B0 and Bayesian optimized ML classifiers. The testing accuracy value

of this model on data generated from two distinct databases has been revealed as 95.82%.

To detect COVID-19, Panwar et al. [21] presented a Transfer Learning (TL) based nCOVnet Deep Learning approach and obtained an accuracy score of 88.10% in training and testing of the CT scan images. Asif et al. [22] designed CNN model to diagnose COVID-19 using CXR images. Without applying any preprocessing techniques, input images have been fed into the Inception-V3 model and accomplished classification accuracy score of 96%.

COVID-19 and Pneumonia detection has been performed using 3-stage model based on CXR images (Bhattacharya et al. [23]). To start with, the affected lung region has been separated from the CXR images by deploying the Conditional Generative Adversarial Network (C-GAN). The characteristics from segmented lung pictures have been extracted in the stage two by using DNN based feature extraction model. Afterwards, different ML classifiers have been deployed to categorize the CXR images according to the extracted features. The proposed combination of VGG19 with Binary Robust Invariant Scale Key-points (BRISK) yielded the best classification accuracy score of 96.6%.

Kaur et al. [24] proposed a classier fusion model using ResNet50 to diagnose COVID-19 using CXR images. Kumar et al. [25] designed a model with the blended features of MobileNetV2 and DarkNet19 model based on CT scan dataset. This was one of the earliest attempts of using open-source DNNs for COVID-19 detection from CXR images. An automated tool for COVID-19 diagnosis, COVID-Net (Linda Wang et al. [26]) has been proposed and testing of same along with VGG19 and ResNet50 has been performed on COVIDx (an open-access dataset formed from collection of five different datasets). This experimental setup attained 93.3% accuracy. The performance of seven deep learning models have been examined (Khalid El Asnaoui et al.[27]) in order to identify and categorize COVID-19 and Pneumonia. The process of image preprocessing has been applied on the raw CT and X-ray images to enhance their quality and an accuracy score of 92.18% has been obtained by using Inception-ResNetV2.

A transfer learning based COVID-19 detection from CXR and CT images has been proposed by Afshar Shamsi et al. [28]. To accomplish the classification objective, four different pretrained models (namely, VGG16, ResNet50, DenseNet121 and InceptionResNet) have been used for feature extraction, and the extracted features were then fed into ML models. With an accuracy score of 87.9%, the combination of ResNet50 model and SVM classifier produced the best results. Arellano et al. [29] proposed a modified pretrained DenseNet121 model, with relearning of last layer for identification of COVID-19 from CXR images. A publicly available COVID-19 database has been used for training the model with an accuracy value of 94.7%.

Most of studies in the existing literature used raw images directly or combined with computationally intensive image preprocessing functions. The vital image features such as texture, color, shape, etc. play a significant role in the process of image classification, and it is performed by the DNN model. Henceforth, this tri-modular approach towards the proposal of optimized hybrid DNN-ML model encompasses first module for feature processing task, which includes the use of proposed image preprocessing techniques for image enhancement. Second module uses pretrained DNN models for feature extraction from enhanced images and third uses optimized ML for classification from the extracted image features.

III. Materials and Methods

This section provides details about various encompassing the proposed hybrid DNN-ML model. The datasets used for training of the proposed model have also been discussed in this section.

A. Description of the Dataset Used

For verification and validation of the proposed optimized hybrid DNN-ML framework, two different and original CXR (Chest X-Rays) and CT scan datasets have been utilized. The first dataset, i.e, CXR dataset, is a multiclass dataset, consists of COVID-19, viral pneumonia, and normal CXR image classes. The second dataset, comprises CT scan images, is a binary class dataset consisting of only COVID and non-COVID (COVID negative) image classes. First dataset in original form has been collected from different sources ([30], [31], [32]). The reason behind choosing these sources is as they contain original or non-processed CXR images. Because of continuous and ongoing research on COVID-19 disease, the newer version of this dataset may contain processed images which may lose their originality with the passage of time. The collected CXR image dataset is a balanced image repository containing 364 images for COVID-19 and other classes in order to avoid the problem of overfitting. Similarly, the second dataset related to lung disease patients comprises 349 CT scan images for both the classes (i.e., COVID-19 and non-COVID-19) [33].

The CXR images of patients suffering from COVID-19, Viral Pneumonia diseases, and Normal CXR images, and CT scan images of COVID-19 and COVID-19 negative patients have been presented in Fig. 1(a) and (b) respectively. CXR images of patients infected with COVID-19 show hyperlucent lung area denoting hyperinflation of lungs due to hindrance of small airways. CXR images with no abnormality detected show chest wall with normal shape and size plus trachea with normal appearance and zero opacity. CXR images of patients infected with viral pneumonia represent diffused bilateral GGO showing disperse alveolar. For training of the proposed hybrid model, enhanced version of both datasets have been prepared by using image preprocessing techniques, as presented in the next section. In order to study the effects of applied image preprocessing technique, results for the same have been verified using both original and enhanced CXR and CT scan image datasets.

B. Image Preprocessing

A variety of proficient and non-invasive medical imaging systems such as MRI, PET, Endoscopy, CT scan and X-Rays can be used to capture images internal infected parts of various organs in the human body. The two most commonly used medical imaging modalities for diagnosis of infectious lungs diseases are CT scan and CXR images. Both raw CXR and

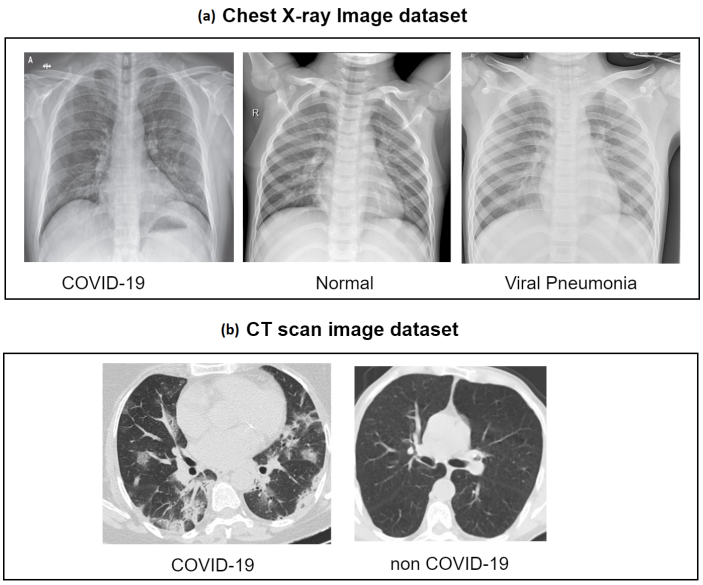


Fig. 1. Sample images from the original CXR and CT scan image datasets.

CT scan images can contain different types of noise patterns, including unrelated blobs, soft lung tissues and fine blood vessels. Henceforth, classification results can be misguided by soft lung tissues and minute blood vessels found in CXR images and CT scans. Thus, in order to improve the model accuracy, it is imperative to preprocess these medical images so that pertinent information can be extracted from the same after eliminating various types of noise patterns [34]. Also, to enhance the image quality of the raw medical data, image preprocessing can be applied to digitized images for enhancement of visual information contained in them [35].

The proposed image preprocessing approach applied to raw CXR images and CT scan images for the purpose of image enhancement comprises the following steps:

1) Image Segmentation (Using Edge Detection)
2) Image Filtration (Median filter)
3) Image Enhancement (Intensity improvement using CLAHE's method)

1) Proposed Image Segmentation Technique: **Image Segmentation** performs segmentation or masking of meaningful region or *Region of Interest (ROI)* area from an image so that there is no effect of noise patterns on the accuracy of image classification process [36].

For detection of COVID-19 using CXR images and CT scans, the target ROI is lungs (i.e. both left and right lung regions). The training of DNN model based on masked lung area, i.e., ROI will always provide with a more accurate disease prediction. The process of image segmentation used to perform ROI extraction from CXR and CT scan images is a very time-consuming and tedious process due to large size of COVID-19 image datasets; therefore, it can't be performed manually. It has been also shown in previous studies that, AI based U-NET model as the conditional Generative Adversial Networks (GANs) [14] [23] has been used for the extraction of ROI regions. The inclusion of DNNs for image segmentation process in various previous studies has contributed to additional overhead in terms of overall training time and computational complexity.

Originally, both CT scan and CXR images are *grayscale* in nature. But, the available CT scan and CXR images in the used dataset may contain some blue colored biomarkers. Therefore, for generation of accurate binary mask in the image segmentation step, it is important to convert these existing original images into grayscale. This study devises an *adaptive edge based image segmentation technique* has been devised as an intermediate and transitional solution for fast and accurate ROI extraction from the original images. This proposed technique uses *Canny Edge Detection* and a set of *Morphological operations* for automatic extraction of lungs ROI. It consists of three major steps.

1) *Image smoothing*, which is performed in the first step by using *Gaussian filter* function $G_f(x,y)$[1], (equation 1);
2) The second step involves calculation of *gradient magnitude* $G_m(x,y)$ and gradient direction $\theta(x,y)$ (equations 2 and 3);
3) Application of *edge point detection and connection* function $E_p(x,y)$ has been performed in step 3 (equation 4).

$$G_f(x,y) = [G_x(x,y) * \ G_y(x,y)] * \ I(x,y) \tag{1}$$

$$G_m(x,y) = \sqrt{G_x^2(x,y) + G_y^2(x,y)} \tag{2}$$

$$\theta(x,y) = arctan\left[\frac{G_y(x,y)}{G_x(x,y}\right] \tag{3}$$

where,

- $G_x(x,y) = (-P_1 + P_2 - P_3 + P_4)/2$;
- $G_y(x,y) = P_1 + P_2 - P_3 - P_4)/2$;
- $I(x,y)$ is an original grayscale CXR/CT scan image;
- P_1, P_2, P_3, P_4 are the pixel values of the coordinates (x,y), $(x+1,y)$, $(x,y+1)$, $(x+1,y+1)$, respectively of an original grayscale image;
- x and y represent the corresponding row and column of an original grayscale image.

The proposed image segmentation technique uses *canny edge detection method* combined with *Hough Transform (HT)* [37] for connection of edge points. Hough Transform method overcomes the boundary leakage problem that arises during the process of edge detection by finding all the collinear points (along the directions of an edge) by joining them as edge points, thereby producing efficient results for low intensity images as well.

[1]Gaussian filter is based on Gaussian function, and it can be defined as

$$G(x,y) = \frac{1}{2\pi\sigma^2}\exp\left[-\frac{(x)^2+(y)^2}{2\pi\sigma^2}\right]$$

$$E_p(x,y) = \begin{cases} \text{mark (x,y) as edge points in } G_f(x,y)^2 & \text{if } G_m(x.y) \geq t \\ \text{mark (x,y) as candidate edge points,} & \text{otherwise} \end{cases} \quad (4)$$

where, t is the threshold value, obtained by applying Ostu method [38]. The resultant HT Canny Edge-detected binary image $BM(x,y)$ is further preprocessed by using 3-step application of widely acceptable *dilation, fill and erosion* morphological operators. First step involves application of dilation operator(dil) for edge dilation to $BM(x,y)$ resulting in $BM_1(x,y)$ (Eq. 5) The second step comprises use of fill operation ($fill$) (Eq. 6) on dilated image $BM1(x,y)$ for filling the holes based object intensity of displayed region resulting in a filled image $BM_2(x,y)$. The third and final step deploys the erosion operator ($erode$) (Eq. 7) on $BM_2(x,y)$ for removal of the connected components on the edge boundary resulting in computation of a final binary image $BM_{final}(x,y)$.

$$BM_1(x,y) = dil(G_f(x,y)); \quad (5)$$

$$BM_2(x,y) = fill(BM_1(x,y)); \quad (6)$$

$$BM_{final}(x,y) = erode(BM_2(x,y)); \quad (7)$$

$$I_s(x,y) = BM_{final}(x,y) * \ I(x,y) \quad (8)$$

where

- $BM_1(x,y)$- A binary mask obtained after edge detection;
- $BM_2(x,y)$- A binary mask obtained after edge dilation;
- $BM_{final}(x,y)$- The final binary mask obtained after removal of connected components;
- $I_s(x,y)$ - The final segmented image.

The final segmented image $I_s(x,y)$ with the extracted ROI is procured by segmenting an original grayscale image with the computed binary image $BM_{final}(x,y)$ mask, resulting from the application of edge detection and morphological operators as a binary mask (Eq. 8).

2) Image Filtration Methodology: **Image filtration** is responsible for modification or enhancement of image modification by removing different types of noise[3] existing in them. *Noise* can be classified as *Gaussian, Salt and Pepper, Poisson, impulsive and speckle noise*. The CXR and CT scan images which have been used to validate proposed model are mostly subject to Gaussian, Salt and Pepper, and Poisson noise. *Gaussian noise* can be eliminated by using a *Gaussian filter* (as described in image segmentation phase). And, for the removal of *Salt and Pepper, and Poisson noise* from CXR and CT images, *median filter* has been applied to segmented image that includes the extracted ROI (for lungs region).

$$I_f(x,y) = Med(I_s(x,y)) \quad (9)$$

where, Med is a median filter and $I_f(x,y)$ is a filtered and sharpened image[4] obtained after application of median filter to segmented image $I_s(x,y)$.

The proposed image filtration technique uses a median filter to generate a low-frequency image by replacing the pixel value with a median pixel value in an image, computed over a square area of 8×8 pixels centered at the pixel locations.

3) Proposed Image Enhancement: **Image enhancement** enhances the contrast value of grayscale and colored images, and plays a vital role in medical imaging for improvement of visual perception quality. In case of CXR and CT scan images, the strong contrast in the white area washes out the vital information saved in white pixels of an image [8]. The proposed image enhancement methodology, applies improved version of *AHE (Adaptive Histogram Equalization)*, termed as *CLAHE (Contrast Limited Adaptive Histogram Equalization)* [39] to filtered images (Eq. 10) for *intensity enhancement, improvement of local contrast and edge definitions;* and produces final enhanced images used for training the proposed hybrid DNN-ML framework.

$$I_e(x,y) = Clahe(I_f(x,y)) \quad (10)$$

where $I_e(x,y)$ is a final enhanced image and $CLAHE$ is an image enhancement technique that enhances images by evenly spreading intensity level in small regions throughout the images and setting up the maximum contrast limit [40].

The step-wise functional implementation for preprocessing of CXR and CT scan images has been presented as Algorithm 1 and the corresponding set of transformations have been demonstrated graphically as in Fig. 2(a)-2(h) and Fig. 3(a)-3(h), respectively.

Algorithm 1 The Proposed Image Preprocessing Methodology

Input: Infolder ⇐ Original Grayscale Image Repository
Output: Outfolder ⇐ Enhanced Image Repository
1: $N \Leftarrow$ Length of Infolder
2: **for** each image I in Infolder $= 1$ to N **do**
3: $IR \Leftarrow imread(infolder(I))$ ▷ imread is a function used to read images from Original Grayscale Image Repository
4: $IG \Leftarrow im2gray(IR)$ ▷ Convert retrieved image (IR) into a grayscale image (IG)
5: $BM \Leftarrow Cannyedge(IG)$ ▷ Canny edge detection and morphological operator to obtain a Binary mask (BM)
6: $maskIG \Leftarrow cast(IG, BM)$ ▷ Apply BM on original grayscale image to obtain segmented image ($maskIG$)
7: $FIG \Leftarrow medfilt2(maskIG)$ ▷ Apply median filter to segmented image to obtain filtered image FIG
8: $CIG \Leftarrow CLAHE(FIG)$ ▷ Apply CLAHE on filtered image to obtain final image CIG
9: $Outfolder \Leftarrow Save(CIG)$ ▷ Save final image in an Enhanced Image Repository
10: **end for**

[3]Noise refers to errors occurring at the time of image acquisition process and in case of CXR, imaging of ribs can be considered as noise

[4]The process of median filtering improves the sharpness of CXR and CT scan images by eliminating the impacts of noise and blurriness during image acquisition.

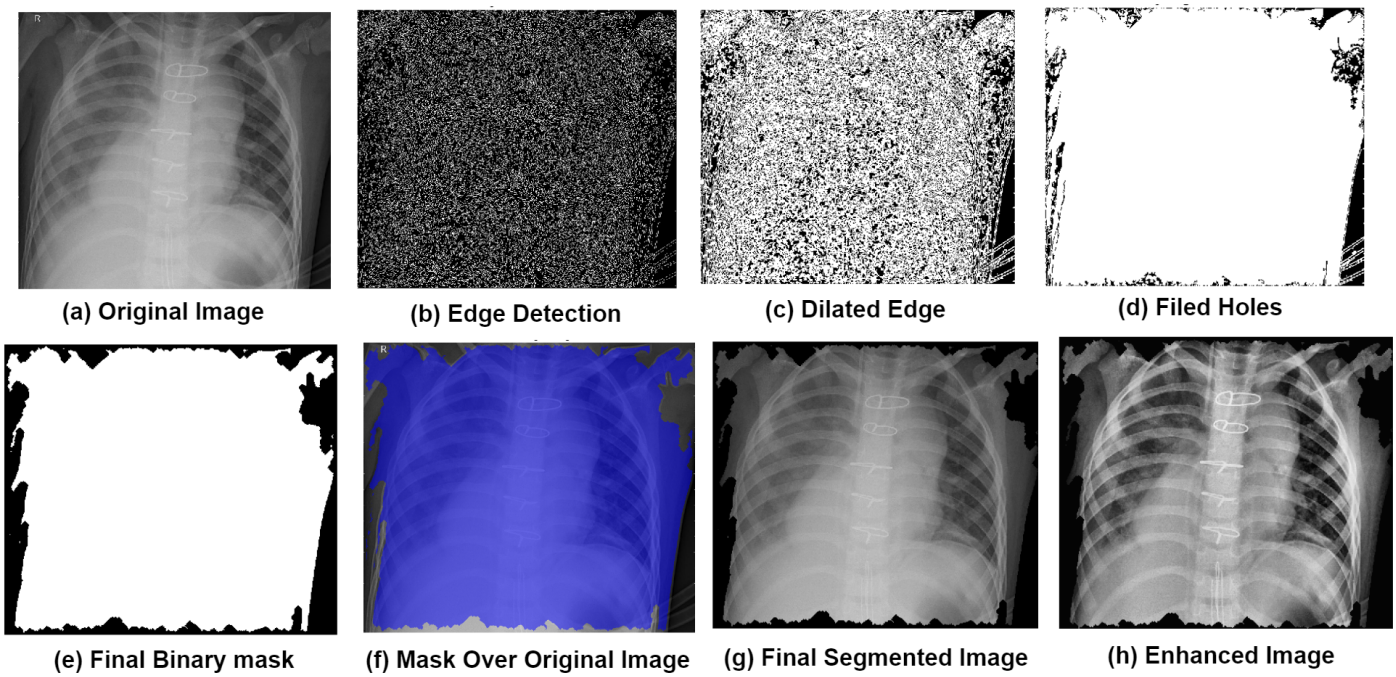


Fig. 2. Application of the proposed segmentation and image enhancement technique on Chest X-ray images.

C. Feature Extraction Process

After the feature enhancement phase (which includes image segmentation, filtration and intensification steps), **feature extraction** process is applied on the processed, enhanced and obtained segmented CXR and CT scan images, from the previous step. Feature extraction is the typical and challenging task, which involves capturing of various significant attributes of an image such as *texture, color, intensity, edge information* etc. This work uses advanced *Deep Neural Networks (DNNs)* to extract features of enhanced images.

Convolutional Neural networks (CNNs) are a sub-type of DNNs, widely used in the domain of computer vision, such as object detection and image classification etc. CNNs consist of input, convolution, max/average pooling, fully connected and classification layers. The combination of these layers with *different techniques and topologies* has resulted in various variations of CNNs such as VGG19, ResNet50, GoogleNet and many more. The layers used by the proposed model for performing the process of feature extraction can be summarized as follows. Firstly, input images are loaded into image / input layer and further fed to convolutional layers for convolution operation and for production of the final output feature map. The pooling layer also called down sampling downsizes the feature map, i.e, reduces the feature map without any signification feature loss related to input image. Lastly, a fully connected layer collects the final output comprising features extracted by CNN model and passes them to classification layer.

This work uses three pretrained DNN models namely, *GoogleNet, ResNet50 and EfficientNetB0* as a feature extractor tool to extract various significant characteristics from CXR and CT scan images. These prominent extracted features are further used by the feature classifiers in order to classify COVID-19 images. Each of these models has a different architecture and specific input/image layer size. It is important to resize the input image according to the input layer size for each model and this study uses data augmentation approach for image resizing.

1) GoogleNet: GoogleNet is a pretrained CNN consisting of 144 layers including convolution, ReLU, average and max pooling, concatenation and fully connected. Inception modules are the building blocks of GoogleNet which comprises convolution with filters in vary size (1×1, 3×3 and 5×5) and *performs convolution operation in parallel*. Each inception block contains max pooling layer which reduces feature dimensions while retaining most significant features simultaneously [41]. The fully connected layer in GoogleNet called "loss3-classifier", performs storage and retrieval of extracted features, and it extracted almost 1000 features in this work.

2) EfficientNetB0: EfficientNetB0 is based on the principle of compound scaling which balances the network length, width and resolution to increase the feature extraction efficiency. It consists of convolution layers, 2D global average pooling layer, batch normalization , fully connected layer, and sixteen depth separable *convolutional blocks (mobile inverted bottleneck convolution)* [42]. The "$dense \mid matmul$" layer has been used in this work for performing feature extraction process, and it extracted almost 1000 significant features.

3) ResNet50: ResNet50 is a variant of ResNet model also called "Residual Network" with deep 50 layers. And, the residual connection existing in this network can be termed as distinguishable feature which helps grasping residual functions and mapping the input with desired output. The two major components of ResNet50 are *convolutional and identity blocks* that further comprise several convolutional layers followed by *batch normalization and activation function (ReLU)*. These

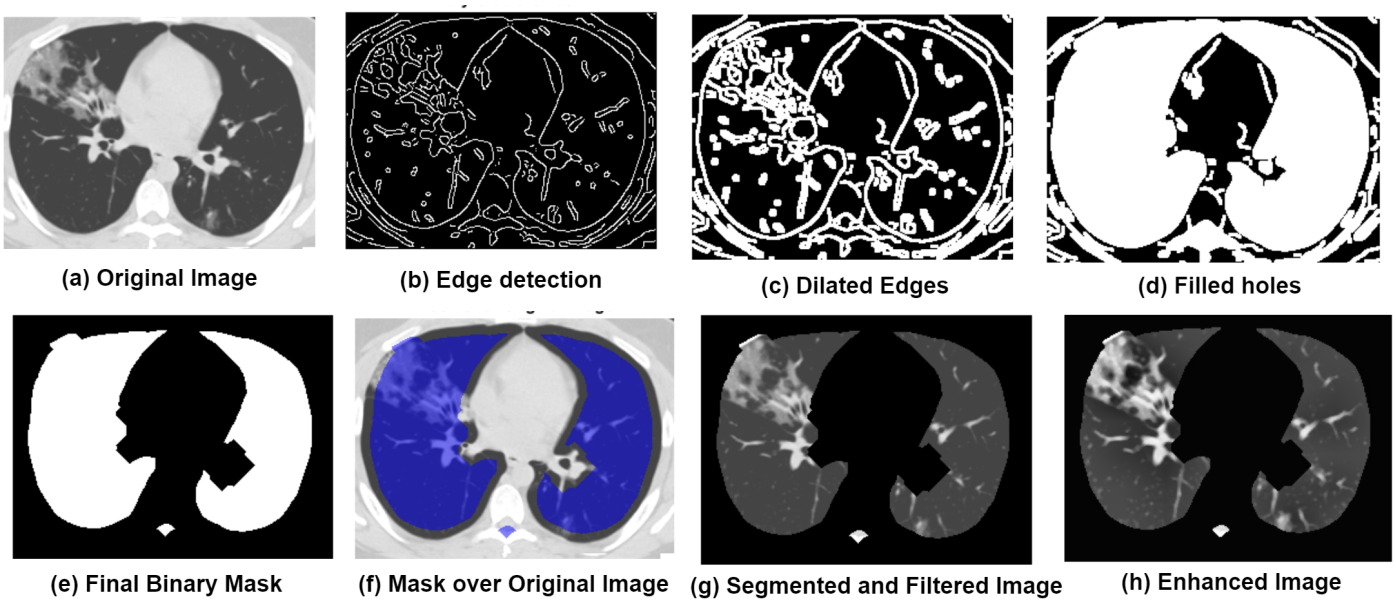


Fig. 3. Application of the proposed segmentation and image enhancement technique on CT scan images.

layers are responsible for capturing the features such as color, shape, texture and edge information [43]. This work uses ResNet50 with "fc1000" fully connected layer for extraction of 1000 features.

For the purpose of validation and benchmarking, all these three pretrained DNN architectures have been trained using enhanced CXR dataset. Each model has been initially operated for 30 epochs and each epoch further consists of 330 iterations, with batchsize of 64 and learning rate of 0.0001. For the training of aforementioned pretrained CNN models the enhanced CXR dataset has been splitted into training and testing set ratio of 70:30 respectively. It has been observed that when these three DNNs, i.e., GoogleNet, EfficientNetB0 and ResNet50 trained on enhanced CXR dataset have been compared for accuracy, they performed with 96.94%, 95.72% and 92.97% respectively. The performance metrics for all three implemented pretrained networks have been shown in Table I. *The highest accuracy has been achieved by ResNet50 (96.94%) with total training time of 514 minute*s, which outperformed the other implemented models to detect COVID-19 using enhanced CXR dataset. GoogleNet stayed close with 95.72% accuracy, but training time (612 min) taken to achieve the same has been more as compared. The training time (674 min.) taken by EfficientNetB0 and achieve accuracy (92.97%). Henceforth, for performing feature extraction in this work, ResNet50 can serve as the best choice. The sequential process of feature extraction has been mentioned in Algorithm 2.

TABLE I. RESULTS OBTAINED FOR PRETRAINED DNNS USING ENHANCED CXR DATASET

DNN model	Accuracy (%)	Recall (%)	Precision (%)	F-Score (%)	Time (min)
GoogleNet	95.72%	95.2	95.8	95.75	612
EfficientNetB0	92.97%	93.4	93.1	92.96	674
ResNet50	96.94%	96.92	96.93	96.92	514

Algorithm 2 FEATURE EXTRACTOR

Input: EDS ← Image Repository for Enhanced images
Output: f_{ext} ← Set of extracted features

1: IMD ⇐ Load(EDS) ▷ Retrieve images from the enhanced image data store
2: Net=ResNet50 or GoogleNet or EfficientNetB0 ▷ Call to pretrained CNN model and initialize as Net
3: COVNET ⇐ Train(Net, IMD) ▷ Train proposed COVNET with network training options and enhanced dataset
4: Calculate and compare the classification accuracy of pretrained models ▷ trained on CXR and CT datasets
5: Select the model with maximum accuracy, $COVNET_{max}$ ▷ as Feature Extractor (FE)
6: $f_{ext} \Leftarrow COVNET_{max}(FC_{layer})$ ▷ Extract set of features from fully connected layer FC_{layer} of extractor

D. Feature Classification Process

The features extracted by the "fc" fully connected feature layer of ResNet50 model are fed to ML classifiers for performing **feature classification**. This study utilizes Bayesian optimized ML classifiers(i.e, *Decision Tree (DT), K-Nearest Neighbor (kNN), Naive Bayes (NB), Discriminant Analysis (DA) and Support Vector Machine (SVM)*) for *feature-based image classification* from CXR and CT scan image datasets.

- *Decision Tree (DT)* classifier [44] is a tree based decision-making model for classification of image features where internal nodes and branches of a tree represent and rules, respectively. DT classifier works

on *entropy and information gain* parameters (Eq. 11).

$$entropy(D) = \sum_{i=1}^{|c|} P_r(C_i) log_2 P_r(C_i),\ where \sum_{i=1}^{|c|} P_r(C_i = 1) \tag{11}$$

On the basis of these parameters, each node of DT is further splitted until the traversal of final leaf node, which represents the final output class. The vital *hyperparameters* that can be considered for fine tuning of DT classifier are *maximum number of splits* and *maximum depth*.

- *K-Nearest Neighbor (kNN)* [45] classifies the new data point according to similarity with its neighboring points. To find similarity, distance between the new and neighboring data points is computed using various methods (i.e., Euclidean, Manhattan, Spearman, Murkowski, etc.), This study computes *euclidean distance* D (equation 12) with k-NN for image classification.

$$D = \sqrt{(x - x_i)^2 - y - y_i)^2} \tag{12}$$

k-NN classifier can be refined by optimal values of *neighbor size (k)* and *distance method* as vital hyperparameters for accurate results.

- *Naive Bayes(NB)* [46] is a simple probabilistic algorithm based on *Bayes' theorem* (Eq. 16) and it classifies images identical to the corresponding disease type (COVID-19, Viral Pneumonia, Normal images) with the largest posterior probability. The objective function $O(x)$ of NB can be defined as (Eq. 13)

$$\widehat{y} = argmax_y P(y) \pi_{i=1}^{n} P(x_i \mid y) \tag{13}$$

where, $P(x_i \mid y)$ is the posterior probability of x_i for given values of y.
In NB classifier, the smoothing hyperparameter (α), which is continuous in nature is the only one that needs to be fine-tuned for refinement of the former.

- *Discriminant Analysis (DA)* [47] is a statistical approach based on Bayes' theorem (16) and estimates the probability of new data point with respect to each class. And, the class having highest probability will be the class for a new point. The objective function $O(x)$ for DA can be defined by (Eq. 14)

$$\widehat{y} = argmax_{y=1...k} \sum_{k=1}^{k} \widehat{P}(k \mid x) C(y \mid k) \tag{14}$$

- *Support vector machine (SVM)* [48], is a prominent nonlinear classifier which divides the dataset into two parts by using a hyperplane. It can efficiently handle both linear and non-linear data. Hyperplane learning is facilitated by Kernel function f(x) (where type of kernel functions are Linear, Polynomial, Radial Basis Function(RBF), Sigmoid). The objective function $O(x)$ for SVM can be defined by Eq. 15.

$$argmin(\frac{1}{n} \sum_{i=1}^{n} max(0, 1 - y_i f(x_i) + E w^T . w)) \tag{15}$$

where, w denotes normalization vector and E represents classification error rate. The most crucial hyperparameter which can be optimized for best results in SVM model is *kernel type*.

1) Hyperparameter Tuning process for ML Classifiers: **Hyperparameter Optimization (HPO)** [49] process involves finding the most appropriate set of hyperparameter values before training phase of ML classifier that results in the best performance in a finite duration of time for a particular dataset. The major objective of HPO is to obtain optimal performance for ML models by finetuning their hyperparameters under time constraints. Since, *training time* is one of the crucial factors for HPO of ML models, therefore, for every new set of hyperparameter values, the entire model is retrained and performance of the latter is evaluated [50], [51]. It is important to determine the optimal values for the relevant hyperparameters in a ML classifier, in order to maximize the value of classification accuracy. Most commonly used hyperparameter tuning methods are *Grid Search (GS), Random Search (RS),* and *Bayesian optimization (BO)* .

Grid search is a time-consuming optimization technique, as it iterates for all the possible values of the selected hyperparameters and henceforth,is an infeasible approach. On the other hand, random search, randomly selects various possible values for the selected set of hyperparameters and computes the results, and may miss the best suitable combination for set of hyperparameter values, concluding it to be an inefficient approach. As compared to GS and RS, BO, is a probabilistic optimization technique that uses prior information (previous outcomes) of various experimented hyperparameter values to compute the next ones and avoids unnecessary iterations. This increases the computational power of BO, and it finds optimal hyperparameter values using fewer iterations as compared [52]. Therefore, this study has adopted BO as a hyperparameter tuning method that makes it as heart and soul of the proposed hybrid DNN-ML model for chest disease diagnosis. The implementation of BO along with its various outcomes, has been discussed in the upcoming sections.

2) Bayesian Optimization: Bayesian optimization (BO) [53] has been derived from Bayes' theorem. And, it states that for a given information I, the posterior probability ($P(M|I)$) of a given model M is directly proportional to to the product of likelihood $P(I|M)$ and marginal or prior probability $P(M)$ defined in Equation 16.

$$P(M|I) = P(I|M) \times \ P(M) \tag{16}$$

Bayesian approach uses the information retrieved from data(considered as prior knowledge) along with the factors that improve existing knowledge from the derived posterior knowledge. BO based Hyperparameter optimization problem can be defined as in equation 17) and its goal is to obtain maximize or minimize the value of objective function $O(x)$.

$$x^* = argmin_{x \in X}\ O(x),\ or\ argmax_{x \in X}\ O(x) \tag{17}$$

where,

- X denotes the search space comprising x samples, expressed as x_1, x_2, x_3,.....x_n with size of search space $= n$.
- x^* represents hyperparameter configuration that generates the maximum/minimum value of $O(x)$.

The values of x_1, x_2,.....x_n can be estimated using the objective function $O(x)$. The sequential combination of these samples and their evaluations forms a set S expressed as $S = x_1, O(x_1), ...x_n, O(x_n)$. The set S is further used to define the surrogate model (SM) for generating the value of posterior probability.

This study uses Gaussian Process (GP) as a surrogate model because of its stochastic behavior and normal distribution property. GP is defined as the function of mean $\mu(x)$ and a covariance matrix $K(x, x')$ (Eq. 18).

$$O \sim GP(\mu(x), K(x, x')) \tag{18}$$

The optimal values for $O(x)$ can be determined using the posterior probability value, according to the acquisition function(α). The minimum value of Bayesian optimized objective function $O(x)$ can be derived by from set X selected on the basis of α.

$$x^+ = argmin_{x \in X}\ \alpha(x \mid X),\ or\ argmin_{x_i \in X1:t}\ O(x_i) \tag{19}$$

where, x^+ represent the position for which objective function $O(x)$ obtains maximum value after t sample points. Acquisition function uses Expected improvement (EI) method (equation 20), to evaluate the degree of improvement for the objective function $O(x)$.

$$EI(x) = max\{0, O_{t+1}(x) - O(x^+)\} \tag{20}$$

The step-wise implementation of proposed Bayesian Optimization for hyperparameter optimization of ML classifiers has been shown as Algorithm 3. The set of computed hyperparameters using BO with ML classifiers, for chest disease diagnosis from CXR and CT scan image datasets are shown in Table II.

TABLE II. THE SET OF OPTIMAL HYPERPARAMETER VALUES FOR DEPLOYED BO-ML CLASSIFIERS

ML classifier	Hyperparameters	CXR images	CT scan images
DT	Minimum leaf Size	1	10
	Maximum Splits	287	13
	Split Criteria	deviance	deviance
	TET(s)[1]	41.58	26.9735
	TOFET(s)[2]	27.523	13.3903
kNN	Neighbors size	8	9
	Distance	euclidean	spearman
	Distance weight	inverse	squaredinverse
	TET(s)	38.9375	32.7209
	TOFET(s)	19.011	17.1105
NB	Distribution Nature	Kernel	Kernel
	Kernel width	1.0826	0.322
	Kernel Type	Normal	Normal
	TET(s)	2338.4096	1150.12
	TOFET(s)	2324.1023	1136.294
DA	Delta	0.01723	0.0188
	Gamma	0.5991	0.8467
	Discriminant type	Linear	Linear
	TET(s)	59.7759	39.8835
	TOFET(s)	41.4394	22.867
SVM	Box Constraint	1.2984	0.1012
	Kernel Scale	92.227	91.017
	Kernel Function	Gaussian	linear
	Coding technique	onevsall	onevsone
	TET(s)	1317.2887	
	TOFET(s)	1285.26	

Algorithm 3 Psuedocode for applied Bayesian optimization technique

Input: X ← Hyperparameter Space of size n, O(x)← Objective function
Output: $x_{optimal}$ ← optimal hyperparameter configuration, $y_{optimal}$ ← optimal objective function value
1: Assign initial value x_0 and calculate objective function value $y_0 = O(x_0)$ ▷ initial hyperparameter configuration
2: Set $x_{optimal} = x_0$ and $y_{optimal} = y_0$ with Training set $T_0 = x_0, y_0$
3: **for** i in range n **do**
4: Obtain new values for hyperparameter configuration by optimizing acquisition function α_i
5: $x_i = argmin_{x \in X}\ \alpha(x \mid T_{i-1})$
6: Calculate objective value $y_i = O(x_i)$
7: Update the training set $T_i = T_{i-1} \cup \{x_i, y_i\}$
8: Update the surrogate model
9: **if** $y_i < y_{optimal}$ **then**
10: $\{x_{optimal}, y_{optimal}\} = \{x_i, y_i\}$
11: **end if**
12: **end for**
13: return $\{x_{optimal}, y_{optimal}\}$ ⇐ optimal set of hyperparameter configuration and objective function value

E. Evaluation Metrics

The four major performance metrics, i.e., *Accuracy, Precision, Recall* and *F-Score* have been utilized in this study to assess the effectiveness of proposed optimized hybrid DNN-ML model.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \times 100\% \tag{21}$$

$$Precision = \frac{TP}{TP + FP} \times 100\% \tag{22}$$

$$Recall = \frac{TP}{TP + FN} \times 100\% \tag{23}$$

$$F - score = \frac{2TP}{2TP + FP + FN} \times 100\% \tag{24}$$

where, *TP, TN, FP, FN* are values of *True Positive, True Negative, False Positive and False Negative* scores, respectively. These values can be calculated from the resultant confusion matrix.

F. Proposed Framework

This section describes the proposed framework for classification of COVID-19 images using two different datasets that

vary according to size, mode and labels. This work proposes a *three-layered optimized hybrid DNN-ML framework*. The three different layers are, feature enhancement, feature extraction and classification layer performing three distinguished tasks. The framework of proposed approach has been presented in Fig. 4. *Phase-I* (feature enhancement layer) performs the image preprocessing task for enhancing the image quality and ROI extraction with significant features of an image. This Phase comprises the proposed image segmentation, median filter and CLAHE methods for capturing image features that denote abnormal/disease region, for passing to Phase-II (feature extraction layer). *Phase-II* has deployed ResNet50 as a feature extractor tool because of its strengths such as lesser training time and higher classification accuracy. A sum total of almost 1000 distinguishable features has been extracted by the Phase-II, further dividing them into training and testing feature set in ratio 70:30. The training features have been given to *Phase-III* (Feature classification layer) as an input for training of ML classifiers. And, the test features have been employed to validate the classification accuracy of Bayesian optimized ML classifiers for CXR and CT scan image datasets. The Algorithm with all sub-procedures for the proposed hybrid BO DNN-ML model has been presented as 4.

Algorithm 4 Sequential Implementation for the proposed hybrid BO DNN-ML model

Input: DS← Image repository
Output: Image classification and accuracy using proposed model

procedure IMAGE PREPROCESSING(DS) ▷ Apply various Image Preprocessing techniques
  N ← Number of images in each class
  for i in range N **do**
    $EDS \leftarrow$ IMAGE PREPROCESSING(DS_i) ▷ call image preprocessing method
    return EDS ▷ returns enhanced CXR dataset
  end for
end procedure
procedure FEATURE EXTRACTION AND CLASSIFICATION(EDS)
  $f_{ext} \leftarrow$ FEATURE EXTRACTOR(EDS) ▷ call feature extractor for EDS
  return f_{ext} ▷ return extracted features
  $BO_{ML} \leftarrow$ BAYESIAN OPTIMIZATION(f_{ext}) ▷ Call optimization to hyperoptimize the ML classifiers
  $f_c, Accuracy \leftarrow$ FEATURE CLASSIFIER(BO_{ML}, f_{ext})
end procedure
return Image class ← f_c with associated accuracy score ← Accuracy

IV. Experimental Results

In this study, three experiments have been conducted for the performance analysis of the proposed approach to detect COVID-19 on two different types of datasets: Chest Xray and CT scan image dateset. Firstly, the images have been segmented and enhanced with the proposed adaptive edge based segmentation algorithm. Afterwards, the proposed hybrid model described in Section 3 has been applied to enhance images for the next processes of feature extraction and classification. This section presents experimental specifications and results obtained in detail.

A. Experimental Setup

The hardware and software specifications of experimental setup include a computer system with 8 GB RAM, 7th Generation Intel Corei7 processor, Windows 10 as an operating system and MATLAB R2023b installed on it. The experiments have been carried out using two different datasets (CXR & CT scan images) and preprocessing applied to them before training of the proposed model. Various Morphological operators predefined in MATLAB [54] have been used for preparation of an enhanced dataset.

B. Experiment 1: Performance Evaluation of ResNet50 using Original CXR and CT Scan Datasets

The experiment 1 investigates the impact of image preprocessing for the improvement of *classification accuracy*. Numerous research studies have shown disease detection models with high classification accuracy but all of them used computationally intensive pretrained models to perform image preprocessing. Moreover, some research studies also utilized ANN models to perform image segmentation which introduces an additional overhead in overall computational time. This experiment illustrates the use of edge detection technique for image segmentation which improves both classification accuracy and efficiency of light weight ResNet50 model. Various performance metrics used for performance evaluation and comparison of ResNet50 with both original and enhanced datasets have been mentioned in Eq. 21, 22, 23 and 24. A summary of comparative results (Table III) show that ResNet50 achieved highest accuracy values of 96.94% and 96.67% with enhanced CXR and CT scan datasets. These results depict that no matter what type of images have been used for the training of DNN model, *feature enhancement always leads to the accuracy improvement*.

TABLE III. Performance Metrics for Comparison of Original and Enhanced Image Datasets with ResNet50

Dataset Modality	Dataset Type	Accuracy (%)	Class	Recall (%)	Precision (%)	F−Score (%)
CXR	Original	95.11%	COVID-19	95.4	100	95.04
			Normal	93.6	91.9	
			Pneumonia	96.3	93.8	
	Enhanced	96.94%	COVID-19	99.08	99.08	96.99
			Normal	99.08	93.10	
			Pneumonia	99.01	92.66	
CT	Original	95.71%	COVID-19	97.1	94.4	95.77
			Non COVID-19	94.3	97.1	
	Enhanced	96.67%	COVID-19	97.1	96.2	96.7
			Non COVID-19	96.2	97.1	

C. Experiment 2: Performance Analysis of the Proposed hybrid DNN-ML Model

Firstly, the proposed hybrid DNN-ML model has been trained and evaluated for CXR images dataset. ResNet50 has been employed as a feature extractor for the proposed model in experiment 2 and features extracted served as a input to disparate ML classifiers(viz., SVM, KNN, NB, DT, and DA). SVM classifier has been able to achieve the highest accuracy value of 97.25% with AUC value as 99.27% (Table IV). The

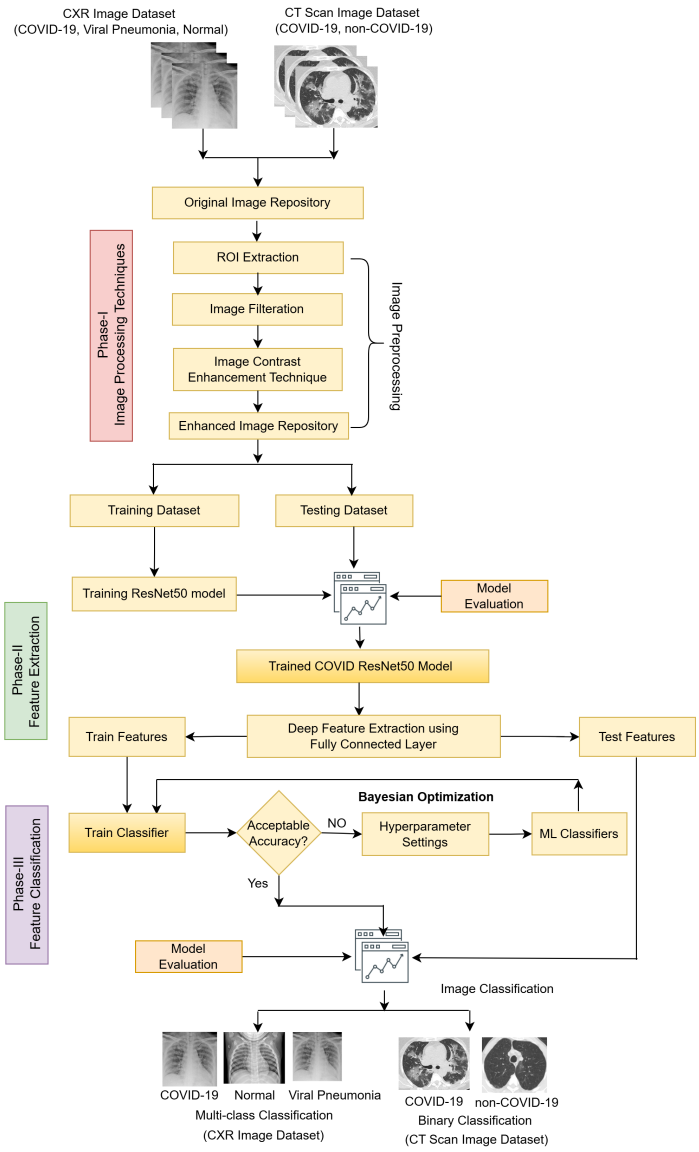


Fig. 4. Conceptual framework of the proposed optimized hybrid DNN-ML model.

confusion matrix obtained for various ML classifiers that have been trained using features extracted by ResNet50 has been shown in Fig. 5.

