

Integrated Ensemble Model for Diabetes Mellitus Detection

Abdulaziz A Alzubaidi, Sami M Halawani, Mutasem Jarrah

Faculty of Computing and Information Technology, King Abdulaziz University, Jeddah, Saudi Arabia

Abstract—Diabetes Mellitus, commonly referred to as (DM), is a chronic illness that affects populations worldwide, leading to more complications such as renal failure, visual impairment, and cardiovascular disease, thus significantly compromising the individual's well-being of life. Detecting DM at an early stage is both challenging and a critical procedure for healthcare professionals, given that delayed diagnosis can result to difficulties in managing the progression of the disease. This study seeks to introduce an innovative stacking ensemble model for early DM detection, utilizing an ensemble of machine learning and deep learning models. Our proposed stacking model integrates multiple prediction learners, including Random Forest (RF), Convolutional Neural Network (CNN) with Long Short-Term Memory networks (CNN-LSTM), and Sequential Dense Layers (SDLs) as base learner models, with the Extreme Gradient Boosting model (XGBoost) serving as the Meta-Learner model. Findings demonstrate that our proposed model achieves a 99% accuracy on the Pima dataset and 97% accuracy on the DPD dataset in detecting diabetes mellitus disease. In conclusion, our model holds promise for developing a diagnostic tool for DM disease, and it is recommended to conduct further testing on the types of diabetes mellitus to enhance and evaluate its performance comprehensively.

Keywords—Diabetes mellitus; machine learning; deep learning; stacking; ensemble learning; RF; CNN-LSTM; SDLs; XGBoost

I. INTRODUCTION

Diabetes mellitus carries a significant health risk and increases the likelihood of getting cardiovascular disease and more complications, which makes our lives suffer [1]. Insulin is essential for controlling blood glucose work, regulating the anabolism of carbohydrates, promoting physical growth, supervising cell division, and monitoring the anabolic activities of proteins and fats [2]. As a result, DM significantly impairs people's daily life and increases their risk of getting chronic illnesses such as cardiovascular disorders, renal failure, and sightlessness [3]. These disorders raise death rates [4]. In 2019, it was estimated that 463 million individuals worldwide suffered from diabetes [5].

There are two main types of diabetes mellitus (DM): Type-1 Diabetes, an autoimmune sickness that destroys the pancreatic beta cells that produce insulin, and the second type of Diabetes, a chronic that often raises blood sugar levels [6]. It can be challenging to differentiate between these kinds and choose the best course of action because doctors sometimes dispute about the best way to diagnose a patient [7]. Globally, diabetes is becoming more common, especially in countries with middle incomes [8]. Therefore, research on diabetes

prediction through machine learning (ML) techniques is needed to help specialists build the best possible treatment plans. By 2030, the Global Sustainable Development Group wants to eradicate diabetes-related early death [8]. Consequently, scholars are consistently investigating various facets of diabetes mellitus. A range of machine learning methods, including the Random Forest and XGBoost algorithms, are utilized in this endeavor, each providing special benefits for classification procedures [9-10-11-12-13]. Convolutional Neural Networks (CNNs) with Long Short-Term Memory architecture and Multilayer perceptron's (MLPs) are two popular deep learning (DL) techniques that offer strong frameworks for managing sequential data and classification tasks [14-15-16].

In the healthcare sector, detecting DM at an early stage is challenging. Patient data is collected, including their ages, the mass of the body, the thickness of the skinfold in the triceps, insulin in the blood, plasma glucose level, the diastolic-blood-pressure, and other variables. Patients then turn to doctors for specialist care. The doctor's medicating process becomes harder because of the lengthy, weeks-long decision-making process that depends on the doctor's expertise and experience [17].

Healthcare science research is currently supported by a wide range of publicly available medical databases. But managing such massive volumes of data by humans is frequently difficult, if not infeasible. Deep learning techniques deliver a solution because they grow and mimic how humans thinking. That happens by providing data at several tiers and successfully resolves the selectivity-invariance issue [18]. In the discipline of medicine, deep learning algorithms have several uses, especially in the diagnosis sector. Much research continually demonstrates the superior performance of deep learning approaches over conventional machine learning techniques. These algorithms are superior to other methods in terms of performance and their ability to lower classification error rates [19].