The same experiment is repeated CT scan dataset and as mentioned before, SVM outshined the rest of ML classifiers by achieving highest accuracy value 97.62% with AUC value as 99.84%. All the performance metrics and confusion matrix related to same have been presented as Table V and Fig. 6, respectively.

D. Experiment 3: Performance Analysis of BO based hybrid DNN-ML Model

This experiment involves training, validation and evaluation of the proposed Bayesian optimized hybrid DNN-ML model using enhanced CXR and CT scan datasets. The results obtained have been shown in Tables VI and VII. The maximum values of classification accuracy when the proposed Bayesian

TABLE IV. RESULTS OF VARIOUS PERFORMANCE METRICS FOR A COMBINATION OF RESNET50 AND ML CLASSIFIERS FOR AN ENHANCED CXR DATASET

Model	**Accuracy** (%)	**Class**	**Recall** (%)	**Precision** (%)	**F−Score** (%)	**AUC** (%)
DT	88.07%	COVID-19	82.6	92.8	88.01	92.15
		Normal	98.2	89.2		
		Viral Pneumonia	83.5	82.7		
k-NN	94.5%	COVID-19	100	97.3	94.4	99.31
		Normal	100	87.9		
		viral pneumonia	83.5	100		
NB	91.13%	COVID-19	82.6	98.9	91.16	99.28
		Normal	98.2	93.9		
		Viral Pneumonia	92.7	82.8		
DA	97.86%	COVID-19	99.1	100	97.87	99.95
		Normal	100	94.0		
		Viral Pneumonia	94.5	100		
SVM	**97.25%**	COVID-19	100	99.1	97.23	99.27
		Normal	100	93.2		
		Viral Pneumonia	91.7	100		

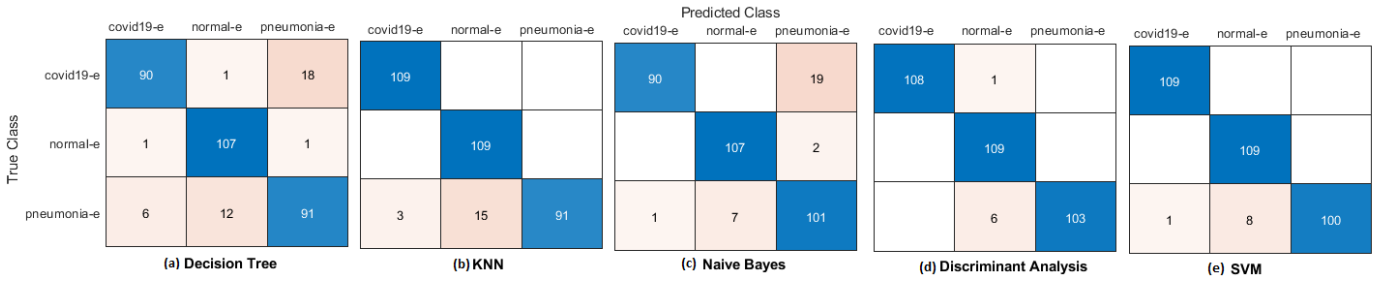


Fig. 5. Confusion matrix obtained with a combination of ResNet50 and ML classifiers for an enhanced CXR dataset.

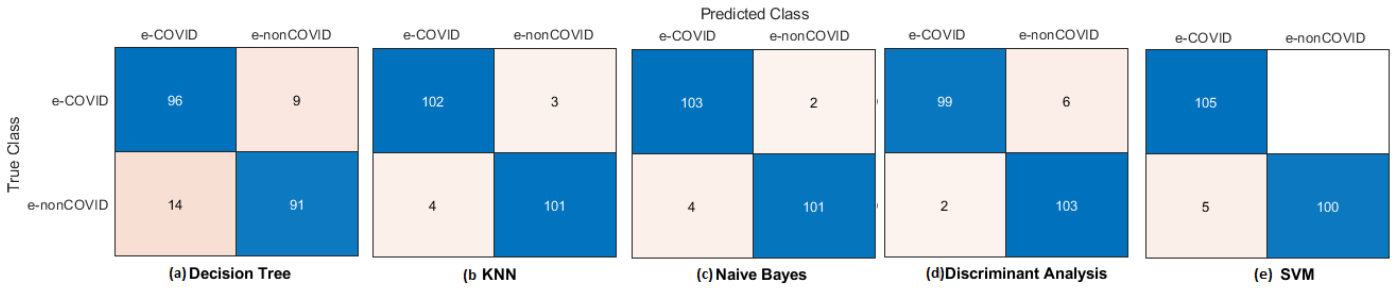


Fig. 6. Confusion matrix obtained with a combination of Resnet50 and ML classifiers for an enhanced CT scan dataset.

TABLE V. RESULTS FOR VARIOUS PERFORMANCE METRICS OBTAINED FOR RESNET50 AND ML CLASSIFIERS FOR AN ENHANCED CT SCAN DATASET

Model	Accuracy (%)	Class	Recall (%)	Precision (%)	F−Score (%)	AUC (%)
DT	89.05%	COVID-19	91.4	87.3	88.97	89.05
		Non COVID-19	86.4	91.0		
k-NN	96.67%	COVID-19	97.1	96.2	96.64	96.67
		Non COVID-19	96.2	97.1		
NB	97.14%	COVID-19	98.1	96.3	97.19	98.98
		Non COVID-19	96.2	98.1		
DA	96.19%	COVID-19	94.3	98.0	96.19	98.49
		Non COVID-19	98.1	94.5		
SVM	**97.62%**	COVID-19	100	95.5	97.62	99.84
		Non COVID-19	95.2	100		

optimized hybrid model has been trained with SVM for CXR and kNN classifier for CT scan image dataset, are 98.78% and 99.05% respectively. The confusion matrix and ROC curve after combining ResNet50 with the proposed BO-ML classifiers using CXR images have been presented in Fig. 7 and 8, respectively. Similarly, the confusion matrix and ROC curve for the proposed ResNet50 based BO-hybrid model trained on CT scan dataset have been presented in Fig. 9 and 10 respectively. The objective function presenting number of evaluations for BO-SVM validated with CXR and BO-KNN for CT scan image datasets have been presented in Fig. 8 and 10, respectively.

The performance comparison of ResNet50 model on original dataset containing raw images and enhanced dataset (using proposed image preprocessing methods) for CXR and CT scan images has been shown in Fig. 11. It can be concluded from the results, that image preprocessing techniques, i.e., primarily image segmentation and enhancement has helped in

TABLE VI. RESULTS OF VARIOUS PERFORMANCE METRICS OBTAINED USING RESNET50 FEATURES AND BO-ML CLASSIFIERS FOR AN ENHANCED CXR DATASET

Model	Accuracy (%)	Class	Recall (%)	Precision (%)	F−Score (%)	AUC (%)
DT	88.38%	COVID-19	86.2	93.1	88.3	93.59
		Normal	96.3	86.1		
		Viral Pneumonia	82.6	86.5		
k-NN	95.72%	COVID-19	96.3	98.1	95.1	99.47
		Normal	100	90.8		
		Viral Pneumonia	90.8	99.0		
NB	95.41%	COVID-19	97.2	96.4	95.37	99.73
		Normal	99.1	92.3		
		Viral Pneumonia	89.9	98.0		
DA	98.17%	COVID-19	96.3	100	98.16	99.95
		Normal	100	98.2		
		Viral Pneumonia	98.2	0.964		
SVM	**98.78%**	COVID-19	99.1	100	98.77	99.97
		Normal	100	97.3		
		Viral Pneumonia	97.2	99.1		

TABLE VII. RESULTS OF VARIOUS PERFORMANCE METRICS OBTAINED USING RESNET50 FEATURES AND BO-ML CLASSIFIERS FOR AN ENHANCED CT SCAN DATASET

Model	Accuracy (%)	Class	Recall (%)	Precision (%)	F−Score (%)	AUC (%)
DT	85.71%	COVID-19	86.7	85.0	85.72	90.80
		Non COVID-19	84.8	86.4		
k-NN	**99.05%**	COVID-19	99.0	98.9	99.01	99.92
		Non COVID-19	90	98.9	99.0	
NB	97.14%	COVID-19	98.1	96.3	98.58	97.97
		Non COVID-19	96.2	98.1		
DA	98.54%	COVID-19	99.0	98.1	98.51	99.7
		Non COVID-19	98.1	99.0		
SVM	98.57%	COVID-19	99.1	98.12	98.55	99.83
		Non COVID-19	98.12	99.1		

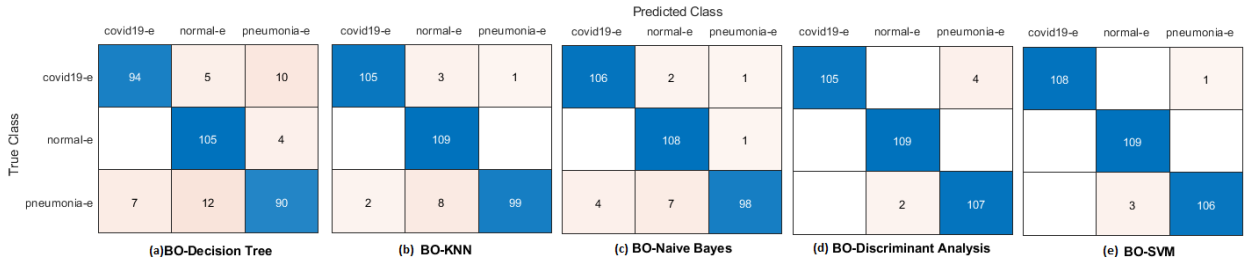


Fig. 7. Confusion matrix obtained by using combination of ResNet50 and BO-ML classifiers for an enhanced CXR dataset.

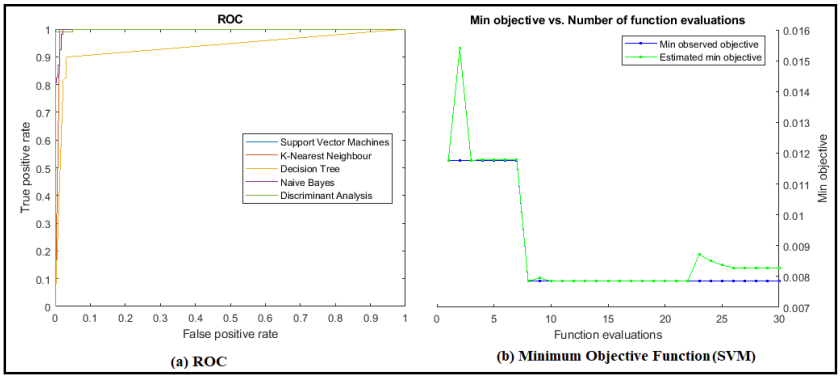


Fig. 8. ROC and Expected Improvement function(BO-SVM) for the proposed optimized hybrid DNN-ML model for CXR images.

tremendous improvement of ResNet50 model's performance for COVID-19 detection. And, it has considerable affect on classification accuracy score. The performance of Bayesian optimized ML classifiers combined with ResNet50 (DNN) is much better in terms of classification accuracy score as compared to non-optimized versions (Fig. 12). The final output in the form of tuple (image class, accuracy score) for CT scan images using the proposed hybrid BO DNN-ML model has been presented in Fig. 13.

E. Discussions

The work presented in this article has been compared with various contemporary and previous research studies for COVID -19 diagnosis using disparate medical imaging datasets. The *comparative analysis* includes parameters such as type of *dataset, classification genre, applied image preprocessing, classification techniques and performance metrics (Accuracy)* as discussed in Table VIII. The research studies that have specifically utilized CXR and CT scan image datasets for the validation of the proposed and deployed COVID-19 diagnosis framework, have only been considered in this comparative analysis. Arman et al. 2022 [13] had utilized a hybrid model that showed higher value of classification accuracy for CXR images but it used complex combination of Bayesian optimization and Xception model resulting in higher training time value. As compared to the proposed study, that used both CXR and CT scan datasets for training and testing of the optimized hybrid model, all the contemporary and previous studies have only focused on the use of single dataset, i.e., either CXR [9], [11], [14], [18], [23] or CT scan datasets [12], [24] for the training and testing of the applied ML models. Moreover, none of these have used adaptive edge based segmentation and devised image preprocessing techniques for image enhancement that contributes to a more accurate and efficient feature extraction process.

V. Conclusions and Future Work

This work has presented the utilization of image preprocessing techniques for dataset enhancement without any additional overhead in computation cost for automated chest disease diagnosis. The proposed image preprocessing technique has used HT-Canny based edge detection method with morphological operators(dilation, fill and erosion) for image segmentation and median filter with CLAHE construct for noise removal respectively. As shown in experimental results of this article, there has been significant improvement in classification accuracy because of image segmentation, filtration and enhancement of CXR(Accuracy score = 96.94%) and CT scan (Accuracy score = 96.67%) images. While comparing the performance of three DNN models, namely, ResNet50, GoogleNet and EfficientNetB0 for feature extraction using CXR dataset, ResNet50 dominated in performance as compared to the rest of models by expending least training time(514 minutes) and best accuracy score(96.94%). Furthermore, the present study combined disparate ML classifers(namely DT, kNN, NB, DA and SVM models) with ResNet50 to formulate the proposed hybrid DNN-ML model. The proposed model showed that SVM classifier outperformed the rest of ML classifiers with an accuracy score of 97.25% and 97.62% for enhanced CXR and CT scan images respectively. Whereas, when Resnet50 was combined with Bayesian optimized ML classifiers to formulate the optimized hybrid DNN-ML model, BO-SVM(Accuracy

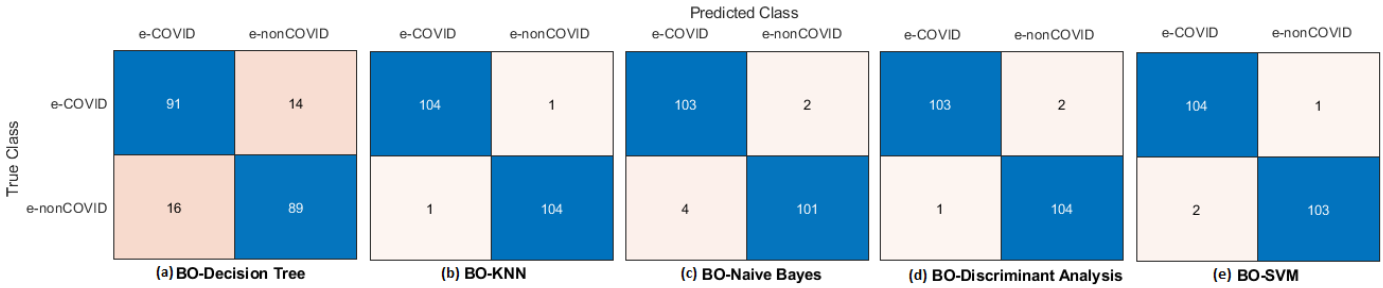


Fig. 9. Confusion matrix generated with ResNet50 and BO-ML classifiers for an enhanced CT scan dataset.

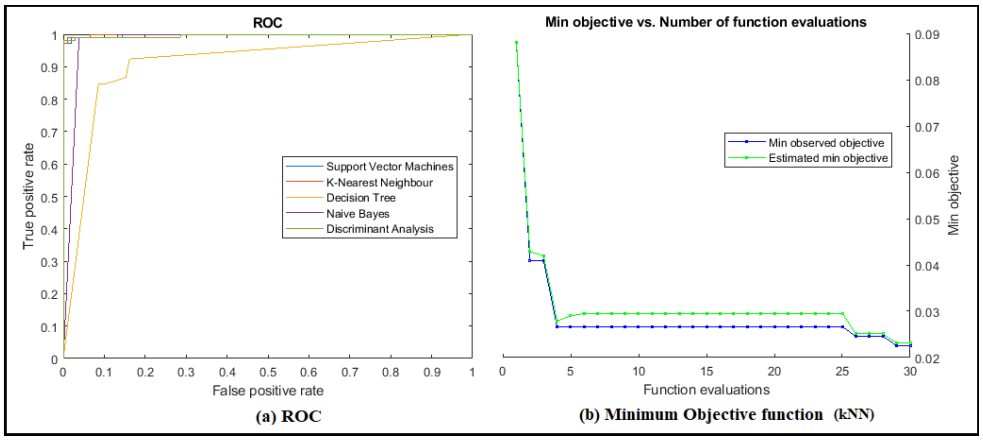


Fig. 10. ROC and Expected Improvement function (BO-kNN) for the proposed optimized hybrid DNN-ML model using CT scan image dataset.

TABLE VIII. COMPARISON OF THE PROPOSED BAYESIAN OPTMIZED HYBRID DNN-ML APPROACH WITH PREVIOUS STUDIES FOR COVID-19 DETECTION

Author (Year)	Image type	Classification Count	Image Preprocessing technique used	Implemented Approach	Accuracy
Kesav et al.,2023 [9]	CXR images	Multiclass	Image resizing	GoogleNet and Bayesian optimized SVM classifier	98.31%
Arman et al.,2022 [11]	CXR images	Multiclass	Image resizing	Bayesian optimized Xception model	99.4%
Canayaz et al.,2022 [12]	CT scan images	Binary	Image resizing	Bayesian optimized kNN with ResNet50	96.42%
Aslan et al.,2022 [14]	CXR images	Multiclass	ANN based Image segmentation	DenseNet and Bayesian optimized SVM classifier	96.29%
Ezzat et al.,2021 [18]	CXR images	Multiclass	Image Augmentation	GSA based DenseNet	98.38%
Bhattacharya et al.,2022 [23]	CXR images	Multiclass	Image segmentation using GAN	VGG-19 with BRISK	96.6%
Kaur et al.,2022 [24]	CT scan images	Binary	Image resizing	Classifier fusion with ResNet50	98.35%
The Proposed Optimized hybrid DNN-ML model	CXR images & CT scan images	Multiclass, Binary	Image segmentation, filtration and CLAHE enhancement	BO-hybrid model	98.78%(CXR), 99.05%(CT)

score = 98.78 %) and BO-kNN (Accuracy score = 99.05 %), outperformed their non-optimized versions, when trained with CXR and CT scan images respectively. It can also be noticed from the experimental results that, the proposed Bayesian optimized hybrid DNN-ML model performed much better when trained and tested on CT scan images than on the CXR dataset. As a future research direction, the use of meta-heuristic algorithms has been proposed to perform feature selection, so as size of feature dimension set can be reduced. This will help in minimizing computation time and improve accuracy score. Also, the proposed Bayesian optimized hybrid DNN-ML approach can be validated using other medical imaging datasets.

REFERENCES

[1] European Centre for Disease Prevention and Control. (2020). COVID-19 situation update worldwide. European Centre for Disease Prevention and Control.

[2] COVID, W. (19). Coronavirus Pandemic: https://www. worldometers. info/coronavirus/(2020). Accessed on December, 12, 2020.

[3] R. Singh, "Corona Virus (COVID-19) Symptoms Prevention and Treatment: A Short Review", JDDT, vol. 11, no. 2-S, pp. 118-120, Apr. 2021.

[4] A.M Al-Awadhi, K. Alsaifi, A. Al-Awadhi, & S. Alhammadi, "Death and contagious infectious diseases: Impact of the COVID-19 virus on stock market returns", Journal of behavioral and experimental finance, 27, 100326,2020.

[5] M.Y. Ng, E.Y. Lee, J. Yang, F. Yang, X. Li, Wang, & M.D Kuo, "Imaging profile of the COVID-19 infection: radiologic findings and

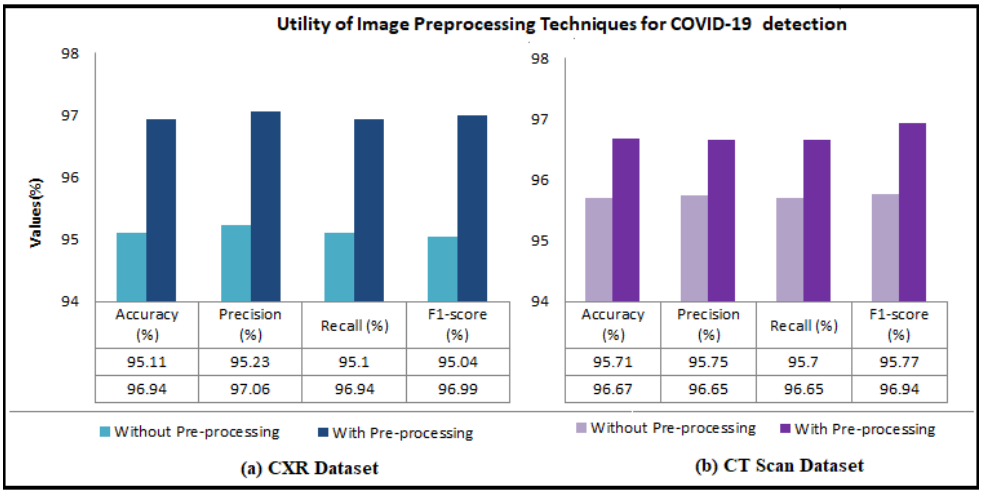


Fig. 11. Performance analysis of ResNet50 with image preprocessing techniques used in proposed approach on COVID-19 detection (CXR images).

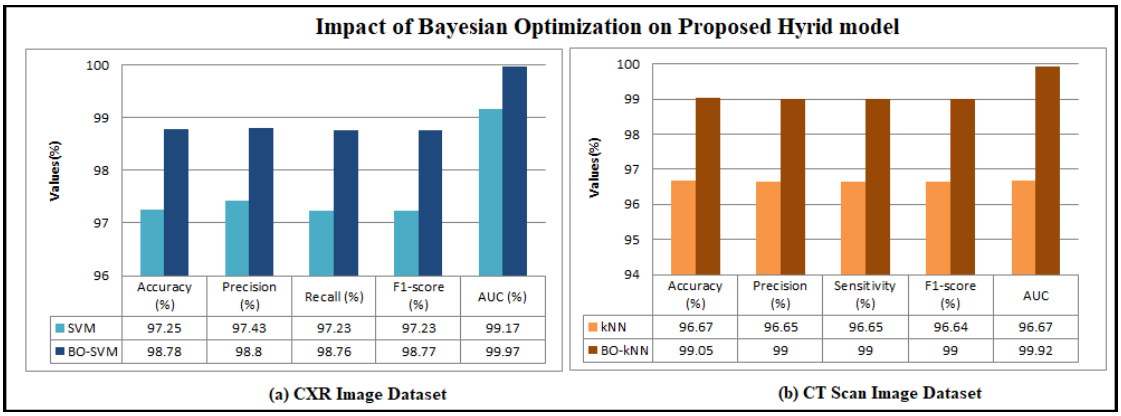


Fig. 12. Impact of bayesian optimization on SVM in the proposed approach for COVID-19 detection (CXR images).

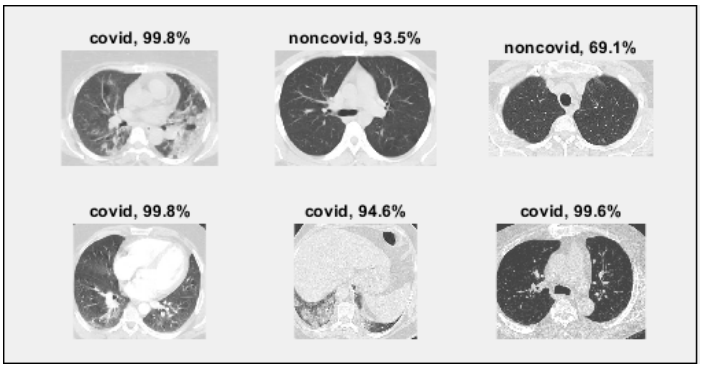


Fig. 13. Final output for classification (Image class, Accuracy) of CT scan images with the proposed hybrid DNN-ML model.

literature review", Radiology: Cardiothoracic Imaging, 2(1), e200034, 2020.

[6] H. Y. F. Wong, H. Y. S. Lam, & A. H. Fong, "Frequency and distribution of chest radiographic finding and distribution in COVID-19 positive patients", Radiology, 296, E72-E78, 2019.

[7] R. Yamashita, M. Nishio, R. K. G. Do, & K. Togashi, "Convolutional neural networks: an overview and application in radiology", Insights into imaging, vol. 9, pp. 611-629,2018.

[8] K. U. Ahamed, M. Islam, A. Uddin, A. Akhter, B.K. Paul, M. A. Yousuf & M. A. Moni, "A deep learning approach using effective preprocessing techniques to detect COVID-19 from chest CT-scan and X-ray images", Computers in biology and medicine, 139, 105014, 2021.

[9] N. Kesav & J. MG, "A deep learning approach with Bayesian optimized Kernel support vector machine for COVID-19 diagnosis", Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization, Vol.11, No. 3, pp 623-637, 2023.

[10] A. Hamza, M. Attique Khan, S. H. Wang, M. Alhaisoni, M. Alharbi, H. S. Hussein, & J. Cha, "COVID-19 classification using chest X-ray images based on fusion-assisted deep Bayesian optimization and Grad-CAM visualization", Frontiers in Public Health, 10, 1046296, 2022.

[11] S. E. Arman, S. Rahman, & S. A. Deowan, "COVIDXception-Net: A Bayesian optimization-based deep learning approach to diagnose COVID-19 from X-Ray images", SN Computer Science, Vol. 3, No. 2, 115, 2022.

[12] M. Canayaz, S. Sehribanoglu, R. Ozdag, & M. Demir, "COVID-19 diagnosis on CT images with Bayes optimization-based deep neural networks and machine learning algorithms", Neural Computing and Applications, Vol. 34, No. 7, 5349-5365, 2022.

[13] M. A. Awal, M. Masud, M. S. Hossain, A. A. M. Bulbul, S. H. Mahmud & A. K. Bairagi, "A novel bayesian optimization-based machine learning framework for COVID-19 detection from inpatient facility data" IEEE Access, 9, 10263-10281, 2021.

[14] M. F. Aslan, K. Sabanci, A. Durdu, & M. F. Unlersen "COVID-19 diagnosis using state-of-the-art CNN architecture features and Bayesian Optimization", Computers in biology and medicine, 142, 105244, 2022.

[15] M. Nour, Z. Comert, & K. Polat, "A novel medical diagnosis model for COVID-19 infection detection based on deep features and Bayesian optimization", Applied Soft Computing, 97, 106580, 2020.

[16] A. Jaiswal, N. Gianchandani, D. Singh, V. Kumar, & M. Kaur, "Classification of the COVID-19 infected patients using DenseNet201 based deep transfer learning", Journal of Biomolecular Structure and Dynamics, 39(15), 5682-5689, 2021.

[17] F. Ucar, & D. Korkmaz, "COVIDiagnosis-Net: Deep Bayes-SqueezeNet based diagnosis of the coronavirus disease 2019 (COVID-19) from X-ray images", Medical hypotheses, 140, 109761, 2020.

[18] D. Ezzat, A. E. Hassanien, & H. A. Ella, "An optimized deep learning architecture for the diagnosis of COVID-19 disease based on gravitational search optimization" Applied Soft Computing, 98, 106742, 2021.

[19] A. K. Das, S. Kalam, C. Kumar, & D. Sinha, "TLCoV-An automated COVID-19 screening model using Transfer Learning from chest X-ray images", Chaos, Solitons & Fractals, 144, 110713, 2021.

[20] M. M. A. Monshi, J. Poon, V. Chung, & F. M. Monshi, "CovidXrayNet: Optimizing data augmentation and CNN hyper-parameters for improved COVID-19 detection from CXR", Computers in biology and medicine, 133, 104375, 2021.

[21] H. Panwar, P. K. Gupta, M. K. Siddiqui, R. Morales-Menendez, & V. Singh, "Application of deep learning for fast detection of COVID-19 in X-Rays using nCOVnet", Chaos, Solitons & Fractals, 138, 109944, 2020.

[22] S. Asif, Y. Wenhui, H. Jin, Y. Tao, & S. Jinhai, "Automatic detection of COVID-19 using X-ray images with deep convolutional neural networks and machine learning", 2020

[23] A. Bhattacharyya, D. Bhaik, S. Kumar, P. Thakur, R. Sharma, & R. B. Pachori, "A deep learning based approach for automatic detection of COVID-19 cases using chest X-ray images", Biomedical Signal Processing and Control, 71, 103182, 2022.

[24] T. Kaur & T. K. Gandhi, "Classifier fusion for detection of COVID-19 from CT scans", Circuits, systems, and signal processing, 41(6), 3397-3414, 2022.

[25] N. Kumar, M. Gupta, D. Gupta, & S. Tiwari, "Novel deep transfer learning model for COVID-19 patient detection using X-ray chest images", Journal of ambient intelligence and humanized computing, 14(1), 469-478, 2023.

[26] L. Wang, Z. Q. Lin, & A. Wong, "Covid-net: A tailored deep convolutional neural network design for detection of COVID-19 cases from chest x-ray images", Scientific reports, 10(1), 19549, 2020.

[27] K. El Asnaoui, & Y. Chawki, "Using X-ray images and deep learning for automated detection of coronavirus disease", Journal of Biomolecular Structure and Dynamics, 39(10), 3615-3626, 2021.

[28] A. Shamsi, H. Asgharnezhad, S. S. Jokandan, A. Khosravi, P. M. Kebria, D. Nahavandi, & D. Srinivasan, (2021), "An uncertainty-aware transfer learning-based framework for COVID-19 diagnosis", IEEE transactions on neural networks and learning systems, 32(4), 1408-1417, 2021.

[29] M. C. Arellano & O. E. Ramos, "Deep learning model to identify COVID-19 cases from chest radiographs" in 2020 IEEE XXVII International Conference on Electronics, Electrical Engineering and Computing (INTERCON) (pp. 1-4). IEEE, 2020.

[30] J.P. Cohen, P. Morrison, L. Dao, COVID-19 image data collection, (2020) arXiv: 2003.11597.

[31] M.E.H. Chowdhury, T. Rahman, A. Khandakar, R. Mazhar, M.A. Kadir, Z. B. Mahbub, K.R. Islam, M.S. Khan, A. Iqbal, N. Al-Emadi, M.B.I. Reaz, Can AI help in screening viral and COVID-19 pneumonia? IEEE Access 8, 2020.

[32] D. Kermany, K. Zhang, M. Goldbaum, Labeled Optical Coherence Tomography (OCT) and Chest X-Ray Images for Classification, Mendeley Data, 2018, p. v2,

[33] Zhao J, Zhang Y, He X, Xie P (2020) COVID-CT-Dataset: a CT scan dataset about COVID-19.

[34] T. Rahman, A. Khandakar, Y. Qiblawey, A. Tahir, S. Kiranyaz, S. B. A. Kashem, & M. E. Chowdhury, "Exploring the effect of image enhancement techniques on COVID-19 detection using chest X-ray images", Computers in biology and medicine, 132, 104319, 2021.

[35] S. M. Islam & H. S. Mondal, "Image enhancement based medical image analysis" in 2019 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT) (pp. 1-5). IEEE,2019.

[36] L. O. Teixeira, R. M. Pereira, D. Bertolini, L. S. Oliveira, L. Nanni, G. D. Cavalcanti, & Y. M. Costa, "Impact of lung segmentation on the diagnosis and explanation of COVID-19 in chest X-ray images", Sensors, 21(21), 7116, 2021.

[37] E. A. Murillo-Bracamontes, M. E. Martinez-Rosas, M. M. Miranda-Velasco, H. L. Martinez-Reyes, J. R. Martinez-Sandoval & H. Cervantes-de-Avila, "Implementation of Hough transform for fruit image segmentation. Procedia Engineering", 35, 230-239, 2012.

[38] M. Huang, W. Yu, & D. Zhu, "An improved image segmentation algorithm based on the Otsu method", in 2012 13th ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing pp. 135-139, IEEE, 2012.

[39] R. A. Manju, , G. Koshy, , & P. Simon, "Improved method for enhancing dark images based on CLAHE and morphological reconstruction", Procedia Computer Science, 165, 391-398, 2019.

[40] Laksmi, T. V., Madhu, T., Kavya, K., & Basha, S. E. Novel image enhancement technique using CLAHE and wavelet transforms. International Journal of Scientific Engineering and Technology, 5(11), 507-511, 2016.

[41] R. Anand, T. Shanthi, M.S. Nithish, & S. Lakshman, "Face recognition and classification using GoogleNET architecture" in Soft Computing for Problem Solving: SocProS 2018, Volume 1 (pp. 261-269), Springer Singapore, 2020.

[42] X. Chen, X. Pu, Z. Chen, L. Li, K. N. Zhao, H. Liu, & H. Zhu, "Application of EfficientNetB0 and GRU based deep learning on classifying the colposcopy diagnosis of precancerous cervical lesions", Cancer Medicine, 12(7), 8690-8699, 2023.

[43] K. He, X. Zhang, S. Ren & J. Sun, "Deep residual learning for image recognition" In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 770-778), 2016.

[44] De Ville, B. Decision trees. Wiley Interdisciplinary Reviews: Computational Statistics, 5(6), 448-455, 2013.

[45] W. Zuo, D. Zhang, & K. Wang, "On kernel difference-weighted k-nearest neighbor classification. Pattern Analysis and Applications", 11, 247-257, 2008.

[46] I. Rish, I. "An empirical study of the naive Bayes classifier", In IJCAI 2001 workshop on empirical methods in artificial intelligence, Vol. 3, No. 22, pp. 41-46, 2001.

[47] A. J. Izenman, "Linear discriminant analysis. In Modern multivariate statistical techniques: regression, classification, and manifold learning" (pp. 237-280). New York, NY: Springer New York, 2013.

[48] Suthaharan, S., & Suthaharan, S. Support vector machine. Machine learning models and algorithms for big data classification: thinking with examples for effective learning, 207-235, 2016.

[49] J. Bergstra, R. Bardenet, Y. Bengio, & B. Kegl, "Algorithms for hyper-parameter optimization"' Advances in neural information processing systems, 24, 2011.

[50] J. Wu, X. Y. Chen, H. Zhang, L. D. Xiong, H. Lei, & S. H. Deng, "Hyperparameter optimization for machine learning models based on Bayesian optimization", Journal of Electronic Science and Technology, 17(1), 26-40, 2019.

[51] J. Snoek, H. Larochelle & R. P. Adams, "Practical bayesian optimization of machine learning algorithms", Advances in neural information processing systems, 25, 2012.

[52] L. Zahedi, F. G. Mohammadi, S. Rezapour, M. W. Ohland & M. H. Amini, "Search algorithms for automated hyper-parameter tuning"' arXiv preprint arXiv:2104.14677, 2021.

[53] P. I. Frazier, "Bayesian optimization"' in Recent advances in optimization and modeling of contemporary problems (pp. 255-278). Informs, 2018.

[54] A. McAndrew, "An introduction to digital image processing with MATLAB", Course Technology Press, 2004.

Preprocessing Techniques for Clustering Arabic Text: Challenges and Future Directions

Tahani Almutairi, Shireen Saifuddin, Reem Alotaibi, Shahendah Sarhan, Sarah Nassif
Department of Information Technology, King Abdulaziz University, Jeddah, Saudi Arabia

***Abstract*—Arabic is a complex language for text analysis because of its orthographic features, rich synonyms, and semantic style. Thus, Arabic text must be prepared more carefully in the preprocessing stage for the analyzer to improve the quality of the results. Moreover, many preprocessing steps have been proposed to improve the text analyzer quality by reducing high dimensionality, selecting the proper features to describe the text, and enhancing the process speed. This paper deeply investigates and summarizes the use of Arabic preprocessing techniques in Arabic text in general and focuses in-depth on clustering. Moreover, it focuses on seven preprocesses that are now used to prepare Arabic and provides the available tools for each of them; the seven preprocess are tokenization, normalization, stopword removal, stemming, vectorization, lemmatization, and feature selection. In addition, this paper investigates any work that uses synonyms and semantic techniques for preprocessing to prepare the text or reduce the dimensionality of the clustering algorithm. Therefore, this survey investigated nine techniques for Arabic text preprocessing to identify the challenges in this area. Finally, this study aims to serve as a reference for researchers interested in this area, and ends with potential future research directions.**

***Keywords*—Arabic preprocessing; Arabic language; survey; clustering; Arabic analysis**

I. Introduction

Arabic is the fourth most spoken language by native speakers, and the fifth most widely used language [1], [2], [3]. Arabic differs from other languages in that it is written and read from right to left with specific grammar, spelling, vocabulary, punctuation marks, and no case-sensitive letters. Moreover, sentences consist of verbs, subjects, and objects [4]. Arabic spoken by Arabs is known as classical Arabic (CA). Over time, CA progressed and developed into modern standard Arabic (MSA). Both versions have the same morphology and syntax but disagree grammatically and stylistically. Arabic dialects developed over time, adding to the diversity of Arabic. Dialects are informal versions of the language spoken by friends and family. Moreover, the dialects differ among Arabic countries [4], [3].

Arabic has recently attracted the attention of researchers, using various technologies and techniques. Over the past decade, Arabic and its dialects have gained attention in natural language processing (NLP) research [5], [6] , which has helped the preprocessing stage in dealing with text in different applications that benefit diverse areas, especially text analysis, such as text classification, clustering, and sentiment analysis [7]. Text clustering is an unsupervised learning technique that discovers and groups text or documents into clusters. Document clustering groups similar documents without prior knowledge of their labels. Thus, every cluster formed contains similar texts or documents. Clustering techniques identify and group texts or documents into clusters. Clustering is significant in many applications such as document retrieval, summarization, recommendation, marketing, customer analysis, document clustering, output detection, agriculture, pharmacy, and image processing [1], [8].

Many studies have used text preprocessing techniques to handle the high dimensionality of data before clustering to obtain good text clustering results [9]. Text preprocessing techniques are typically used to dramatically decrease the document size, facilitate feature selection, and enhance processing speeds. Arabic text pre-processing does not involve straightforward steps. Arabic text requires a morphological analysis that adds more challenges but is required for many reasons. The first is common orthographic variations, in which words are derived from a trilateral or quadrilateral origin system. The original system often obscures Arabic terms. Another reason for this is that Arabic synonyms within a language are extensive [1], [10], [6]. Moreover, preprocessing algorithms and techniques are limited and require further research [11].

Several Arabic preprocessing techniques, similar to other languages, have been proposed. However, some are specific and related to Arabic NLP techniques, such as tokenization, stemming, normalization, stopword removal, and lemmatization. Most Arabic preprocessing surveys focus on the classification, categorization, or specific types of Arabic fields. Only one study [12] focused on clustering to demonstrate the significant impact of term turning with stemming preprocess when vectorizing text based on TF-IDF. The authors examined and compared different text-preprocessing techniques for clustering Arabic documents. It investigated the effectiveness of techniques such as term pruning, term weighting using TF-IDF, and morphological analysis methods, including root-based stemming, light stemming, and raw text normalization. Furthermore, this study evaluates the impact of clustering algorithms, including the widely used partitional algorithm.

In contrast, [5] comprehensively classified works on three Arabic varieties: MSA, CA, and Dialect, focusing on Arabic and Arabizi types. The authors endeavored to associate each work with publicly available resources, wherever possible, to facilitate further research and development in this field. In another study [13], the authors discussed the challenges posed by Arabic in the field of NLP and provided a brief history of Arabic NLP. The tools and resources available for Arabic NLP can be broadly classified into enabling technologies that are not user-facing, and advanced user-targeting applications.

In addition, [10] proposed the extraction of information from Arabic social media text by addressing data collection, cleaning, enrichment, and availability challenges. The

objective is to enhance information quality and reliability, contributing to a more effective and accurate analysis of Arabic social media content. Finally, according to [14], preprocessing is required to improve the accuracy and efficiency of Arabic information. Therefore, the author analyzed the existing techniques and identified the limitations and challenges of both types. They emphasized the importance of stemming and the need for further research to improve Arabic information retrieval.

This work explores and summarizes the application of Arabic preprocessing methods for clustering. Only one survey paper in the literature examines preprocessing methods for clustering Arabic text; other survey papers address different tasks or focus on one or two steps of the preprocessing process. This paper focus on tokenization, Arabic normalization, stop word removal, stemming, and lemmatization. It also discusses feature selection and vectorization as preparation steps for dimensionality reduction. In addition, it investigates Arabic synonyms and semantic solutions to determine whether any researcher used them as a preprocessing step to reduce the dimensionality of the clustered data.

In addition, this survey can also serve as a resource for scholars who are generally interested in Arabic preprocessing in general by answering the following questions: which preprocessing steps are used for clustering in particular, and to the Arabic language in general? Which tools are available for each preprocessing step's algorithms and techniques? Furthermore, can preprocessing be utilized to minimize dimensions and does it address the richness of synonyms in Arabic? And last, is one of the main features of Arabic that preprocessing addresses semantics?

To the best of our knowledge, this survey is one of the few that thoroughly examines clustering-focused Arabic preprocessing methods. The several preprocessing phases are covered in this paper, which also explores the field by offering resources for each Arabic language preprocessing step. While it primarily focuses on text clustering, it will also produce scientific information about Arabic preprocessing that can serve as a foundation for other domains.

The remainder of this paper is organized as follows. Section II presents the methodology used for this survey. Section III presents the Arabic preprocessing techniques. Recent studies on clustering preprocessing techniques are discussed in Section IV. Finally, the discussion and conclusions are presented in Sections V and VI, respectively.

II. Methodology

This survey was based on the PRISMA framework, a standard research strategy that searches online databases. PRISMA framework is used to systematically and carefully select high-quality and dependable references. The framework is shown in Fig. 1 and consists of four phases: identification of the research and search process, paper selection process, quality assessment, and extraction and synthesis phases.

In the first phase, the research process focuses on identifying and searching for high-quality resources that may be beneficial. This survey conducted an efficient search of libraries and journal databases, including IEEE Xplore, Springer, and ScienceDirect, to find high-quality resources. Books, survey papers, and articles that contained important information for the survey were found via Google Scholar. Also, this survey looked up pertinent articles and references using a variety of keywords. Following their collection, the documents were sorted according to a set of criteria in order to identify the most pertinent and appropriate ones.

The following criteria were used to choose and assess the research publications in the paper selection process and quality assessment phases. Initially, between 2017 and 2024, possibly pertinent papers on Arabic preprocessing had to be presented at prestigious conferences or in scholarly journals. Additionally, the study needed to be relevant to the scope of this survey's Arabic preprocessing processes, with an emphasis on or discussion of the preprocessing tool that yields excellent results for Arabic language preparation. The titles and abstracts of the gathered papers were checked to make sure they fit the survey's scope. Lastly, the remaining studies were carefully reviewed. Only the most appropriate and high-quality papers valuable to this survey were used as references in the extraction and synthesis phase, resulting in 60 research papers.

III. Arabic Preprocessing Techniques

This paper outlines the main Arabic preprocessing techniques shown in Fig. 2. The order of steps may vary depending on the model and research goals. The following subsection discusses these techniques and suggests helpful resources for each step.

A. Tokenization

Tokenization is an essential preprocessing step for any language because it is the first step that splits the long text into a small part called a "token," making it easy to use in text analysis algorithms [15], [16]. These words were then separated into numerous delimiters, including white spaces, tabs, and punctuation marks [15]. Although there are multiple tokenization algorithms, there is only one possible way to split a dataset into words. Moreover, a sub-word tokenizer allows splitting a dataset into units called "sub-words" [16], [12]. Therefore, depending on the depth of linguistic analysis, different Arabic tokenizer levels can be developed.