Numerous deep-learning methods are promising in the medical field. The CNN-LSTM architecture is an effective structure suited for handling classification problems and time-series processing of data. Convolutional neural networks (CNNs) and long short-term memory (LSTM) networks are coupled in this design to make use of their respective advantages in processing sequential input and extracting pertinent information [14]. A crucial aspect of deep learning is the Multilayer Perceptron (MLP), a conventional neural network architecture. Because of its historical relevance and fundamental role in the evolution of neural networks, this

artificial neural network with feedforward algorithms is frequently regarded as a (classical) model [15]. Under deep learning techniques, the Keras framework is excellent at developing sequential stacking models, such as dense layered models, which are well-known for their ability to forecast time-series data. These thick layers are typically fully connected, with each node in one layer connecting to every node in the next, resulting in a chain-like structure [20].

Given the extensive information on diabetes mellitus and the diverse techniques employed for its prediction, we propose an enhanced stacking ensemble model that combines machine and deep learning models for DM prediction.

The contributions for this study are as follows:

- Developing an innovative stacking ensemble model for detecting Diabetes Mellitus by integrating machine learning and deep learning models, utilizing an ensemble of RF, CNN-LSTM, and SDLs models as base learners, with the XGBoost model serving as a meta-learner.
- Merging a combination of ML and DL techniques, including ADASYN, RFECV, GridSearchCV, and Optuna, aimed at refining the performance of the proposed stacking model in detecting DM.

The format of this document is as follows: Section II provides an overview of related work in the area. Section III details the materials and methods used in our research. Section IV presents the experimental setup for our ensemble model. In Section V, we discuss the performance measures. Our results and discussion are covered in Sections VI and VII respectively. In the end, the conclusions are addressed in Section VIII.

II. RELATED WORK

Ensemble learning is a mathematical and analytical methodology that simulates human learning by combining diverse machine learning models to yield more accurate predictions [21-22]. Dutta et al. highlight the importance of ML-based ensemble models in diabetes prediction, advocating the exploration of deep learning techniques alongside ensemble learning, particularly through the stacking method [23]. Ganie and Malik [24] address ensemble methods for detecting Type-2 Diabetes Mellitus, including Bagging, which emphasizes aspects of lifestyle and uses the SMOTE technique that is; (artificial minority oversampling) for dataset rebalancing, validated using cross-validation techniques.

Laila and colleagues [25] investigated effective ensemble algorithms for early-stage prediction of diabetic risks, employing 17 features sourced from the UCI library encompassing diverse datasets. The study employed predictive models, like Ada-Boost, bootstrap aggregation, and RF, and evaluated the proposed model's accuracy and other performance measures. The Random Forest ensemble approach outperformed AdaBoost and Bagging concerning accuracy, scoring 97 percent.

Prasad and Geetha [26] propose an ensemble model utilizing ensemble approaches like bootstrap aggregation, RF, and Ada-boost, together with classification techniques such as

(Naive Bayes). Joshi et al. [27] used the logistic regression model and the decision tree to predict diabetes type-2 in the Pima dataset, with an accuracy reaching 78%.

Javale and Desai [28] delved into elevating healthcare information analytics through the application of an ensemble methodology using machine learning, specifically addressing challenges posed by unbalanced datasets. Their approach incorporated SMOTE and adaptive ADASYN oversampling methods. Various performance evaluation approaches were employed, like train-test split and K-folding. The diabetes dataset underwent an ensemble strategy utilizing the average Stacking-C technique, encompassing classifiers such as K-Nearest Neighbors, Random Forest, and others.

Early DM detection can save human lives and help healthcare workers to control the illness. Many individuals diagnosed with diabetes are unaware of the risk aspects they may be exposed to before the diagnosis happens [25]. Patil et al. [29] introduced an approach for predicting Type-2 Diabetes Mellitus (T2DM) utilizing a stacking ensemble model. The primary aim is to minimize the period between diabetes disease detection and medical checkups. Proposed non-dominated sorting genetic algorithm (second version) stacking model compared against Boosting, bootstrap aggregation, RF, and Random Subspace techniques. Results demonstrate that the proposed ensemble model outperforms the traditional ensemble models, achieving an accuracy of 81 percent.

Zhou et al. [30] introduce an enhanced deep neural network algorithm for diabetes prediction, concentrating on type-1 & type-2 diabetes. Deep learning algorithms, such as Dense Layer Neural Networks [31], MLP models [32], and CNN-based architectures, have demonstrated success in various aspects of diabetes-related activities. Zhu et al.'s [33] comprehensive study emphasizes the superiority of deep learning over traditional machine learning in diabetes diagnosis, glucose management, and complication diagnosis. CNNs are particularly praised for clinical imaging issues, offering feature extraction capabilities. CNN-based architectures are employed to analyze clinical scans, diagnose complications, and assess food images for individuals with diabetes. The emerging field of AI and deep learning holds promising prospects for advancing diabetes applications [16].

Sainte et al. [34] explore new techniques for diabetes prediction, incorporating a wide range of DL methods. A CNN-LSTM model emerges as the most accurate, recording a 95% accuracy in predicting diabetes. The study compares the accuracy of DL models (95%) with that of ML models (68–74%), showcasing the superior performance of deep learning. Kim et al. [16] utilize various deep-learning models, including RNNs, for blood glucose predictions. Gupta et al. [31] offered a Deep Dense Layer Neural Network model for diabetes prediction utilizing the Pima dataset with 768 samples. It was getting an accuracy of 84%. Dense layers are densely connected, meaning each neuron acquires input from all the neurons in the previous layer (each neuron is linked to all neurons of the last layer [31]). The proposed DDLNN model was evaluated by the cross-validation technique to optimize the model performance. Majority voting was utilized to select the best outcomes among the models [31]. Deep learning

techniques have shown remarkable success in various fields, including disease prediction and diagnosis [31]. Table I summarizes the most notable research based on their limits and advantages, as well as the data sources.

TABLE I. SIGNIFICANT INVESTIGATIONS IN DETECTING DIABETES MELLITUS

Ref.	DATASET SOURCES	Advantages	Limitations
M.Gollapalli et al, 2022, [35]	Healthcare institution (KFUH), Saudi Arabia	Applying the Cross-validation method in the training process substantially improves the performance of the ML models.	Insufficient utilization of deep learning models for enhancing outcomes.
A. Dutta et al, 2022,[23]	DDC dataset from Bangladesh	Employing the Grid Search-CV technique to optimize the model's performance through fine-tuning its hyperparameters.	Need for more clinical data to enhance the outcomes.
A. Singh et al, 2021, [36]	PIMA Indian diabetes, USA	Implementing the Recursive Feature Elimination technique minimizes the dataset's range of features, making the model more reliable with accurate results.	Implementation of the suggested approach in medical life assessment.
A. Syed & T. Khan, 2020, [37]	PIMA Indian diabetes, USA	Utilizing the SMOTE technique to balance classes within the dataset, thereby preventing overfitting.	Limited variety in the medical dataset under examination.
Chou et al, 2023, [38]	Taipei Municipal medical center, Taiwan	Utilizing Microsoft Machine Learning Studio platform for training the models.	Insufficient utilization of deep learning models for enhancing outcomes.

III. MATERIALS AND METHODS

This research introduces an ensemble stacking for detecting DM disease, comprising two crucial construction levels. The first level, termed base learners, involves the preparation, training, and initial predictions by a combination of ML and DL models. Predictions from these initial learners are used as inputs for a new model called meta-learner located at the second level and train from the information provided to come up with the final prediction. For our base learners, we have chosen Random Forest, CNN-LSTM, and SDLs due to their distinctive capabilities in classification problems. The XGBoost model is employed at the second level (meta-learner), which assists significantly in managing unbalanced dataset classes' via lowering the loss function and boosting the weight of incorrectly categorized categories. To optimize our base learners, we apply GridSearchCV and Optuna technologies, aiming to achieve the best possible results for Random Forest, CNN-LSTM, and SDLs. Our proposed stacking model incorporates the cross-validation method with several iterations to obtain best findings. Additionally, an adaptive oversampling technique (ADASYN) is implemented

to balance the classes of the Pima dataset investigated in our study and increase the size of the dataset in a way that does not include overfitting issues. Fig. 1 outlines the schema behind our suggested stacking model.