Because tokenization is essential, many research publications have discussed and used this technique. Similar to the authors in [15], the three tokenizers included stochastic, disjoint-letter, and morphological. The authors also discussed three other tokenizers: characters, words, and sentences. After evaluating these six tokenizers through applied unsupervised and supervised approaches, the authors provided the benefits and drawbacks of each Arabic tokenizer technique based on studying the effects of various tokenizers in Arabic for diverse Arabic classification techniques. Moreover, they created a tokenizer framework as an open-source library [1].

In addition, [10], [17] studied many Arabic preprocessing techniques on social media, including tokenization. The authors of [10] provided a combined solution for Arabic social media text preprocessing challenges at different stages: data collection, cleaning, enrichment, and availability. The proposed

[1] https://github.com/ARBML/tkseem

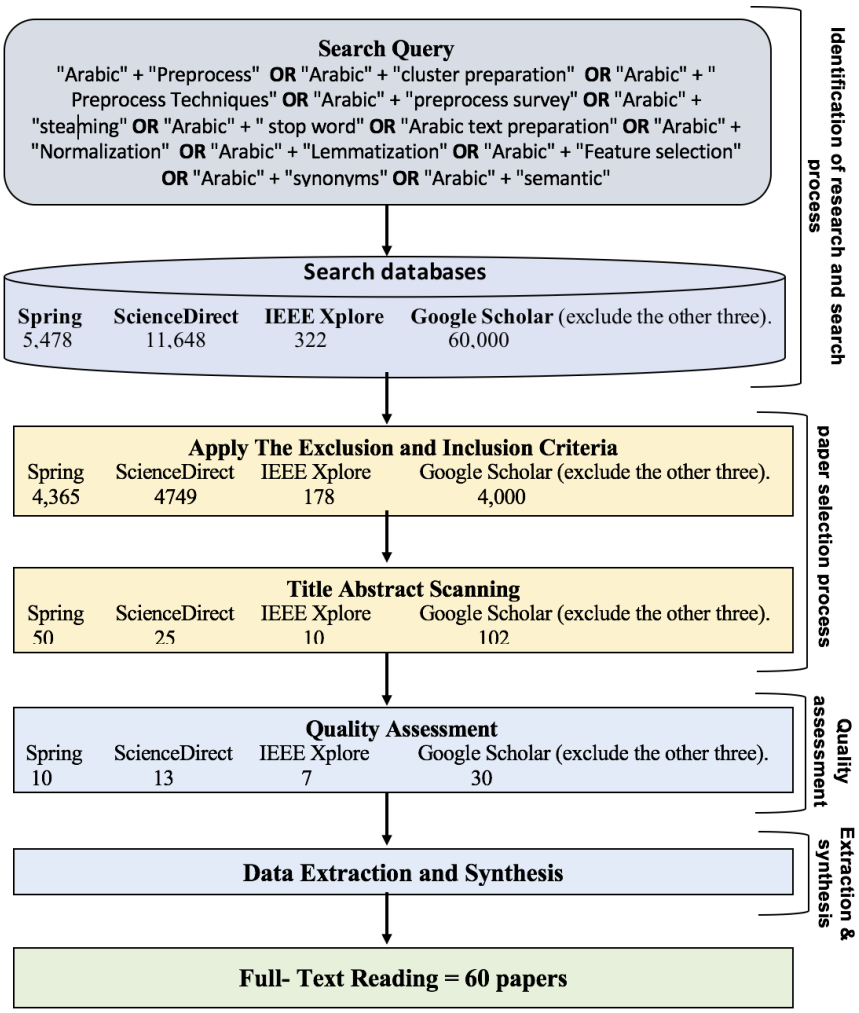


Fig. 1. The PRISMA flow diagram.

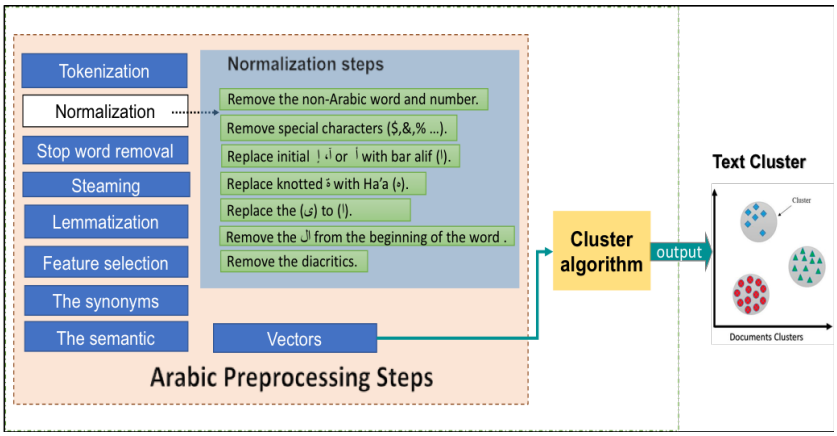


Fig. 2. Main Arabic preprocessing techniques discussed in this survey for clustering.

approach applies several NLP tools, including tokenization, stemming, and morphological analyses. The authors in [17] also investigated the impact of the 26 preprocesses applied to Arabic tweets by training a classifier to identify health-related tweets, studying the tokenization that divides the text into specific units with other preprocessing steps, where the text is divided into units in the tokenization process, and typically, those units are words.

Moreover, [12] and [18] studied Arabic document preprocessing. The authors of [12] examined and compared text preprocessing techniques in Arabic document clustering to study their effectiveness. However, they first prepared Arabic document text by tokenizing it based on words and removing stop words. In addition, [18]investigated the impact of stemming methods, an essential preprocessing step in classification. The main preprocessing steps used by the authors of [18] for Arabic include tokenization, normalization, stopword removal, feature extraction, and stemming. White space was used to tokenize the Arabic dataset.

Finally, tokenization is an essential Arabic preprocessing step because it splits long texts into small parts that are easy to use with any text analysis algorithm. However, tokenization in Arabic is not a straightforward step, according to [10]. This is complicated because of the structural sophistication of the Arabic language. Nevertheless, many tools are available for solving tokenization problems in Arabic, such as NLTK (Python) and PyArabic (Python). Finally, all reviewed resources tokenized the Arabic dataset-based words.

B. Stopword Removal

Stopword removal is essential for preprocessing all languages because it removes unimportant words and reduces the dimensionality of the texts used in the text analyzer. Stopwords include prepositions, pronouns, and articles that appear frequently and do not help distinguish documents [19]. Thus, the stopword benefits only from a syntactic procedure and does not distinguish the subject matter [20]. Furthermore, it is essential to filter out noise from vital text that is extraneous and irrelevant to the tasks [21]. Therefore, stop words are typically released with the support of a predefined list that includes common stop words.

Moreover, the most general strategy for stopword removal is based on frequency calculation and the removal of the most frequent words in all documents, which is known as dataset dependency [11], [20]. Another method to remove stop words is to use artificial patterns and perform entropy calculations. However, some general stop word lists can be used and are available online in many languages, including Arabic [20].

Stopword removal is applied to all languages as an essential step towards reducing the number of unimportant words. Many publications have discussed and provided solutions to this step in Arabic. For instance, [22] built a list containing 11,403 Arabic stop words for stop word removal. In contrast, the authors in [23] created an Arabic stopword list available online, consisting of three lists: a general list (1,377 stopwords), a corpus list that manually checked for the most repeated words appearing over 25,000 times (235 stopwords), and a combined list from the previous two. However, the authors of [18] removed stopwords based on the prepared list, and used the same stopword list as [23].

In contrast, the authors of [21] proposed a solution to the Arabic stopword construction problem by providing a comprehensive list called the Arabic stopword list (ASL) with a stopword analyzer. The analyzer connected the ASL list with machine learning to discover the most probable stop words, divided into three classes: native particles, special nouns, and special verbs for 3,931 stop words. The authors then derived a complex stop word list containing 67,153 words by adding all possible clitics to the simple stop word list. Finally, the authors tested the proposed list against the available list through a quantitative evaluation, revealing that the ASL exceeded other lists in terms of coverage. Compared with the proposed solution, the main stopword lists are Khoja [24], El-Khair [23], and the stopword project [2] and the Ranks NL[3]. Table I summarizes these stopword lists.

TABLE I. SUMMARY OF AVAILABLE STOPWORD LISTS

Stopword list name	Description	Limitation
Khoja	It contains 168 stopwords and was developed using statistical and rule-based techniques.	The list contains not all criticized forms of every stopword and corpus dependent.
El-Khair	It created three stopword lists: The first is a general stopwords list based on the Arabic language syntactic classes, consisting of 1,377 stopwords. The second is Corpus-based: manual checking of the words that appear more than 25,000, verifying this condition, and containing 235 words. Finally, the third list consists of 1,529 stopwords by combining the general list and the corpus-based lists	The list was not comprehensive and did not contain discretized.
Stopwords project	The Arabic list comprises 162 stopwords created under the GNU GPL v3 license.	The list corpus dependency contains some stopwords that may do not be correctly classified as stopwords, such as «مليار، قوة ، اعلنت» (billion, force, announced).
Ranks NL	The Ranks NL project was search engine optimization, but the project suggests an Arabic stopword list containing 102 stopwords.	Corpus was dependent.
ASL	The general list contained 3,931 stopwords. Then, the authors added all possible clitics to get the complex list, which contained 67,153.	

The authors of [12] used a general strategy to determine a stopword list by calculating the term frequency for all terms and then removing the most frequent words after sorting the terms. Nevertheless, the authors first prepared the Arabic text through the tokenization step and removed stop words; thus, they mainly examined and documented clustering. Similarly, the authors of [17] used a general strategy to remove stopwords and compared text preprocessing techniques in Arabic based on the frequency of words in the dataset. Consequently,

[2]http://code.google.com/p/stop-words
[3]http://www.ranks.nl/stopwords/arabic

they studied stopword removal and eliminated frequently used unwanted words.

Furthermore, the authors of [22] conducted scoping reviews over the past two decades (2000–2020) for Arabic topic identification, following the PRISMA-ScR guidelines. One stage that they focused on was the preprocessing step with feature extraction, which included stopword removal, stemming lemmatization, and feature selection. The authors provide an overview of the field to make it current. Finally, they recommended future work to enhance topic identification performance by working on preprocessing phases or implementing other algorithms.

Finally, stopword removal reduces text dimensionality because it removes words that frequently appear in the text without affecting text analysis, such as prepositions and pronouns. Moreover, stop-word elimination is recommended in most cases [22]. There are two strategies for stopword removal based on the reviewed papers: a general strategy based on word frequency or building, and using a general stopword list. Many general lists can be used and are available online, such as those published in 2006 by [25] containing 1,377 words. Moreover, a recent list was published in 2019 by [21] called the ASL list, consisting of three categories: native particles, particular nouns, and special verbs. Thus, the list contains 3,931 words.

C. Normalization

Text normalization is essential for text classification and analysis. Some normalization steps were applied to all languages, such as removing numbers and special characters. However, other normalization steps are language-dependent, because there is a specific way to normalize each language. For example, one step might include normalizing uppercase letters to lowercase letters in English, as there are no lowercase or uppercase letters in Arabic. Thus, normalization in Arabic related to normalizing letters includes normalizing different forms of alif (أ إ ا) to (ا) and removing diacritics that are not used in English [26].

Moreover, several normalization steps are typically executed to reduce the number of extracted terms. The significant normalization applied to Arabic is presented in the following steps, based on the reviewed papers (Fig. 2).

Thus, normalization differs from stopword removal. Normalization is related to formalizing the shape and form of words and other goals, whereas stopword removal focuses on removing unimportant words from the text. Because Arabic requires a unique normalizing process, many research publications have focused on normalization steps. Table II lists the most commonly used Arabic normalizations.

Normalization is a significant and essential step in most text analysis models. For example, the authors in [17] investigated the impact of 26 preprocesses applied to Arabic social media, particularly tweets, by training a classifier to identify health tweets. Fourteen of the 26 preprocesses focused on variants to normalize the Arabic letters. The authors studied a normalization step that converts a list of words into a more uniform sequence, such as removing punctuation, diacritics, repeating, and duplicating letters, including noise removal that seeks to destroy unwanted characters from the text, such as non-Arabic

TABLE II. THE MOST ARABIC NORMALIZATION THAT USED

#	Normalization step
1	Remove non-letters and special characters, such as $, &, and %
2	Remove non-Arabic letters
3	Replace initial (إ، آ، أ) with bar alif (ا)
4	Replace Ta'a marbota (ة) with Ha'a (ه)
5	Remove the (ال) from the beginning of the word
6	Replace the final (ى) with (ا)
7	Remove the diacritics.

letters and numbers. They also described Arabic normalization steps related to Arabic letters, and the researchers normalized many letters, including two letters, five letters, and six letters, such as "Hamza," "alif" type, and "ta'a marbota."

Similarly, [10] provided an integrated solution for Arabic preprocessing in social media challenges by cleaning a dataset and normalizing Arabic social media text. The cleaning step focuses on cleaning noisy text by selecting only the required text instead of many processes to remove different data noise. The algorithm transforms each letter into its standard form during cleaning. For example, "alif" " ا" has several forms, which are " آ ، إ، أ". In contrast, the normalization step changes the text into its standard form, with the algorithm changing a non-normal word into a normal word by eliminating duplicate characters and using a set of common non-normal words.

Furthermore, the authors of [18] normalized Arabic letters to help downgrade the various character shapes and produce a uniform shape representing these shapes. Therefore, the authors investigated the impact of stemming methods on feature reduction and classification accuracy.

Normalization is another essential step in preparing text for analysis. Some normalization processes are applied to all languages, whereas the other steps are language-dependent. Since Arabic has many characteristics related to letter shape, it requires normalization, including different forms of alif (أ إ ا) to (ا), Ta'a marbota (ة) to Ha'a (ه), and the final (ى) with (ا). Moreover, the normalization of Arabic involves removing the diacritics and the (ال) from the beginning of the word. Although several normalization steps are usually conducted to decrease the number of extracted terms, many available tools can be used for normalization problems, such as normalizing Arabic words using camel tools (Python) and PyArabic (Python).

D. Stemming

Stemming refers to the reduction of words to their stems or roots that are used to fit different word variants. Moreover, stemming is an essential preprocessing step for preparing text for the analyzer model, regardless of language type. There are three types of Arabic stemming, based on [27], [28], [6].

The first is root-based stemming, where the primary goal is to extract the roots of words using a stemmer. For example, in the Arabic language, several words can be reduced to one root, such as (العلماء ،العلم التعليم), which means "the education,"

"the scientists," and "the teacher," respectively, and are reduced to the root (علم), which means "science" [11], [28], [22].

Second, the stemmer eliminates additional suffixes and prefixes from words using light-stemming. Thus, it does not extract the original root; it only removes prefixes and suffixes from words. For example, the Arabic word (المعلم), which means "the teacher," is reduced to (معلم), which means "teacher." Hence, the semantics of the words are not affected by light stemming. Moreover, it normalizes words with diacritics and stretching characters (Tatweel ex: خبـــــــــــر converts to خبر) [11], [28], [22].

Finally, using a stem-based approach, the stemmer determines the stem that is part of the word to which grammatical prefixes and suffixes are added [28], [22]. The stem-based method does not attempt to determine the word root; it typically removes only clitics from a given word [28]. This technique preserves and represents the meaning of the text content well because it grammatically regroups inflected and related words [28], [22].

Stemming is a vital step in preparing text for analysis in most languages. Therefore, many studies have been conducted on Arabic stemming. For instance, a study [27] published in 2019 compared the stemmer ARLSTem with two versions (ARLSTem v1.0 and ARLSTem v1.1) to other light stemmers: the ISRI stemmer, Soori's stemmer, Assem's stemmer, and the Light10 stemmer. The authors then proposed an improved version of the ARLSTem stemmer by reordering some steps while adding others to improve performance. Next, the authors compared the improved version with the original and other stemmers, such as Light10. The results showed that the two versions of ARLSTem algorithm outperformed other algorithms based on understemming errors and overstemming indices.

Similarly, the authors in [10] provided an integrated solution for Arabic preprocessing of social media challenges by applying their new stemming algorithm to social media containing standard Arabic and dialects. The new Arabic stemmer module generates word stems, word roots, and specific word identifiers. The proposed model solved Arabic social media challenges that help understand a word's meaning by providing its root and determining whether the word is a noun, stop stopword, non-standard Arabic word, dialect, error, or non-Arabic word.

In comparison, [18] investigated the impact of stemming methods, an essential preprocessing step in the classification accuracy, and the number of features used. The primary preprocesses used were tokenization, normalization, stopword removal, feature extraction, and stemming. First, the authors used the Light10 stemmer for stemming and represented documents as vectors. They then conducted experiments to test the impact of the stemming algorithms on the K-nearest neighbor, naive Bayes, and decision tree classification algorithms. The results showed the effectiveness of the stemming algorithm in reducing half of the features used while improving accuracy.

Similarly, in [28], the authors compared Arabic topic identification to investigate different stemming algorithms. The main objective was to study the effects of stemmers: root-based, light stem, and stem-based. The authors used primary and available steaming approaches for the root-based selection of the Khoja and Tashaphyne root approaches. In contrast, light stem used Light10 and Tashaphyne light approaches, whereas for other types, they selected Farasa and Alkhalil1. The authors experimented using latent Dirichlet allocation (LDA) to identify the topics with different Arabic stemming algorithms. Hence, the authors determined which forms of words—root, stem, and light—were affected by the preprocessing step in topic identification. Indeed, Light10 identified topics better than the other stemmers.

Furthermore, [12] examined and compared text-preprocessing techniques in Arabic document clustering. The study indicated the significant impact of term pruning with stemming and term weighting as preprocessing steps in studying text preprocessing effectiveness. Notably, the authors experimented with different preprocessing techniques with term pruning combinations using the three systems. The first system combined term pruning with term weighting and light stemming (Light10). The second system combines term pruning with term weighting and normalization, whereas the last system combines term pruning with term weighting, normalization, and Khoja root-based stemming. The results were based on precision, recall, and F-measures, as the evaluation measures showed that the system using light stemming achieved better results than the others.

Similarly, [17] investigated the impact of the 26 preprocesses applied to Arabic health tweets by training four classifiers: MNB, logistic regression, linear SVC, and KNN. They discussed the steaming process and classified steaming into three types—root, light, and lemmatization—and found that stemming techniques increased the accuracy of the classifier.

Additionally, [22] investigated the effectiveness of Arabic preprocesses for text classification to demonstrate that properly selecting preprocessing techniques leads to a positive Arabic classification outcome. The preprocess techniques mentioned and discussed are stopword removal, stemming, and lemmatization. Moreover, they studied the effects of different combinations of these techniques to determine the best performance using ARLSTem v1.08. It was the best stemmer based on the surveyed papers that demonstrated positive outcomes in Arabic text classification, whereas the lemmatizer used MADAMIRA v2.1. The authors concluded that the best performance occurred when applying the three preprocess techniques were applied: stopword, stemming, and lemmatization.

Finally, the authors in [29] conducted scoping reviews over the past two decades (2000–2020) for Arabic topic identification following PRISMA-ScR guidelines, focusing on the preprocessing step. The preprocessing step and feature extraction included stopword removal, stemming lemmatization, and feature selection. Based on the authors' reviewed papers, the best algorithms were recommended for the steaming process. Common stemmers successfully used in Arabic topic identification are ARLSTem, Tashaphyne light stemmer, Farasa, Khoja stemmer, Light10, Al Khalil Morph Sys, Assem's stemmer, Soori's stemmer, and the ISRI stemmer. Finally, the authors suggest future research to enhance the topic identification performance by working on the preprocessing phases and implementing other algorithms.

Thus, stemming algorithms aim to remove suffixes, infixes, and other letters to grammatically reduce the multiforms of the same word in texts while reducing features. Moreover, stemming is an essential preprocessing step, regardless of the language type. Table III summarizes the main stemming algorithms used in Arabic.

TABLE III. SUMMARY OF AVAILABLE STEMMING TOOLS

Stemming name	Stemming type	Resource
Khoja	Root-based	[30]
ISRI	Root-based	[31]
Light10 (there is the previous version of Light 1 until 9)	Light stemming	[32]
ARLSTem	Light stemming	[33], [27]
Assem's stemmer	Light stemming	[34]
Tashaphyne	Light stemming	[35]
Farasa	Stem-based	[36]

Based on the reviewed papers, two Arabic stemming types are widely used: light stemming and root-based stemming. According to [28]. Moreover, it is more efficient than topic identification. Hence, the best stemming algorithm for clustering is light stemming, whereas the two best available stemming algorithms are ARLSTem and Light10 stemming, based on reviewed papers, such as [28], [18].

Additionally, based on reviewed papers, root-based stemming has been used by some authors. The Khoja stemmer, which removes the longest suffix and prefix, is a well-known root-based stemmer. Subsequently, the root dictionary matches the remainder with verbal and noun patterns [28]. At the same time, the Tashaphyne root stemmer is another famous root-based stemmer for Arabic. This stemmer identifies the root for removing prefixes and suffixes based on a default or two customized lists of prefixes and suffixes [28]. Moreover, the Tashaphyne stemmer identifies the light stems of words.

E. Vectors

Feature extraction (vectorization) is an essential preprocessing step, and the last step is performed before the analyzer algorithm. The feature extraction process transforms the text into vectors [26]. Moreover, the two most commonly used methods for extracting features from text are term frequency-inverse document frequency (TF-IDF) and Bag of Words (BoW) [26], [37]. These methods are among the most popular for computing term weights [22], [28]. Table IV summarizes the main differences between the extraction methods.

TABLE IV. SUMMARY OF VECTOR TECHNIQUES

	BoW	TF-IDF
Focus	It constructs a collection of vectors, including the word count of occurrences in the document.	TF-IDF includes information on the more and less important words ones.
Main information	BoW vectors are straightforward to interpret.	Usually acts better in the machine learning approach.
Drawbacks	Most avoid BoW and TF-IDF techniques when understanding the context of words most involved.	
Widely used for text analysis	Used but less than TF-IDF	Yes

The TF-IDF method comprises the term frequency (TF) and inverse document frequency (IDF). The TF is related to calculating the token frequency occurrence in a document, which propagates proportionally with the document size. Therefore, tokens occur more frequently in long documents than in shorter ones. The second part, the IDF, focuses on weighing down frequent tokens while scaling up rare tokens by calculating the importance of a word in all documents [17], [29], [37].

Second, the BoW is the most straightforward representation of text in terms of numbers. This method represents a sentence as a BoW vector; therefore, the sentence is presented as words with a number to indicate the occurrence in a sentence, such as TF. For example, when applying a BoW to documents (text), the result is a matrix based on all terms as columns and appears in each document as a row. Therefore, the sparse matrix result contains many 0s because it combines all the vectors [22], [29], [37].

The main drawback of using the BoW model is that the vocabulary size increases if new sentences contain new words. Hence, the lengths of the vectors also increase. Finally, BoW does not retain information related to the grammar of a sentence or the ordering of words in the text [22], [29].

Vectorization is vital for preparing text for text analysis in most languages. Many studies on Arabic feature extraction (vectors) related to this step have been published. For example, the authors in [12] investigated and compared text preprocessing techniques in Arabic document clustering and studied the effectiveness of the following text preprocessing techniques: term pruning, steaming, and term weighting based on TF-IDF. The authors tested the preprocessing combinations using these three systems. The first system included term pruning with term weighting (TF-IDF) and light stemming. The second combines term pruning with term weighting (TF-IDF) and normalization. The last system combines term pruning with term weighting (TF-IDF), normalization, and root-based stemming. Thus, this study demonstrates the significant impact of term pruning with light stemming and TF-IDF. Furthermore, the authors in [17] investigated the impact of 26 preprocesses applied to the Arabic healthcare tweet classifier and mentioned that the most extracted feature methods used were BoW and TF-IDF. However, they applied TF-IDF as a feature-extraction preprocess to transform the text into vectors.

The authors of [22] explored the significance of Arabic preprocessing for text classification to prove that proper selection of preprocessing techniques leads to a positive Arabic classification result. Moreover, the authors investigated the consequences of different combinations of preprocessing techniques to determine the best performance. They applied feature extraction TF-IDF and, as feature selection, used chi-square, then ran the classifiers. Additionally, the authors in [18] investigated the impact of stemming methods on classification accuracy and the number of features used. Feature extraction was one of the main preprocesses used in this study. The authors represented the documents as vectors using TF-IDF.

Finally, [29] followed the PRISMA-ScR guidelines to conduct scoping reviews over the past two decades (2000–2020) of Arabic topic identification. The authors focused on the preprocessing step, and they mentioned that the most stan-

dard techniques used for feature extraction included BoW, Bag of Concepts (BoC), count vectorizer, TF-IDF vectorizer, and chi-square test ($\tilde{\chi}^2 test$). Finally, the authors recommend future research to enhance topic identification performance by working on the preprocessing phases and implementing other algorithms.

Thus, vectorization is vital for converting textual features into numerical vectors. Vectorization is the final step in text-analysis models, and classification and clustering are essential. The TF-IDF method is the most widely used method for Arabic clustering and document analyses. Based on surveys and literature review, this method has proven to be the most effective.

F. Feature Selection

Feature selection methods have been verified as valuable for text classification and clustering. There were three feature types: irrelevant, powerfully relevant, and weakly relevant. Moreover, the clustering approach requires a suitable feature selection method that reduces the selection of irrelevant features to represent the text data [38]. The feature selection method removes irrelevant and duplicate features from datasets and retains features that include reliable and helpful information to reduce the dimensionality of the text in clustering techniques [28], [38]. Consequently, high-dimensional data affect the efficiency and performance of clustering approaches, dramatically decreasing the efficiency and increasing the execution time [38]. Feature selection algorithms can be classified into two primary categories: filtering and wrapping [9], [38].

The most well-known methods used for feature selection are chi-square and information gain (IG). The chi-square test is a filtering method for selecting features. Chi-square is a straightforward and computationally fast method. It can deal with a sizable dimensional feature and has proven its efficiency mainly when applied with the TF-IDF extracted feature technique [29], [38].

Similar to the work of [38], the researchers extracted features from the training datasets by calculating the TF-IDF score for each feature and then applied the chi-square test to select relevant features and eliminate irrelevant features. The chi-squared test was efficient. The chi-square test tests the independence between the occurrence of a specific feature and the occurrence of a specific type. The null hypothesis of the chi-square test was that no relationship existed between the two variables. Therefore, they are independent of each other.

The authors of the review paper [29] discussed the chi-square test as a feature selection method used for Arabic topic identification. The authors conducted scoping reviews for Arabic topic identification following the PRISMA-ScR guidelines over the past two decades (2000–2020), focusing on four phases of feature selection.

IG was used to discover the most relevant features in the label dataset. The IG ranks these features based on their entropy values, which reflect the impact of a unique feature on deciding the type label. The most relevant features are selected based on a predefined threshold value. As in [26], the authors first built the TF-IDF matrix by applying the feature extraction method and then used the IG method as a filtration step to select the most relevant Arabic features based on their ranks. This filtering step is essential for reducing the spatial dimensionality of the classifier by eliminating all features with IG ranks below a given threshold value. In [39], the authors used an IG feature ranking technique to remove irrelevant features. Furthermore, IG was used to reduce the size of the features and select the top-ranked features to train the classifier.

Term pruning is a feature selection that provides a helpful step based on a survey that eliminates words with counts less or greater than a typical threshold. For example, an experimental survey [18] showed that the best practical pruning factor is a minimum of three. Thus, pruning aims to remove any word that appears fewer than three times and is considered unimportant, thereby reducing its features.

The authors of [12] examined and compared text preprocessing techniques in Arabic document clustering and demonstrated the significant impact of term pruning on different preprocessing techniques. They experimented with different combinations of preprocessing techniques with term pruning using three, five, seven, and nine terms in the three systems. The first system combined term pruning with term weighting and light stemming (Light10), whereas the second system combined term pruning with term weighting and normalization. The last system combined term pruning with term weighting, normalization, and Khoja root-based stemming. The results were based on the precision, recall, and F-measure as evaluation measures. They showed that term pruning with a minimum of three combinations of term weighting-based TF-IDF and light stemming achieved better results than the others.

G. Lemmatization

Text lemmatization seeks to regroup semantically associated words. Unfortunately, lemmatization is limited as a preprocessing task in Arabic text classification, and no study has used lemmatization for clustering based on reviewed and surveyed papers because it is a complicated level of text processing. Moreover, most Arabic lemmatizers are royal and not publicly available, unlike Arabic stemmers. Nevertheless, some studies have reported lemmatization efficiency, particularly for information retrieval, text summarization systems, and text indexation [22].

Typically, document classification is straightforward when the meaning of the content is well represented. Thus, lemmatization regroups equivalent written words semantically in various syntactic structures and relates them to their ecclesiastical base representation called a “lemma” (i.e., a dictionary reference form). Thus, applying lemmatization to text classification as a preprocessing task is particularly beneficial [22].

The authors of [29] used text lemmatization plans to regroup semantically associated words written in various syntactic forms and associate them with their lemma. Unlike stemming, lemmatization is reasonably limited to Arabic preprocess because most Arabic lemmatizers are royal and not publicly available. However, Farasa[4] and MADAMIRA[5] are tools available for lemmatizers.

[4]https://alt.qcri.org/farasa/

[5]http://innovation.columbia.edu/technologies/cu14012_arabic-language-disambiguation-for-natural-language-processing-applications?license=108

Thus, lemmatization preprocessing steps may be helpful, based on survey papers, similar to [22], [17]. For instance, the authors of [22] investigated the effectiveness of Arabic preprocesses for text classification and found that appropriate preprocessing techniques lead to an optimistic Arabic classification outcome. Hence, they studied the effects of different combinations of the following techniques to determine the best performance: stopword removal, stemming, and lemmatization. The authors used MADAMIRA v2.1 as a lemmatizer because it was the best lemmatizer based on surveyed papers that demonstrated positive outcomes in Arabic text classification. The authors identified the best performance among the three preprocess techniques together.

Similarly, [17] investigated the impact of the 26 preprocesses applied to Arabic healthcare tweets by training four classifiers: MNB, logistic regression, linear SVC, and KNN. The authors discussed the effect of preprocessing on the classifier, one of which was lemmatization. They found that the lemmatization stemming type performed well for all four classifier models.

Furthermore, the authors in [29] conducted scoping reviews for the past two decades for Arabic topic identification based on the PRISMA-ScR guidelines to provide new researchers in the field and advanced practitioners with a closer look at improvements in Arabic topic identification in the last decade, while suggesting several recommendations for better Arabic topic identification. The authors focused on several stages, one of which was preprocessing. The preprocessing step and feature extraction mentioned in [29] involve stopword removal, stemming lemmatization, and feature selection. Based on the authors and reviewed papers, a few Arabic lemmatizers are available for free. The best algorithms recommended for lemmatization by the authors are Farasa and MADAMIRA. Finally, the authors suggest future research to improve topic identification performance by performing preprocessing phases or implementing other algorithms.

Finally, some preprocessing steps may be helpful based on the survey papers, but they are not essential for preparing the Arabic text. One of these steps is lemmatization. The lemmatization step aims to regroup semantically associated words written in different syntactic forms, associate them with their lemmas in the text, and reduce their features. Unlike stopword removal and stemming, the use of lemmatization for Arabic preprocessing is reasonably limited because most Arabic lemmatizers are proprietary and not publicly available [29]. However, the lemmatization process improves performance as a preprocess but is rarely used by researchers because research in this area has not been published or is freely available online. The only two tools available in Arabic are Farasa and MADAMIRA. Thus, this area requires further research and proposed solutions public to the research community.

H. Synonyms

One Arabic language characteristic is morphological richness, wherein the same verb can have thousands (literally) of different forms [40]. Table V summarizes these methods, and the following paragraphs summarize the most widely available methods and tools for finding and creating synonyms for Arabic.

TABLE V. SUMMARY OF SYNONYM TECHNIQUES

	Sources	**Techniques**	**Resource**
WordNet [41]	The AWN was created based on Princeton WordNet (PWN) strategy and contents.	Translation and manually checked.	Open
word2Vec [42]	Twitter and Wikipedia.	CBoW and Skip-Gram have different n-gram and unigram features.	Open
FastText [43]	Common Crawl and Wikipedia.	CBoW with sub-wording techniques.	Open

The first method to find and create synonyms for Arabic is Word2Vec, an efficient solution to synonym problems that leverages the context of the target words proposed by Google. There are two types of Word2Vec: skip-gram and a Continuous Bag of Words (CBoW) [40], [41], [44]. The following paragraphs briefly describe these two methods.

In skip-grams, the input is the center word (target), whereas the outputs are the words surrounding the target words. For example, in the sentence "I have a pretty cat," assuming the window size is 5, the input for the neural network would be "a" whereas the output would be "I," "have," "pretty," and "cat." The network contained one hidden layer with dimensions equal to the embedding size, which was smaller than the input/output vector dimension. The output layer is a softmax activation function so that each part of the output vector represents how a specific word will probably occur in the context. Word embedding for the center words can be obtained by extracting the hidden layers after providing a one-hot expression of that word into the network [40], [42], [45].

Furthermore, the vectors are more "significant" in describing word relationships. For example, vectors obtained by removing two related words sometimes represent meaningful concepts such as gender or verb tense. Finally, for all input and output data, most studies used exact dimensions and one-hot encoding [40], [42], [45].

While CBoW is similar to skip-gram, the inputs are the words surrounding the center words, whereas the output is the center word (target). The idea is that given a context, they like to know which word is most likely to occur. Thus, it aims to discover embeddings by indicating the center word in a context that provides other words in the context without respect to their order in the sentence [40], [42], [45].

The most significant distinction between skip grams and CBoW is the manner in which word vectors are generated. The CBoW model involves all samples with the target word in the neural network, and then takes the average of the extracted hidden layer. For example, assume only two sentences in a dataset: "He is a friendly man" and "She is a smart princess." To compute the word representation for the word "a," two examples are needed: "He is a friendly man" and "She is a smart princess." These are placed in a neural network that takes the average value of the hidden layer [40], [42], [45].

The second method for determining synonyms is FastText, an extension of Word2Vec, as suggested by Facebook in 2016. FastText breaks words into several n-grams (subwords) instead of providing one word to the neural network. Infrequent words are then adequately represented, because it is highly likely that some of their n-grams will also appear. For instance, the

tri-grams for the word fruit are "fru," "rui," and "uit." The word-embedding vector for the fruit was the sum of these n-grams. After training the neural network, the results are word embeddings for all n-grams in the training dataset [40], [46], [43].

Finally, WordNet is used for constructing lexical resources for SA. The Arabic WordNet (AWN) is based on the universally accepted Princeton WordNet (PWN) strategy and content. It enables translation into English and dozens of other languages at the lexical level. In addition, it encodes language-specific concepts and links as required or preferred. The results are called core word nets for Arabic with the essential system embedded in a solid semantic framework [41].

Moreover, the enrichment of AWN was published by [47], and the latest version was constructed by semi-automatically expanding its content, adapting and using existing approaches and resources generated for other languages to expand AWN [32]. Thus, AWN is a lexical dictionary or database utilized to discover synonyms and determine different relations among Arabic words containing several elements, including nouns, verbs, adjectives, and adverbs, which vary into sets of cognitive concepts (i.e., synsets) [48].

Because Arabic is rich in synonyms, a possible way to improve clustering performance is to reduce synonym words to a single word. Thus, one of the main goals of this survey is to determine whether anyone has used the synonym method as a preprocessing step to reduce the number of synonyms to help reduce the features. Based on the reviewed papers, this study found that none of these methods were used as preprocessing steps. Nevertheless, synonym methods and techniques have been used to build lexical or corpora [49], which are used to build medical datasets, or to build a corpus for some approaches such as sentiment analysis, similar to the authors in [17]. This idea may need more research in the future to satisfy and clarify whether synonym methods can be used to reduce the features by replacing the synonym with one word, thereby demonstrating that this technique reduces the feature's dimensionality and improves clustering quality.

I. Semantics

In this study, Arabic semantics are related to Arabic semantic ambiguity, mainly Arabic word sense disambiguation (WSD), a task that seeks to determine the meaning of a word given its context [50]. Arabic semantic obscurity depends on the meaning of a word or many words that can be misconstrued. Obscurity can come from the order of words or word types as verbs or nouns [51]. Moreover, Arabic, which uses diacritics, such as the Arabic read word (علم), can take many meanings based on how diacritic (عُلِمَ: understood), (عَلَمٌ: flag), (عَلَّمَ: teach), and (عِلْمٌ: science).

Semantics and meaning can be classified as branches of linguistics. Semantics and meaning are associated; some use semantics to serve meaning, whereas others use meaning to serve semantics. The first states that science investigates the requirements provided by linguistic symbols to carry out meaning. In contrast, the second emphasizes the importance of overt semantics in hidden meanings. However, another definition of semantics is the branch of linguistics that attempts to investigate changes in meaning via the analysis of linguistic structure phonetically, morphologically, lexically, and syntactically, while considering changes in usage over time [51].

Many studies have investigated this issue. For example, [50] published a 2019 review paper discussing Arabic word meaning disambiguation to inspire readers to solve Arabic words with morphological and semantic ambiguities. The morphological challenge of the Arabic language is still lacking since more than ten variations can be assigned to a non-vocalized Arabic word, for example, the words «شَعْر» (hair) or «شِعْر» (poetry). Moreover, the authors investigated numerous studies in this field by analyzing their assessment methods and linguistic resources (e.g., corpora and lexicons), recommending solutions to current semantic problems, and suggesting future directions to improve research in this area.

Thus, the Arabic language has a problem related to semantic obscurity because semantic ambiguity is associated with the meaning of a word, which can be misunderstood. Obscurity can be reached from the order of words or word types as verbs or nouns [42]. Therefore, one way to improve the clustering quality and performance is to use semantic solutions as a helpful method to solve ambiguous words when the word has the same form but different meanings based on sentences and semantics. Similar to [52], the CBOW model was used to capture the semantic relationship between the terms (word tokens), followed by using K-means to identify essential documents for Arabic summarization.

This is similar to the results of [53], who proposed a solution for Arabic Word Sense Induction (WSI) in NLP, mainly when applied to sentiment analysis. It presents a two-stage approach; the first stage uses a Transformer-based encoder like BERT or DistilBERT to encode the input sentence into context representations. In the second stage, k-means clustering and agglomerative hierarchical clustering (HAC) were applied to the embedded corpus obtained in the first stage. The proposed approach improves existing methods and revolutionizes WSI research.

Hence, one of the main goals of this survey was to determine whether anyone had used the semantic method as a preprocessing step to reduce word ambiguity and improve clustering quality. However, based on the reviewed papers, no study has used this feature as a preprocessing step. Moreover, semantic methods have been used in separate research fields using a domain-based approach. Semantic ambiguity in Arabic has recently become an active research topic. However, perhaps using the semantic ambiguity solution to improve Arabic clustering quality requires more research to satisfy and clarify whether a semantic method and word ambiguity list can improve clustering quality.

IV. Recent Arabic Clustering Publication

This section reviews the recent Arabic clustering research papers and articles published between 2019 and 2024. It summarizes the preprocessing steps used by the authors to prepare the Arabic text, as seen in Table VI. The following section discusses the main points extracted from Table VI below.

All reviewed papers used the stemming and converting the results into TF-IDF vectors. Most of the reviewed documents (91%), used the word tokenization step and removal of stopwords, while (66%) used Arabic normalization. All reviewed papers used TF-IDF as a feature representation for clustering Arabic documents. Only two reviewed papers used bigram to combine two terms as a collocation technique to better understand semantic and meaning problem—only one reviewed paper used term pruning to eliminate low-frequency words.

Regarding the feature selection method, only two studies used IG; however, they focused on the Arabic classification-based bio-inspired method. In addition, based on the reviewed papers, only one study used dimension-reduction methods as a preparation step for Arabic clustering. None of the reviewed documents applied semantic methods for Arabic clustering, or even importing, to overcome contextual meaning. None of the reviewed studies applied synonymous methods to Arabic clustering.

Fig. 3 summarizes the percentages of used main preprocess steps related for the Arabic normalization, stemming, and feature selection extracted from Table VI. Fig. 3(A) shows that more than half of the reviewed Arabic research papers 67% applied normalization preprocess for Arabic text before applying a text analyzer. In addition, Fig. 3(B) shows that all of the reviewed Arabic research papers applied the stemming method to prepare Arabic texts. Most stemming methods used root stemming (54%), whereas 46% used light stemming. Also, one research paper applied both types of stemming. Finally, Fig. 3(C) shows that 17% of the review papers applied feature selection methods exactly applied IG. In contrast, the remaining 83% did not apply feature selection methods to prepare the Arabic texts.

V. Discussion

This survey intensively investigated and focused on many Arabic preprocessing steps as references for researchers engaged in Arabic preprocessing for clustering. The survey was concluded by answering the questions raised at the beginning of this paper based on the results obtained. What preprocessing steps are used for the Arabic language in general and the cluster, especially? Preprocessing researchers have used tokenization, normalization, stopword removal, stemming, and vectorization. Some researchers have used feature selection to prepare text for a cluster to reduce the number of dimensions. In contrast, lemmatization, synonyms, and semantics were not used as preprocessing for clustering.

The second question, What tools are available for the algorithms and techniques for each preprocessing step? Many resources have been proposed for each preprocessing step; however, some preprocesses, such as lemmatization, require more investment and improvement. In addition, as a future recommendation, you can propose a stopword removal list for many fields because each domain has its own list.

In addition, the last two questions, Does preprocessing address the richness of synonyms in Arabic and can it be used to reduce dimensions? Does preprocessing address semantics as a primary characteristic of Arabic? This survey found that no one had used the synonym method as a preprocessing step to reduce features. In addition, no one has used the semantic method as a preprocessing step to reduce word ambiguity and improve clustering quality. Additionally, this survey suggests conducting further research to address the richness of synonyms in Arabic to reduce the dimensions by combining many synonyms with the same meaning. In addition, the researcher may deal with semantics to improve clustering quality as a preprocessing process to improve step. Therefore, future research should investigate the effectiveness of using synonymous techniques and semantic solutions to reduce the dimensions of complexity or improve the quality of cluster results.

Therefore, this study concludes that preprocessing steps are preferable because they provide benefits and improve the quality of an analysis model based on the surveyed papers. Based on the reviewed research papers, Arabic preprocessing is an active research area. Nevertheless, this field requires further research, especially clustering, to provide more solutions and publicly available tools to benefit the Arabic research community.

VI. Conclusion

Arabic is a complex language because of its unique orthography, grammar, and punctuation. A sentence contains a verb, a subject, or an object. Words are derived from trilateral or quadrilateral systems, making them challenging for algorithms and computational systems to understand. Moreover, Arabic has extensive synonyms and the meanings of Arabic words depend on the semantics of the sentences. For all these reasons, the analysis algorithms require specific Arabic preprocessing techniques.

Arabic text preprocessing methods are used in clustering to reduce the size of documents, simplify feature selection, and enhance processing speed. Hence, the primary purpose of a text preprocessing task in clustering is to address the considerable dimensionality problems. Therefore, several preprocessing techniques have been proposed.

This survey intensely investigated and focused on many Arabic preprocessing steps as a reference for researchers engaging in Arabic preprocessing for clustering. These preprocesses include tokenization, normalization, stopword removal, stemming, vectorization, lemmatization, feature selection, synonyms, and semantics. Moreover, this survey classified these preprocesses as essential or preferable steps.

Therefore, this paper concludes that preprocessing steps are preferable since they may benefit some models and improve the quality based on the surveyed papers. Additionally, this survey suggests conducting further research to deal with a richness of synonyms for Arabic to reduce the dimension and deal with semantics to improve the clustering quality as a preprocessing process. Indeed, based on the surveyed papers, no one to date has dealt with these characteristics to improve clustering quality.

The Arabic preprocessing step is an active research area based on the reviewed research papers. Nevertheless, this field needs more research, especially for clustering, to provide more solutions and publicly available tools to benefit the Arabic research community.

TABLE VI. SUMMARY OF ARABIC CLUSTERING RESEARCH PAPERS.

Ref	Year	Tokenization	Stopword removal	Normalization	Stemming	Feature selection	Synonyms	Collection: Bigram, trigram	Vectors	Final output	Addition step
[54]	2019	Yes (word)	Yes	No	ISRI (Root)	No	No	Sentence level	TF-IDF	vectors	Sentence segmentation.
[9]	2019	Yes (word)	Yes	No	Yes (Light Stemming)	No	No	No	TF-IDF	VSM	Segmentation
[39]	2019	Yes (word)	Yes	Yes	Yes (Root)	Yes (IG)	No	No	TF-IDF	Matrix vectors	
[55]	2020	Yes (word)	Yes	Yes	Yes (Root)	No	No	Yes (Bigram)	TF-IDF	VSM	
[26]	2020	Yes (word)	Yes	Yes	Yes (Root)	Yes (IG)	No	Yes (Bigram)	TF-IDF	VSM	
[56]	2020	Yes (word, phrase)	Yes	Yes	Yes (ISRI)	No	No	No	TF-IDF	Matrix vectors.	Dimensional reduction.
[57]	2020	Yes (word)	Yes	No	stemming (Light10)	No	No	No	TF-IDF	document-term matrix	
[58]	2021	No	Yes	Yes	Yes (Light)	No	No	No	TF-IDF	VSM	Eliminate insignificant words
[59]	2022	Yes (word)	Yes	No	Yes (root and light)	No	No	No	TF-IDF	VSM	
[60]	2023	Yes (word)	Yes	Yes	Yes (Light Stemming)	No	No	No	TF-IDF	vectors	Arabic-BERT model
[61]	2023	Yes(word)	NO	Yes	Yes (ISRI)	No	No	No	TF-IDF and CountVectorizer	vectors	
[62]	2024	Yes	Yes	Yes	Yes (light stemmer)	No	No	No	TF_IDF	vectors	Aravec pre-trained word embedding.

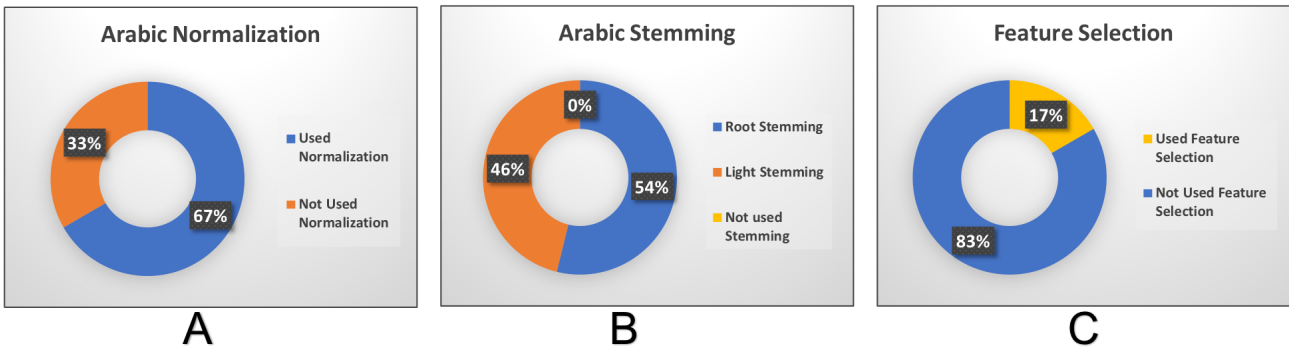


Fig. 3. (A) Percentage of reviewed Arabic papers that used the normalization step to prepare Arabic text. (B) Percentage of reviewed Arabic papers that used the stemming method with different types to prepare Arabic text. (C) Percentage of reviewed Arabic papers that used the feature selection step to prepare Arabic text.

REFERENCES

[1] W. Hadi, Q. A. Al-Radaideh, and S. Alhawari, "Integrating associative rule-based classification with Naïve Bayes for text classification," *Applied Soft Computing*, vol. 69, pp. 344–356, 2018.

[2] S. Bahassine, A. Madani, and M. Kissi, "Arabic text classification using new stemmer for feature selection and decision trees," *Journal of Engineering Science and Technology*, vol. 12, pp. 1475–1487, June 2017.

[3] S. L. Marie-Sainte, N. Alalyani, S. Alotaibi, S. Ghouzali, and I. Abunadi, "Arabic Natural Language Processing and Machine Learning-Based Systems," *IEEE Access*, vol. 7, pp. 7011–7020, 2019.

[4] S. L. Marie-Sainte and N. Alalyani, "Firefly Algorithm based Feature Selection for Arabic Text Classification," *Journal of King Saud University - Computer and Information Sciences*, vol. 32, no. 3, pp. 320–328, March 2020.

[5] I. Guellil, H. Saâdane, F. Azouaou, B. Gueni, and D. Nouvel, "Arabic natural language processing: An overview," *Journal of King Saud University - Computer and Information Sciences*, vol. 33, no. 5, pp. 497–507, June 2021.

[6] A. Alothman and A. Alsalman, "Arabic morphological analysis techniques," *International Journal of Advanced Computer Science and Applications*, vol. 11, no. 2, 2020. [Online]. Available: http://dx.doi.org/10.14569/IJACSA.2020.0110229

[7] A. Hotho, A. Nürnberger, and G. Paass, "A Brief Survey of Text Mining," *LDV Forum - GLDV Journal for Computational Linguistics and Language Technology*, vol. 20, pp. 19–62, 01 2005.

[8] M. Alhawarat and M. Hegazi, "Revisiting K-Means and Topic Modeling, a Comparison Study to Cluster Arabic Documents," *IEEE Access*, vol. 6, pp. 42 740–42 749, 2018.

[9] A. K. Sangaiah, A. E. Fakhry, M. Abdel-Basset, and I. El-henawy, "Arabic text clustering using improved clustering algorithms with dimensionality reduction," *Cluster Computing*, vol. 22, no. 2, pp. 4535–4549, March 2019.

[10] M. O. Hegazi, Y. Al-Dossari, A. Al-Yahy, A. Al-Sumari, and A. Hilal, "Preprocessing Arabic text on social media," *Heliyon*, vol. 7, no. 2, p. e06191, February 2021.

[11] A. S. Daoud, A. Sallam, and M. E. Wheed, "Improving Arabic document clustering using K-means algorithm and Particle Swarm Optimization," in *2017 {Intelligent} {Systems} {Conference} ({IntelliSys})*, September 2017, pp. 879–885.

[12] M. A. Alhanjouri, "Pre Processing Techniques for Arabic Documents Clustering," *International Journal of Engineering and Management Research (IJEMR)*, vol. Volume: 7, Number: 2, no. Volume: 7, Number: 2, 2017.

[13] K. Darwish, N. Habash, M. Abbas, H. Al-Khalifa, H. T. Al-Natsheh, H. Bouamor, K. Bouzoubaa, V. Cavalli-Sforza, S. R. El-Beltagy, W. El-Hajj, M. Jarrar, and H. Mubarak, "A panoramic survey of natural language processing in the Arab world," *Commun. ACM*, vol. 64, no. 4, pp. 72–81, mar 2021.

[14] A. A. Al-Khulaidi and S. M. Yaseen, "Comparative Analysis and Evaluation of Stemming and Preprocessing Techniques for Arabic Text," *Sana'a University Journal of Applied Sciences and Technology*, vol. 1, no. 4, 2023.

[15] Z. Alyafeai, M. S. Al-shaibani, M. Ghaleb, and I. Ahmad, "Evaluating various tokenizers for Arabic text classification," *Neural Processing Letters*, vol. 55, no. 3, pp. 2911–2933, 2023.

[16] M. A. Attia, "Arabic tokenization system," in *Proceedings of the 2007 {Workshop} on {Computational} {Approaches} to {Semitic} {Languages} {Common} {Issues} and {Resources} - {Semitic} '07*. Prague, Czech Republic: Association for Computational Linguistics, 2007, p. 65.

[17] Y. Albalawi, J. Buckley, and N. S. Nikolov, "Investigating the impact of pre-processing techniques and pre-trained word embeddings in detecting Arabic health information on social media," *Journal of Big Data*, vol. 8, no. 1, p. 95, December 2021.

[18] J. Atwan, M. Wedyan, Q. Bsoul, A. Hammadeen, and R. Alturki, "The Use of Stemming in the Arabic Text and Its Impact on the Accuracy of Classification," *Scientific Programming*, vol. 2021, pp. 1–9, November 2021.

[19] H. Ahonen, O. Heinonen, M. Klemettinen, and I. Verkamo, "Applying Data Mining Techniques in Text Analysis," Citeseer, Tech. Rep., 1997.

[20] A. Alajmi, E. Saad, and R. R. Darwish, "Toward an ARABIC Stop-Words List Generation," *International Journal of Computer Applications*, vol. 46, pp. 8–13, January 2012.

[21] D. Namly, K. Bouzoubaa, R. Tajmout, and A. Laadimi, "On Arabic Stop-Words: A Comprehensive List and a Dedicated Morphological Analyzer," in *Arabic {Language} {Processing}: {From} {Theory} to {Practice}*, K. Smaïli, Ed. Cham: Springer International Publishing, 2019, pp. 149–163.

[22] E. kah Anoual and I. Zeroual, "The effects of Pre-Processing Techniques on Arabic Text Classification," *International Journal of Advanced Trends in Computer Science and Engineering*, vol. 10, pp. 41–48, February 2021.

[23] I. A. El-Khair, "Effects of Stop Words Elimination for Arabic Information Retrieval: A Comparative Study," February 2017.

[24] S. Khoja, "APT: Arabic part-of-speech tagger," in *Proceedings of the Student Workshop at NAACL*, 2001, pp. 20–25.

[25] I. A. El-Khair, "Effects of Stop Words Elimination for Arabic Information Retrieval: A Comparative Study," *International Journal of Computing & Information Sciences*, vol. 4, pp. 119–133, January 2006.

[26] A. Alzaqebah, B. Smadi, and B. H. Hammo, "Arabic Sentiment Analysis Based on Salp Swarm Algorithm with S-shaped Transfer Functions," in *2020 11th {International} {Conference} on {Information} and {Communication} {Systems} ({ICICS})*, April 2020, pp. 179–184.

[27] K. Abainia and H. Rebbani, "Comparing the Effectiveness of the Improved ARLSTem Algorithm with Existing Arabic Light Stemmers," in *2019 {International} {Conference} on {Theoretical} and {Applicative} {Aspects} of {Computer} {Science} ({ICTAACS})*, vol. 1, December 2019, pp. 1–8.

[28] M. Naili, A. H. Chaibi, and H. H. B. Ghezala, "Comparative Study of Arabic Stemming Algorithms for Topic Identification," *Procedia Computer Science*, vol. 159, pp. 794–802, January 2019.

[29] A. E. Kah and I. Zeroual, "Arabic Topic Identification: A Decade Scoping Review," *E3S Web of Conferences*, vol. 297, p. 01058, 2021.

[30] S. Khoja, "Shereen Khoja - Research," 1999.

[31] K. Taghva, R. Elkhoury, and J. Coombs, "Arabic stemming without a root dictionary," in *International Conference on Information Technology: Coding and Computing (ITCC'05)-Volume II*, vol. 1. IEEE, 2005, pp. 152–157.

[32] M. Alshomary, "light10stemmer," 2015.

[33] K. Abainia, S. Ouamour, and H. Sayoud, "A novel robust Arabic light stemmer," *Journal of Experimental & Theoretical Artificial Intelligence*, vol. 29, no. 3, pp. 557–573, May 2017.

[34] A. Chelli, "Assem's Arabic Stemmer," 2018.

[35] T. Zerrouki, "Tashaphyne, Arabic light stemmer," 2012.

[36] A. Abdelali, K. Darwish, N. Durrani, and H. Mubarak, "Farasa: A Fast and Furious Segmenter for Arabic," in *15th Annual Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 2016, pp. 11–16.

[37] M. Masadeh, M. A, S. B, H. J, H. K, C. Chola, and A. Y. Muaad, "Investigating the impact of preprocessing techniques and representation models on arabic text classification using machine learning," *International Journal of Advanced Computer Science and Applications*, vol. 15, no. 1, 2024. [Online]. Available: http://dx.doi.org/10.14569/IJACSA.2024.01501110

[38] H. Tang, L. Zhou, X. Chengjie, and Q. Zhu, "A Method of Text Dimension Reduction Based on CHI and TF-IDF," in *Proceedings of the 4th International Conference on Mechatronics, Materials, Chemistry and Computer Engineering 2015*. Atlantis Press, 2015/12, pp. 1854–1857.

[39] M. Tubishat, M. A. M. Abushariah, N. Idris, and I. Aljarah, "Improved whale optimization algorithm for feature selection in Arabic sentiment analysis," *Applied Intelligence*, vol. 49, no. 5, pp. 1688–1707, May 2019.

[40] K.-H. H. (Steeve), "Word2Vec and FastText Word Embedding with Gensim," 2018.

[41] C. Fellbaum, M. Alkhalifa, W. Black, S. Elkateb, A. Pease, H. Rodriguez, and P. Vossen, "Introducing the arabic wordnet project," in *Proceedings of the 3rd Global Wordnet Conference, Jeju Island, Korea, South Jeju, January 22-26, 2006*, P. Sojka, K.-S. Choi, C. Fellbaum, and P. Vossen, Eds., 2006, p. 295–299, proceedings of the 3rd Global Wordnet Conference, Jeju Island, Korea, South Jeju, January 22-26, 2006; Third Global Wordnet Conference ; Conference date: 22-01-2006 Through 26-01-2006.

[42] A. B. Soliman, K. Eissa, and S. R. El-Beltagy, "AraVec: A set of Arabic Word Embedding Models for use in Arabic NLP," *Procedia Computer Science*, vol. 117, pp. 256–265, January 2017.

[43] T. Mikolov, E. Grave, P. Bojanowski, C. Puhrsch, and A. Joulin, "Advances in Pre-Training Distributed Word Representations," in *Proceedings of the International Conference on Language Resources and Evaluation (LREC 2018)*, 2018, pp. 52–55.

[44] N. Al-Twairesh, "The Evolution of Language Models Applied to Emotion Analysis of Arabic Tweets," *Information*, vol. 12, no. 2, 2021. [Online]. Available: https://www.mdpi.com/2078-2489/12/2/84

[45] M. M. Fouad, A. Mahany, N. Aljohani, R. A. Abbasi, and S.-U. Hassan, "ArWordVec: efficient word embedding models for Arabic tweets," *Soft Computing*, vol. 24, no. 11, pp. 8061–8068, June 2020.

[46] A. Joulin, E. Grave, P. Bojanowski, M. Douze, H. Jégou, and T. Mikolov, "FastText.zip: Compressing text classification models," December 2016.

[47] L. Abouenour, K. Bouzoubaa, and P. Rosso, "On the evaluation and improvement of Arabic WordNet coverage and usability," *Language Resources and Evaluation*, vol. 47, no. 3, pp. 891–917, September 2013.

[48] V. Samawi, S. A. Yousif, and Z. Sultani, "Utilizing Arabic WordNet Relations in Arabic Text Classification: New Feature Selection Methods," *IAENG International Journal of Computer Science*, vol. 46, pp. 750–761, November 2019.

[49] R. H. AlMahmoud and B. H. Hammo, "SEWAR: A corpus-based N-gram approach for extracting semantically-related words from Arabic medical corpus," *Expert Systems with Applications*, vol. 238, p. 121767, 2024.

[50] B. Elayeb, "Arabic word sense disambiguation: a review," *Artificial Intelligence Review*, vol. 52, no. 4, pp. 2475–2532, December 2019.

[51] T. I. Ababneh, S. M. Ramadan, and I. M. Abu-Shihab, "Perspectives on Arabic Semantics," *International Journal of Humanities and Social Science*, vol. 7, no. 7, p. 8, 2017.

[52] S. Abdulateef, N. A. Khan, B. Chen, and X. Shang, "Multidocument Arabic Text Summarization Based on Clustering and Word2Vec to Reduce Redundancy," *Information*, vol. 11, no. 2, 2020. [Online]. Available: https://www.mdpi.com/2078-2489/11/2/59

[53] R. Saidi and F. Jarray, "Sentence Transformers and DistilBERT for Arabic Word Sense Induction." in *ICAART (3)*, 2023, pp. 1020–1027.

[54] R. Z. Al-Abdallah and A. T. Al-Taani, "Arabic Text Summarization using Firefly Algorithm," in *2019 {Amity} {International} {Conference} on {Artificial} {Intelligence} ({AICAI})*, February 2019, pp. 61–65.

[55] R. H. AlMahmoud, B. Hammo, and H. Faris, "A modified bond energy algorithm with fuzzy merging and its application to Arabic text document clustering," *Expert Systems with Applications*, vol. 159, p. 113598, November 2020.

[56] S. Al-Saqqa and G. Al-Naymat, *Unsupervised Sentiment Analysis Approach Based on Clustering for Arabic Text*. International Business Information Management Association (IBIMA), August 2020.

[57] A. A. Mohamed, "An effective dimension reduction algorithm for clustering Arabic text," *Egyptian Informatics Journal*, vol. 21, no. 1, pp. 1–5, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1110866518301579

[58] A. R. Alharbi, M. Hijji, and A. Aljaedi, "Enhancing topic clustering for Arabic security news based on k-means and topic modelling," *IET Networks*, vol. 10, no. 6, pp. 278–294, 2021, _eprint: https://ietresearch.onlinelibrary.wiley.com/doi/pdf/10.1049/ntw2.12017.

[59] A. H. Aliwy, K. Aljanabi, and H. A. Alameen, "Arabic text clustering technique to improve information retrieval," in *AIP Conference Proceedings*, vol. 2386, no. 1. AIP Publishing, 2022.

[60] R. H. Al Mahmoud, B. H. Hammo, and H. Faris, "Cluster-based ensemble learning model for improving sentiment classification of arabic documents," *Natural Language Engineering*, pp. 1–39, 2023.

[61] S. Larabi-Marie-Sainte, M. Bin Alamir, and A. Alameer, "Arabic Text Clustering Using Self-Organizing Maps and Grey Wolf Optimization," *Applied Sciences*, vol. 13, no. 18, 2023. [Online]. Available: https://www.mdpi.com/2076-3417/13/18/10168

[62] R. H. AlMahmoud and M. Alian, "The effect of clustering algorithms on question answering," *Expert Systems with Applications*, vol. 243, p. 122959, 2024.

Impact of Emojis Exclusion on the Performance of Arabic Sarcasm Detection Models

Ghalyah Aleryani[1], Wael Deabes[2], Khaled Albishre[3], Alaa E. Abdel-Hakim[4]
Department of Computer Science in Jamoum
Umm Al-Qura University, Makkah, 25371, Saudi Arabia[1]
Department of Computational, Engineering Mathematical Sciences (CEMS),
Texas A&M University-San Antonio, San Antonio, 78224, USA[2]
Computers and Systems Engineering Department, Mansoura University, Mansoura, 35516, Egypt[2]
Department of Computer Science in Jamoum, Umm Al-Qura University, Makkah, 25371, Saudi Arabia[3]
Department of Computer Science in Jamoum, Umm Al-Qura University, Makkah, 25371, Saudi Arabia[4]

***Abstract*—The complex challenge of detecting sarcasm in Arabic speech on social media is exacerbated by the language's diversity and the nature of sarcastic expressions. There is a significant gap in the capability of existing models to effectively interpret sarcasm in Arabic, necessitating more sophisticated and precise detection methods. In this paper, we investigate the impact of a fundamental preprocessing component on sarcasm detection. While emojis play a crucial role in mitigating the absence of body language and facial expressions in modern communication, their impact on automated text analysis, particularly in sarcasm detection, remains underexplored. We examine the effect of excluding emojis from datasets on the performance of sarcasm detection models in social media content for Arabic, a language with a super-rich vocabulary. This investigation includes the adaptation and enhancement of AraBERT pre-training models by specifically excluding emojis to improve sarcasm detection capabilities. We use AraBERT pre-training to refine the specified models, demonstrating that the removal of emojis can significantly boost the accuracy of sarcasm detection. This approach facilitates a more refined interpretation of language, eliminating the potential confusion introduced by non-textual elements. The evaluated AraBERT models, through the focused strategy of emojis removal, adeptly navigate the complexities of Arabic sarcasm. This study establishes new benchmarks in Arabic natural language processing and offers valuable insights for social media platforms.**

***Keywords*—Arabic language; AraBERT; sarcasm detecting; data preprocessing; emojis impact; social media content**

I. Introduction

The evolution of social media platforms has transformed them into spaces for free speech, allowing users to express their ideas and opinions openly. While this open environment encourages meaningful discussions, it can also lead to problems when individuals use expressions or statements that may offend others due to differences in beliefs, backgrounds, gender, or race. Although such speech is protected under the latest version of the Communications Decency Act (CDA 230) [1], many social media platforms are actively working to enhance user protection against hateful and offensive content. A significant challenge, however, arises from their reliance on human monitoring and user reports [2], [3].

In recent years, there has been a significant increase in the prevalence of sarcastic speech on social media, giving rise to serious social problems, including social conflicts, racist crimes, and the spread of negative social influence. To mitigate this issue, many social media platforms have implemented word filters based on NLP techniques. Sarcastic speech includes exchanges of sarcastic or offensive remarks between individuals and extends beyond text comments to include multimedia content such as videos and audio [4]. Existing research efforts have primarily focused on proposing models for the automatic detection of sarcastic speech within text comments on platforms like Twitter or YouTube [5], [6]. These models typically involve several stages, including data processing, representation, and classification. However, these stages present several challenges, particularly concerning the Arabic language [7], [8]:

1) Its rich vocabulary, and dialectical variations.
2) Arabic sarcasm often relies heavily on contextual indications and cultural references.
3) Collecting and annotating a large dataset for Arabic sarcasm detection produces unique difficulties.
4) Time-consuming and requires extensive data learning.
5) The multimodal nature of social media content is integrating information from various sources such as text, images, videos, and emojis.

Generally, sarcasm detection heavily relies on textual data classification. However, textual data lacks crucial expressive features such as facial expressions, body language, and tone variations. To address this limitation, social media communities have introduced emojis as non-traditional vocabulary to compensate for this deficiency in information. Emojis have demonstrated their effectiveness in partially bridging the gap between textual and vocal/visual communication [9].

Nevertheless, rich languages like Arabic possess their own tools that can better address this gap than emojis, including a vast and diverse lexicon. Table I provides a comparison of the number of roots between Arabic and other languages. Additionally, Arabic speakers use numerous dialects with significant dialectical variations. Moreover, sarcasm in Arabic heavily relies on contextual cues, puns, and euphemisms. This results in an extensive textual information space generated solely using textual vocabulary, decreasing the contribution of the limited emojis dictionary.

These factors raise questions regarding the added value of emojis to classifier performance. While it is intuitive that

TABLE I. COMPARISON BETWEEN ARABIC AND SOME OTHER LANGUAGES IN TERMS OF THE NUMBER OF ROOTS

Language	Approximate Number of Roots	Notes
Arabic	23090 [11]	Roots of 3 letters By applying 73 triliteral patterns and 18 affixes produced around 27.6M words.
English	8,400 [12]	This study highlights the significant growth in root word vocabulary during the primary school years.
Russian	450 [13]	The study highlights the significance of understanding root words for mastering Russian vocabulary.

adding redundant data should not degrade performance, excess data has been shown to harm classifiers' performance in certain instances [10]. Hence, we hypothesize that using emojis for sarcasm detection in rich languages like Arabic may either reduce accuracy or offer no improvement. The rationale behind this approach is that focusing purely on the textual elements allows the models to concentrate on linguistic and semantic analysis without the potential confounds introduced by emojis.

The successful development of an Arabic sarcasm detection model, enhanced by transfer learning methods and refined by the strategic removal of emojis from the dataset, will have a significant social impact. This approach will enable social media platforms to more effectively detect and moderate sarcastic speech, thereby mitigating its potentially harmful effects and fostering a more positive and respectful online discourse. By focusing on the textual content and reducing noise in the data, the model's precision in identifying sarcasm will be improved, making the moderation process more efficient and reliable. Additionally, this research will lay the groundwork for further advancements in the creation of language-specific models for sarcasm detection, equipping diverse language communities with the tools needed to address such speech more effectively.

In this work, we investigate the following research questions: 1.

1) Does including emojis in the data improve pre-trained models' ability to detect sarcasm in the Arabic language?
2) How can the performance of the AraBERT pre-trained models for sarcasm detection in the Arabic language on social media platforms be improved by removing emojis from the data?
3) How accurately can it identify and classify sarcasm in Arabic speech on social media?

In the next section, we discuss related work on detecting textual aggressions on social media. In Section III, we describe the methodology used in this study. Section IV presents the experiments and analysis of the results. Finally, Section V concludes the study and discusses the future direction of sarcasm speech in the Arabic language.

II. RELATED WORK

User-generated content online, especially on social media platforms, can sometimes contain harmful language and hate speech, which can have detrimental effects on the online community and potentially lead to hate crimes. Recently, there has been a marked interest in smart algorithms designed to automatically identify and flag such offensive language and hate speech. Nonetheless, NLP research concerning the Arabic language is typically limited [14], as is the investigation into sarcasm detection. In this section, we aim to highlight previous efforts focused on detecting Arabic sarcasm on social media.

In the field of NLP, large-scale pre-trained models have become the standard approach for a wide range of tasks. Models like Bidirectional Encoder Representations from Transformers (BERT) are trained on massive datasets, allowing them to generalize effectively to various downstream tasks [6]. The BERT model has been employed for extensive Arabic datasets, resulting in the creation of AraBERT [15]. In [16], an automated approach to detect offensive language and detailed hate speech in Arabic tweets is proposed. The BERT model is utilized alongside two traditional machine learning methods: (i) Support Vector Machine (SVM) and (ii) Logistic Regression (LR). Additionally, the authors explore the integration of sentiment analysis and emojis descriptions as supplementary features to the textual content of tweets. Experimental results indicate that the BERT-based model outperforms existing benchmark systems in three key areas: (a) detecting offensive language with an F1-score of 84.3%, (b) identifying hate speech with an 81.8% F1-score, and (c) discerning detailed categories of hate speech (such as race, religion, social class, etc.) with a 45.1% F1-score. While sentiment analysis marginally boosts the model's efficiency in detecting offensive language and hate speech, it does not enhance the model's capability in classifying specific hate speech types.

Moreover, a universal language-agnostic approach is proposed in [17] to gather a substantial proportion of tweets with offensive and hate content, regardless of their subjects or styles. The authors gathered a significant collection of offensive tweets by leveraging the non-verbal cues in emojis. They then applied the proposed methodology to Arabic tweets and compared the results with English tweets, highlighting notable cultural variances. A consistent pattern emerged with these emojis signifying offensive content over various periods on Twitter. They hand-labeled and made publicly available the most extensive Arabic dataset encompassing offensive language, detailed hate speech, profanity, and violent content. As a result, the authors found that even advanced transformer models can overlook cultural contexts, backgrounds, or the precision inherent in authentic data, such as sarcasm.

As highlighted by [18], identifying offensive language within Arabic content is intricate. Several challenges arise, such as: (i) The colloquial language frequently used on social media platforms. These posts often contain abbreviations and slang, making it challenging for classifiers to understand and process them semantically. (ii) The diverse dialects and versions of the Arabic language add another layer of complexity to discerning offensive content. The text may require extensive preprocessing before being fed into a classification model. To combat the issue of colloquialism, the researchers processed each tweet by translating emoticons and emojis into their Arabic textual equivalents and breaking down hashtags into individual words separated by spaces. To address the issue of dialect variation, the texts were transformed from regional dialects to Modern Standard Arabic (MSA). The study tested various classifiers, including traditional machine learning models such as SVM, LR, Decision Tree (DT),

Bagging, AdaBoost, and Random Forest (RF). The results show that, among traditional machine learning models, SVM topped the list with an F1-score of 82%, followed by LR at 81% and DT at 69%. For ensemble models, Bagging led with an F1-score of 88%, followed by RF at 87%, and AdaBoost at 86%.

In the study [19], features were derived from textual descriptions of emojis. Depending on the class size and the specific emojis, these features either enhanced or hindered the model's performance. The authors introduced a transformer-based technique to tackle the problem of detecting offensive language. Their approach utilized variations of the CAMeL-BERT model and was tested using a combination of four benchmark Arabic Twitter datasets, all annotated for hate speech detection, including the dataset from the OSACT5 2022 workshop shared task. The model demonstrated proficiency in identifying offensive content in Arabic tweets, achieving an accuracy of 87.15% and an F1-score of 83.6%.

In [20], the word embedding (word2vec) feature was applied in tandem with part-of-speech and/or emojis to detect such language in Indonesian tweets on Twitter. They also experimented with combining unigrams with part-of-speech and/or emojis. Classification for this study was performed using a Support Vector Machine, Random Forest Decision Tree, and Logistic Regression methods. The highest accuracy attained was 79.85%, with an F-Measure of 87.51%, using the fusion of unigram features, part-of-speech, and emojis. Furthermore, the work in [21] investigates the impact of combining emojis-based elements, which are increasingly common on social media, with multiple textual factors for the sentiment analysis of casual Arabic tweets. The authors employed four methods to extract textual features: Bag-of-Words (BoW), Latent Semantic Analysis, and two Word Embedding variants. The study evaluates the impact of merging emojis with these textual elements using the SVM classifier, considering both scenarios: with and without feature selection. Results indicate that models incorporating emojis with word embedding and optimal feature selection produce better outcomes. The implications of amalgamating emojis-based aspects from Arabic tweets with diverse textual elements are explored.

In this study, we hypothesize that removing emojis from the training dataset will enhance the ability of AraBERT pre-trained models to discern the subtleties of sarcastic language in Arabic text, as it will encourage the model to focus more on context.

III. Proposed Work

We use transfer learning with three pre-trained models: AraBERT-v2, AraBERTv02-twitter, and Multi-dialect-BERT-base-Arabic. The models are evaluated using datasets from three different sources: SemEval 2020, YouTube, and L-HSAB, preprocessing. The evaluation of the models is performed with and without emojis. This suggests that the role of emojis in understanding the text is considered a variable in the model's performance. Finally, the performance of these models is evaluated using standard performance metrics: accuracy, precision, recall, and F1-score.

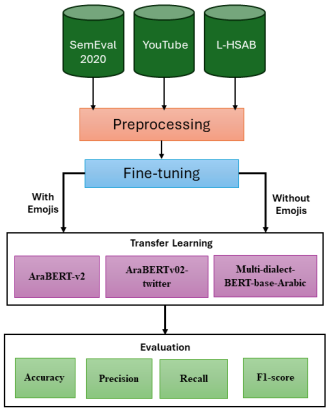


Fig. 1. Workflow of the methodology.

A. Classification Model

The core of this study revolves around the selection and application of advanced machine learning models, specifically tailored for processing Arabic text, the workflow of this study is presented in Fig. 1.

These models are critical in accurately detecting sarcasm in Arabic social media content. The selected models are:

1) Arabert_v2: This model is a BERT-based framework specifically adapted for Arabic text. AraBERT's architecture allows it to understand the contextual nuances of the Arabic language, making it an ideal choice for tasks like sarcasm detection where context plays a crucial role. Moreover, it comprises 12 transformer layers and 768 hidden units in each layer [22].

2) AraBERTv02-twitter: Building upon the foundation laid by AraBERT, Arabert_v2 is an advanced version that offers improved capabilities. Its enhanced features include a better understanding of dialectal variations and more refined contextual analysis. This version is particularly effective in handling the intricate aspects of sarcasm in various forms of Arabic language and dialects, maintaining the same architecture layer of Arabert_v2 [23].

3) Multidialect Bert base AraBERT: Recognizing the diversity of the Arabic language, this model is designed to handle multiple dialects. Its architecture is tailored to adapt to the linguistic variations found across different Arabic-speaking regions and the model is trained on the entire Wikipedia for each language, which is crucial for a comprehensive sarcasm detection tool that can operate effectively across diverse social media platforms and content. The architecture is composed of 12 encoder blocks, each equipped with 12 self-attention heads, and is followed by hidden layers that have a size of 768 [24], [23].

Table II presents the key hyperparameters for the three models, which have been determined based on empirical.

B. Datasets

The success of transfer learning models in NLP tasks heavily relies on the quality and relevance of the datasets used

TABLE II. MAIN HYPERPARAMETERS FOR FINE-TUNING THE THREE MODELS

Parameter	Value
Adam optimizer	1e-8
Learning Rate	5e-5
Batch Size	16
Maximum Sequence Length	256
Epochs	15



Fig. 2. Snapshot of SemEval dataset.

for training and testing. In this study, three distinct datasets have been carefully selected to ensure comprehensive coverage of the various facets of Arabic sarcasm and related nuances in social media contexts. These datasets are:

1) SemEval 2020 task 12 dataset: This dataset is specifically designed for sarcasm detection, making it highly relevant for this research. Fig. 2 shows a sample of the SemEval dataset. It comprises a collection of social media posts annotated for sarcasm, providing a foundational basis for training and testing the models. Including this dataset is crucial as it offers direct insights into the textual characteristics and markers indicative of sarcasm in social media content [25].

2) YouTube dataset for anti-social behavior detection: Recognizing the importance of context in sarcasm detection, this dataset from YouTube, comprising comments and posts in Arabic, is employed to understand and identify patterns of anti-social behavior, as shown in Fig. 3 [26]. This dataset provides a broader perspective on how sarcasm might be intertwined with or differentiated from other forms of communication, particularly those that are anti-social or negative.

3) L-HSAB (Levantine hate speech and abusive language) dataset: The L-HSAB dataset is instrumental in understanding the landscape of hate speech and abusive language in Arabic,



Fig. 3. Snapshot of YouTube dataset.



Fig. 4. Snapshot of L-HSAB dataset.

particularly in the Levantine dialect. Fig. 4 shows a snapshot of the L-HSAB dataset. Its inclusion aids in refining the detection algorithms to discern sarcasm from potentially similar but contextually different expressions like hate speech or abusive remarks.

C. Preprocessing

Preprocessing plays a crucial role in the machine learning workflow for classification tasks. This phase requires the modification and improvement of input data to make it suitable and beneficial for use with transfer learning models. The study implemented the following steps:

1) Data cleaning is a pivotal step in data analysis, emphasizing the importance of removing noise and irrelevant information to enhance data quality. Data cleaning is a pivotal step in data analysis, emphasizing the importance of removing noise and irrelevant information to enhance data quality [27]. This step was applied to the three datasets by removing unnecessary columns, handling missing values, and text data normalization
2) Text data preprocessing is a fundamental step in NLP, enabling the transformation of raw text into a clean, organized format suitable for analysis. This process involves: (i) using regular expressions to identify and remove unwanted characters, spaces, or patterns within text data, (ii) tokenization is the process of breaking down the text into smaller units, such as words or phrases, (iii) removal of stop words, and (iv) performing Stemming which is reducing words to their root form.
3) Splitting the dataset into training, validation, and test sets in two distinct versions: one version includes emojis in the text data, and the other version with emojis excluded.

D. Performance Metrics

The evaluation of the models is conducted using a set of standard performance metrics in transfer learning:

1) Accuracy: this metric measures the proportion of correctly identified instances (both sarcastic and non-sarcastic) among the total instances. It provides an overall effectiveness of the model.

$$Accuracy = \frac{\text{Number of Correct Predictions}}{\text{Total Number of Predictions}} \quad (1)$$

2) Precision and recall: these metrics evaluate the accuracy of the model in identifying sarcasm. Precision measures the proportion of correctly identified sarcastic instances among all instances identified as sarcastic, while recall measures the proportion of correctly identified sarcastic instances among all actual sarcastic instances.

$$Precision = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}} \quad (2)$$

$$Recall = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}} \quad (3)$$

3) F1 score: it is the harmonic mean of precision and recall, providing a single metric that balances both. It is particularly useful when the class distribution is imbalanced.

$$F1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (4)$$

IV. Results and Analysis

In this work, the investigation primarily focuses on the influence of emojis in classifying offensive language in Arabic. Additionally, it presents a comprehensive evaluation and interpretation of the outcomes, highlighting which machine learning model is most effective in accurately identifying offensive speech with or without emojis. The results are critical for understanding the nuances and accuracy of language processing techniques in a linguistically diverse and complex region like the Arab world.

The comparative analysis, which is presented in Fig. 5v to 7, assesses the influence of emojis on the performance of Arabic language models in classifying offensive content within the SemEval, YouTube, L-HSAB datasets, respectively. This evaluation is segmented into four distinct metrics: accuracy, recall, precision, and F1-score. The performance of the three considered, AraBERT_v2 (v2), AraBERT_v2_Twitter (TW), and multi_dialect_bert_base_arabert (MD) models are investigated for all of these datasets.

A. Accuracy

In Fig. 5(a), 6(a), and 7(a), the accuracy metrics are evaluated across multiple models, including AraBERT_v2 (v2), AraBERT_v2_Twitter (TW), and multi_dialect_bert_base_arabert (MD). Across these figures, the accuracy rates for each model are displayed both with and without the inclusion of emojis in the dataset. In all cases, except for L-HSAB with the TW model, the accuracy tends to be higher when emojis are excluded. This trend suggests that emojis may introduce ambiguity or noise that degrades the models' ability to accurately classify text.

B. Recall

The recall results across the three models: V2, TW, and MD, are examined through Fig. 5(b), 6(b), and 7(b), respectively. For all the datasets, it observed that excluding emojis leads to an increase in recall for TW and MD, while in the V2 models with the YouTube and SemEval dataset, recall shows a slight improvement when emojis are included. This suggests that emojis may enhance detection for some models but could potentially disrupt others, resulting in missed detection. The V2 model, however, exhibits negligible differences in both cases.

C. Precision

Analyzing precision across different models in Fig. 5(c), 6(c), and 7(c) yields the following observations. In Fig. 5(c), the precision is higher when emojis are excluded for all three models. Particularly, there is a significant increase in precision observed for the V2 and TW models when emojis are excluded, suggesting that emojis may have a notable impact on lowering precision, especially for these models. For Fig. 6(c), the precision outcomes show improvements for all models when emojis are removed. The V2 and MD models demonstrate significant enhancement in precision without emoji, indicating a reduction in false positives. The TW model shows no significant difference in both cases. In Fig. 7(c), the MD model shows improvement in precision without emoji, indicating potential complications that emojis introduce in precision tasks across different models. However, the V2 and TW models show a minor increase in precision with emoji, suggesting a better ability to identify positive instances when emojis are present.

D. F1-score

In Fig. 5(d) and 6(d), a consistent trend is observed where excluding emojis benefits the F1-score across all models. This indicates that emojis do not significantly contribute to the classification process and could potentially even restrict it. The same thing almost exists in Fig. 7(d). For V2, the F1-score slightly drops when emojis are excluded. The difference in the MD case is almost negligible. These findings imply that emojis exclusion mitigates the trade-off between precision and recall for this particular model.

Based on these results, the proposed framework concludes that excluding emojis consistently improves accuracy, recall, precision, and F1-score across various models, indicating its beneficial impact on sarcasm speech classification. Emojis introduce ambiguity or noise that degrades classification accuracy, while their exclusion leads to increased recall, particularly for the TW and MD models. Precision notably increases, especially for the V2 and MD models, when emojis are removed, reducing false positives. Similarly, the F1-score improves across all models when emojis are excluded, except for a slight drop in V2's case, implying a better balance between precision and recall. Therefore, these results support our main hypothesis (RQ1, RQ2, and RQ3) that removing emojis during the preprocessing stage has a positive, or at least non-negative, impact on improving sarcasm speech classification performance.

Reflecting these findings, we believe that excluding emojis in sarcasm detection increases focus on an important aspect of NLP. Although emojis are commonly used in digital communication, their influence on language processing can vary significantly depending on the linguistic context. In the case of Arabic, a language rich in vocabulary and dialects, emojis may introduce more noise than benefit, as demonstrated by the performance improvements observed in our models when these non-textual elements were omitted.

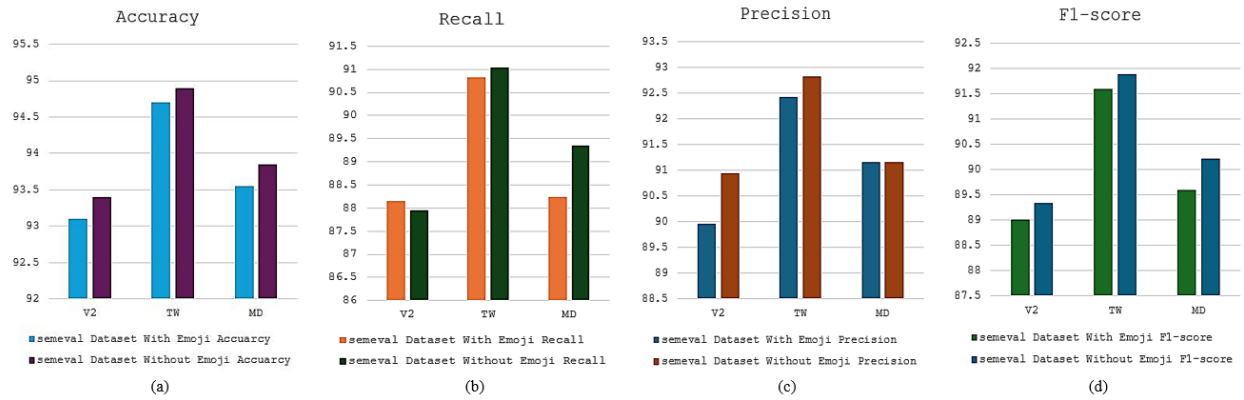


Fig. 5. Comparison results of classification with and without emojis on the SemEval dataset.

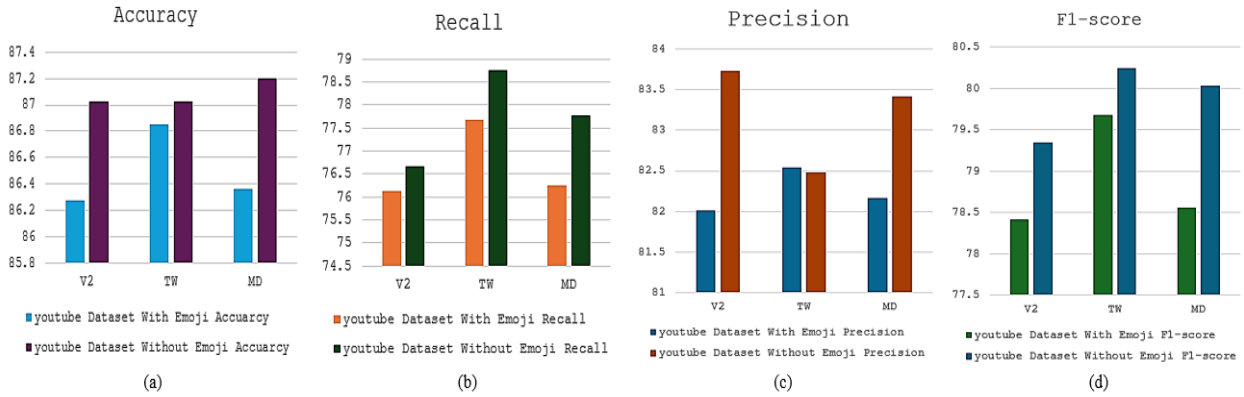


Fig. 6. Comparison results of classification with and without emojis on the YouTube dataset.

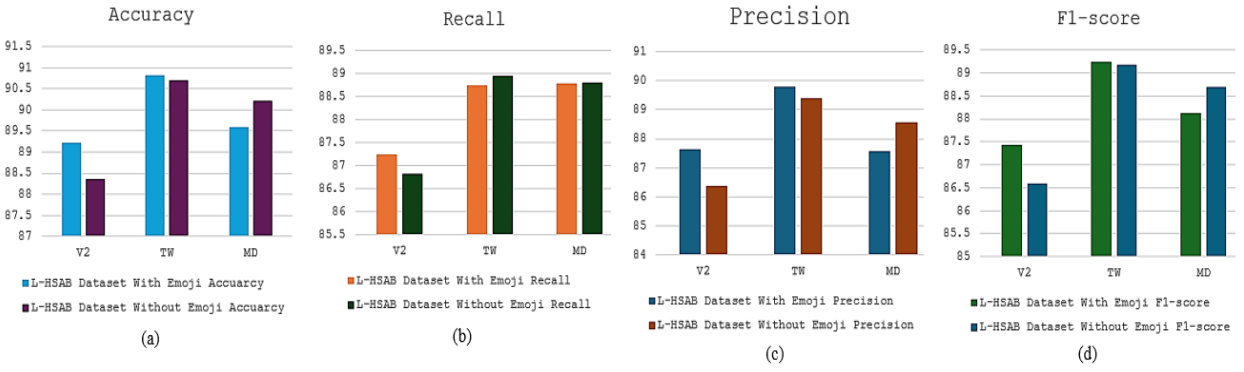


Fig. 7. Comparison results of classification with and without emojis on the L-HSAB dataset.