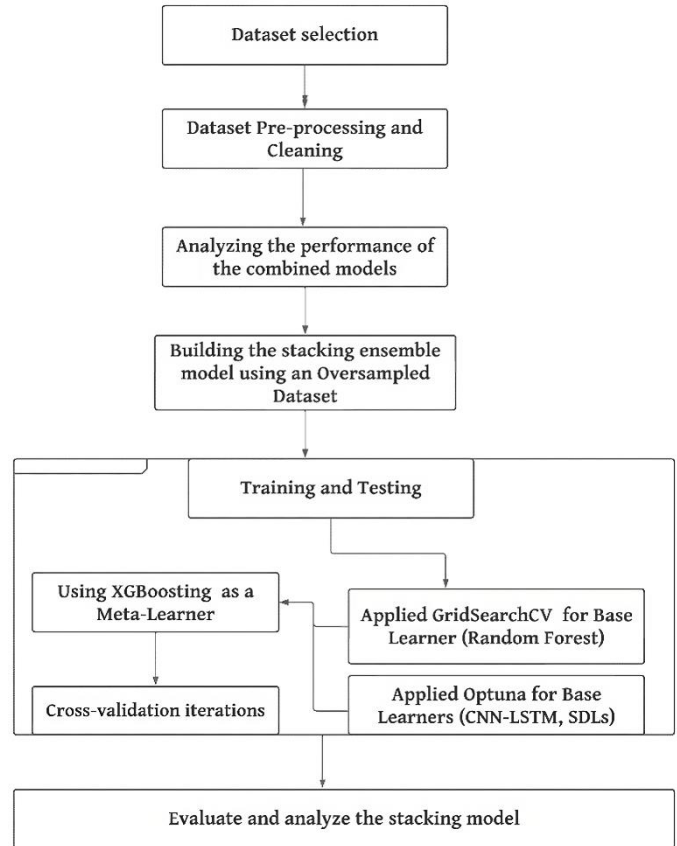


Fig. 1. Proposed stacking model.

A. Stacking

Stacking is an ensemble technique that leverages a "meta-model" to enhance predictive performance by integrating multiple base learners' predictions. In the stacking methodology, there are two distinct levels of model building. At level 1, a set of diverse base learners, each trained on a specific split from the dataset, then produces separated predictions. These predictions serve as input for the meta-learner at level 2 [35].

The goal of the second-level meta-learner is to produce predictions that are reliable and accurate by efficiently combining them [35]. Our proposed model is constructed using the stacking ensemble method, incorporating additional contributions, such as employing cross-validation techniques and leveraging the GridSearchCV hyperparameter tuning method for the Random Forest (RF) base learner. Additionally, Optuna is employed for the CNN-LSTM and SDLs deep learning learners.

The ensemble stacking methodology, as depicted in Fig. 2, involves multiple k-folds cross-validations (m*n) traversing the training dataset and each base learners models. Subsequently, the predictions (m*M) from multiple base learners are entered as inputs to the meta-learner, which learns

from this collective information to generate the final prediction. This approach enhances the model's efficiency for generalization and produces accurate predictions by leveraging the diverse insights from individual base learners and optimizing their combination through the meta-learner. It outperforms other ensemble models in prediction performance, so we have chosen to apply it in our study. The stacking approach aims to provide us with the concept of meta-learning, which can minimize ML model generalization errors [29].

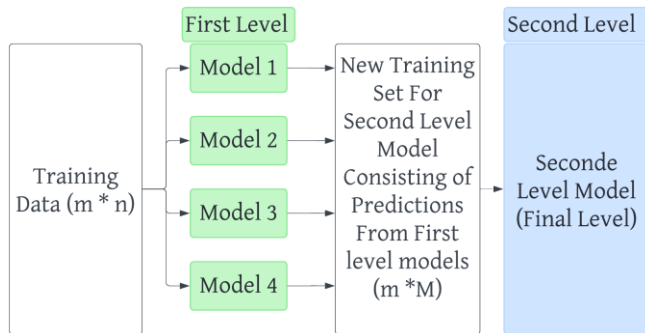


Fig. 2. Stacking ensemble methodology.