V. Conclusions and Future Work

In this paper, we investigated the impact of excluding emojis during the preprocessing stage on the performance of Arabic sarcasm detection models. Detecting sarcasm in Arabic social media speech presents a multifaceted challenge due to linguistic diversity and the intricacies of sarcastic expressions. By evaluating the accuracy, recall, precision, and F1-score metrics across various models and datasets, we demonstrated that emojis exclusion enhances sarcasm detection accuracy. This research provides a more precise interpretation of language by eliminating potential confusion introduced by non-textual elements, ultimately contributing to the advancement of language processing techniques in linguistically diverse regions. Moreover, our findings offer valuable insights for social media platforms and natural language processing research. By highlighting the positive impact of emojis exclusion on sarcasm detection model performance, new benchmarks are established in Arabic natural language processing.

Nevertheless, there are some limitations to this study, which need to be mentioned. First, such research specifies the

focus of sarcasm detection on textual information rather than considering that social media content often contains data in the form of images, videos, and so on in addition to text data. Second, although the given datasets offered a wide variety of texts, it should be noted that they might not include all the different types, styles, and details of the Arabic language used across Arabic-speaking countries. This limitation implies that when generalizing the findings of this study, the conclusions may be restricted to the contexts captured by the datasets.

Future research could focus on developing advanced machine-learning tools that interpret emojis alongside the text, reducing confusion from emojis misuse or overuse. This study provides the way for further advancements, particularly in enhancing the robustness and applicability of sarcasm detection models for Arabic and other languages, as follows:

- Integrating text with other data types like images, videos, and audio could enhance the accuracy of sarcasm detection. For instance, facial expressions in images accompanying sarcastic text might provide additional context that is not captured by textual analysis alone. The ability to process multiple data formats in real time could significantly improve the performance of the model.
- Real-time monitoring tools on social media platforms could be significantly enhanced to detect and report sarcasm in live conversations, which can help prevent the propagation of harmful comments and improve automated customer service responses. Integrating these models into platforms like Twitter and Facebook could also offer deeper insights into public opinion trends.
- A key challenge in Arabic sarcasm detection is the availability of datasets where the sarcasm is annotated. Future efforts could enhance this by using more effective schemes of data annotation like crowdsourcing or semi-supervised learning, as well as in collecting datasets that include various dialects and cultural settings to improve the model's transferability and reliability.

ACKNOWLEDGMENT

The authors would like to thank the Deanship of Scientific Research at Umm Al-Qura University for supporting this work by Grant Code: (24UQU4350534DSR02).

REFERENCES

[1] V. Dumas, "Enigma machines: Deep learning algorithms as information content providers under section 230 of the communications decency act," *Wis. L. Rev.*, p. 1581, 2022.

[2] A. O. Marwa Khairy, Tarek M. Mahmoud and T. A. El-Hafeez, "Comparative performance of ensemble machine learning for arabic cyberbullying and ofensive language detection," *Language Resources and Evaluation, Springer*, 2023.

[3] V. Sukhavasi and V. Dondeti, "Effective automated transformer model based sarcasm detection using multilingual data," *Multimedia Tools and Applications*, pp. 1–32, 2023.

[4] S. Mihi, B. Ait Ben Ali, I. El Bazi, S. Arezki, and N. Laachfoubi, "Automatic sarcasm detection in dialectal arabic using bert and tf-idf," in *The Proceedings of the International Conference on Smart City Applications*. Springer, 2021, pp. 837–847.

[5] J. A. García-Díaz, S. M. Jiménez-Zafra, M. A. García-Cumbreras, and R. Valencia-García, "Evaluating feature combination strategies for hate-speech detection in spanish using linguistic features and transformers," *Complex & Intelligent Systems*, vol. 9, no. 3, pp. 2893–2914, 2023.

[6] M. Koroteev, "Bert: a review of applications in natural language processing and understanding," *arXiv preprint arXiv:2103.11943*, 2021.

[7] I. A. Farha and W. Magdy, "From arabic sentiment analysis to sarcasm detection: The arsarcasm dataset," in *Proceedings of the 4th Workshop on Open-Source Arabic Corpora and Processing Tools, with a Shared Task on Offensive Language Detection*, 2020, pp. 32–39.

[8] A. Rahma, S. S. Azab, and A. Mohammed, "A comprehensive review on arabic sarcasm detection: Approaches, challenges and future trends," *IEEE Access*, 2023.

[9] J. Subramanian, V. Sridharan, K. Shu, and H. Liu, "Exploiting emojis for sarcasm detection," in *Social, Cultural, and Behavioral Modeling: 12th International Conference, SBP-BRiMS 2019, Washington, DC, USA, July 9–12, 2019, Proceedings 12*. Springer, 2019, pp. 70–80.

[10] A. E. Abdel-Hakim and W. A. Deabes, "Impact of sensor data glut on activity recognition in smart environments," in *2017 IEEE 17th International Conference on Ubiquitous Wireless Broadband (ICUWB)*. IEEE, 2017, pp. 1–5.

[11] M. T. B. Othman, M. A. Al-Hagery, and Y. M. El Hashemi, "Arabic text processing model: Verbs roots and conjugation automation," *IEEE Access*, vol. 8, pp. 103 913–103 923, 2020.

[12] A. Biemiller and N. Slonim, "Estimating root word vocabulary growth in normative and advantaged populations: Evidence for a common sequence of vocabulary acquisition." *Journal of educational psychology*, vol. 93, no. 3, p. 498, 2001.

[13] G. Z. Patrick, "Roots of the russian language," *(No Title)*, 1989.

[14] M. El-Melegy, A. Abdelbaset, A. Abdel-Hakim, and G. El-Sayed, "Recognition of arabic handwritten literal amounts using deep convolutional neural networks," in *Pattern Recognition and Image Analysis: 9th Iberian Conference, IbPRIA 2019, Madrid, Spain, July 1–4, 2019, Proceedings, Part II 9*. Springer, 2019, pp. 169–176.

[15] W. Antoun, F. Baly, and H. Hajj, "Arabert: Transformer-based model for arabic language understanding," *arXiv preprint arXiv:2003.00104*, 2020.

[16] M. J. Althobaiti, "Bert-based approach to arabic hate speech and offensive language detection in twitter: Exploiting emojis and sentiment analysis," *International Journal of Advanced Computer Science and Applications*, vol. 13, no. 5, 2022.

[17] H. Mubarak, S. Hassan, and S. A. Chowdhury, "Emojis as anchors to detect arabic offensive language and hate speech," *Natural Language Engineering*, vol. 29, no. 6, pp. 1436–1457, 2023.

[18] F. Husain and O. Uzuner, "Investigating the effect of preprocessing arabic text on offensive language and hate speech detection," *Transactions on Asian and Low-Resource Language Information Processing*, vol. 21, no. 4, pp. 1–20, 2022.

[19] S. Al-Dabet, A. ElMassry, B. Alomar, and A. Alshamsi, "Transformer-based arabic offensive speech detection," in *2023 International Conference on Emerging Smart Computing and Informatics (ESCI)*. IEEE, 2023, pp. 1–6.

[20] M. O. Ibrohim, M. A. Setiadi, and I. Budi, "Identification of hate speech and abusive language on indonesian twitter using the word2vec, part of speech and emoji features," in *Proceedings of the 1st International Conference on Advanced Information Science and System*, 2019, pp. 1–5.

[21] S. Al-Azani and E.-S. M. El-Alfy, "Combining emojis with arabic textual features for sentiment classification," in *2018 9th International Conference on Information and Communication Systems (ICICS)*. IEEE, 2018, pp. 139–144.

[22] H. Elfaik *et al.*, "Combining context-aware embeddings and an attentional deep learning model for arabic affect analysis on twitter," *IEEE Access*, vol. 9, pp. 111 214–111 230, 2021.

[23] M. A. Humayun, H. Yassin, and P. E. Abas, "Dialect classification using acoustic and linguistic features in arabic speech," *IAES International Journal of Artificial Intelligence*, vol. 12, no. 2, p. 739, 2023.

[24] A. S. Alammary, "Bert models for arabic text classification: a systematic review," *Applied Sciences*, vol. 12, no. 11, p. 5720, 2022.

[25] M. Zampieri, P. Nakov, S. Rosenthal, P. Atanasova, G. Karadzhov, H. Mubarak, L. Derczynski, Z. Pitenis, and Ç. Çöltekin, "Semeval-2020 task 12: Multilingual offensive language identification in social media (offenseval 2020)," *arXiv preprint arXiv:2006.07235*, 2020.

[26] A. Alakrot, L. Murray, and N. S. Nikolov, "Dataset construction for the detection of anti-social behaviour in online communication in arabic," *Procedia Computer Science*, vol. 142, pp. 174–181, 2018.

[27] T. Nguyen, C. Van Nguyen, V. D. Lai, H. Man, N. T. Ngo, F. Dernoncourt, R. A. Rossi, and T. H. Nguyen, "Culturax: A cleaned, enormous, and multilingual dataset for large language models in 167 languages," *arXiv preprint arXiv:2309.09400*, 2023.

A Configurable Framework for High-Performance Graph Storage and Mutation

Soukaina Firmli, Dalila Chiadmi, Kawtar Younsi Dahbi
SIP Research Team-Rabat IT Center-EMI, Mohammed V University in Rabat, Morocco

***Abstract*—In the realm of graph processing, efficient storage and update mechanisms are crucial due to the large volume of graphs and their dynamic nature. Traditional data structures such as adjacency lists and matrices, while effective in certain scenarios, often suffer from performance trade-offs such as high memory consumption or slow update capabilities. To address these challenges, we introduce CoreGraph, an advanced graph framework designed to optimize both read and update performance. CoreGraph leverages a novel segmentation method and in-place update techniques, along with configurable memory allocators and synchronization mechanisms, to enhance parallel processing and reduce memory consumption. CoreGraph's update throughput (with up to 20x) and analytics performance exceed those of several state-of-the-art graph structures such as Teseo, GraphOne and LLAMA, while maintaining low memory consumption when the workload includes updates. This paper details the architecture and benefits of CoreGraph, highlighting its practical application in traffic data management where it seamlessly integrates with existing systems providing a scalable and efficient solution for real-world graph data management challenges.**

***Keywords*—Data structures; concurrency; graph processing; graph mutations; high-performance computing; traffic management**

I. Introduction

Research in graph processing technology continues to evolve and prosper due to the unique ability of graphs to represent complex relationships and inter-dependencies within data, making them suitable for many modern applications. Graph applications include Knowledge Graphs (KG) [1] used in search engines, personal assistants, and recommendation systems; Graph Neural Networks (GNNs) [2] for AI tasks like node classification, link prediction, and graph classification; and real-time graph analysis for streaming data from sources like Twitter and financial transactions.

As a result of the widespread use of graphs, there is a growing need to efficiently manage and analyze them. This has led to the development of graph processing systems and graph databases like Neo4j [3], which are well-suited for storing, analyzing, and streaming large graph data due to their scalability, real-time processing capabilities, and ability to handle complex relationships. However, these systems face multiple challenges inherent to graph processing and streaming that limit the effective use of graphs in the real world due to varying graph characteristics, memory-intensive algorithms, and access patterns that cause latency, as well as the continuous evolution of graph topology and properties [4].

To understand and address these challenges, it is important to recognize that at the heart of each graph system lies the graph data structure, which stores the vertices and their edges, and whose performance largely contributes to the overall performance of the system. However, classic graph data structures typically trade some characteristics for others [5]. For instance, Compressed Sparse Row (CSR) representations are memory efficient with good read-only performance but have poor update performance. The challenge, therefore, is to design graph data structures that ideally offer excellent read-only performance, fast mutations (i.e., vertex or edge insertions and deletions), and low memory consumption with or without mutations.

To address these problems with classical data structures, efforts have been made to improve the updated friendliness of CSR. These include in-place update techniques [6], [7], batching techniques [8], [9], changeset-based updates with delta maps [10], and multi-versioning [11]. However, systems still struggle to offer the best tradeoff between read and, update performance and memory consumption.

To this aim, we present in this paper CoreGraph, a highly configurable end-to-end graph framework designed to address these challenges. CoreGraph builds on our previous work CSR++ [12], to offer a high-performance concurrent data structure for graph topology and properties storage. The framework is designed for in-place graph mutations, achieving superior analytics and update performance with significantly lower memory requirements compared to state of art solutions like LLAMA [11].

The main contributions of our framework can be summarized as follows:

1) Efficient in memory architecture: CoreGraph implements an in-memory architecture that allows for efficient graph loading, mutation, and analysis, making it suitable for a wide range of applications.

2) Optimized storage and update protocol: CoreGraph offers a storage and update protocol based on a novel segmentation method that optimizes graph storage for both read and write operations while minimizing memory space.

3) High configurability: CoreGraph is highly configurable by integrating advanced synchronization mechanisms, including Hardware Transactional Memory (HTM) and adaptable locking strategies, to enhance parallel performance while minimizing contention. Furthermore, the framework supports configurable memory allocators to optimize resource utilization and performance based on specific application needs.

We also demonstrate the practical applicability and portability of CoreGraph as a lightweight graph storage and mutation framework through a case study on traffic data manage-

ment, a domain characterized by high-frequency updates and large data volumes. CoreGraph not only improved the performance of graph mutations and analytics but also seamlessly integrated with an existing framework for traffic data management. This integration highlights CoreGraph's potential as a vital component in real-world data discovery and exploitation chains.

The rest of this paper is organized as follows: Section II presents related works, Section III gives an overview of the proposed approach for graph storage and mutation, and Section V presents the use case related to the traffic data management domain, which aims to integrate CoreGraph with an existing framework for graph analysis and updates. The study concludes in Section VI.

II. Related Work

In this section, we present some notable state-of-the-art research work that proposes optimization to classic data structures for graph storage and mutation for graph processing and streaming. We give a brief overview of each work in terms of i) graph storage and ii) graph updates, and then draw their main limitations.

A. Graph Storage

First, their exists different techniques available for researchers to optimize data structures, among which we found: optimizing the memory layout of the data structures [27] [9] [25], using compression algorithms [26] [24] and using special memory allocator [19] [18].

SSTGraph [26] use a compression algorithm based on the tinyset parallel dynamic set data structure, which implements set membership using sorted packed memory arrays. This allows for logarithmic time access and updates, as well as optimal linear time scanning, by minimizing serialization overhead.

To improve the cache performance of dynamic graph data structures, many researchers [18] [19][20] use bucketing technique where buckets are used to group edges from the same source vertex together, or using linked lists to group edges from different source vertices together. As we note from the evaluation results of works in the literature, there are only a few works that provide a thorough sensitivity analysis of different variations of their solutions like [25].

Another approach used by systems [19] [9] [18] in literature is using a special memory allocator, either internal to their system or external, to minimize the memory fragmentation caused by frequent reallocations on dynamic data structures. For instance, to efficiently perform memory reclamation and manage space, Hornet's [19] internal memory management uses a B+ tree for insertions and deletions to keep track of the available blocks of edges. However, when deletions are not frequent, which is the case in most real-world scenarios, the overhead of the memory reclamation makes the update performance slower.

B. Update Protocols

The ingestion and storage of the new updates play a crucial role in the overall performance of the systems. The methods used to implement them, that we refer to as update protocols, vary in the literature, therefore, we propose our classification for these protocols. First, update storage can be either in-place or using deltas, while update ingestion can be in bulk or concurrently with analytic workloads.

1) Update storage: Techniques that use in-place updates employ the static data structures in a way that allows for in-place digestion of sets with insertions and deletions of vertices and edges, without requiring the expensive rebuilding of the data structure. Systems [28] [20] [25], develop a variant of CSR based on Packed Memory Arrays (PMAs), that provides efficient in-place updates by leaving space at the end of each adjacency list. Moreover, when the number of gaps is too small or too large, systems are required to perform a re-balancing of the tree to rearrange the gaps in the array. This may slowdown the update performance and delay the analytic workloads.

Dynamic Arrays are used by systems like NetworKit [6] to perform in-place edge insertions by directly storing the new edges in dynamic growable edge arrays and reallocating twice the initial array size if there is no memory space for the new edge. The same method is employed by the Madduri et al. [9]. The author in [11] with the exception that the size of the new edge array is defined in terms of a customizable factor rather than a fixed factor of two. Subsequently, when employing dynamic arrays, the amortized cost for updates is O(1) for insertions and O(deg V) for deletions. However, the memory footprint can be quite substantial as the reallocations leave unused space, when there are no updates.

On the other hand, systems [10] [29] [23] [30] employ additional data structures to store the new updates referred to as *deltas*. GraphOne [23] implements a hybrid store for deltas using adjacency lists (AL) store and an edge list (EL). The AL keeps track of a linked list of vertex degrees at various points in time using timestamps. However, the performance of GraphOne suffers because of the indirection layer and the multiple levels of data in the adjacency list as [12] points out, making it not reliable for read-intensive workloads.

2) Update ingestion: The ingestion of updates can be performed in bulk, with alternating phases: pending updates wait for currently executing queries to finish before they start executing, and pending queries wait for currently executing updates to finish before they start executing. Update ingestion can also be concurrent with query execution, but in that case, the system needs to guarantee consistency on the data and at user level [11].

Moreover, the updates can be ingested as single operations or in batches. First, supporting single updates can be challenging. In fact, depending on the availability of memory, systems need to allocate new blocks to store the new edges [18]. Consequently, the frequent checks for memory availability and reallocations cause a large overhead, making the single updates very slow. To remediate the slow single update performance, systems [25] [11] [23], opt for batch updates where the batch of edge updates is pre-processed to allow for parallel updates, and to reduce the system calls for frequent allocations.

C. Discussion

The related works provide a background for optimizing graph representations for graph storage and updates. Table I

TABLE I. SUMMARY OF REVIEWED SYSTEMS AND THEIR SUPPORTED CAPABILITIES. INFRA.: SINGLE MACHINE (SM) OR DISTRIBUTED (DIST); DS: DATA STRUCTURES; SU: SINGLE UPDATES; BATCH: BATCH UPDATES; MV: MULTIVERSIONING; COMPACT: COMPACTION; UP. STORE: IN-PLACE (IP) OR DELTA (D) STORAGE; SCANS: PERFORMANCE IN READ WORKLOADS; MEM: MEMORY CONSUMPTION; UP. PERF.: PERFORMANCE IN MUTATION

Systems	Infra.	DS	SU	Batch	MV	Compact.	Up. Store	Scans	Mem	Up. Perf.
BGL[13]	SM	AL	+	-	-	+	IP	-	-	+
PGX.SM[14]	SM	CSR	-	+	+	-	D	+	+	-
Ligra[15]	SM	CSR	-	-	-	-	X	+	+	-
GraphLab[16]	Dist	CSR	-	-	-	-	X	-	+	-
PGX.D[17]	Dist	CSR	-	+	+	-	D	+	+	-
STINGER[18]	SM/Dist	AL	+	+	-	-	IP	-	-	+
Hornet[19]	GPU	AL	+	+	-	-	IP	-	+	+
Compact[9]	SM	AL	-	+	-	+	IP	-	-	+
PCSR[20]	SM	CSR	+	-	-	+	IP	+	-	+
LLAMA[21]	SM	CSR	-	-	+	+	D	+	-	-
Metall[22]	SM	AL	-	+	+	+	D	-	+	+
GraphOne[23]	SM	AL + EL	+	+	+	+	D	-	+	+
Aspen[24]	SM	Tree	-	+	+	+	IP	-	+	+
Teseo[25]	SM	Tree	+	+	+	+	IP	+	-	+
SSTGraph[26]	SM	AL	-	+	-	-	IP	-	-	+

presents a summary of our related work and their limitations, which we discuss below.

1) Achieving optimal performance trade-off: As discussed and highlighted in Table I, most systems tend to improve on an aspect of performance, either read performance, updates throughput or memory consumption, and to barely achieve the best trade-off between all three of these aspects. For instance, the majority of the systems struggle to support several versions of the graph when storing updates in deltas, because of high memory cost, and graph compaction [11] is necessary to reduce the amount of space needed for changes. However, the performing frequent compaction requires expensive computation and slows down the performance of the system.

2) Lack of configuration and hybrid representation: systems still exhibit limited configurability in their data structures. This limitation underscores the need for more adaptable data structures to accommodate diverse workload requirements, especially by considering different sizing of buckets within structures facilitating in-place updates. A noticeable absence of configurability in current literature highlights the need for refining existing designs to better suit diverse workloads tailored to specific application requirements.

3) Lack of comprehensive experimental study: There is a notable emphasis on protocols for updating graph topology. However, it is infrequent to find comprehensive implementations and evaluations of performance when considering other graph data elements, such as graph properties, reverse edges, or intermediary results in graph algorithms, and a data storage-focused approach is vital for achieving high performance in real-world applications. We also highlight the lack of validation of these graph processing systems in real-world scenarios, as most are only evaluated against generic data sets or synthetic data data sets that are not diversified.

Moreover, The performance of graph systems is significantly influenced by the dynamic memory allocator employed. To our knowledge, there are only few studies [31] that provide experimental analysis on how memory allocation impacts high-performance query engines, but not for graph storage and mutation systems.

Therefore, in this work, we aim to address these challenges by proposing a generic and domain-independent system for graph storage and mutation.

III. PROPOSED FRAMEWORK

In this section, we present our highly configurable end-to-end graph framework CoreGraph that allows the loading, storage and mutation of graphs, as well as their analysis, while providing high performance, low memory consumption and high update throughput. The architecture of CoreGraph is shown in Fig. 1.

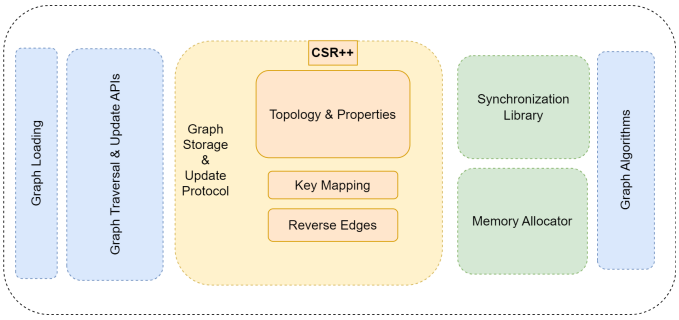


Fig. 1. CoreGraph framework.

CoreGraph enables fast concurrent accesses to the main graph data (vertex and edge tables) and stores additional graph data, such as reverse edges, user-defined keys, and vertex and edge properties. Additionally, it supports graph analytics kernels based on Green-Marl implementation [32], which are optimized to run in parallel (Section III-A).

Moreover, CoreGraph offers the main capabilities to load graphs in-memory using smart allocation and mutate graphs by exposing storage and update APIs, as well as a high performance and low memory footprint storage (Section III-B),

Finally, CoreGraph offers high configurability based on specific requirements and workloads by providing i) a configurable reallocation size for edge arrays to reduce the memory footprint when there are less frequent edge insertions (Section III-B), ii) configurable synchronization mechanism to enhance parallelism and reduce the overhead of contention (Section III-C), and iii) configurable memory allocators to optimize performance and resource utilization (Section III-D).

The main elements of our framework are represented in Fig. 1, and a detailed description will be presented below.

A. Graph Storage

The storage component of CoreGraph is a concurrent multimap that maintains the graph's topology and properties, based on the segmentation method where a fixed number of keys are grouped in an array referred to as segment. The segmentation method enables a less bulky, cache-friendly layout. Additionally, CoreGraph maintains additional graph data in addition to the basic graph data (vertex and edge tables), allowing for rapid algorithms. By storing extra graph data like reverse edges and user-defined keys, it can handle directed graphs and external IDs.

1) Topology: We store vertices in *segments* and the number of vertices stored in each segment of a graph is determined by the global configuration NUM_V_SEG. When there are N vertices in a graph, the number of segments is: num_segment = N / NUM_V_SEG + 1, as shown in Fig. 2(A). As for the structure of vertices, CoreGraph stores i) the degree ii) a pointer to their neighbor list if the degree is more than 1 and iii) optionally a pointer to the edge properties as shown in Fig. 2(B). This structure is similar to a combination of CSR and adjacency lists; However, CoreGraph's storage performance is superior to other adjacency-inspired approaches for skewed graphs due to the *inlining* of the single edges in the vertex structure, meaning if a vertex has only one neighbor, then the edge corresponding to said neighbor is stored within the structure of the vertex instead of a pointer, which minimizes the memory consumption and the cache misses when accessing the edge. Finally, if a vertex has more than one neighbor, then it's adjacency is stored as an array of edges; each edge contains the vertex ID corresponding to the index of the neighbour in the segment, and segment ID of its corresponding neighbor.

A complete sensitivity analysis [12] of the segment size shows that setting the value depends on the workload, since 1) for read intensive workloads, cache performance decreases if the size of segments is small, and the frequently allocated segments are not necessarily be allocated contiguously in memory. However, for update intensive workloads, contention between threads updating vertices in the same segment may increase, if the NUM_V_SEG is set to a big value, causing a decrease in the update throughput. Therefore we expose this option in the configuration of CoreGraph for users to set the The NUM_V_SEG so that they get the best performance depending on their workload. when configuring the segment layout. CoreGraph sets the default segment size to a value that is a multiple of 4 to enable cache alignment [4].

2) Properties: CoreGraph allows the user to configure the storage to store extra pointers for the vertex and edge properties. If the to-be-loaded graph data set does not include vertex/edge properties, then property support can be disabled to save memory. If activated, CoreGraph keeps a vector of pointers to vertex-property arrays which are parallel to the vertex array. Similarly, if edge properties are activated, CoreGraph stores a pointer to an array of edge-property values within each vertex structure and we use the same segmentation approach as for edges. In case of multiple properties, we allocate an array that stores the values for different edge properties in a cache-aligned manner. The edge properties are stored separately from the edge arrays to enable copy-on-write functionality for the edge-property arrays of only the updated vertices, unlike CSR which requires rebuilding the edge properties for the entire graph. This approach also makes it easier to maintain the property values in the same order as the edges if sorting is needed after an update. While this design incurs a moderate memory overhead, it offers significant benefits for performance, as the performance of the insertion of new edges with an edge properties of CoreGraph is an order of magnitude faster than other state-of-the-art in-place update systems, namely, Teseo [25] and STINGER [18].

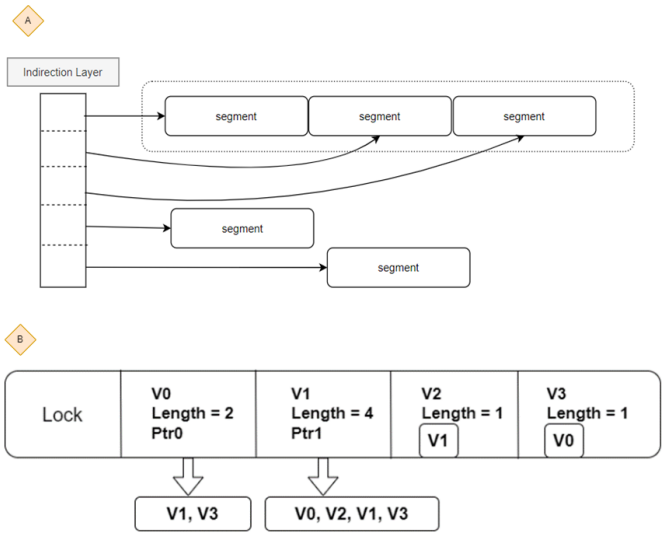


Fig. 2. Representation of Topology in CoreGraph. A) An indirection layer stores pointers to the segments to allow adding new ones and the segments are initially stored contiguously upon the loading of the graph to enhance the read performance. B) The segment structure stores a fixed number of vertices which in turn store information about their adjacencies.

B. Update Protocols

For fast, low-memory-usage graph mutation, CoreGraph offers update APIs as abstractions for an update protocol. The update protocol includes creating new vertices and edges, removing existing ones, and modifying the vertex and edge properties. When compared to CSR, which must reallocate and copy edge and vertex arrays whenever the graph's topology is changed, our segmented approach offers significantly higher update throughput due to the fact that each vertex and edge can be updated independently during graph mutation. Vertices in CoreGraph are kept in indexed arrays (see Fig. 3). If there is free space in a segment, adding a new vertex costs O(1), otherwise CoreGraph creates a new segment if necessary. However, in CoreGraph, adding a new vertex is faster reaching an update throughput of 9M vertices/s due to the segmented

nature of the vertex array; that is, we do not need to replicate the entire vertex array when adding or removing entries.

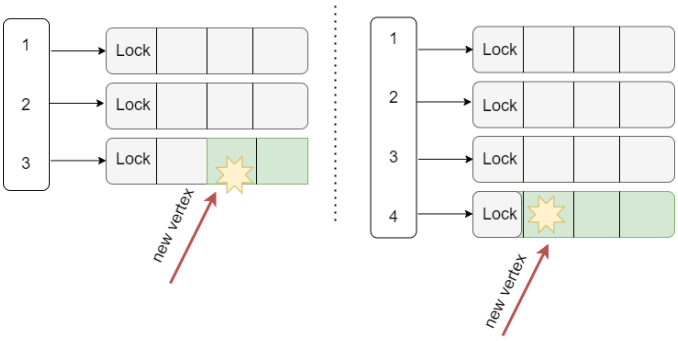


Fig. 3. Update Protocol for vertex insertions. If there is space in the segment, CoreGraph sets the length value of the first non-valid vertex from -1 to the degree of the new vertex (left), otherwise, it allocates a new segment and sets the new vertex (right).

Moreover, adding a new edge only necessitates reallocating the edge array of the source vertex because edges are stored in an array per vertex as shown in Fig. 4. When employing dynamic arrays, the amortized cost for updates is O(1) for insertions and O(deg V) for deletions. However, the memory footprint can be quite substantial as the reallocations leave unused space, when there are no updates. CoreGraph addresses this problem in two ways.

First, we use a smart memory manager can help keep track of unused space and perform a better strategy for pre-allocation while maintaining a memory footprint comparable to static graph data structures.

Second, CoreGraph exposes the reallocation factor as a configuration option for users to set it at compile time. The reason for this is that most system only test on specific use cases such as the SNAP [9] data sets and assuming they're analyzing power-law graphs, thus they set the reallocation factor to double the size by default. However, today users can take advantage of the power of machine learning [33] to estimate the factor by which we can grow our dynamic arrays, using statistical variables in a stream for specific types of graphs. For instance, we can learn the patterns of the creation of new relationships in social networks, e.g., a post that goes viral on social media might bring many followers in a short amount of time This means that for a celebrity account, there are higher chances of gaining more followers than a normal account, therefore we can choose a higher factor (x4, x5) to pre-allocate the edge lists while keeping an x2 factor for normal accounts to maintain a low memory footprint of our dynamic graph. Therefore, by allowing the configuration of the reallocation factor of edge arrays, CoreGraph adapts to different types of graphs and workloads, and minimizes the overall memory consumption of its dynamic graph data structure.

With respect to deletions, CoreGraph includes a mechanism for removing entities (both physically and virtually). The system reallocates the edge array by shifting the other edges to the left to physically delete an edge. In case of virtual deletion, where an edge or vertex is logically deleted from the CoreGraph, the deleted flag is set to indicate this.

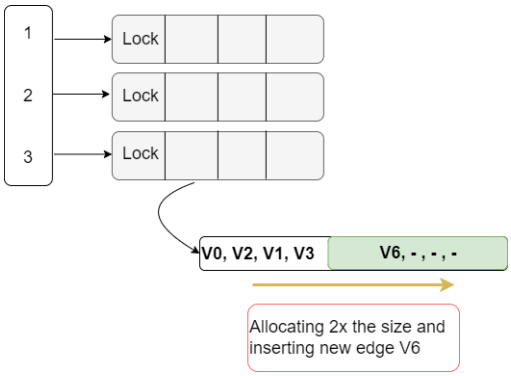


Fig. 4. Update protocol for edge insertion.

Algorithm 1 Batch Update Protocol in CoreGraph for Vertex and Edge Insertions

Input: Set of new edges E_{new}
Output: Updated CoreGraph with new edges
//Step 1: Grouping and Conversion
for all edge $e \in E_{new}$ **do**
  Group e by its source vertex $src(e)$
  Convert source and destination user keys to internal keys
end for
Insert new vertices in CoreGraph and assign new internal IDs (Sequential)
//Step 2: Sorting and Insertion
for all source vertex v in parallel **do**
  Sort new edges originating from v
  Insert sorted edges into direct and reverse maps
end for
//Step 3: Final Sorting and Reallocation
for all modified segment s in parallel **do**
  Merge old edges and new edges for each source vertex in s
  Sort the final edge arrays
  Reallocate edge properties according to the new order of edges
end for

Last but not least, CoreGraph uses a batch insertion protocol to insert vertices and edges in parallel, greatly enhancing the update's performance over a singular update. CoreGraph uses a bulk update mode, wherein updates and analytical procedures are executed sequentially rather than simultaneously. Using a fast in-place update protocol, CoreGraph is resilient enough to handle frequent small updates. Algorithm 1 shows the steps to achieve the batch insertions of vertices and edges in CoreGraph.

C. Synchronization Library

CoreGraph as a customizable graph storage and mutation framework, offers different configurations for synchronization to ensure consistency when running parallel graph analytics or parallel graph mutations. CoreGraph allows the user to configure the synchronization mechanism by offering different synchronization primitives to cater for different types of workloads and graph characteristics. CoreGraph allows locking at the segment level, however this may lead to high contention

when updates are targeting the same segment can cause a slowdown , such as multiple threads inserting edges to the same vertex in high skew graphs.

To remediate to this problem, CoreGraph allows two main configurations for low and high contention workloads. This is done by 1) using adaptable locking mechanism to switch to different lock algorithm depending on the contention level and 2) using HTM for high contention workloads to speed up the parallel execution without incurring extra memory overhead.

First, CoreGraph integrates the library GLS [34] for two reasons. It offers adaptiveness, which is designed to switch the synchronization depending on the workload, allowing it to adapt to different types of workloads and therefore give the best performance in most cases. Moreover, in debug mode it gives us information about the level of contention thanks to its built-in profiler lock (as shown in Fig. 5) which allows us to switch to different synchronization mechanism.

```
Prof Lock + RTM
Threads | Stats
1   |   [lock stats] avg spin: 0          | avg queue: 0.00  | avg lat: 24          | n_acq: 5649 @ (0x2aaaacc81298)m
2   |   [lock stats] avg spin: 16         | avg queue: 0.04  | avg lat: 72          | n_acq: 137324 @ (0x2aaaacc81298)m
4   |   [lock stats] avg spin: 47         | avg queue: 0.22  | avg lat: 560         | n_acq: 677173 @ (0x2aaaacc84298)m
6   |   [lock stats] avg spin: 94         | avg queue: 0.92  | avg lat: 1650        | n_acq: 1483192 @ (0x2aaaacc83298)m
8   |   [lock stats] avg spin: 227        | avg queue: 2.85  | avg lat: 4887        | n_acq: 1935047 @ (0x2aaaacc83298)m
12  |    [lock stats] avg spin: 707       | avg queue: 7.86  | avg lat: 15899       | n_acq: 2002555 @ (0x2aaaacc84298)m
24  |    [lock stats] avg spin: 2022      | avg queue: 20.12 | avg lat: 48325       | n_acq: 2003008 @ (0x2aaaacc8f298)m
```

Fig. 5. Real-time monitoring of contention using profiler lock in GLS and HTM. The average queue allows to quantify the contention level.

CoreGraph also allows swapping synchronization primitives automatically depending on the level of conflict: it prioritizes queue-based locks for high contention workloads such as edge insertions on vertices in the same segment and spinlocks like ticket locks for low contention workloads.

Furthermore, CoreGraph offers high parallel performance through the use of Hardware Transactional Memory (HTM) in Intel Restricted Transactional Memory (RTM) as a synchronization mechanism to simulate fine-grain locking at the vertex level without incurring extra memory cost. Each segment uses a single lock as a backup locking mechanism in case of aborts, i.e., lock elision.

We run an experiment to compare the performance of CoreGraph with HTM compared to fine-grained synchronization. We implemented of variant of CoreGraph with fine-grained synchronization where instead of storing one lock for individual segments, we store one lock per individual vertex. Update times for several segments and multithreaded scalability are shown in Fig. 6. The NUM_V_SEGMENT is modified to store the graph in different number of segments, and we insert 100K edges using a random uniform distributions on the final segments.

Fig. 6 demonstrates that even with 24 threads, the quickest solution is fine-grain locking, made possible by OpenMP scheduling. When utilizing several threads, CoreGraph with coarse locks stops scaling, and performance degrades as segment sizes increase. This is due to the significant conflict among threads updating adjacent segment vertices. On the other hand, HTM performs much better when there are many conflicts created as we use only two segments to store a whole graph.

Finally, the memory requirements of several locking mechanisms are shown in Table II. The cost of implementing HTM

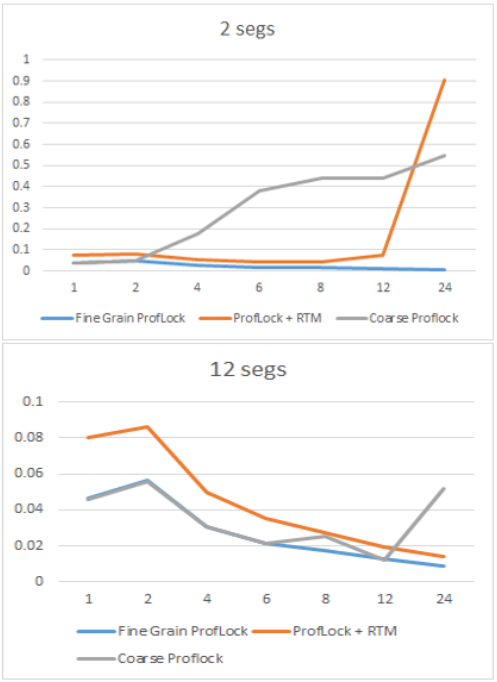


Fig. 6. Update times for several segments and multithreaded scalability. X-axis: number of threads, Y-axis: time (ms).

is comparable to that of a coarse lock since it does not need additional memory. However, fine-grained locking requires much more storage space since it adds S = sizeof(lock) * num_of_vertices, to the overall size of the structure.

Users may opt for HTM for suitable cases such to take most advantage of its performance, namely in cases where there are many conflicts. Moreover, it does not cause an increase in memory consumption.

D. Memory Allocation

Rather than using standard memory allocation functions for storing and updating large skewed graphs in-memory, which would be highly inefficient, CoreGraph optimizes this process with two main components : 1) an internal memory pool that pre-allocates the memory for loading the adjacency lists in a contiguous space and 2) a configurable allocation strategy using three state-of-the-art dynamic memory allocators, namely Glibc, Jemalloc [35] and TCMalloc [36]

First, we implement a smart loading procedure to optimize the physical memory layout of CoreGraph and therefore CoreGraph's analytic performance. Here, we describe the loading protocol using the memory manager in CoreGraph.

At the start of a graph loading process, memory of size S = sizeof(edge) * number_of_edges, is allocated. Then, we utilize that area to store the entire graph's edges in a continuous space within the same primary block. While we still keep track of pointers to edge lists for each vertex, the smart loader ensures that the edges are first stored in contiguous space. The memory compression made possible by this optimization brings us very near within 10% to the CSR in terms of read performance.

TABLE II. CONFIGURATIONS PROVIDED IN COREGRAPH

Configuration	Implementation	Impact
Reallocation Factor	Resizing the edge array when inserting new edges based on factor in config.	Low memory footprint, less memory allocations
Enabling Vertex/Edge Properties	Enabling storage for extra pointers for vertex/edge properties	Lower memory footprint, high cache performance
Synchronization Mechanism	Spinlock, Ticket/Profiler Lock, Intel HTM	Less contentions, high parallel performance
Memory Allocation	Glibc, Jemalloc, TCMalloc	Less fragmentation, high parallel performance

Secondly, CoreGraph employs efficient dynamic memory allocators, which allow to reduce memory fragmentation that occurs when memory is constantly being reallocated and so has chunks of unused space, or fragments, left over. In fact, in addition to the standard Glibc memory allocator, users can configure CoreGraph to use two state-of-the-art libraries for dynamic memory allocation: Jemalloc and TcMalloc. To demonstrate the performance gain from using these libraries, we performed a sensitivity analysis on CoreGraph with the three libraries for memory allocation [12] by running the same workload on three different allocator configurations of CoreGraph. As expected, Jemalloc and TCMalloc allow for a better performance than Glibc's performance for edge insertions. This is due to the parallel and thread-safe malloc implementation in both Jemalloc and TCMalloc, which gives better scalability when using multiple threads as shown in the plots.

IV. EVALUATION

We evaluated our framework on popular reference benchmarks [12] using readily available real-world and synthetic data sets that are popular in the graph research community, using two internal and external benchmark suites, Green-Marl [32] and GFE [37]. We evaluate the performance using real-world graph data sets [38], namely Twitter (1.4 billion edges), LiveJournal (68 million edges), and synthetic dataset [39] such as Uniform-24 (260 million edges), and Graph500-22 (69 million edges). We ran the experiments on a two-socket, 36-core machine with 384 GB of RAM. Its two 2.30 Ghz Intel Xeon E5-2699 v3 CPUs have 18 cores (36 hardware threads) and 32 KB, 256 KB and 46 MB L1, L2, and LLC caches, respectively.

CoreGraph outperforms the state of the art systems, reaching an update throughput of up to 24 Meps (million edges per second) compared to STINGER [18] (10 Meps), Teseo [25] (2.5 Meps) and GraphOne [23] (2 Meps). It also performs within 10% of CSR in graph analytics, while having a moderate memory overhead of 33% compared with CSR.

We highlighted the performance gains from using our framework in terms of high read performance, high update throughput while still maintaining a low memory footprint.

We demonstrated that CoreGraph offers high degree of configurability and we summarize these findings in Table II.

V. COREGRAPH FOR TRAFFIC DATA MANAGEMENT

In this section, we present our case study. We deemed important to test our solution on a more domain-specific dataset that represents real-world use case and can be exploited to extract insights for researchers and industry practitioners. The use case focuses on the traffic data management, which is critical for improving urban and highway traffic conditions to optimize traffic flow, detect congestion, and enhance overall transportation efficiency.

In this context, we extended the framework DIKCC, developed by our research team [40] [41] for traffic data management. The framework extracts knowledge from heterogeneous data sources and construct a knowledge graph characterized by high volume (300M edges and 4M edges) and high frequency of updates (10M updates/s).

Building on this foundational work, we integrate our framework CoreGraph into DIKCC framework, to provide high performance storage necessary in a context of high volume data and constantly evolving graphs. Our framework ensures efficient updates which are crucial in environements with real-time data stream. It also enables advanced analytic capabilities to address complex issues in traffic management. Moreover, CoreGraph outperforms graph databases or other state-of-the-art systems, that offer resource-consuming features that are not relevant to the traffic data discovery use case studied by our research team.

We utilized the same dataset as in the previous study [40], which consists of continuous views of three tables: Event, Route, and Intersection. The result of the knowledge construction from this data through is a property graph where the nodes are the intersections and the edges are the roads that link these intersections Fig. 7 shows the resulting graph.

We use the created mapping from relational-to-graph model, and we pre-process the datasets by converting them from CSV format as generated by the SQL queries, into file format that is supported by our framework, i.e, adjacency files, which does not require any additional schema definition and is consumed out of the box. Using this graph, our framework allows to recommend the optimal route to users to reach their destination by executing path traversal algorithms, which provide information about the state of the roads.

The updates to the graph are triggered from changes in the relational data by the creation of a new congestion event, that we process by updating the property value of edge of the corresponding road. Our framework supports topological mutation in addition to the updates to property values. Therefore, we extend the update workloads to the ones triggered by the creation of new roads between intersections, which we process as the insertion of new nodes and new edges. As for the new incoming updates, CoreGraph consumes continuous views to get the updates from the database as logs of edge lists, and performs the updates on the congestion status of roads (i.e. the edge property values) using our parallel batch updates API, and using our single update API for the insertions of new intersections and new roads (i.e. topological updates) since they are infrequent.

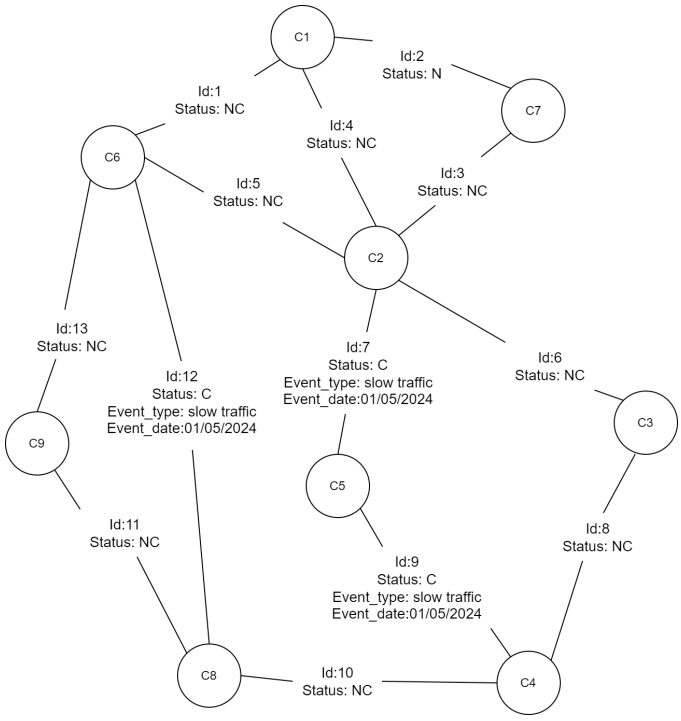


Fig. 7. Knowledge Graph for Traffic [40]. CoreGraph finds the shorted not congested routes by running BFS and setting the property values of edges accordingly.

We write a script to load the graph and measure the memory consumption. We find that the in-memory size is about 12 GB, which in principal can be supported on the same machine where the relational database resides.

After loading the graph, we run the path traversal algorithm namely BFS to find the shortest non-congested route from source to destination as depicted in Fig. 7. We then compare the output results to a validation data set for the top 10 nodes. After validating the correctness of our results, we measure the performance of the analytic workload which takes 2s to traverse the graph 300M of edges, which introduces minimal overhead to the end-to-end value chain process.

As for updates, since we store the adjacency of vertices in sorted order by id of intersections, we are able to locate the edge defined for the road by doing a binary search over the neighbour list using the identifier of the destination node which is the intersection. We use the index in that list to update the property value and set the congested state of the road. Moreover, for our extension to support high throughput topological updates, we use the prediction ML model from DIKCC to set the reallocation factor of the edge arrays, by predicting how much space we need to allocate for newly added intersections giving us an average throughput of 24M updates/sec.

We showed that our framework can be used as a vital component of the data discovery chain by demonstrating how our lightweight framework is easily integrated into a pre-existing framework with low overhead for traffic data management. We concluded that the performance gains from our integration are significantly higher than integration with a graph database system especially, and our framework can be easily compiled as a C++ library and shipped with the end-to-end framework without requiring extra configuration.

VI. Conclusion

In conclusion, this paper introduced CoreGraph, a highly configurable graph framework designed to address the challenges of efficient graph storage, mutation, and analysis. CoreGraph's storage and update protocol designs offer superior performance in terms of memory efficiency and update throughput compared to traditional graph data structures while maintaining read performance within 10% of CSR. The framework's integration of advanced synchronization mechanisms and configurable memory allocators enhances its adaptability to diverse workloads, making it suitable for real-time applications such as traffic data management. Our case study demonstrated that CoreGraph not only improves update throughput and analytic performance but also seamlessly integrates with existing frameworks, highlighting its potential as a critical component in modern data discovery and management systems. This study lays the groundwork for future research and development in optimizing graph processing systems, emphasizing the importance of balancing performance, memory consumption, and update efficiency in real-world scenarios.

References

[1] C. Peng, F. Xia, M. Naseriparsa, and F. Osborne, "Knowledge graphs: Opportunities and challenges," *Artificial Intelligence Review*, vol. 56, no. 11, pp. 13 071–13 102, 2023.

[2] C. Gao, Y. Zheng, N. Li, Y. Li, Y. Qin, J. Piao, Y. Quan, J. Chang, D. Jin, X. He *et al.*, "A survey of graph neural networks for recommender systems: Challenges, methods, and directions," *ACM Transactions on Recommender Systems*, vol. 1, no. 1, pp. 1–51, 2023.

[3] "Neo4j," http://www.neo4j.org.

[4] S. Firmli, V. Trigonakis, J.-P. Lozi, I. Psaroudakis, A. Weld, D. Chiadmi, S. Hong, and H. Chafi, "Csr++: A fast, scalable, update-friendly graph data structure," in *24th International Conference on Principles of Distributed Systems (OPODIS'20)*, 2020.

[5] S. Firmli and D. Chiadmi, "A review of engines for graph storage and mutations," in *EMENA-ISTL*, 2020.

[6] C. L. Staudt, A. Sazonovs, and H. Meyerhenke, "NetworKit: a tool suite for large-scale complex network analysis," *Network Science*, vol. 4, no. 4, p. 508–530, 2016.

[7] B. Wheatman and H. Xu, "Packed compressed sparse row: a dynamic graph representation," in *HPEC*, 2018.

[8] R. Cheng, E. Chen, J. Hong, A. Kyrola, Y. Miao, X. Weng, M. Wu, F. Yang, L. Zhou, and F. Zhao, "Kineograph: taking the pulse of a fast-changing and connected world," in *EuroSys*, 2012.

[9] K. Madduri and D. A. Bader, "Compact graph representations and parallel connectivity algorithms for massive dynamic network analysis," in *IPDPS*, 2009.

[10] A. Kyrola, G. Blelloch, and C. Guestrin, "GraphChi: large-scale graph computation on just a PC," in *OSDI*, 2012.

[11] P. Macko, V. J. Marathe, D. W. Margo, and M. I. Seltzer, "LLAMA: efficient graph analytics using large multiversioned arrays," in *ICDE*, 2015.

[12] S. Firmli and D. Chiadmi, "A scalable data structure for efficient graph analytics and in-place mutations," *Data*, vol. 8, no. 11, p. 166, 2023.

[13] "Boost adjacency list documentation," https://www.boost.org/doc/libs/1_67_0//libs/graph/doc/adjacency_list.html.

[14] R. Raman, O. van Rest, S. Hong, Z. Wu, H. Chafi, and J. Banerjee, "PGX.ISO: parallel and efficient in-memory engine for subgraph isomorphism," in *GRADES*, 2014.

[15] J. Shun and G. E. Blelloch, "Ligra: a lightweight graph processing framework for shared memory," in *Proceedings of the 18th ACM SIGPLAN symposium on Principles and practice of parallel programming*, 2013, pp. 135–146.

[16] Y. Low, J. E. Gonzalez, A. Kyrola, D. Bickson, C. E. Guestrin, and J. Hellerstein, "Graphlab: A new framework for parallel machine learning," *arXiv preprint arXiv:1408.2041*, 2014.

[17] N. P. Roth, V. Trigonakis, S. Hong, H. Chafi, A. Potter, B. Motik, and I. Horrocks, "PGX.D/Async: a scalable distributed graph pattern matching engine," in *GRADES*, 2017.

[18] D. Ediger, R. McColl, J. Riedy, and D. A. Bader, "Stinger: High performance data structure for streaming graphs," in *2012 IEEE Conference on High Performance Extreme Computing*. IEEE, 2012, pp. 1–5.

[19] F. Busato, O. Green, N. Bombieri, and D. A. Bader, "Hornet: an efficient data structure for dynamic sparse graphs and matrices on GPUs," in *HPEC*, 2018.

[20] B. Wheatman and H. Xu, *A Parallel Packed Memory Array to Store Dynamic Graphs*, pp. 31–45. [Online]. Available: https://epubs.siam.org/doi/abs/10.1137/1.9781611976472.3

[21] P. Macko, V. J. Marathe, D. W. Margo, and M. I. Seltzer, "Llama: Efficient graph analytics using large multiversioned arrays," in *2015 IEEE 31st International Conference on Data Engineering*. IEEE, 2015, pp. 363–374.

[22] K. Iwabuchi, K. Youssef, K. Velusamy, M. Gokhale, and R. Pearce, "Metall: A persistent memory allocator for data-centric analytics," *Parallel Computing*, vol. 111, p. 102905, 2022.

[23] P. Kumar and H. H. Huang, "GraphOne: a data store for real-time analytics on evolving graphs," *ACM Trans. Storage*, vol. 15, no. 4, 2020. [Online]. Available: https://doi.org/10.1145/3364180

[24] L. Dhulipala, G. E. Blelloch, and J. Shun, "Low-latency graph streaming using compressed purely-functional trees," in *Proceedings of the 40th ACM SIGPLAN Conference on Programming Language Design and Implementation*, ser. PLDI 2019. New York, NY, USA: Association for Computing Machinery, 2019, p. 918–934. [Online]. Available: https://doi.org/10.1145/3314221.3314598

[25] D. De Leo and P. Boncz, "Teseo and the analysis of structural dynamic graphs," *Proc. VLDB Endow.*, vol. 14, no. 6, p. 1053–1066, 2021. [Online]. Available: https://doi.org/10.14778/3447689.3447708

[26] B. Wheatman and R. Burns, "Streaming sparse graphs using efficient dynamic sets," in *2021 IEEE International Conference on Big Data (Big Data)*. IEEE, 2021, pp. 284–294.

[27] A. Kyrola, G. Blelloch, and C. Guestrin, "{GraphChi}:{Large-Scale} graph computation on just a {PC}," in *10th USENIX symposium on operating systems design and implementation (OSDI 12)*, 2012, pp. 31–46.

[28] M. A. Bender, E. D. Demaine, and M. Farach-Colton, "Cache-oblivious B-trees," *SIAM Journal on Computing*, vol. 35, no. 2, pp. 341–358, 2005. [Online]. Available: https://doi.org/10.1137/S0097539701389956

[29] M. Haubenschild, M. Then, S. Hong, and H. Chafi, "ASGraph: a mutable multi-versioned graph container with high analytical performance," in *GRADES*, 2016.

[30] M. Paradies, W. Lehner, and C. Bornhövd, "GRAPHITE: an extensible graph traversal framework for relational database management systems," in *SSDBM*, 2015.

[31] D. Durner, V. Leis, and T. Neumann, "On the impact of memory allocation on high-performance query processing," in *Proceedings of the 15th International Workshop on Data Management on New Hardware*, ser. DaMoN'19. New York, NY, USA: Association for Computing Machinery, 2019. [Online]. Available: https://doi.org/10.1145/3329785.3329918

[32] "Green-Marl code," https://github.com/stanford-ppl/Green-Marl.

[33] A. Bifet, R. Gavalda, G. Holmes, and B. Pfahringer, *Machine learning for data streams: with practical examples in MOA*. MIT press, 2023.

[34] J. Antić, G. Chatzopoulos, R. Guerraoui, and V. Trigonakis, "Locking made easy," in *Proceedings of the 17th International Middleware Conference*, 2016, pp. 1–14.

[35] J. Evans, "A scalable concurrent malloc(3) implementation for FreeBSD," in *BSDCan*, 2006.

[36] A. H. Hunter, C. Kennelly, D. Gove, P. Ranganathan, P. J. Turner, and T. J. Moseley, "Beyond malloc efficiency to fleet efficiency: a hugepage-aware memory allocator," in *OSDI*, 2021.

[37] "GFE driver code," https://github.com/cwida/gfe_driver.

[38] "SNAP (2014). Stanford Network Analysis Platform," http://snap.stanford.edu/snap.

[39] M. Capotă, T. Hegeman, A. Iosup, A. Prat-Pérez, O. Erling, and P. Boncz, "Graphalytics: A big data benchmark for graph-processing platforms," in *Proceedings of the GRADES'15*, 2015, pp. 1–6.

[40] S. Yousfi, D. Chiadmi, and M. Rhanoui, "Smart big data framework for insight discovery," *Journal of King Saud University-Computer and Information Sciences*, vol. 34, no. 10, pp. 9777–9792, 2022.

[41] S. Yousfi, M. Rhanoui, and D. Chiadmi, "Towards a generic multimodal architecture for batch and streaming big data integration," *arXiv preprint arXiv:2108.04343*, 2021.

Detecting Malware on Windows OS Using AI Classification of Extracted Behavioral Features from Images

Nooraldeen Alhamedi, Kang Dongshik
University of the Ryukyus, Japan, Okinawa, Japan

Abstract—In this research, using dynamic analysis ten critical features were extracted from malware samples operating in isolated virtual machines. These features included process ID, name, user, CPU usage, network connections, memory usage, and other pertinent parameters. The dataset comprised 50 malware samples and 11 benign programs, providing a data for training and testing the models. Initially, text-based classification methods were employed, utilizing feedforward neural networks (FNN) and recurrent neural networks (RNN). The FNN model achieved an accuracy rate of 56%, while the RNN model demonstrated better performance with an accuracy rate of 68%. These results highlight the potential of neural networks in analyzing and identifying malware based on behavioral patterns. To further explore AI's capabilities in malware detection, the extracted features were transformed into grayscale images. This transformation enabled the application of convolutional neural networks (CNN), which excel at capturing spatial patterns. Two CNN models were developed: a simple model and a more complex model. The simple CNN model, applied to the grayscale images, achieved an accuracy rate of 70.1%. The more complex CNN model, with multiple convolutional and fully connected layers, significantly improved performance, achieving an accuracy rate of 88%. The findings from this research underscore the importance of dynamic analysis. By leveraging both text and image-based classification methods, this study contributes to the development of more robust and accurate malware detection systems. It provides a comprehensive framework for future advancements in cybersecurity, emphasizing the critical role of dynamic analysis in identifying and mitigating threats.

Keywords—Malware analysis; dynamic analysis; image classification; malware behavior extraction; text

I. INTRODUCTION

Recently, the number, severity, sophistication of malware attacks, and cost of malware inflicts on the world economy have been increasing exponentially. Attacks with these kinds of software have a disastrous effect and cause considerable material damage to individuals, private companies, and governments' assets. Thus, malware should be detected before damaging the important assets in the company [1].

The primary motivation for this research stems from the need to enhance existing detection mechanisms to keep pace with the constantly changing threat landscape with traditional analysis methods, we aim to significantly improve the detection and classification accuracy of malicious software.

One of the key advantages of our approach is the combination of dynamic malware analysis with AI-driven techniques. This allows for a more comprehensive understanding of malware behavior. This hybrid approach not only improves detection rates but also enhances the ability to accurately classify and understand the nature of malware. There are two main techniques for analyzing malware - static and dynamic analysis. Static analysis examines the malware code without actually executing it. This by integrating advanced artificial intelligence (AI) techniques can provide information about suspicious functions, network activity, impacted files, etc. Dynamic analysis executes the malware code in an isolated environment to observe its runtime behavior. This provides insight into the full impact of the malware. A key benefit of static analysis is the ability to thoroughly inspect malware code using techniques like disassembly and decompilation to identify suspicious functions related to replication, propagation, payload activation, and more [2].Static techniques help reveal overall structure, dependencies, triggers for malicious events, and obfuscation attempts. However, lacking runtime behavior, static analysis cannot confirm real impact of suspected capabilities. Complex packing or encryption techniques also limit code inspection. Other hand, dynamic analysis provides direct observation of malware behavior in action by executing it and monitoring resulting activity.

Dynamic analysis confirms suspected functions based on static clues and captures full infection chains showing progression and end objectives of malware according to case studies by [3]. Dynamic monitoring of memory access, networks calls, system API usage, and more creates a comprehensive picture. Additionally, dynamic analysis is particularly effective in identifying and analyzing newly emerging malware strains. As it focuses on the runtime behavior, it is better equipped to handle polymorphic and metamorphic malware that may change its form to evade static analysis techniques. Leveraging AI models for the analysis of malware code or the study of malware behavior has significantly contributed to the detection of malware in recent years. Numerous AI models have been integrated into static or dynamic approaches to augment both the malware detection rate and feature extraction processes. Despite the notable progress in the field of AI, these models still face various challenges. This research will use many model of AI in order to detect malware.

A. Difficulties in Detecting Malware

Robust malware analysis faces numerous obstacles. The sheer volume of malware proliferating at a rapid pace presents a formidable challenge in comprehensively examining this ever-expanding threat landscape. Additionally, malware authors employ sophisticated obfuscation tactics, such as code interchange, amalgamation, register reassignment, null insertion, and subroutine reordering [3], purposefully designed to evade detection by anti-malware systems. Despite decades of development, these security solutions still exhibit high false positive rates, undermining their accuracy.

Moreover, certain malware strains possess the ability to identify virtualized environments, resulting in altered or ceased execution, hindering effective analysis. The evasion techniques employed by malware necessitate lengthy detection times, potentially ranging from minutes to hours depending on the specific malware variant, during which systems remain vulnerable to compromise. Furthermore, the ambiguity surrounding API calls, as both malicious and benign software may legitimately invoke common APIs, complicates the process of distinguishing malware based on API usage patterns.

These factors, including the immense scale, obfuscation methods, virtual environment detection capabilities, delayed identification timelines, and the dual usage of APIs, collectively contribute to the arduous nature of robust malware analysis, necessitating the development of advanced techniques to overcome these challenges effectively, juxtaposition of text classification and image classification in the analysis of extracted behavior. It underscores that a nuanced understanding of program nature, distinguishing between benign and malicious entities, can be achieved through thorough behavior analysis. The model primarily relies on the extraction of malware features. Within the developed script, two distinct observers play a crucial role. The first observer extracts the entirety of the process, encompassing its characteristics, as well as details related to internet connections. The second observer is tasked with monitoring any file creation specifically linked to the malware. While previous research has explored the use of AI in malware detection, there remains a need for a more robust and adaptive framework capable of effectively handling the diverse and rapidly evolving nature of modern malware threats. Our study aims to fill this gap by developing a hybrid model that can effectively detect and classify malicious software, even in the face of obfuscation techniques and emerging threats.

II. Related Work

Artificial Intelligence (AI) has emerged as a powerfully tool in this ongoing struggle to detect and classify malwares offering advanced capabilities in identifying and mitigating malware threats.

In study [4], the third paper analyzes different classical machine learning algorithms for malware detection - Random Forest, Support Vector Machine (SVM), grid search optimized SVM, and K-Nearest Neighbors (KNN). The goal is to validate the effectiveness of these models for detecting zero-day malware attacks. The dataset from Kaggle contained 19,611 PE files, with 14,599 malicious samples and 5,012 benign files with 77 numeric features. Three training/test splits were used. Various accuracy metrics were calculated: accuracy, F1-score, confusion matrix, precision, recall and Type I/II errors. Random Forest performed the best with 96% accuracy and 93% F1score, with low errors and fastest training time. Optimized SVM improved results significantly but slowed down execution. KNN also performed decently with simpler implementation. Analysis showed Random Forest has good prospects for realtime zero-day malware detection. The model can process 25,000 files per second. For deployment, more diverse input data covering different malware families is needed.

In study [5], the authors used convolutional neural networks (CNNs) for malware classification by visualizing malware programs as grayscale images. The images are generated from the bytecode of malware programs and classified using CNN architectures. They evaluate several well-known CNN models like AlexNet, ResNet, and VGG16 using transfer learning on a malware image dataset. They also propose a custom shallow CNN architecture that achieves 96% accuracy, but is faster to train than the other complex models. The customized CNN and transfer learning models are also tested as feature extractors, with the features fed into SVM and KNN classifiers. This achieves even better performance up to 99.4% accuracy. They set a new benchmark on the public BIG 2015 malware dataset. The proposed system combining CNN feature extraction + SVM classifier obtains state-of-the-art 99.4% accuracy in distinguishing between nine malware classes. Visualization and CNN-based classification is shown to be effective for malware detection. The approach is computationally efficient compared to static/dynamic analysis. Fusing different CNN model predictions can further improve performance.

In study [6], the authors used Support Vector Machines (SVMs) for malware analysis and classification. SVMs are supervised learning models that can analyze high-dimensional, sparse data and recognize patterns. The authors collect a heterogeneous malware dataset from a real threat database. The data has features like time, format, domain, IP address. They visualize the dataset using techniques like scatter plots and radius visualization to understand correlations and structure before classification. A SVM model with polynomial kernel is trained on the dataset to classify malware vs normal software. The model is validated using cross-validation, leave-one-out and random sampling. The SVM classifier achieves 93-95% accuracy, 97-98% sensitivity and 86-90% specificity on the malware dataset. Validation shows the model generalizes very well. The high performance highlights that SVMs can effectively classify heterogeneous malware data gathered from computer networks and security systems.

In study [7], the paper proposes a deep learning framework for malware visualization and classification using convolutional neural networks (CNNs). The key aspects are: Malware files are converted into three image types - grayscale, RGB color, and Markov images. Markov images help retain global statistics of malware bytes. A Gabor filter approach is used to extract textures and discriminative features from the malware images. Two CNN models are used for classification – a custom 13-layer CNN and a pretrained 71-layer Xception CNN fine-tuned for malware images. The framework is

evaluated on two public Windows malware image datasets, a custom Windows malware dataset, and a custom IoT malware dataset. Markov images provide the best results, with the fine-tuned Xception CNN achieving over 99% accuracy on multiple datasets. The computational efficiency is also better compared to prior works. The approach demonstrates effectiveness for real-time malware recognition and classification. The visualization and deep learning framework extracts features automatically without extensive feature engineering. The framework's resilience against adversarial attacks is also analyzed by adding noise to test images. Some drop in accuracy is noticed, indicating scope for improvement. The current landscape underscores the significance of AI models as powerful tools for the analysis, classification, and detection of malware. These models can seamlessly integrate with both static and dynamic analysis, yielding noteworthy results that underscore their pivotal role in shaping the future of this field.

Arabo et al. [8] analyzed CPU and RAM usage patterns as potential indicators for detecting ransomware processes. Their findings suggested that while not the primary factors, monitoring CPU and RAM could complement other behavioral characteristics in identifying malicious processes. Regarding CPU usage, they observed variations that showed potential for distinguishing ransomware activities. Specifically, for the ViraLock ransomware sample, the maximum CPU usage peaked at 25% [1]. Such CPU spikes could potentially signify the initiation of encryption or other malicious operations by the ransomware. As for RAM consumption, the study found that ransomware samples generally exhibited low and relatively stable memory usage patterns. In the case of ViraLock, the maximum RAM usage was only around 2% [1]. However, the authors noted that while low RAM usage alone may not be a definitive indicator, it could be considered in combination with other behavioral factors. The researchers highlighted that while CPU and RAM usage showed some differences between ransomware and benign processes, the most significant distinguishing factor was abnormally high disk read/write activity [1]. Nonetheless, incorporating CPU and RAM monitoring alongside disk usage analysis could potentially enhance the accuracy and robustness of ransomware detection systems based on process behavior analysis.

In study [9], the Integrated Malware Classification Framework (IMCFN) converts malware binaries into grayscale and color images for classification using CNNs. It outperforms models like VGG16 and ResNet50, particularly with color images, and proves effective on the IoT-Android dataset, highlighting its potential for improving malware detection in diverse environments.

In study [10], this research proposes a novel malware classification method using CNNs. Malware programs are converted into grayscale images and fed into various CNN architectures. Experimental results show impressive accuracy (up to 99.4%) using CNN-extracted features with SVM. The approach demonstrates robustness across different malware categories, offering a significant contribution to cybersecurity.

III. Comparative Analysis with Previous Works

Our approach employs relatively simple models, which positively impacts the time and resource efficiency of the learning process. This streamlined design not only enhances the speed of model training and inference but also optimizes resource usage, making our method more suitable for environments with limited computational power. This balance between simplicity and effectiveness further contributes to the practical applicability of our approach in real-world malware detection scenarios. Particularly, our use of CNNs for image-based classification demonstrates strong potential, achieving competitive accuracy with less reliance on extensive feature engineering. Additionally, our approach does not depend on information provided by external monitoring frameworks like Cuckoo. Instead, it leverages a native program developed using Python libraries, which enhances the reliability and accuracy of feature extraction. This self-contained method ensures more consistent and precise data collection, further strengthening the effectiveness of our malware detection process.

IV. Methodology

The current investigation is centered on the behavioral analysis within an isolated Windows environment in virtual machine for the purpose of detecting malware. To achieve this, a combination of Recurrent Neural Network (RNN) for text classification and Convolutional Neural Network (CNN) for image classification is employed to analyze the extracted data. Diverging from the methodologies outlined in previous studies [3], [6], and [7], the classification approach adopted here focuses on the inherent characteristics of the malware file itself. This is achieved through a comprehensive analysis of the malware binary file and, notably, by representing the malware file as an image utilizing various visualization techniques. In this research, the emphasis is on visualizing the malware's behavior and subsequently conducting analyses based on these extracted features. Visual representations and also analyze the extracted features as a text. The presented model offers a juxtaposition of text classification and image classification in the analysis of extracted behavior. It underscores that a nuanced understanding of program nature, distinguishing between benign and malicious entities, can be achieved through thorough behavior analysis. The model primarily relies on the extraction of malware features. Within the developed script, two distinct observers play a crucial role. The first observer extracts the entirety of the process, encompassing its characteristics, as well as details related to internet connections. The second observer is tasked with monitoring any file creation specifically linked to the malware. The experimental framework involves the extraction of 10 distinct features through the monitoring of behaviors within an isolated Virtual Machine. Python libraries such as psutil, subprocess, wmi, watchdog, time, json, and os were employed to develop functions responsible for observing malware behavior and subsequently extracting pertinent information to a JSON file. The extracted features encompassed critical aspects such as process ID, process name, username, CPU percentage. The modules for this research were developed using TensorFlow and Keras, leveraging the Sequential model architecture. These tools enabled efficient tools enable construction and training of neural networks for malware detection, facilitating both text-based and image-based classification with enhanced accuracy through deep learning techniques. Fig. 1 illustrates the flow chart of the methodology of this research.

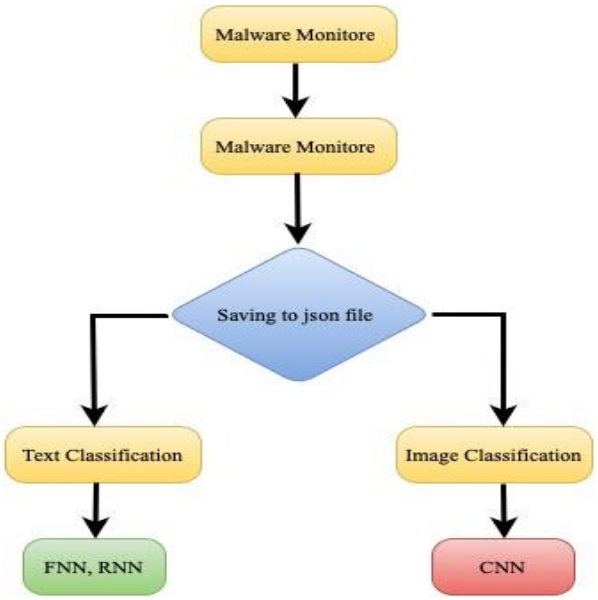


Fig. 1. Provides a visual representation of the proposed idea.


```
0:
   label:          0
   pid:            48872
   name:           "4f97a7f893939680bf36ccc03af19cc2d9ae3e4c7696fefc79ff5750ace15bae.exe"
   username:       "WINDOWS-10\\vboxuser"
   cpu_usage:      "none"
   connections:    "[pconn(fd=-1, family=<AddressFamily.AF_INET: 2>, type=<SocketKind.SOCK_STREAM: 1>,
                   laddr=addr(ip='10.0.2.15', port=50603), raddr=addr(ip='34.117.59.81', port=443),
                   status='ESTABLISHED'), pconn(fd=-1, family=<AddressFamily.AF_INET: 2>,
                   type=<SocketKind.SOCK_STREAM: 1>, laddr=addr(ip='10.0.2.15', port=50602),
                   raddr=addr(ip='194.169.175.113', port=50500), status='ESTABLISHED')]"
   parent:         "none"
   child:          "[{'ExecutablePath
                   \\r\\r'}, {'C:\\\\Users\\\\vboxuser\\\\Desktop\\\\mal-
                   DB\\\\4f97a7f893939680bf36ccc03af19cc2d9ae3e4c7696fefc79ff5750ace15bae.exe'}]"
   execution:      "none"
   filecreated:
      0:           '{"file_path": "C:\\\\Users\\\\vboxuser\\\\AppData\\\\Local\\\\Microsoft\\\\Edge\\\\User
                   Data\\\\Cookies"}\n{"file_path": "C:\\\\Users\\\\vboxuser\\\\PycharmProjects
                   \\\\pythonProject\\\\venv\\\\Scripts\\\\mal-file_created00"}\n'
```


Fig. 2. Sample of JSON file content connections details, parent process, child process, execution path, and created files.

Following the extraction of these features, the gathered information is stored in a JSON file for further next step. Fig. 2 shows a sample of JSON file.

A. Text Analysis

The analytical process for the extracted features unfolded across two phases. Initially, the data underwent textual analysis, leveraging a simple feedforward neural network (FNN) model designed for binary classification using the Keras library to create a fully connected dense layer with 128 nodes.

The output layer has 1 node and uses 'sigmoid' activation for binary classification.

Subsequently, a recurrent neural network (RNN) model was employed to classify the same textual data, creates an embedding layer that transforms integer word indices to dense word vector representations.

B. Image Analysis

By transforming data into images, researchers can leverage the vast body of knowledge and advancements in image processing techniques, readily applicable to the analysis of the transformed data. This data-to-image transformation unlocks the power of CNNs for a wider range of analysis tasks, promoting deeper insights into complex datasets. So this research implement the power of CNN alongside with the behavior analysis Subsequent to the behavioral analysis, the extracted features underwent further evaluation through an image classification paradigm. A dedicated function was developed to transform these feature data into grayscale

images. This transformative process involved the removal of associated labels, conversion of the data into binary numerical representations, subsequent transformation of these binary values into hexadecimal equivalents, and, finally, depiction of these hexadecimal values onto a 30*30 grayscale canvas.

The 30x30 size was empirically determined to balance information preservation and computational efficiency. Representing features as images enabled the utilization of convolutional neural networks (CNNs), which excel at capturing spatial patterns. The extracted features underwent further evaluation through an image classification paradigm. This visual representation approach offered several key advantages. Firstly, it enabled leveraging powerful deep learning techniques like convolutional neural networks, adept at capturing spatial patterns invaluable for malware characterization. Secondly, transforming features into images facilitated uncovering intrinsic relationships and patterns obfuscated in the original data's raw representation. Thirdly, the image domain allowed seamless integration of transfer learning and pre-trained models, expediting the analysis process. Lastly, the visually interpretable nature of images could provide insights into the discriminative characteristics learned by the models, aiding explain ability. By combining dynamic monitoring with visual analytics, this multi-pronged approach offered a potent framework for comprehensive malware analysis and classification.

The dataset employed for experimentation comprised 50 instances of .EXE malware sourced from diverse families, obtained from the Malware Bazaar database, a freely accessible online repository. Additionally, 11 benign programs were included for comparative analysis. The monitoring process lasted three seconds for every malware instance, during which the monitoring code ran in the background, observing the processes and file creation activities of the malware. After the monitoring period, the code produced a JSON file containing the captured information. The dataset has been divided into 40 malware behavior and 6 benign program behavior for the training and 10 malware behavior and 5 benign program behavior for testing Fig. 3 provides visual representation of the converting text to image. Fig. 4 is a sample of obtained image.

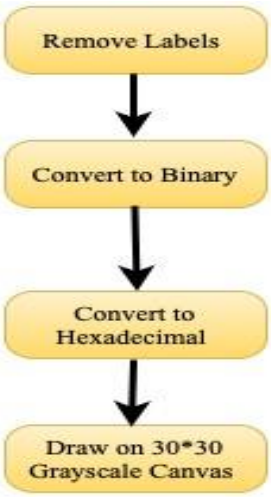


Fig. 3. Provides visual representation of the converting text to image.

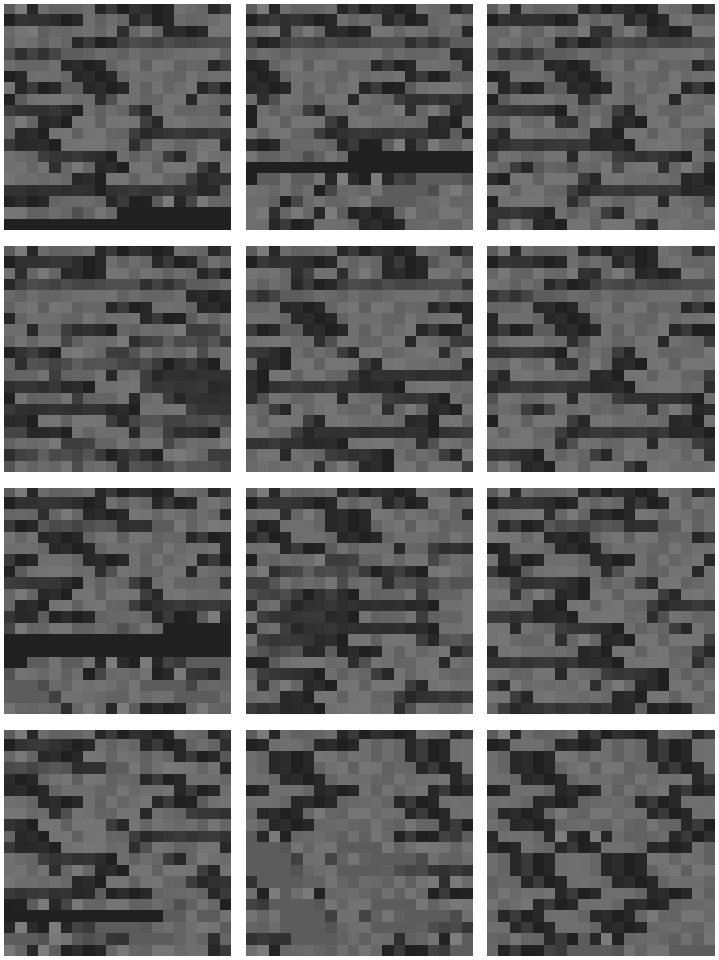


Fig. 4. Provides a sample representation of the resultant images, offering a glimpse into their visual characteristics.

V. The Experiment

A. Text Analysis

The described FNN model exhibited an accuracy rate of 56% with a corresponding loss rate of 0.78.

For the RNN model: It takes the vocabulary size equal to 32 and output dimensionality as arguments. Also LSTM layer models the sequential nature and long-range context of text. The output dense layers act as classifiers on top of LSTM representations. The model is compiled with binary cross entropy loss, adam optimizer and accuracy metric.

With epoch 100, yielding an improved accuracy rate of 68% with a reduced loss rate of 0.67.

B. Image Analysis

Convolutional Neural Networks (CNNs) have revolutionized image analysis due to their ability to extract intricate spatial features. However, their power can be extended to non-image data by transforming it into a suitable image representation. This approach offers several advantages:

CNNs excel at automatically learning relevant features from images, circumventing the need for manual feature engineering, a time-consuming and potentially error-prone step in traditional

analysis. Data transformation allows for the visualization of complex relationships between data points within the image domain. This empowers CNNs to identify subtle patterns that might be obscured in the raw data format.

The experiment has done using two suggested model. The first model is simple and the second model is more complex both models are based on CNN.

The simple model consists of:

- Conv2D layer: Performs 2D convolution with 32 filters and 3x3 kernel. Extracts spatial features from input image.
- MaxPool2D: Max pooling layer reduces dimensions to summarize the features detected by convolution layer.
- Flatten: Flattens the pooled feature map into a 1D vector to prepare for fully-connected layers.
- Dense layers: Fully-connected layers that act as classifier on top of the extracted features. 64 nodes in first dense layer.

Output layer contains single node with 'sigmoid' activation for binary classification. This model takes input images of shape (30, 30, 1) indicating 30x30 grayscale images. With epoch 30

Fig. 5 represent the structure of the first CNN model.

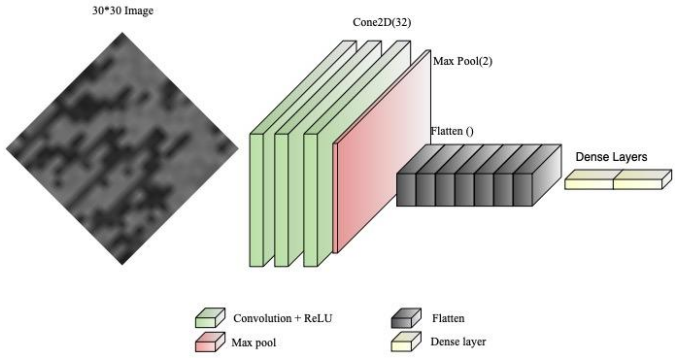


Fig. 5. The structure of the first model.

Using this simple Model over these grayscale pictures gives accuracy rate 70.1% with loss 0.67.

The second model also based on CNN with more complex architecture shown in Fig. 6.

The model then uses several convolutional layers (Conv2D) to extract features from the image. These layers apply filters (also called kernels) that slide across the image, detecting patterns and edges.

The first Conv2D layer has 256 filters, each of size 3x3. As the filter slides across the image, it performs element-wise multiplication between the filter weights and the corresponding pixel values in the image. The results are then summed and passed through an activation function (relu in this case) to introduce non-linearity. This process helps identify low-level features like edges, corners, and simple shapes.

The subsequent Conv2D layers follow the same principle but with a different number of filters (128 and 64 in this example). These layers extract progressively more complex features based on the lower-level features detected earlier.

MaxPooling2D layers are inserted after some convolutional layers. These layers downsample the feature maps by taking the maximum value within a specific window (2x2 in this example). This helps reduce the number of parameters and computational cost while potentially capturing the most important features.

The Dropout layer (commented out) randomly drops a certain percentage (25% in this example) of activations during training. This helps prevent the model from overfitting to the training data by forcing it to learn more robust features.

After the convolutional and pooling layers, the model uses a Flatten layer to convert the 3D feature maps into a 1D vector. This allows the fully-connected layers to process the extracted features. The model then uses several fully-connected layers (Dense) to classify the image. These layers work similarly to traditional neural networks, where each neuron receives input from all neurons in the previous layer, performs weighted sums, and applies an activation function.

The first three fully-connected layers (4096, 2048, and 1024 neurons) are responsible for learning complex, high-level representations based on the extracted features. The relu activation allows these layers to learn non-linear relationships between the features. The final dense layer has only one neuron with a sigmoid activation function. This neuron outputs a value between 0 and 1, representing the probability of the image belonging to a specific class.

As a summary for this model the convolutional layers act as feature detectors, extracting progressively more complex features from the input image. The pooling layers reduce the dimensionality of the data while retaining important information. The dropout layer helps prevent overfitting. The fully-connected layers learn high-level representations and produce the final classification probability.

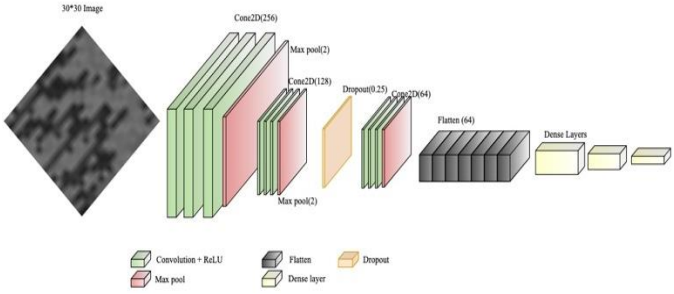


Fig. 6. The structure of the second CNN model.

Using this complex Model over these grayscale pictures gives accuracy rate 88% with loss 0.31.

VI. Results and Discussion

Comprehensive performance evaluation through bar charts illustrates accuracy and loss metrics for both text and image classification. The findings suggest that combining behavioral analysis with AI models, particularly in the image domain,

holds promise for effective malware detection. This multimodal approach provides a holistic understanding of malware behavior, potentially enhancing overall detection capabilities in the evolving cybersecurity landscape. The study contributes to advancing malware detection methodologies by leveraging the synergy between static and dynamic analyses, bolstered by AI integration, and offers insights into the promising potential of image-based classification for improved accuracy in identifying malicious behavior. The bar chart for accuracy and loss is given in Fig. 7.

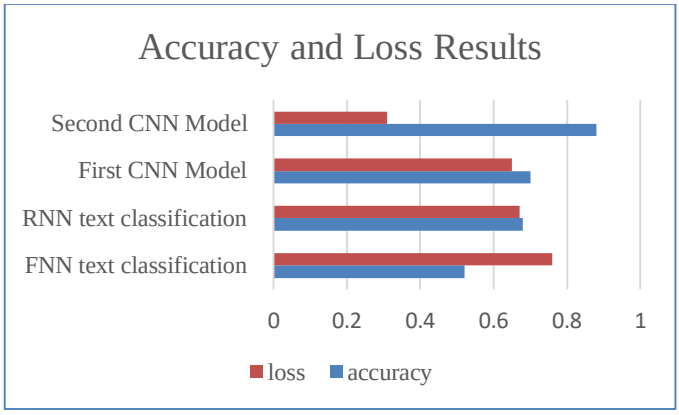


Fig. 7. Bar chart for accuracy and loss.

The Second Model with numerous convolutional and fully-connected layers grants high capacity for learning intricate features. While advantageous for complex datasets, it can lead to overfitting, particularly with limited training data. The model memorizes training data too well, hindering performance on unseen examples. Furthermore, training and running this deep model can be computationally expensive due to the high number of parameters. This translates to significant processing power and memory requirements, potentially limiting its use in resource-constrained environments. The results from the text classification and image classification shows that these methods of analyzing malware might be a good way to detect the malware using the extracted behavioral features.

Our primary objective was to enhance malware detection capabilities by combining dynamic analysis with AI techniques. The results of our image-based CNN model, achieving 88% accuracy, demonstrate significant progress towards this goal. Our image-based classification approach achieved 88% accuracy, which is comparable to the 96% accuracy reported by Sharma et al. [7] using a similar CNN-based method. However, our approach differs in that we focus on behavioral features rather than binary code visualization. The superior performance of our image-based CNN model (88% accuracy) compared to the RNN text classification model (68% accuracy) suggests that the spatial relationships captured in the image representation of behavioral features are particularly informative for malware detection. The superior performance of our image-based CNN model (88% accuracy) compared to the RNN text classification model (68% accuracy) suggests that the spatial relationships captured in the image representation of behavioral features are particularly informative for malware detection.

VII. Conclusion

This study successfully employs dynamic analysis within a virtual machine (VM) to extract crucial behavioral features from Windows malware. Integrating these features with advanced text and image classification models (RNN and CNN) shows promise for malware detection. Image classification, based on transformed feature data, achieves a superior accuracy of 88% compared to 68% in text classification. This multi-modal approach, combining behavioral analysis with AI models, provides a nuanced understanding of malware behavior.

One significant limitation of this study is the relatively small dataset used, consisting of only 50 malware samples and 11 benign programs. This limited sample size may not fully represent the vast diversity of malware in the wild, potentially affecting the generalizability of our results. Future work should involve a substantially larger dataset, encompassing a wider range of malware families and benign software to validate and potentially improve the model's performance. Another limitation is the brief 3-second monitoring period used for each malware instance. While this duration was chosen to balance efficiency and data collection, it may not capture the full range of behaviors exhibited by more sophisticated malware that employs delayed execution or other evasion techniques. Extended monitoring periods in future studies could provide more comprehensive behavioral data, potentially improving detection accuracy. To enhance the robustness and generalizability of our model.