B. ADASYN (Adaptive Synthetic sampling)

ADASYN is a data augmentation strategy that addresses unbalanced classes in ML datasets, notably for classification problems. Unbalanced classes happen when the number of samples in specific classes is quite different, resulting in biased model performance [39]. This strategy focuses on the minority class by creating synthetic samples for underrepresented cases. Unlike classic oversampling methods, ADASYN tailors the synthetic sample creation procedure to the local density of minority class instances. This adaptive strategy tries to reduce the risk of overfitting while also improving the model's generalization to previously hidden data [39]. By dynamically altering the creation of synthetic samples, ADASYN aims to improve models' learning ability in the presence of unbalanced data sets, resulting in more robust and accurate classification results [39].

C. Random Forest

Random Forest is an approach to machine learning used for both detection and forecasting problems and plays an important role in optimizing forecasting models [40]. This technique is based on an assembly of trees of decisions, where each tree divides the inputs into categories using a sequence of possibilities [40]. Random Forest adds randomization to the building of decision trees (DT) by:

- 1) For each decision tree, samples are chosen randomly from the initial dataset (training part) with replacement.
- 2) To construct each tree, a subset of variables is randomly chosen.

After generating several result trees, Random Forest combines each of the tree forecasts via votes or averages to reach an overall prediction result. Random Forest can avoid overfitting during the model training. It also uses significance statistics to analyze the effect of every variable on categorization, providing helpful findings [40]. RF is

commonly used in many applications, such as recognizing images, recognition of words and pricing forecasting [40].

D. CNN-LSTM

The CNN-LSTM architecture combines two types of networks: convolutional neural networks and long short-term memory networks. Which is designed to understand the forms and patterns in data and how these shapes evolve over time [14-15-41].

CNNs use convolutional layers to apply filters to input data and extract features hierarchically. These layers detect patterns at various levels of abstraction, gradually learning representations with increasing complexity as data flows through the network. CNNs help in comprehending the patterns and interactions between distinct data components by first scanning the data and assessing spatial arrangements and connections. For example, in diabetes, this application could involve monitoring blood sugar levels throughout the day, analyzing a patient's reactions to medicines or lifestyle modifications, and recording how these factors change over weeks or months [14-15-41].

Subsequently, the LSTM, a form of recurrent neural network, specializes in processing and comprehending sequential input such as time series, text, and speech. LSTM is particularly effective at remembering patterns across time and assessing changes and advancements within a sequence. In the case of diabetes, this could entail tracking changes in blood sugar levels throughout the day, analyzing patient reactions to therapies or lifestyle changes, and understanding how these aspects play out over weeks or months [14-15-41]. The combination of these two networks enables the model to understand the spatial positioning of pieces in the data and their time evolution [17-21].

E. Sequential Dense Layers (SDLs)

The Sequential Dense Layers Model, built with the Keras framework, is a modern neural network structure that layered stack [20]. This model is based on tightly connected layers, which form a chain-like structure with each node in one layer connecting to every node in the next layer [20]. This design, similar to an ordered manufacturing line, allows for the systematic flow of information from one processing unit to the next, enabling methodical data analysis [32]. The SDLs model excels in terms of clarity and efficiency, establishing a wall layer by layer, with each layer executing specialized computations on the data [20]. Its sequential technique facilitates thorough data processing and feature extraction, and it is adept at discovering nuanced patterns in large datasets [32]. Furthermore, the model excels at extracting relevant temporal features, which improves its capacity to detect significant patterns in time series data. This skill is essential for making good forecasts in time-series forecasting jobs [20-32].

F. GridSearchCV

GridSearchCV is a well-known method of ML applications for its ability to identify optimal hyperparameter values for a certain model. These hyperparameters, like the number of batches in neural networks, can control the configuration and behavior of the model. GridSearchCV involves a systematic exploration of various hyperparameter combinations and

assessing the performance using a Cross-validation method. It entails establishing a collection of potential hyperparameter parameters, training and evaluating that model through every combination, and finally picking the best combination that produces the most accurate results [23]. This systematic approach improves the model's performance and mitigates the overfitting occurrences.