We recommend several areas for future exploration. First, increasing the diversity and quantity of the training data by including a wider range of malware families and benign samples is crucial. Additionally, exploring additional features, such as registry changes, could provide valuable insights into malware behavior. Second, experimenting with different visualization techniques for image generation and testing more complex CNN architectures or pre-trained models with fine-tuning could further improve accuracy and efficiency. Third, addressing the threat of adversarial attacks is essential. Incorporating noise resilience mechanisms into the model can help mitigate the impact of such attacks and ensure the model's reliability in real-world scenarios. By pursuing these enhancements, we can contribute to advancing malware detection methodologies and ensuring their adaptability in the ever-evolving cybersecurity landscape.

References

[1] Aslan, Ö., & Samet, R. (2019). A comprehensive review on malware detection approaches. IEEE Access, Advance online publication. https://doi.org/10.1109/ACCESS.2019.2963724.

[2] Roundy, K.A. and Miller, B.P., 2013, August. Binary-code obfuscations in prevalent packer tools. In Proceedings of the 2013 ACM workshop on Software PROtection (pp. 3-14).M. Young, The Techincal Writers Handbook. Mill Valley, CA: University Science, 1989.

[3] Rossow, C., Dietrich, C. J., Grier, C., Kreibich, C., Paxson, V., Pohlmann, N., ... & van Steen, M. (2012). Prudent practices for designing malware experiments: Status quo and outlook. In 2012 IEEE Symposium on Security and Privacy (pp. 65-79). IEEE..

[4] Nafiiev, A., Kholodulkin, H., & Rodionov, A. (2022). Comparative analysis of machine learning methods for detecting malicious files. Algorithms and Methods of Cyber Attacks Prevention and Counteraction

[5] V. S. P. Davuluru, B. N. Narayanan and E. J. Balster, "Convolutional Neural Networks as Classification Tools and Feature Extractors for Distinguishing Malware Programs," 2019 IEEE National Aerospace and Electronics Conference (NAECON), 2019, pp. 273-277,

[6] M. Kruczkowski and E. Niewiadomska-Szynkiewicz, "Support Vector Machine for malware analysis and classification," 2014 IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT), 2014, pp. 415-420,

[7] Sharma, O., Sharma, A., & Kalia, A. (2022). Windows and IoT malware visualization and classification with deep CNN and Xception CNN using Markov images. Journal of Intelligent Information Systems. Advance online publication.

[8] Arabo, A., Dijoux, R., Poulain, T., & Chevalier, G. (2020). Detecting Ransomware Using Process Behavior Analysis. Procedia Computer Science, 168, 289-296.

[9] Vasan, D., Alazab, M., Wassan, S., et al. (2020). "IMCFN: IMage-based ClassiFication using Neural Networks for Malware Detection." Computer Networks, 171, 107138

[10] V. S. P. Davuluru, B. N. Narayanan and E. J. Balster, "Convolutional Neural Networks as Classification Tools and Feature Extractors for Distinguishing Malware Programs," 2019 IEEE National Aerospace and Electronics Conference (NAECON), 2019, pp. 273-277

The Impact of Virtual Collaboration Tools on 21st-Century Skills, Scientific Process Skills and Scientific Creativity in STEM

Nur Atiqah Jalaludin[1], Mohamad Hidir Mhd Salim[2]*, Mohamad Sattar Rasul[3],
Athirah Farhana Muhammad Amin[4], Mohd Aizuddin Saari[5]

STEM Enculturation Research Centre, Faculty of Education, Universiti Kebangsaan Malaysia, Bangi, Malaysia[1, 3, 5]
Institute of Visual Informatics, Universiti Kebangsaan Malaysia, Bangi, Malaysia[2]
ExxonMobil Malaysia, Kuala Lumpur, Malaysia[4]

***Abstract*—Virtual collaboration tools have become increasingly important in STEM education, especially after the COVID-19 pandemic. These tools offer many benefits, including developing 21st-century skills and fostering scientific process skills and scientific creativity. However, there are concerns regarding their effectiveness across different genders and regions. This study evaluates the impact of the ExxonMobil Young Engineers (EYE) program, which uses the Zoom application, on enhancing 21st-century skills, scientific process skills, and scientific creativity among secondary school students in Malaysia. The participants primarily consist of 520 secondary school students, with teachers acting as facilitators and professional engineers from ExxonMobil serving as instructors. A pre-test survey was conducted to assess students' initial skill levels. The program consisted of three phases: briefing, breakout room activities, and final reflections. After the program, a post-test survey was conducted to evaluate changes in student skills. Data analysis was analyzed using SPSS software by employing descriptive statistics, MANOVA with Wilks' lambda, one-way ANOVA, and partial eta squared to measure the program's impact and the influence of gender and regional factors. The results showed significant improvements in all three skill areas post-intervention: 21st-century skills, scientific process skills, and scientific creativity. Gender differences were significant for 21st-century skills, while regional differences significantly affected scientific process skills. The EYE program could enhance students' STEM-related skills using virtual collaboration tools like Zoom. However, regional and gender differences highlight the importance of adapting programs to address specific challenges and ensuring equitable opportunities for all students.**

***Keywords*—Virtual collaboration tools; 21st-century skills; scientific process skills; scientific creativity; STEM education**

I. Introduction

Virtual collaboration tools have become more common in STEM education, offering many benefits for developing 21st-century skills and scientific process and creativity skills. Online collaborative learning through small group discussions has been shown to promote knowledge co-construction and higher-order thinking skills in STEM subjects [1]. Virtual environments provide opportunities for direct interactions, helping students build knowledge and develop the mental processes involved in learning [2]. These tools facilitate collaborative learning and improve the effectiveness of learning experiences, especially when properly supported [3]. During the COVID-19 pandemic, the shift to virtual educational spaces raised concerns, especially in STEM education, where changes to lab work and online teaching practices are actively discussed [4]. Virtual simulation in teacher education has emerged as a method to provide opportunities for teachers to practice essential skills, such as parent-teacher collaboration, in a safe and controlled environment [5]. Even so, the effectiveness of these tools depends heavily on the technological infrastructure available to students and educators and their proficiency in using these technologies. There is a need for training programs to ensure that teachers can effectively integrate these tools into their teaching practices.

To effectively prepare students for success in the 21st century, educators must focus on teaching and look after 21st-century skills. These skills include collaborative problem-solving, critical thinking, creativity, communication, and digital literacy [6], [7], [8]. 21st-century skills include learning and innovation, information technology, and career skills, which are crucial for students' future [9], [10]. Teachers must integrate these skills into their teaching practices to ensure students have the necessary abilities to thrive in their future careers [11], [12]. By emphasizing these skills, educational institutions can also prepare students to meet the demands of the modern workforce requirement [13]. One major issue is the lack of school resources, leading to unequal access to the tools and training necessary for teachers and students. Schools in underprivileged areas may have issues with related infrastructure or funding to provide a conducive environment to develop these skills.

Scientific process skills are fundamental for conducting scientific research and advancing scientific knowledge [14]. These skills include a blend of physical, emotional, and thinking abilities used in scientific work [15]. Key process skills, such as identifying variables, forming hypotheses, and designing experiments, are vital for solving problems and creating new knowledge [16], [17]. Science process skills focus on knowledge transfer and problem-solving in real-life scenarios [18]. Developing these skills through activities like scientific questioning, experimental skills, and data interpretation enhances students' scientific literacy and attitudes toward science-based subjects [19], [20], [21]. Clear instruction helps learn science inquiry skills [22]. Metacognitive abilities support scientific process skills by helping students manage their

*Corresponding Author.

understanding and learning processes [23]. Again, the same issues will happen to schools with limited resources, which may struggle to provide hands-on experiences and experimental activities for developing these skills.

Scientific creativity includes understanding scientific phenomena, developing scientific knowledge creatively, solving complex science problems, enhancing product quality, and designing innovative scientific products [24]. These skills are connected with scientific process skills, where people who think and discuss, like scientists, show better scientific creativity [25]. Scientific creativity involves producing original and valuable outcomes with a specific purpose using available information [26]. It is a higher-order thinking skill essential for scientific thinking and differentiating between typical and exceptional scientific thinkers [27]. Enhancing scientific creativity often involves mastering creative thinking, which stimulates science process skills like observation, prediction, and hypothesis formation [28]. Scientific creativity relies on scientific knowledge and skills, combining a static structure with a developmental one [29]. Encouraging lifelong learning can improve individuals' scientific creativity skills [30]. One major challenge in nurturing scientific creativity is the need for professional development for teachers to help them recognize and nurture scientific creativity in their students. Many educators may lack the training or confidence to implement creative teaching strategies effectively.

Based on the issues related to these three skills. This study aims to evaluate the effectiveness of a virtual collaboration approach in improving 21st-century skills, scientific creativity, and scientific process skills in STEM education. Specifically, it will assess the impact of the ExxonMobil Young Engineers program, delivered through virtual collaboration tools, on enhancing these crucial skills among students. This paper comprises the following sections: a background study, a methodology section, results and discussion, and concludes with conclusions and future directions.

II. Background Study

In STEM education, nurturing 21st-century skills, scientific process skills, and scientific creativity skills has become a focal point for preparing students for the demands of the modern workforce. STEM education, which integrates Science, Technology, Engineering, and Mathematics, cultivates problem-solving abilities in real-world contexts and adopts essential 21st-century skills [31]. These skills include logical reasoning, problem-solving, collaboration, critical thinking, creativity, and communication, which are crucial for success in the current related job market [32]. STEM practices have been increasingly emphasized globally to enhance students' competencies in mathematics and engineering, aiming to equip them with the necessary skills to survive and thrive in today's society [33]. By incorporating project-based learning approaches and activities that involve scientific inquiry and engineering design processes, STEM education can effectively develop 21st-century practices and other related skills [34]. Even so, there is a need for continuous professional development for educators to keep pace with the rapidly evolving STEM fields. Teachers must be proficient in STEM content and pedagogical strategies that promote active learning and critical thinking. Without adequate training and support, educators may struggle to deliver STEM curricula that engage and inspire students effectively.

Virtual meeting tools enhance STEM education by promoting collaboration, engagement, and learning outcomes. Teachers highly value hands-on activities in STEM education [35], and platforms like Zoom have become crucial during the pandemic [36]. Digital tools have made STEM education more accessible [37], and online collaborative tools have been shown to improve learning outcomes and motivation [38]. Many tools are available and effective for virtual teaching [39], [40]. It can be used to enhance online presence whilst improving collaborative learning [41], [42]. The shift to remote instruction due to COVID-19 has led to exploring online resources for self-learning [43]. Universities are encouraged to cultivate self-regulated and peer-collaborative learning skills online [44]. However, the transition to virtual STEM education comes with challenges. One major issue is the digital divide, where students from low-income families may lack access to reliable internet connections and necessary devices, preventing them from fully participating in online learning. This gap worsens educational inequalities and limits the effectiveness of virtual meeting tools in improving STEM education for all students.

Various factors, including gender stereotypes, cultural norms, and personal interests influence gender disparities in STEM education. Studies have shown that gender differences in STEM careers can be traced back to early adolescence and are caused by societal expectations and decision-making processes [45]. Other than that, implicit gender-science stereotypes vary across countries and can contribute to gender differences in STEM achievement and representation [46]. Addressing gender differences in STEM fields is crucial not only for promoting gender equality but also for diversifying the workforce and creating a more competitive environment [47]. Efforts to bridge the gender gap in STEM education should consider the impact of cultural factors, traditional gender role beliefs, and the importance of providing equitable learning opportunities for all students [48]. Effective strategies in one country or community might not be applicable in another. Tailoring interventions to local needs and involving the community in developing and implementing programs can enhance their effectiveness.

Rural students often face challenges in STEM education due to limited access to resources and qualified teachers compared to urban areas [49]. Geographic differences in postsecondary STEM participation are influenced by students' demographics, aspirations, and academic preparation [50]. When given resources, rural educators can create strong systems for advanced STEM talent development [51]. Leadership practices, such as community relationships and empowering STEM teachers, contribute to STEM education success in rural schools [52]. To bridge the gap, it is crucial to engage diverse students in pursuing STEM fields [53]. Implementing STEM programs in rural schools, such as using the STEM Engineering Design Process, can produce positive outcomes [54]. This initiative should focus on inclusivity and diversity, ensuring that all students, including those from diverse urban backgrounds, have fair access and opportunities in STEM fields. This approach helps reduce the educational gap and encourages students to pursue STEM careers.

III. METHODOLOGY

The ExxonMobil Young Engineers Program utilizes the Zoom application as the medium for the virtual collaboration session, which involves 10 schools. The participants primarily consist of secondary school students, with teachers acting as facilitators and professional engineers from ExxonMobil serving as instructors, with an instructor-to-student ratio of 1:10. Each student is provided with STEM kits before the program starts. A pre-test survey was conducted before the program started to identify the students' initial levels of 21st-century skills, scientific process skills, and scientific creativity. The program has three phases: In Phase 1, instructors gather all participants in the main room to explain the program's structure and goals. In Phase 2, participants are divided into breakout rooms, one for each school, each with one instructor, up to five facilitators, and ten students. They conduct three educational modules, each lasting one hour, focusing on real jobs for oil and gas engineers and using a problem-based learning approach. The modules, adapted from energy4me®, are "Getting the Oil Out," "Core Sampling," and "Exploring Oil Seeps," which replicate real challenges in the oil and gas industry. In Phase 3, all participants return to the main room for a final discussion, where instructors and facilitators answer questions and provide final reflections on the program outcomes. This structured approach ensures comprehensive engagement and learning for all participants. After the program was completed, a post-test survey was distributed to evaluate the program's impact on the three main skills, which is the main objective of this study. Fig. 1 shows the overview of the EYE program structure and study phases.

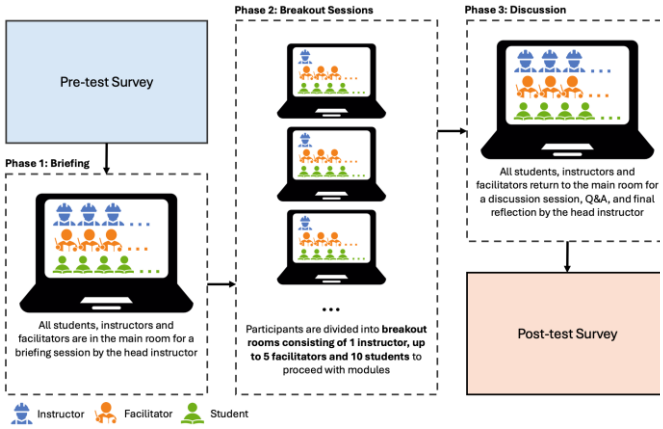


Fig. 1. Overview of the EYE program structure and study phases.

Fig. 2 shows the execution of the EYE program via Zoom. Instructors led the module through Zoom, while facilitators assisted students in following the instructors' instructions. Participants used STEM kits to accomplish the modules by following the instructions provided by the instructors.

A. Instrument

The data collection tool employed in this study was a questionnaire adapted from previous related studies, as shown in Table I. Minor modifications were made to customize it to achieve the research's specific objectives. This instrument was used to gather data to assess the program's impact on STEM literacy by conducting pre and post-surveys before and after the program. The questionnaire comprised two main sections: the first focused on gathering demographic information about the respondents, and the second focused on the study's core constructs. These constructs related to STEM skills were categorized.



Fig. 2. Execution of the EYE program via ZOOM (own sources).

TABLE I. QUESTIONNAIRE ITEMS ON 21ST-CENTURY SKILLS, SCIENTIFIC PROCESS SKILLS, AND SCIENTIFIC CREATIVITY

Construct	Items	Item Source
21st-century skill	I will be optimistic about completing a given activity through Zoom.	[55]
	I always think of different methods from different perspectives in solving a problem.	[56]
	I always come up with something new when studying science.	[55]
	I am always optimistic about performing tasks using Zoom.	[57]
	I always think critically and rationally to complete tasks.	[56]
Scientific process skill	Before starting an activity, I must plan, make hypotheses and identify problems.	[58]
	I get information from the facilitator and the internet to carry out activities	[59]
	I will record all work steps and processes, and evaluate the activities I produce	[58]
	I keep all data, calculations, and activity sketches as evidence of the work.	[58]
	I present and show the results of my activities to the facilitators and friends through Zoom	[59]
Scientific creativity	I like to give views and suggestions to facilitators and friends about the activities carried out through Zoom	[60]
	I like to discuss the design of my activities with other friends through Zoom	[61]
	I always think about the effects and consequences of the activities that will be carried out	[61]
	I like associating the activities with the latest elements of science and technology.	[56]
	I like to follow the example of activities given by the facilitator through Zoom	[60]

B. Analysis

In this study, gender and region are the dependent variables, while 21st-century skills, scientific process skills, and scientific creativity are independent variables. The objective is to evaluate the ExxonMobil Young Engineers (EYE) program's impact on these skills and to explore how gender and regional factors influence skill development. Data were collected through pre- and post-test surveys administered to secondary school students in Malaysia, using a 5-point Likert scale. SPSS software was used for analysis, including descriptive statistics to summarize mean values and standard deviations. Multivariate Analysis of Variance (MANOVA) tested the significance of skill differences between pre-and post-test scores and examined the effects of gender and region. One-way ANOVA assessed the significance of differences by gender and region for each skill, with Partial Eta Squared measuring the effect size. Levene's test was performed to assess the assumption of homogeneity of variances across groups. Levene's test results were insignificant ($p > 0.05$), suggesting that the variances are approximately equal across the groups for the dependent variables. This indicates that the assumption of homogeneity of variances is met, supporting the validity of proceeding with the MANOVA. Therefore, the MANOVA was conducted assuming that the equal variances condition is satisfied." Stratified random sampling was employed to select students from both rural and urban areas.

IV. Findings and Discussion

The table shows the demographic characteristics of the study's participants, giving information about their gender and where they live. The data shows that 196 participants were male, making up 37.7%. The majority, 324 participants, were female, making up 63.2%. 192 participants were from urban areas, accounting for 37.0%. A larger portion, 328 participants, were from rural areas, making up 63.0% of the total. This breakdown of participants by gender and location helps understand the study's sample, which can affect the generalizability of the research findings and any potential impacts of demographics on the study's results. Table II shows the demographic information of the study.

Table III shows the mean score values of the pre-test and post-test for 21st-century skills, scientific process skills, and scientific creativity. The data indicated increased scores in all three skill areas: 21st-century skills improved from a pre-test mean of 4.3304 to a post-test of 4.4423, scientific process skills from 4.2322 to 4.3495, and scientific creativity from 4.0635 to 4.2858. Statistical analyses confirmed these improvements, with significant p-values ($p < 0.05$) indicating the program's effectiveness.

TABLE II. Demographic Information

Demographic Characteristics	Number of Participants
Gender	
Male	196 (37.7%)
Female	324 (63.2%)
Region	
Urban areas	192 (37.0%)
Rural Areas	328 (63.0%)

These results suggest that the ExxonMobil Young Engineers (EYE) program significantly enhances students' STEM-related skills. The improvements in post-test scores across all three areas indicate that virtual collaboration tools, such as the Zoom application used in the EYE program, can effectively support skill development. Zoom's use as a virtual collaboration tool has several advantages that contribute to these positive outcomes. Zoom allows for interactive sessions, where students can engage in real-time discussions, collaborative projects, and hands-on activities, which are crucial for developing 21st-century and scientific process skills. The platform's features, such as breakout rooms, screen sharing, and real-time feedback, facilitate a dynamic and engaging learning environment that can adapt to different teaching styles and learning needs.

While the findings are promising, several limitations must be considered. The sample size, though adequate, may not represent the entire population of secondary school students in Malaysia. Potential biases could arise from self-reported data and using a single program for analysis. The assumptions of the statistical tests, such as the normality of data distribution, were checked but could still affect the results. Comparing these findings with previous research, it is evident that virtual collaboration tools are becoming increasingly important in STEM education. The greater improvement in scientific creativity among female students observed in this study highlights the potential for virtual collaboration tools like Zoom to provide an inclusive learning environment that supports all students. This requires further investigation to understand the underlying factors better and tailor the program to benefit both genders equally.

TABLE III. Mean Score Value of The Pre-Test and Post-Test For 21st Century Skills, Scientific Process Skills, and Scientific Creativity

Construct	Mean Value	
	Pre-test	Post-test
21st century skills	4.3304	4.4423
Scientific process skills	4.2322	4.3495
Scientific creativity	4.0635	4.2858

A. 21st Century Skills

The analysis presented in Table IV demonstrates significant improvements in 21st-century skills post-intervention, with Wilks' lambda = 0.982, F = 9.181, p = 0.003, and partial eta squared = 0.018. These results highlight the effectiveness of the ExxonMobil Young Engineers (EYE) program, which uses virtual collaboration tools, such as Zoom, to enhance students' competencies. The higher post-test scores indicate that integrating virtual collaboration tools effectively improves students' 21st-century skills, including critical thinking, problem-solving, collaboration, and digital literacy.

The analysis also found significant gender differences, with female students showing greater improvement in their 21st-century skills compared to male students (Wilks' lambda = 0.992, F = 3.883, p = 0.049, partial eta squared = 0.008). This suggests that virtual collaboration tools like Zoom may be particularly effective for female students, potentially due to the inclusive and flexible nature of virtual environments. These

platforms can accommodate different learning styles and preferences, which might help female students engage and excel more effectively. Further research should investigate the reasons behind these gender differences to tailor the EYE program more effectively for both genders.

Regional factors did not significantly affect skill development (Wilks' lambda = 0.999, F = 0.753, p = 0.386, partial eta squared = 0.001), indicating that the EYE program's effectiveness is consistent across different regions. This consistency suggests that virtual collaboration tools like the Zoom application can deliver high-quality education regardless of regional disparities [38]. The broad applicability of the EYE program across various geographical locations underscores the potential of virtual collaboration tools to provide equitable learning opportunities to students from diverse backgrounds.

The significant improvement in 21st-century skills emphasizes the potential of virtual collaboration tools as effective mediums for developing essential student competencies. Virtual collaboration tools like Zoom offer several advantages, including accessibility, flexibility, engagement, and collaboration [42]. These collaboration tools allow students to stay engaged with interactive content and real-time feedback and work together with peers and instructors through virtual platforms [40].

TABLE IV. MULTIVARIATE ANALYSIS OF 21ST CENTURY SKILLS, GENDER, AND REGION USING WILK'S LAMBDA

Effects	Wilks' Lambda Value	F	df1	df2	P	Partial Eta Squared
21st Century Skill	0.982	9.181b	1	504	0.003*	0.018
Gender	0.992	3.883b	1	504	0.049*	0.008
Region	0.999	.753b	1	504	0.386	0.001

B. Scientific Process Skills

The analysis in Table V shows a significant improvement in scientific process skills post-intervention, with Wilks' lambda = 0.964, F = 18.916, p = 0.000, and partial eta squared = 0.036. This indicates that the ExxonMobil Young Engineers (EYE) program, which uses virtual collaboration tools like the Zoom application, effectively enhanced students' scientific process skills. The significant increase in scores underscores the impact of virtual collaboration tools in providing a robust educational experience.

Gender differences were found to be insignificant (Wilks' lambda = 0.999, F = 0.273, p = 0.602, partial eta squared = 0.001), suggesting that both male and female students benefited equally from the program. This finding is important as it highlights the inclusive nature of virtual collaboration tools like Zoom, which can cater to diverse groups of students without gender bias. Zoom's flexibility and interactive features may create an equitable learning environment where all students can thrive [36].

However, regional differences significantly affected scientific process skills (Wilks' lambda = 0.992, F = 4.067, p = 0.044, partial eta squared = 0.008). This indicates that the program's effectiveness varied slightly across different regions. These regional variations suggest that while virtual collaboration tools like Zoom are generally effective, there may be differences in how they are implemented or accessed in various areas [49]. Some schools involved in the EYE program have issues with technology infrastructure that affect the program's flow. Other factors such as internet connectivity, availability of digital devices, and local educational practices, could also influence the effectiveness of the EYE program across regions [51].

These findings highlight the potential of virtual collaboration tools, such as Zoom, to enhance scientific process skills among secondary school students. The effectiveness of the EYE program across gender lines suggests that it is a valuable tool for promoting STEM education for all students. The regional differences indicate a need for tailored approaches to address specific challenges that may arise in different areas. Ensuring consistent access to resources and support across regions can help maximize the benefits of virtual collaboration tools.

TABLE V. MULTIVARIATE ANALYSIS OF SCIENTIFIC PROCESS SKILLS, GENDER, AND REGION USING WILK'S LAMBDA

Effects	Wilks' Lambda Value	F	df1	df2	P	Partial Eta Squared
Scientific Process Skill	0.964	18.916b	1	504	0.000*	0.036
Gender	0.999	0.273b b	1	504	0.602	0.001
Region	0.992	4.067b	1	504	0.044*	0.008

C. Scientific Creativity

The analysis in Table VI shows a significant improvement in scientific creativity post-intervention, with Wilks' lambda = 0.905, F = 52.836, p = 0.000, and partial eta squared = 0.095. This substantial effect size indicates that the ExxonMobil Young Engineers (EYE) program, which utilizes virtual collaboration tools such as Zoom, effectively enhances students' scientific creativity. The large improvement in scores underscores the impact of virtual collaboration tools in providing an engaging and stimulating educational experience.

Gender differences were found to be insignificant (Wilks' lambda = 0.992, F = 4.067, p = 0.804, partial eta squared = 0.000), suggesting that both male and female students benefited equally from the program. This finding is important as it highlights the inclusive nature of virtual collaboration tools like Zoom, which can cater to diverse groups of students without gender bias. The interactive features of Zoom, such as breakout rooms, real-time collaboration, and multimedia integration, may contribute to creating an equitable learning environment [38] where all students can develop their scientific creativity.

Regional differences also showed no significant effect on scientific creativity (Wilks' lambda = 0.999, F = 0.512, p = 0.474, partial eta squared = 0.001). This suggests that the EYE program's effectiveness in enhancing scientific creativity is consistent across different regions. The consistency of the program's impact across various geographical locations highlights the potential of virtual collaboration tools like Zoom to deliver high-quality education regardless of regional factors. This result indicates that the program can be broadly applied and

adaptable to various regional contexts, providing learning opportunities to students from diverse backgrounds.

The significant improvement in scientific creativity shows that virtual collaboration tools effectively develop important skills in students. Virtual tools like Zoom offer accessibility, flexibility, engagement, and collaboration. They allow students to access resources from any location, learn at their own pace, stay engaged with interactive content, and collaborate with peers and instructors, which is important in STEM education [35]. The EYE program effectively enhances scientific creativity using Zoom, demonstrating its potential to provide inclusive, high-quality education. These findings highlight the importance of integrating virtual collaboration tools in educational programs to prepare students for future challenges in a rapidly evolving digital world.

TABLE VI. MULTIVARIATE ANALYSIS OF SCIENTIFIC CREATIVITY, GENDER, AND REGION USING WILK'S LAMBDA

Effects	Wilks' Lambda Value	F	df1	df2	P	Partial Eta Squared
Scientific Creativity	0.905	52.836b	1	504	0.000*	0.095
Gender	0.992	4.067b	1	504	0.804	0.000
Region	0.999	0.512 b	1	504	0.474	0.001

V. CONCLUSION AND FUTURE IMPROVEMENT

The ExxonMobil Young Engineers (EYE) program has shown great promise in improving students' 21st-century skills, scientific process skills, and scientific creativity through virtual collaboration tools like Zoom. The program's structured approach, which includes interactive sessions, group projects, and real-time feedback, has effectively engaged students and helped them develop these skills. The overall improvements in all three skill areas highlight the program's success. Enhancing 21st-century skills shows virtual collaboration tools can improve critical thinking, problem-solving, collaboration, and digital literacy. Gender differences were noted, with female students showing greater improvement in their 21st-century skills than male students. This suggests that the inclusive and flexible nature of virtual environments, like those facilitated by Zoom, may be particularly effective for female students. Scientific process skills also showed significant improvement after the intervention. Both male and female students benefited equally, indicating that virtual collaboration tools can provide a fair learning experience. However, regional differences did affect scientific process skills, suggesting some variation in program effectiveness across different areas. The most improvement was in scientific creativity, with both male and female students benefiting equally, and the program's effectiveness was consistent across different regions. This indicates that virtual collaboration tools like Zoom have a broad and consistent impact on improving students' creative thinking and innovation skills. These findings highlight the importance of integrating virtual collaboration tools in educational programs to prepare students for the challenges of the modern workforce. The EYE program's success across gender and regional lines suggests that such initiatives can be scaled and adapted for broader educational contexts, providing high-quality, inclusive education.

While the EYE program has shown considerable success, several areas need enhancement to maximize its impact and ensure all students benefit equally. The regional differences observed in scientific process skills suggest a need for tailored approaches to address specific challenges in different areas. Ensuring consistent access to resources, such as reliable internet connectivity and digital devices, can help bridge these gaps. Additionally, training local educators to use virtual collaboration tools effectively can enhance the program's impact across all regions. The greater improvement observed among female students in 21st-century skills requires further investigation. Understanding the underlying factors contributing to these differences can help tailor the program to support both genders equally. Future research could also explore the elements of virtual collaboration environments suitable for female students to adapt to the program accordingly. Integrating more advanced technologies, such as augmented reality (AR) and virtual reality (VR), could also further enhance the learning experience. These technologies can provide immersive and interactive learning environments that stimulate students' creativity and critical thinking skills.

ACKNOWLEDGMENT

We would like to thank all participants, engineers from ExxonMobil, and teachers from the schools involved in this study. The work was supported by a university research grant PDE52.

REFERENCES

[1] L. Nungu, E. Mukama, and E. Nsabayezu, "Online collaborative learning and cognitive presence in mathematics and science education: case study of University of Rwanda, College of Education," Education and Information Technologies, vol. 28, no. 9, pp. 10865-10884, 2023. https://doi.org/10.1007/s10639-023-11607-w.

[2] L. Dieker, J. Rodríguez, B. Lignugaris, M. Hynes, and C. Hughes, "The potential of simulated environments in teacher education," Teacher Education and Special Education the Journal of the Teacher Education Division of the Council for Exceptional Children, vol. 37, no. 1, pp. 21-33, 2013. https://doi.org/10.1177/0888406413512683.

[3] C. Girvan and T. Savage, "Identifying an appropriate pedagogy for virtual worlds: a communal constructivism case study," Computers & Education, vol. 55, no. 1, pp. 342-349, 2010. https://doi.org/10.1016/j.compedu.2010.01.020.

[4] K. Turner, J. Adams, and S. Eaton, "Academic integrity, STEM education, and COVID-19: a call to action," Cultural Studies of Science Education, vol. 17, no. 2, pp. 331-339, 2022. https://doi.org/10.1007/s11422-021-10090-4.

[5] S. Luke and S. Vaughn, "Embedding virtual simulation into a course to teach parent–teacher collaboration skills," Intervention in School and Clinic, vol. 57, no. 3, pp. 182-188, 2021. https://doi.org/10.1177/10534512211014873.

[6] D. Rizaldi, E. Nurhayati, and Z. Fatimah, "The correlation of digital literation and STEM integration to improve Indonesian students' skills in the 21st century," International Journal of Asian Education, vol. 1, no. 2, pp. 73-80, 2020. https://doi.org/10.46966/ijae.v1i2.36.

[7] E. Hendarwati, L. Nurlaela, and B. Bachri, "Collaborative problem-solving based on mobile multimedia," International Journal of Interactive Mobile Technologies (IJIM), vol. 15, no. 13, p. 16, 2021. https://doi.org/10.3991/ijim.v15i13.23765.

[8] J. Varghese and M. Musthafa, "Why the optimism misses? an analysis on the gaps and lags of teachers' perceptions of 21st-century skills," Shanlax International Journal of Education, vol. 10, no. 1, pp. 68-75, 2021. https://doi.org/10.34293/education.v10i1.4322.

[9] K. Motallebzadeh, F. Ahmadi, and M. Hosseinnia, "Relationship between 21st-century skills, speaking and writing skills: a structural equation

modeling approach," International Journal of Instruction, vol. 11, no. 3, pp. 265-276, 2018. https://doi.org/10.12973/iji.2018.11319a.

[10] G. Widayana, "The influence of technical skills and 21st-century skills on the job readiness of vocational students," 2023. https://doi.org/10.4108/eai.6-10-2022.2327433.

[11] B. Dhakal, "Pedagogical use of 21st-century skills in Nepal," Chintan-Dhara, pp. 1-13, 2023. https://doi.org/10.3126/cd.v17i01.53252.

[12] Z. Hidayatullah, I. Wilujeng, N. Nurhasanah, T. Gusemanto, and M. Makhrus, "Synthesis of the 21st-century skills (4c) based physics education research in Indonesia," JIPF (Jurnal Ilmu Pendidikan Fisika), vol. 6, no. 1, p. 88, 2021. https://doi.org/10.26737/jipf.v6i1.1889.

[13] R. Aporbo, "Effects of 21st century skills to research writing abilities of senior high school students," International Journal of Research Publications, vol. 111, no. 1, 2022. https://doi.org/10.47119/ijrp10011111020224009.

[14] S. Mengistie, "Analysis of the infusion of science process skill contents into the Ethiopian grade nine biology textbook: a content analysis," IPS J. Edu, vol. 1, no. 1, pp. 1-12, 2023. https://doi.org/10.54117/ije.v1i1.7.

[15] [15] N. Husna, A. Halim, E. Evendi, M. Syukri, S. Abdulmadjid, E. Elisa, and I. Khaldun, "Impact of science process skills on scientific literacy," Jurnal Penelitian Pendidikan IPA, vol. 8, no. 4, pp. 2123-2129, 2022. https://doi.org/10.29303/jppipa.v8i4.1887.

[16] T. Andini, S. Hidayat, E. Fadillah, and T. Permana, "Scientific process skills: preliminary study towards senior high school student in Palembang," Jpbi (Jurnal Pendidikan Biologi Indonesia), vol. 4, no. 3, pp. 243-250, 2018. https://doi.org/10.22219/jpbi.v4i3.6784.

[17] J. Juhji and P. Nuangchalerm, "Interaction between science process skills and scientific attitudes of students towards technological pedagogical content knowledge," Journal for the Education of Gifted Young Scientists, vol. 8, no. 1, pp. 1-16, 2020. https://doi.org/10.17478/jegys.600979.

[18] M. Meganita, P. Papilaya, and D. Rumahlatu, "Application of the science model community-based problem-solving technology in improving learning outcomes, science process skills, and students' scientific attitudes," Bioedupat Pattimura Journal of Biology and Learning, vol. 2, no. 1, pp. 10-18, 2022. https://doi.org/10.30598/bioedupat.v2.i1.pp10-18.

[19] B. Khumraksa and P. Burachat, "The scientific questioning and experimental skills of elementary school students: the intervention of research-based learning," Jurnal Pendidikan IPA Indonesia, vol. 11, no. 4, pp. 588-599, 2022. https://doi.org/10.15294/jpii.v11i4.36807.

[20] S. Gültekin and T. Altun, "Investigating the impact of activities based on scientific process skills on 4th-grade students' problem-solving skills," International Electronic Journal of Elementary Education. https://doi.org/10.26822/iejee.2022.258.

[21] E. Ahsani and A. Rusilowati, "Students' process skills and scientific attitude: implementation of integrated science teaching materials based on elementary students' science literacy," Elementary Islamic Teacher Journal, vol. 10, no. 2, p. 325, 2022. https://doi.org/10.21043/elementary.v10i2.17156.

[22] P. Kruit, R. Oostdam, E. Berg, and J. Schuitema, "Effects of explicit instruction on the acquisition of students' science inquiry skills in grades 5 and 6 of primary education," International Journal of Science Education, vol. 40, no. 4, pp. 421-441, 2018. https://doi.org/10.1080/09500693.2018.1428777.

[23] M. Cindiati, S. Suharsono, and D. Diella, "Correlation between metacognition ability and students' science process skills on cellular bioprocess materials," Jurnal Pelita Pendidikan, vol. 9, no. 1, 2021. https://doi.org/10.24114/jpp.v9i1.22440.

[24] M. Sumo, "The influence of the project-based learning model on the scientific creativity of physics education undergraduate students at Madura Islamic University," SEJ (Science Education Journal), vol. 8, no. 1, pp. 19-31, 2024. https://doi.org/10.21070/sej.v8i2.1651.

[25] T. Kaçan and İ. Şahin, "The relationship between scientific creativity and scientific process skills in science education," Journal of Education and Learning, vol. 7, no. 5, pp. 156-167, 2018. https://doi.org/10.5539/jel.v7n5p156.

[26] S. Suyidno, M. Nur, L. Yuanita, and B. Prahani, "Validity of creative responsibility-based learning: an innovative physics learning to prepare the generation of creative and responsibility," IOSR Journal of Research & Method in Education (IOSR-JRME), vol. 7, no. 1, pp. 56-61, 2017. https://doi.org/10.9790/7388-0701025661.

[27] R. Sternberg, R. Todhunter, A. Litvak, and K. Sternberg, "The relation of scientific creativity and evaluation of scientific impact to scientific reasoning and general intelligence," Journal of Intelligence, vol. 8, no. 2, p. 17, 2020. https://doi.org/10.3390/jintelligence8020017.

[28] S. Rizal, A. Putra, Y. Suharto, and Y. Wirahayu, "Creative thinking and process science skill: self-organized learning environment on watershed conservation material," Jurnal Pendidikan IPA Indonesia, vol. 11, no. 4, pp. 578-587, 2022. https://doi.org/10.15294/jpii.v11i4.39571.

[29] M. Devanda, A. Sugilar, and R. Fadillah, "The relationship between the environment-based learning model and students' scientific creativity in biology lessons," Journal of Educational Sciences, vol. 6, no. 2, pp. 258-267, 2022. https://doi.org/10.31258/jes.6.2.p.258-267.

[30] O. Nacaroğlu and F. Mutlu, "Investigating lifelong learning tendencies and scientific creativity levels of prospective science teachers," Acta Educationis Generalis, vol. 13, no. 1, pp. 74-95, 2023. https://doi.org/10.2478/atd-2023-0004.

[31] Y. Zheng, P. Li, X. Yang, Y. Guo, X. Qiu, X. Jin, and T. Zheng, "K–12 science, technology, engineering, and math characteristics and recommendations based on analyses of teaching cases in China," Frontiers in Psychology, vol. 13, 2022. https://doi.org/10.3389/fpsyg.2022.1010033.

[32] S. Xu and S. Zhou, "The effect of students' attitude towards science, technology, engineering, and mathematics on 21st-century learning skills: A structural equation model," Journal of Baltic Science Education, vol. 21, no. 4, pp. 706-719, 2022. https://doi.org/10.33225/jbse/22.21.706.

[33] H. Tuong, N. Pham, M. Nguyen, V. Tien, Z. Lavicza, and T. Houghton, "Utilizing STEM-based practices to enhance mathematics teaching in Vietnam: developing students' real-world problem-solving and 21st-century skills," Journal of Technology and Science Education, vol. 13, no. 1, p. 73, 2023. https://doi.org/10.3926/jotse.1790.

[34] N. Dat, "Arduino-based experiments: leveraging engineering design and scientific inquiry in STEM lessons," International Journal of STEM Education for Sustainability, vol. 4, no. 1, pp. 38-53, 2024. https://doi.org/10.53889/ijses.v4i1.317.

[35] K. Margot and T. Kettler, "Teachers' perception of STEM integration and education: a systematic literature review," International Journal of STEM Education, vol. 6, no. 1, 2019. https://doi.org/10.1186/s40594-018-0151-2.

[36] S. Ohnigian, J. Richards, D. Monette, and D. Roberts, "Optimizing remote learning: leveraging Zoom to develop and implement successful educational sessions," Journal of Medical Education and Curricular Development, vol. 8, 2021. https://doi.org/10.1177/23821205211020760.

[37] N. Hamad, "A review of innovative approaches to STEM education," International Journal of Science and Research Archive, vol. 11, no. 1, pp. 244-252, 2024. https://doi.org/10.30574/ijsra.2024.11.1.0026.

[38] M. H. Mhd Salim, N. M. N. M. Ali, and M. T. M. T. Ijab, "Understanding students' motivation and learning strategies to redesign massive open online courses based on persuasive system development," *International Journal of Advanced Computer Science and Applications*, vol. 10, no. 12, pp. 234–241, 2019.

[39] L. Hill, "Blackboard collaborate ultra: an online, interactive teaching tool," Academy of Management Learning and Education, vol. 18, no. 4, pp. 640-642, 2019. https://doi.org/10.5465/amle.2019.0027.

[40] N. Dahal, B. Luitel, B. Pant, I. Shrestha, and N. Manandhar, "Emerging ICT tools, techniques and methodologies for online collaborative teaching and learning mathematics," Mathematics Education Forum Chitwan, vol. 5, no. 5, pp. 17-21, 2020. https://doi.org/10.3126/mefc.v5i5.34753.

[41] M. Lailiyah and K. Yustisia, "Collaborative concept mapping: a study of group work satisfaction in vocational higher education," Journal of Vocational Education Studies, vol. 5, no. 2, pp. 312-321, 2022. https://doi.org/10.12928/joves.v5i2.6181.

[42] Z. Jeremić, N. Milikic, J. Jovanovic, M. Brkovic, and F. Radulovic, "Using online presence to improve online collaborative learning," International Journal of Emerging Technologies in Learning (IJET), vol. 7, no. S1, p. 28, 2012. https://doi.org/10.3991/ijet.v7is1.1918.

[43] L. Gerard, K. Wiley, A. DeBarger, S. Bichler, A. Bradford, and M. Linn, "Self-directed science learning during COVID-19 and beyond," Journal

of Science Education and Technology, vol. 31, no. 2, pp. 258-271, 2021. https://doi.org/10.1007/s10956-021-09953-w.

[44] S. Khan, M. Uzair-ul-Hassan, and R. Mutalib, "Nurturing self-regulated and peer collaborative learning skills in students within online mode: exploring teachers' perspectives," International Journal of Distance Education and E-Learning, vol. 8, no. 1, pp. 35-46, 2023. https://doi.org/10.36261/ijdeel.v8i1.2650.

[45] M. Wang and J. Degol, "Gender gap in science, technology, engineering, and mathematics (STEM): Current knowledge, implications for practice, policy, and future directions," Educational Psychology Review, vol. 29, no. 1, pp. 119-140, 2016. https://doi.org/10.1007/s10648-015-9355-x.

[46] T. Charlesworth and M. Banaji, "Gender in science, technology, engineering, and mathematics: issues, causes, solutions," Journal of Neuroscience, vol. 39, no. 37, pp. 7228-7243, 2019. https://doi.org/10.1523/jneurosci.0475-18.2019.

[47] N. Hübner, E. Wille, J. Cambria, K. Oschatz, B. Nagengast, and U. Trautwein, "Maximizing gender equality by minimizing course choice options? Effects of obligatory coursework in math on gender differences in STEM," Journal of Educational Psychology, vol. 109, no. 7, pp. 993-1009, 2017. https://doi.org/10.1037/edu0000183.

[48] M. El-Hout, A. Garr-Schultz, and S. Cheryan, "Beyond biology: the importance of cultural factors in explaining gender disparities in STEM preferences," European Journal of Personality, vol. 35, no. 1, pp. 45-50, 2021. https://doi.org/10.1177/0890207020980934.

[49] R. Coltogirone, S. Kuhn, S. Freeland, and S. Bergeron, "Fish in a dish: using zebrafish in authentic science research experiences for under-represented high school students from West Virginia," Zebrafish. https://doi.org/10.1089/zeb.2022.0074.

[50] G. Saw and C. Agger, "STEM pathways of rural and small-town students: opportunities to learn, aspirations, preparation, and college enrollment," Educational Researcher, vol. 50, no. 9, pp. 595-606, 2021. https://doi.org/10.3102/0013189x211027528.

[51] L. Ihrig, S. Assouline, D. Mahatmya, and S. Lynch, "Developing students' science, technology, engineering, and mathematics talent in rural after-school settings: rural educators' affordances and barriers," Journal for the Education of the Gifted, vol. 45, no. 4, pp. 381-403, 2022. https://doi.org/10.1177/01623532221123786.

[52] S. Murphy, "Leadership practices contributing to STEM education success at three rural Australian schools," The Australian Educational Researcher, vol. 50, no. 4, pp. 1049-1067, 2022. https://doi.org/10.1007/s13384-022-00541-4.

[53] B. Jeanpierre and R. Hallett-Njuguna, "Exploring the science attitudes of urban diverse gifted middle school students," Creative Education, vol. 05, no. 16, pp. 1492-1496, 2014. https://doi.org/10.4236/ce.2014.516166.

[54] X. Duong, N. Nguyen, M. Nguyen, and T. Thao-Do, "Applying STEM engineering design process through designing and making of electrostatic painting equipment in two rural schools in Vietnam," Jurnal Pendidikan IPA Indonesia, vol. 11, no. 1, pp. 1-10, 2022. https://doi.org/10.15294/jpii.v11i1.31004.

[55] N. M. Arsad, K. Osman, and T. M. T. Soh, "Instrument development for 21st-century skills in Biology," Procedia - Social and Behavioral Sciences, pp. 1470–1474, 2011. https://doi.org/10.1016/j.sbspro.2011.03.312.

[56] T. R. Kelley, J. G. Knowles, J. Han, and E. Sung, "Creating a 21st Century Skills Survey Instrument for High School Students," American Journal of Educational Research, vol. 7, no. 8, pp. 583–590, 2019. https://doi.org/10.12691/education-7-8-7.

[57] D. B. Boyaci and N. Atalay, "A scale development for 21st Century skills of primary school students: A validity and reliability study," International Journal of Instruction, vol. 9, no. 1, pp. 133–135, 2016. https://doi.org/10.12973/iji.2016.9111a.

[58] C. Gormally, P. Brickman, and M. Lut, "Developing a test of scientific literacy skills (TOSLS): Measuring undergraduates' evaluation of scientific information and arguments," CBE Life Sciences Education, vol. 11, no. 4, pp. 364–377, 2012. https://doi.org/10.1187/cbe.12-03-0026.

[59] W. N. F. Wan Husin, M. Fairuz, M. Syukri, and L. Halim, "Competencies of science centre facilitators," Journal of Turkish Science Education, vol. 12, no. 2, pp. 49–62, 2015. https://doi.org/10.12973/tused.10140a.

[60] E. M. L. Soriano de Alencar, "Creativity in Organizations: Facilitators and Inhibitors," in Handbook of Organizational Creativity, Elsevier, pp. 87–111, 2011. https://doi.org/10.1016/B978-0-12-374714-3.00005-7.

[61] W. Hu and P. Adey, "A scientific creativity test for secondary school students," International Journal of Science Education, vol. 24, no. 4, pp. 389–403, 2002. https://doi.org/10.1080/09500690110098912.

The Impact of E-Commerce Drivers on the Innovativeness in Organizational Practices

Abdulghader Abu Reemah A Abdullah, Ibrahim Mohamed, Nurhizam Safie Mohd Satar
Faculty of Information Science and Technology, Universiti Kebangsaan Malaysia, Bangi, UKM 43600 Malaysia

***Abstract*—Innovation in e-commerce practices has revolutionized the way goods and services are purchased or sold online. This relatively new tool for online transaction provides range of access to wealth of information and knowledge needed to facilitate electronic commerce globally using internet network. The case is not the same in the developing countries where e-commerce innovation is deprived of key the components to drives developing economy. To clearly understand innovation in e-commerce diffusion, 375 quantitative data generated from e-commerce organizations in Libya. Statistically analysis of the key drivers of e-commerce innovations focused on the need for a shift in organizational attitude and knowledge through decision making that are committed to meeting customer's needs. The inter statistical covariance indicated a strong homogeneity between the drivers of e-commerce with mean value range of 4.09 to 4.82 (58.4 % to 68.8% of responses) indicating that 219 to 258 respondents out of 375 are of the same view. There is strong positive correlation between the drivers of e-commerce innovations except for e-commerce management style that has moderate relation and were statistically significant at 0.00 level. This study clearly explained the main factors of interest that are versatile in providing timely delivery of goods, efficient services and in meeting with e-commerce developmental trend.**

***Keywords*—E-commerce innovation; e-commerce drivers; performance management; decision making; management style**

I. Introduction

The progress in e-commerce represents a potential opportunity to enhance the growth of developing nations and to improve the effectiveness of business processes at the organizational level. The Organizations across developing countries can conveniently utilize open-ended opportunities to improve commercial activities globally. Innovation in e-commerce is transforming the developing economy because it drives commercial sectors to benefit from efficient and low-cost transactions [1], [2], [3]. E-commerce online platforms aid in the purchase or sale of goods and services and have contributed to building business networks and partnerships in most developing and developed countries, organizations and businesses have appreciated development innovations and the emergence of transitional changes in e-commerce practices. E-commerce has become an essential pillar off economic growth and internationalization to maximize the advantages of technology-driven commercial activities.

The easy access to e-commerce resources has significantly contributed to the integration of various technologically mediated media service frameworks for the sale of products and services. E-commerce innovations are at the forefront of transforming conventional commerce into an electronic-based business format [4], [5], [6]. Furthermore, the emerging experience of accessing e-commerce resources for online transactions has added to the use of different electronic devices and gadgets for economic activities. The versatility of electronic devices such as hand phones, computers, gadgets, electronic media, and handheld devices has improved the way goods and services are handled over the years [7], [8].

However, there are issues limiting the wider use to e-commerce websites including the cost of electronic systems, internet infrastructure, online security for data and personal information [9], [10]. Other challenges include ease of access to online platforms, customer experience, cost of shipping purchased products, online customers support, sustainability, and frequent updates requiring larger storage for new features to support online sales of products and services [11]. To generate depth of insight into e-commerce advances, these factors were grouped as drivers of e-commerce innovations in the present study. The emerging new features and apps incorporated with e-commerce require specialized skill [12], [13] and are vital for improvinge online businesses [14]. A good understanding of how these drivers influence innovativeness within organizations is very important for enhancinge the competitive edge of the digital marketplace [7].

However, the objectives of the present study are to identify the impact of the e-commerce drivers and to analyze how they influence e-commerce innovation. A clear understanding of the factors that have consistently contributed to the improvement recorded in e-commerce development could improve the competitive edge of e-commerce. This is important, as e-commerce has become a well-known branch of commerce that uses electronic systems for the purchase and sale of goods and services via online channels. The quality of e-commerce services will continue to improve over time as research efforts are targeted at meeting customers' needs [13], [15] As online businesses continue to widen with the inclusion of new innovative changes [16], retailers may be compelled to adapt certain changes to facilitate online transactions. The drivers of e-commerce could help drive e-commerce performance with insight into the potential influences on e-commerce innovations. The drivers of e-commerce innovation at the organizational level could further improve the effectiveness of different services and the competitiveness of online businesses.

II. Literature Review

The present study is based on the influence of the key drivers of e-commerce on innovative features from the literature. These variables have been widely recognized and used separately considering their impact on e-commerce activities at the organizational level. In the context of e-commerce practices, managerial and operational capabilities for organizational innovation have been validated based on management

style, decision making, people's development (also known as workers development), process management and performance management [1], [17]. [18] found that e-commerce development has been affected by technological and organizational attitudes toward innovation, as well as financial concerns. The management framework for assessing disruptive innovations by [19] found that decision making is a key driver of technological innovation in organizations. This could be because the decision to use a certain technology is based on management's decisions and approval. Decision-making has been referred to as an approach to improve technological innovation [20] and among drivers for the adoption of eco-design practices [21].

An empirical study that explored the interplay of managerial and operational capabilities to infuse organizational innovation in SMEs [1] provided a formal reflection on the role of innovation in transforming organizational practices. In [22], it was shown that e-commerce will continue to improve management practices with the emergence of new technological features that make transactions easier and faster. A study by [18] provided a depth of insight into the prevailing issues of concern regarding e-commerce practices, which are related to organizational practices, technological development, financial resources and external factors.

Previous research findings are foundational to the choice of instrument, used in this study, research approach, variables of interest, and the analytical method used in the present study. A literature study by [18], [21] found that e-commerce adoption has been severely affected by technological innovation, financial assets, management practices, decision-making, and organizational factors. Financial and technological factors have demonstrated a consistent influence over the years on organizational factors. E-commerce practices at the organizational level were demonstrated in a study by [1]. Important variables used include process management, management style, decision making, performance management, people's development, and organizational innovations. The study in [23] found that emerging economies have realized the benefits of e-commerce and are investing heavily in technology to fully explore e-commerce potentials. This study's structure is reflective of previous findings that represented a synergetic focus on factors that constrained innovation in e-commerce practices.

III. Research Framework

This study examined contextual factors that have been identfied as drivers of e-commerce innovation. The drivers include organizational management styles (eCMS) that provide insight into the need for a swift response to online customers and feedback to information and requests relative to sales of products and services. Decision-making in e-commerce (DM) provided a clear understanding of the plans to improve e-commerce practices. e-commerce worker development (eCWD) focuse on training and employers' development to cope with innovative changes in technology and the application of e-commerce features. The e-Commerce management process (eCPM) explains activities and tasks to effectively improve e-commerce services. e-commerce performance management (eCPfM) sets clear business goals for e-commerce organizations. The e-Commerce Organization Attitude to Innovation (eCOAI) explained the commitment of thee-commerce management team to adopt new technology to efficiently deliver e-commerce services. Knowledge of e-commerce technology (KeCT) focuses on the ability of innovative features to drive e-commerce activities. Financial concern refers to the capability to effectively manage financial resources to achieve healthy e-commerce services. External factors (extF) entail pressure from external entities, such as customers and suppliers. e-Commerce Innovation (eCI) explains new ways of conducting e-commerce business using improved features. These important e-commerce factors were collectively used in this studt, and their effect on e-commerce innovation was hypothetically validated using statistical measurements. This study sets up a new conceptual and holistic view that contextually explains the variable themes that matter in the theoretical model of e-commerce innovation. The hypothesis of this study asserts that "e-commerce drivers have a positive relationship with e-commerce innovation". This study explores the key drivers of e-commerce practices and their overriding influences on e-commerce innovation. E-commerce innovations are embodied with potential opportunities to earn promising benefits from e-commerce features that support sales of goods and services via online websites [24]. The clear identification of e-commerce drivers and potential barriers in this study is a pathway to improve the effectiveness and versatility of online services. Reflecting on the work of [18], e-commerce has been constrained by knowledge of e-commerce technology, organizational management style, financial concerns, as well as external factors relative to supply and customer push. The research framework constituting the key drivers of e-commerce innovations are as shown in Fig. 1.

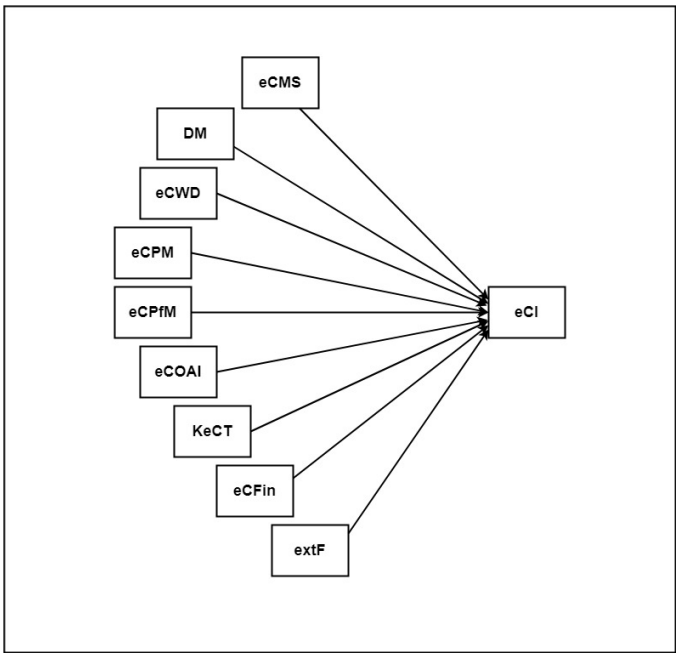


Fig. 1. The research framework of e-commerce drivers' effect on e-commerce innovations.

IV. Research Approach

The population of this study comprised 375 managers who have saved a minimum of five years on the most active 150 e-commerce websites across Libya. Participants ranging

from 20 to 65 years of age were selected to respond to structured research items that focused on the drivers of e-commerce innovations. Data collected from online and e-commerce employees willingly contributed to their practical experience in e-commerce transactions and management involvement to improve e-commerce practices. Employers' age, position and years of experience were major considerations for data collection and 384 data collected were carefully screened of which 375 were considered appropriate for this study. Incomplete questionnaires were excluded to improve the accuracy and reliability of the research findings. Participation in the survey was voluntary, and participant experience and organizational roles contributed to the overall quality of the research findings.

A. Measures

e-commerce organizational practices revolve around e-commerce innovations [19], [24], [25]. To improve e-commerce activities and competitiveness at all levels, this study statistically explored the participants' knowledge of e-commerce innovations and the types of devices used as well as the type of training and the issues that have constrained e-commerce progress. Pearson's correlation statistics were used to measure the strength of the relationship between the relative research variables based on r and p value. Inter-statistical analysis was used to provide a clear explanation on the variance and the mean value in the research items. ANOVA with Tukey's test for statistics was used to estimate the degrees of freedom based on the power of observation and the linear relationship of the hypothesized relationship. Multivariate analysis was used to explain the research hypotheses and address the null hypothesis of the study. A statistical test between the subject effect was used to support the hypothetical result.

V. Contextual Description of Research Variables

The analysis reported in this section was based on numerous factors that support the e-commerce business including the need for timely communication and feedback with customers/clients also referred to as the e-commerce management style (eCMS). e-commerce organizational practices have been challenged by the difficulty of handling sensitive personal information, data and important software and innovative apps that have added value to the competitiveness of online businesses [26]. This is because the management style of e-commerce firms has failed to incorporate trends in technology into management practices. This could be because, decision making (DM) to improve e-commerce practices has not taken into consideration the need for adequate and timely training on e-commerce updates and innovations also referred to as e-commerce worker development (eCWD). In this study, all processes and tasks associated with e-commerce transactions is referred to as e-commerce process management (eCPM) and a clear business set goal for the e-commerce process is referred to as e-commerce performance management (eCPfM). Other important factors such as e-commerce organizations attitude towards innovation (eCOAI) have focused on making e-commerce more attractive to a range of online consumers into an effort to reduce shopping and delivery time. Organizations knowledge of the impact of e-commerce innovation has opened doors to new outlets to make online sales easier and convenient [27], [28], [29]. It becomes imperative that e-commerce competitiveness e-commerce rotates around a transformative shift in innovation [30]. Knowledge of e-commerce technology (KeCT) has become a vital tool for e-commerce development. This is because insufficient knowledge about new developments in e-commerce practices poses greater challenges in positioning developing economies on a fast development path [31]. This is very visible from the emergence of relatively new technological innovations changes that have enabled organizations to tap new knowledge capabilities to foster the production and sales of products by extending service networks and providing timely services to larger numbers of online shoppers [32]. It is noteworthy to consider the financial capability of e-commerce organizations to acquire the necessary infrastructure to handle different services. Financial involvement (eCFin) to sustain e-commerce development is an important factor that determines e-commerce versatility to render efficient and timely services at a relatively lower cost. External pressure is another important factor (extF) in this study that focuses on customer and supplier commitment and respond to the business environment and is discussed in this study relative to new ways of doing online business, also referred to as e-commerce innovation (eCI).

VI. Statistical Analysis

A. Descriptive Statistics of e-Commerce Knowledge Across the Study Population

Proportion of Libyan using e-commerce platform to purchase products are denoted by "Yes and No device used for e-commerce transaction, type of training to adopt e-commerce innovations and problems with purchasing product and vices via e-commerce platform are as shown in Table I.

The mean distribution of the research variables ranged from 4.0880 to 4.8160 (58.4% to 68.8%), indicating that 219 to 258 out of 375 of the study population had the same view about the research outcome, as shown in Table II.

B. Correlation Matrix of the Research Variable

The correlation matrix shows the relationship levels of the variables ranging from moderate (only for eCMS) to strong positive relationships with other drivers of e-commerce innovation (Table III). Inter-item covariance explains the homogeneity of the tested variables. The mean value of the variable item was closely related (4.088 – 4.816) with a mean value of M = 4.431 (Table IV, Fig. 2, 3).

C. ANOVA Statistics with Tukey's Test for Nonadditivity

Tukey's estimated power to which observations must be made to achieve additivity = 2.056. The model analysis revealed the interaction effect between e-commerce drivers and e-commerce innovation at the organizational level (Table V).

D. Multivariate Analysis of the Research Hypotheses

- indent Variable: eCI.
- .581 eCMS(eCOAI)+.564 eCMS (KeCT)+.839 eCMS(eCFin)+.330 eCMS(extF) - 1.315 eCMS (Error).

TABLE I. E-COMMERCE KNOWLEDGE, DEVICES USED AND PRACTICES

				95% **Confidence Interval**		
	F	**%**	**Bias**	**Std. Error**	**L**	**U**
Yes	330	88	-0.1	1.7	84.5	90.9
No	45	12	0.1	1.7	9.1	15.5
Total	**375**	**100**				
Smartphones	306	81.6	-0.1	1.9	77.6	85.1
Laptop	27	7.2	0	1.3	4.8	9.6
Desktop PC	12	3.2	0	0.9	1.3	5.1
Tablets	30	8	0.1	1.4	5.6	10.9
Total	**375**	**100**				
Formal training	165	44	0.1	2.5	39.2	49.3
On-the-job training	108	28.8	-0.1	2.3	24	33.3
Both formal and on-the-job training	102	27.2	0	2.4	22.7	31.7
Total	**375**	**100**				
Location	78	20.8	0	2.1	16.8	24.8
Delivery time	72	19.2	0	2	15.2	23.2
Product quality	59	15.7	0	1.8	12.3	19.2
No problem	108	28.8	0	2.3	24	33.3
Delay	58	15.5	0.1	1.8	12	19.2
Total	**375**	**100**				

Note: F, L, U refer to frequency, lower, and upper, respectively.

TABLE II. DESCRIPTIVE STATISTICS E-COMMERCE INNOVATION DRIVERS

Variables		**Mean**	**Std. Deviation**	**N**
eCMS		4.472	1.68784	375
DM		4.264	1.55201	375
eCWD	4.088	1.86589		375
eCPM	4.224	1.73605		375
eCPfM	4.392	1.62647		375
eCOAI	4.656	1.64709		375
KeCT	4.528	1.57983		375
eCFin	4.464	1.67337		375
extF	4.408	1.66642		375
eCI	4.816	1.53768		375

TABLE III. CORRELATION MATRIX OF RESEARCH VARIABLES

Variable item	***eCMS***	***DM***	***eCWD***	***eCPM***	***eCPfM***	***eCOAI***	***KeCT***	***eCFin***	***extF***	***eCI***
eCMS	1.000	0.430	0.450	0.446	0.435	0.488	0.445	0.462	0.479	0.488
DM	0.430	1.000	0.967	0.978	0.960	0.939	0.951	0.944	0.954	0.941
eCWD	0.450	0.967	1.000	0.972	0.964	0.952	0.947	0.963	0.951	0.962
eCPM	0.446	0.978	0.972	1.000	0.977	0.958	0.966	0.966	0.967	0.962
eCPfM	0.435	0.960	0.964	0.977	1.000	0.961	0.962	0.967	0.965	0.953
eCOAI	0.488	0.939	0.952	0.958	0.961	1.000	0.967	0.966	0.966	0.976
KeCT	0.445	0.951	0.947	0.966	0.962	0.967	1.000	0.969	0.966	0.951
eCFin	0.462	0.944	0.963	0.966	0.967	0.966	0.969	1.000	0.956	0.959
extF	0.479	0.954	0.951	0.967	0.965	0.966	0.966	0.956	1.000	0.956
eCI	0.488	0.941	0.962	0.962	0.953	0.976	0.951	0.959	0.956	1.000

TABLE IV. INTER ITEM STATISTICAL COVARIANCE

	Mean	**Max**	**Range**	**Max / Min**	**Var.**	**N of Items**
Item Means	4.816	0.728	1.178	0.045	10	
Inter-Item Covariances	3.149	2.022	2.795	0.331	10	

TABLE V. ANOVA WITH TUKEY'S TEST FOR NONADDITIVITY

			Sum of Squares	**df**	**Mean Square**	**F**	**Sig**
Between People			8964.85	374	23.97		
Within People	Between People		150.55	9	16.728	42.071	0.000
	Residu	Nonadditivity	20.441a	1	20.441	52.191	0.000
		Balance	1317.91	3365	0.392		
		Total	1338.35	3366	0.398		
	Total		1488.9	3375	0.441		
Total			10453.75	3749	2.788		

Grand Mean = 4.4312

TABLE VI. MULTIVARIATE ANALYSIS AND RELIABILITY STATISTICS

T-Squared	F	df1	df2	Sig	Cronbach's Alpha	Cronbach's Alpha (Standardized Items)
841.244	91.472	9	366	0.000	0.983	0.984

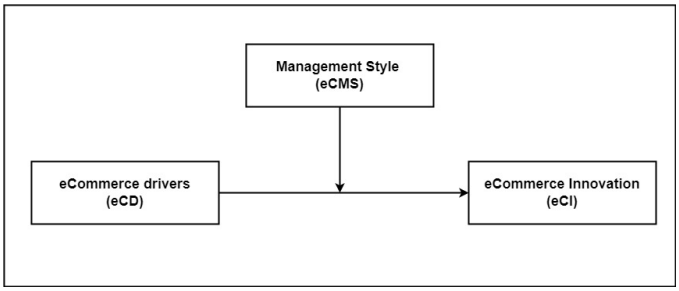


Fig. 2. Linear model of the moderation effect of eCMS on e-Commerce drivers with eCI.

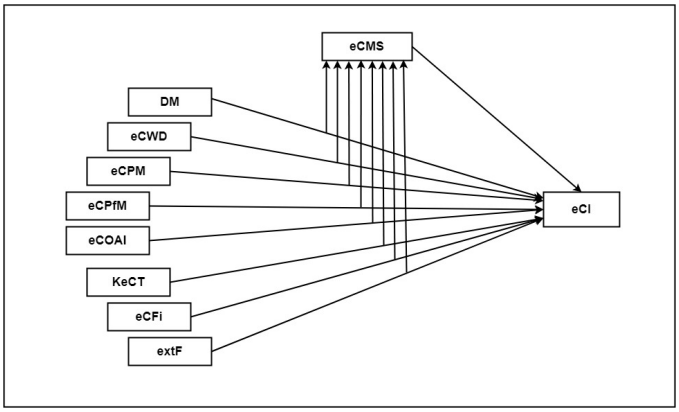


Fig. 3. Hypothetical path of the moderating effect of eCMS on the eCD influence on eCI.

- Cannot compute the appropriate error term using Satterthwaite's method.

With the intention of exploring different innovative ways to improve e-commerce practices at the organizational level, the partial sum of squares at an intercept with e-commerce innovation was .581 eCMS(eCOAI), .564 eCMS (KeCT), .839 eCMS(eCFin), .330 eCMS(extF), .771 eCMS(eCWD), .878 eCMS(eCOAI), .300 eCMS (DM), with - 1.315 eCMS (Error), indicating that e-commerce process management (eCPM) could not compute the appropriate error term. However, the hypothetical statement that eCMS(eCPM) has a positive relationship with eCI at the organizational level and was significant at 0.00 level (Table VII).

- Dependent Variable: eCI

A test of equality of error was conducted to confirm whether the variance of the dependent variable was equal across groups for the null hypothesis (Table VIII).

VII. DISCUSSION

The study population consisted of Libyans aged 20 years and above 65 years old who are eligible e-commerce users and have a minimum of five years working as management staff in e-commerce organizations across Libya. results showed that 330 (88%) out of 375 of the study population purchased products and services via e-commerce websites. 306 (81.6%) users of e-commerce infrastructure accessed the website using smartphones, while others used tablets (30 users constituting 8%), laptops (27 users comprised of 7.2%), and desktop PC covering (12 users constituting 3.2%). Training of e-commerce is mainly by formal training and "on the job training" on-job training using facilities and infrastructures for e-commerce services. Majority of the e-commerce users (28.8%) were satisfied with the services, while others emphasized on the need to use identifiable location addresses (20.8), others focused on improving the delivery time (19.2%), product quality (15.7%), and to address delay in product and service delivery (Table I). Item statistics show that the mean value ranges from 4.09 to 4.82 (58.4 % to 68.8% of responses), indicating that 219 to 258 respondents out of 375 in the study population are of the same view about the research outcome. This finding shows that this study is reliable and can be used to generalize innovations in e-commerce practices across the developing countries, especially in the Middle East. The statistical results (Table II) show that eCI had the highest mean value (M = 4.82, SD = 1.54), followed by eCOAI (M = 4.66, SD = 1.65), KeCT (M = 4.53, SD = 1.58), eCMS (M = 4.47, SD = 1.69), eCFin (M= 4.46, SD = 1.67), extF (M = 4.41, SD = 1.67), eCPfM (M = 4.39, SD = 1.63), DM (M = 4.26, SD = 1.55), eCPM (M = 4.22, SD = 1.74), and eCWD (M = 4.09, SD = 1.86).

The correlation statistical result showed that eCMS has a moderate positive correlation ranging from r .430–.488, p = .000 level, with all e-commerce innovation drivers. Except for eCMS, the other drivers of e-commerce innovation have a very strong positive correlation. DM had a strong positive correlation, ranging from r = .944.978, p = .000. eCWD is strongly positively correlated at r = .951 – 972, p = .000, eCPM at r = .958 – 978, p = .000, eCPfM at r = .953 – 977, p = .000, eCOAI at r = .952 – 976, p = .000, keCT at r = .951 – 967, p = .000, eCFin at r = .944 – 969, p = .000, extF at r = .951 – 967, p = .000 and eCI at r = .941 – 976, p = .000. The strength of the relationship between variables varies. eCMS showed a stronger relationship with eCI and eCOAI, whereas DM showed a stronger relationship with eCPM, eCWD with eCPM, eCPM with DM, eCPfM with eCPM, eCOAI with eCI, KeCT with eCFin, eCFin with KeCT, extF with eCPM, and eCI with eCOAI (Table III).

An ANOVA one-degree Tukey's test for no additivity of freedom was used to test the hypothesized linear interaction across the research variables. The ANOVA model assumed no randomized treatment or additive block across the measured items. The interaction between the items and no additivity was statistically significant. The multivariate multiple responses of the variables were generalized using Hoteling's T-Squared Test. Significant differences between the multivariate means of the data sets shown in Table VI summarize e-commerce the significant relationship between e-commerce divers and

TABLE VII. TESTS OF THE EFFECTS OF E-COMMERCE DRIVERS

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
Intercept	Hypothesis	485.653	1.000	485.653	374.319	
	Error	7.055	5.438	1.297a		
eCMS	Hypothesis	0.000	5.000	0.000	0.000	1.000
	Error	4.500	330.000	.014b		
DM	Hypothesis	0.250	2.000	0.125	9.167	0.000
	Error	4.500	330.000	.014b		
eCWD	Hypothesis	4.625	6.000	0.771	56.528	0.000
	Error	4.500	330.000	.014b		
eCPM	Hypothesis	0.000	0.000	.	.	.
	Error	.	.	.c		
eCPfM	Hypothesis	0.000	4.000	0.000	0.000	1.000
	Error	4.500	330.000	.014b		
eCOAI	Hypothesis	7.500	4.000	1.875	137.500	0.000
	Error	4.500	330.000	.014b		
KeCT	Hypothesis	1.600	4.000	0.400	29.333	0.000
	Error	4.500	330.000	.014b		
eCFin	Hypothesis	0.000	6.000	0.000	0.000	1.000
	Error	4.500	330.000	.014b		
ExtF	Hypothesis	0.000	3.000	0.000	0.000	1.000
	Error	4.500	330.000	.014b		

TABLE VIII. LEVENE'S TEST OF EQUALITY OF ERROR VARIANCE

F	df1	df2	Sig.
16.597	44	330	0.000

e-commerce innovations. The multivariate probability of F-distribution explained the hypothetical statement that the e-commerce drivers positively influenced e-commerce innovation were generalized by the statistics underlying t-distribution.

We reject H0 because α t2 (841.244) is greater than the critical value for F (91.472).

Analysis of variance based on a test of the between-subject effect was conducted to test the ability of the model to clearly explain possible variation. The displayed variable labels indicate that the values of all terms were significantly related at the intercept; eCMS, DM, eCWD, eCPfM, eCOAI, KeCT, eCFin, and extF computed similar error terms. However, larger values of the sum of squares, mean squares, and F values indicate a greater amount of variation accounted for by the model error term.

The p-values for the hypothetical relationship between eCMS, eCPfM, eCFin, and extF and eCI, while DM, eCWD, eCOAI, and KeCT were statistically significant at 0.00 level. The practical significance of each term was based on the ratio of the variation in the sum of squares accounted for by the term to the sum of the variation accounted for by the term, and the variation left to error. The tendencies of bias in standard errors, as well as t-statistics or F–statistics in drawing inferences, especially during model misspecification, were addressed by conducting heteroscedasticity tests to check for structural breaks. The heteroscedasticity test revealed that no single observation was dominant (Table VI). This has provided a clear understanding of the main factors requiring close attention in meeting the developmental innovations in e-commerce and how newly added features in online web shops have improved existing practices. It is clear that e-commerce innovations are versatile in providing timely delivery of goods and efficient services [33], [34].

This intriguing shift in e-commerce innovation has transformed digital commerce and provided multiple choices for online customers to buy products that were previously available only in physical retail shops. This relatively new business opportunity is embodied in specific skills that require successive developments and updates. The adoption of e-commerce innovation is part of the digital transformation from remote sales, and the delivery of customer service is an open-ended business network [35]. With the recent surge in online shopping, e-commerce stands out as a potential tool for enhancing entrepreneurial capability and competitiveness in organizations.

VIII. CONCLUSION AND RECOMMENDATION

Recent innovations have widened e-commerce knowledge and its role in a developing economy. Various drivers have been identified as key factors in enhancing e-commerce performance at the organizational level. Identifying the key drivers of e-commerce innovation makes the future of e-commerce brighter, as they contribute to addressing the backdrop in e-commerce performance at the organizational level. E-commerce innovation has been acknowledged as the main factor driving online transactions. The development of e-commerce workers is an issue of concern that needs to be addressed to meet the growing trend in innovation. The management styles of e-commerce firms incorporate development trends in e-commerce innovations into the training of workers and services to improve workers' knowledge and e-commerce performance. This is because knowledge of e-commerce at the organizational level is affected by attitudes towards innovation. Decision-making at the management level should incorporate funding of e-commerce infrastructures, since formal training and on-the-job training are prevalent ways of equipping employees in e-commerce organizations.

The drivers of e-commerce innovation model design showed that, at the intercept (eCMS + DM + eCPD + eCPM + eCPfM + eCOAI + KeCT + eCFin + extF), the drivers of e-commerce innovations have a strong positive relationship. Multivariate analysis has quantified the moderating effect of e-commerce management style (eCMS) on the relationship between e-commerce drivers (eCD) and e-commerce innovation (eCI). eCD had a strong positive relationship with eCI. Management styles at various e-commerce organizations aim

to invest more effort and resources to improve the performance of e-commerce innovation drivers.

E-commerce companies selling and delivering goods and services across multiple online platforms have indicated that the potential of e-commerce is promising. e-commerce has made shopping easier and positioned e-Shop strategically to carter customers using fast and easy-to-handle electronic systems. Therefore, for further improvement, customer interest should be considered first. To facilitate online transactions, it is important to provide access to different product links and a clear description of product quality and price to potential customers.

E-commerce organizations should also partner with firms that can provide a reliable target for different customers, especially those that rate preferences and choice of online customers, to improve sales of products and services via e-commerce sites. This can be achieved by fine-tuning or providing a descriptive explanation of available products and services to potential customers using different media platforms that are actively used by many people. Efforts in this direction will further e-commerce processes and attract potential customers.

Declaration of Competing Interest

The authors declare no conflicts of interest regarding the data and information reported in this study.

Acknowledgment

The authors acknowledge the funding support provided by the Faculty of Information Science and Technology, Universiti Kebangsaan Malaysia (under the research grant TAP). The support from a few professionals such as Associate Prof. Dr Nurhizam Safie Mohd Satar, Ts. Dr Ibrahim Mohamed and Prof. Dr Charles Ahamefula Ubani who enriched the paper and constructively enhanced the presentation. The Centre for Software Technology Management (SOFTAM) and Bani Waleed University of Libya for their mentorship and invaluable contribution. The commitment of the editor-in-chief, associate editor, and reviewers' comments for improvement is among the efforts to put this paper in the best shape it deserves.

This work was supported by funding from the TAP and the Faculty of Information Science and Technology Universiti Kebangsaan Malaysia. under Grant TAP.

References

[1] Z. Ali, I. M. Zwetsloot, and N. Nada, "An empirical study to explore the interplay of Managerial and Operational capabilities to infuse organizational innovation in SMEs," *Procedia Computer Science*, vol. 158, pp. 260–269, 2019. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S1877050919312104

[2] V. Anderson, "Crafting and labor: An investigation of the e-commerce giant, etsy," Ph.D. dissertation, 2022, hak cipta - Database copyright ProQuest LLC; ProQuest does not claim copyright in the individual underlying works; Terakhir diperbarui - 2023-03-08. [Online]. Available: https://www.proquest.com/dissertations-theses/crafting-labor-investigation-e-commerce-giant/docview/2674829587/se-2

[3] A. A. R. A. Abdullah, I. Mohamed, N. S. M. Satar, A. S. Madaki, and H. S. Hawedi, "Innovations in E-Commerce Development and The Potential Disruptive Features," in *2023 International Conference on Electrical Engineering and Informatics (ICEEI)*. Bandung, Indonesia: IEEE, Oct. 2023, pp. 1–6. [Online]. Available: https://ieeexplore.ieee.org/document/10346640/

[4] G. Agag, "E-commerce Ethics and Its Impact on Buyer Repurchase Intentions and Loyalty: An Empirical Study of Small and Medium Egyptian Businesses," *Journal of Business Ethics*, vol. 154, no. 2, pp. 389–410, Jan. 2019. [Online]. Available: http://link.springer.com/10.1007/s10551-017-3452-3

[5] W. Aslam, A. Hussain, K. Farhat, and I. Arif, "Underlying Factors Influencing Consumers' Trust and Loyalty in E-commerce," *Business Perspectives and Research*, vol. 8, no. 2, pp. 186–204, Jul. 2020. [Online]. Available: http://journals.sagepub.com/doi/10.1177/2278533719887451

[6] A. M. Al-Adamat, M. KassabAlserhan, L. S. Mohammad, D. Singh, S. I. S. Al-Hawary, A. A. S. Mohammad, and M. F. A. Hunitie, "The Impact of Digital Marketing Tools on Customer Loyalty of Jordanian Islamic Banks," in *Emerging Trends and Innovation in Business and Finance*, R. El Khoury and N. Nasrallah, Eds. Singapore: Springer Nature Singapore, 2023, pp. 105–118.

[7] Y. Ismail, "E-commerce in the World Trade Organization: History and latest developments in the negotiations under the Joint Statement," Tech. Rep., 2020.

[8] E. E. Izogo and C. Jayawardhena, "Online shopping experience in an emerging e-retailing market," *Journal of Research in Interactive Marketing*, vol. 12, no. 2, pp. 193–214, May 2018. [Online]. Available: https://www.emerald.com/insight/content/doi/10.1108/JRIM-02-2017-0015/full/html

[9] J. Guo, W. Zhang, and T. Xia, "Impact of Shopping Website Design on Customer Satisfaction and Loyalty: The Mediating Role of Usability and the Moderating Role of Trust," *Sustainability*, vol. 15, no. 8, p. 6347, Apr. 2023. [Online]. Available: https://www.mdpi.com/2071-1050/15/8/6347

[10] F. Cui, D. Lin, and H. Qu, "The impact of perceived security and consumer innovativeness on e-loyalty in online travel shopping," *Journal of Travel & Tourism Marketing*, vol. 35, no. 6, pp. 819–834, Jul. 2018. [Online]. Available: https://www.tandfonline.com/doi/full/10.1080/10548408.2017.1422452

[11] X. Lin, X. Wang, and N. Hajli, "Building E-Commerce Satisfaction and Boosting Sales: The Role of Social Commerce Trust and Its Antecedents," *International Journal of Electronic Commerce*, vol. 23, no. 3, pp. 328–363, Jul. 2019. [Online]. Available: https://www.tandfonline.com/doi/full/10.1080/10864415.2019.1619907

[12] F. H. Kristanto, H. Wimanda Rahma, and M. Nahrowi, "Factors Affecting E-Commerce Customer Loyalty In Indonesia," *Jurnal Syntax Transformation*, vol. 3, no. 09, pp. 1150–1164, Sep. 2022. [Online]. Available: https://jurnal.syntaxtransformation.co.id/index.php/jst/article/view/613

[13] S.-H. Chen, H. Xiao, W.-d. Huang, and W. He, "Cooperation of Cross-border E-commerce: A reputation and trust perspective," *Journal of Global Information Technology Management*, vol. 25, no. 1, pp. 7–25, Jan. 2022. [Online]. Available: https://www.tandfonline.com/doi/full/10.1080/1097198X.2021.2022396

[14] A. Ansari, F. Stahl, M. Heitmann, and L. Bremer, "Building a Social Network for Success," *Journal of Marketing Research*, vol. 55, no. 3, pp. 321–338, Jun. 2018. [Online]. Available: http://journals.sagepub.com/doi/10.1509/jmr.12.0417

[15] P.-L. Sheu and S.-C. Chang, "Relationship of service quality dimensions, customer satisfaction and loyalty in e-commerce: a case study of the Shopee App," *Applied Economics*, vol. 54, no. 40, pp. 4597–4607, Aug. 2022. [Online]. Available: https://www.tandfonline.com/doi/full/10.1080/00036846.2021.1980198

[16] Y. Ismail, "E-commerce in the World Trade Organization: History and latest developments in the negotiations under the Joint Statement," Tech. Rep., 2020. [Online]. Available: https://www.iisd.org/publications/report/e-commerce-world-trade-organization-history-and-latest-developments

[17] F. Herzallah, M. M. Ayyash, and K. Ahmad, "The Impact of Language on Customer Intentions to Use Localized E-Commerce Websites in Arabic Countries: The Mediating Role of Perceived Risk and Trust," *The Journal of Asian Finance, Economics and Business*, vol. 9, no. 1, pp. 273–290, 2022. [Online]. Available: https://koreascience.kr/article/JAKO202200567709465.page

[18] S. Dahbi and C. Benmoussa, "What Hinder SMEs from Adopting E-commerce? A Multiple Case Analysis," *Procedia Computer Science*, vol. 158, pp. 811–818, 2019. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S1877050919312888

[19] Q. Jiang, C. W. Phang, C.-H. Tan, and J. Chi, "Retaining Clients in B2B E-Marketplaces: What Do SMEs Demand?" *Journal of Global Information Management*, vol. 27, no. 3, pp. 19–37, Jul. 2019. [Online]. Available: http://services.igi-global.com/resolvedoi/resolve.aspx?doi=10.4018/JGIM.2019070102

[20] L. Abdullah and H. Lim, "A Decision Making Method with Triangular Fuzzy Numbers for Unraveling the Criteria of E-Commerce," *WSEAS Transactions on Computers*, vol. 17, pp. 126–135, 2018. [Online]. Available: https://www.wseas.com/journals/articles.php?id=2229

[21] S. Buzuku and T. Kässi, "Drivers and Barriers for the Adoption of Eco-design Practices in Pulp and Paper Industry: a Case Study of Finland," *Procedia Manufacturing*, vol. 33, pp. 717–724, 2019. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S2351978919305694

[22] M. Rashidin, D. Gang, S. Javed, and M. Hasan, "The Role of Artificial Intelligence in Sustaining the E-Commerce Ecosystem: Alibaba vs. Tencent," *Journal of Global Information Management*, vol. 30, no. 8, pp. 1–25, Jun. 2022. [Online]. Available: https://services.igi-global.com/resolvedoi/resolve.aspx?doi=10.4018/JGIM.304067

[23] E. Wagner Mainardes, C. M. De Almeida, and M. de Oliveira, "e-Commerce: an analysis of the factors that antecede purchase intentions in an emerging market," *Journal of International Consumer Marketing*, vol. 31, no. 5, pp. 447–468, Oct. 2019. [Online]. Available: https://www.tandfonline.com/doi/full/10.1080/08961530.2019.1605643

[24] G. Lăzăroiu, O. Neguriţă, I. Grecu, G. Grecu, and P. C. Mitran, "Consumers' Decision-Making Process on Social Commerce Platforms: Online Trust, Perceived Risk, and Purchase Intentions," *Frontiers in Psychology*, vol. 11, p. 890, May 2020. [Online]. Available: https://www.frontiersin.org/article/10.3389/fpsyg.2020.00890/full

[25] M. O. H. K. Alzaabi, R. B. Omar, and A. R. B. Romle, "Organizational Factors and E-Commerce Adoption in SMEs of United Arab Emirates: Mediating Role of Perceived Strategic Value," *International Journal of Entrepreneurship*, vol. 25, no. 1S, pp. 1–18, Jun. 2021. [Online]. Available: https://www.abacademies.org/abstract/organizational-factors-and-ecommerce-adoption-in-smes-of-united-arab-emirates-mediating-role-of-perceived-strategic-valu-11371.html

[26] V. Jain, B. Malviya, and S. Arya, "An Overview of Electronic Commerce (e-Commerce)," *Journal of Contemporary Issues in Business and Government*, vol. 27, no. 3, Apr. 2021. [Online]. Available: https://cibg.org.au/article_10898.html

[27] B. Requena, G. Cassani, J. Tagliabue, C. Greco, and L. Lacasa, "Shopper intent prediction from clickstream e-commerce data with minimal browsing information," *Scientific Reports*, vol. 10, no. 1, p. 16983, Oct. 2020. [Online]. Available: https://www.nature.com/articles/s41598-020-73622-y

[28] S. M. Ahmed, I. Ahmad, S. Azhar, and S. Arunkumar, "Current State and Trends of E-Commerce in the Construction Industry: Analysis of a Questionnaire Survey," *Revista Ingeniería de la Construcción*, pp. 47–58, 2002.

[29] A. A. R. A. Abdullah, A. S. Madaki, I. Mohamed, N. S. Bin Mohd, and K. Ahmad, "The Impact of IT on Knowledge Sharing Environment and Management Practice," in *2022 International Conference on Cyber Resilience (ICCR)*. Dubai, United Arab Emirates: IEEE, Oct. 2022, pp. 1–7. [Online]. Available: https://ieeexplore.ieee.org/document/9995990/

[30] L. Abdullah, R. Ramli, H. O. Bakodah, and M. Othman, "Developing a causal relationship among factors of e-commerce: A decision making approach," *Journal of King Saud University - Computer and Information Sciences*, vol. 32, no. 10, pp. 1194–1201, Dec. 2020. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S1319157818309558

[31] N. Li, Q. Zhang, and Z. Zhao, "Performance Evaluation System Based on Online Monitoring of Internet of Things," *Security and Communication Networks*, vol. 2022, pp. 1–8, Apr. 2022. [Online]. Available: https://www.hindawi.com/journals/scn/2022/7745227/

[32] A. A. Ahi, N. Sinkovics, and R. R. Sinkovics, "E-commerce Policy and the Global Economy: A Path to More Inclusive Development?" *Management International Review*, vol. 63, no. 1, pp. 27–56, Feb. 2023. [Online]. Available: https://link.springer.com/10.1007/s11575-022-00490-1

[33] E. V. Zenkina, "About current trends in global e-commerce," *BENEFICIUM*, no. 1, pp. 68–73, Apr. 2022.

[34] Y. Wang, J. Anderson, S.-J. Joo, and J. R. Huscroft, "The leniency of return policy and consumers' repurchase intention in online retailing," *Industrial Management & Data Systems*, vol. 120, no. 1, pp. 21–39, Nov. 2019. [Online]. Available: https://www.emerald.com/insight/content/doi/10.1108/IMDS-01-2019-0016/full/html

[35] S. Sulhoff, "A qualitative study examining consumer personality traits in the online paid knowledge market," Ph.D. dissertation, 2023, hak cipta - Database copyright ProQuest LLC; ProQuest does not claim copyright in the individual underlying works; Terakhir diperbarui - 2023-10-10. [Online]. Available: https://www.proquest.com/dissertations-theses/qualitative-study-examining-consumer-personality/docview/2798582769/se-2