G. Optuna

Optuna, introduced in 2019 by Akiba et al., is a free hyperparameter tuning framework designed to streamline the trial-and-error process in optimizing model training accuracy [42-43]. It employs a targeted API-based strategy, allowing the automatic optimization of hyperparameter values for various machine learning algorithms within a specified trial limit. Versatile and 'pythonic' in operation, Optuna makes no distinction between machine learning and deep learning frameworks [42-43]. In this research, Optuna was utilized to tune hyperparameters such as dense layers units, batch size, activation function, loss function and others. The optimized values, including "sigmoid" activation function, three Dense layers units, batch size, reduction factor, and min early stopping rate, resulted in the best validation accuracy with the "binary cross entropy" loss function. Optuna's approach significantly enhances hyperparameter tuning efficiency, contributing to improved model performance and accuracy.

H. Recursive Feature Elimination with Cross-Validation

RFECV is a wrapping approach that eliminates unwanted features, enhances model generalization by preserving independent and effective features while eliminating duplicate and weak ones with minimal impact on training error. Employing an iterative feature ranking method, it conducts backward feature reduction. Initially, the model is built with the entire feature set, ranking each feature based on relevance. The least significant feature is then eliminated, and the process repeats iteratively. The sequence number, T , serves as the feature ranking, and T_i represents the top-ranked features used in each iteration. The final model incorporates the best-performing features, and the optimal value of T_i is selected [44], [45]. In our research, this technique was applied, resulting in the identification of the best training features.

I. Extreme Gradient-Boosting

XGBoost, derived from the gradient-boosting decision tree developed by Tianqi Chen et al. [46], is a well-known machine learning model known for its adaptability and efficiency. Unlike GBDT, XGBoost employs regularization methods to minimize model complexity and reduce overfitting. The algorithm employs an approximation approach to enhance gradient boosting, focusing on finding the optimal split for improved expandability and efficiency. XGBoost introduces features like parallel operations and early stopping to expedite model execution, with the added advantage of increased classification accuracy [9]. According to Zhao et al. [10], XGBoost effectively prevents overfitting in training models. Additionally, its built-in parallel processing ability allows for higher training speeds.

Furthermore, the XGBoost model can gain insight from unbalanced learning data by adjusting the weights of classes.

XGBoost is among the best models for dealing with unbalanced datasets, particularly if the class distribution has low variance [11].

XGBoost works with a number of weak learners and enhances their performance via an enhancement strategy.

In conclusion, XGBoost stands out as a potent and widely embraced tool in the field of machine learning. Its effectiveness extends to solving intricate problems and significantly enhancing the performance of predictive models.

J. Cross-validation

Cross-validation (CV) in machine learning is a common resampling data approach to verify the generalization of a prediction model without going overboard. It entails partitioning the dataset across folds throughout the training and testing phases, with each "fold" being a subset generated for analysis. The dataset's samples will be allocated to the previously mentioned (the folds) randomly with no repetition. Throughout each iteration, the $k-1$ subset serves as the training set employed to train the model, while the remaining subset, known as the "unseen dataset," is used to assess the model's performance. This iterative method will continue until all k -subsets have been used as validation sets [47]. Fig. 3 depicts the cross-validation method [48]. To improve the results, our stacking model enabled cross-validation using 5-fold splits.

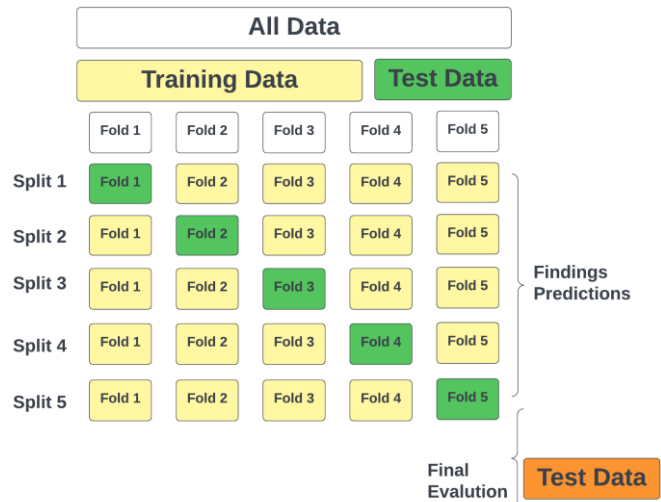


Fig. 3. Cross-validation technique.

K. Study Data

The Pima dataset is a popular dataset for machine learning and data analytics. It belongs to the Pima indigenous community from Arizona, USA. This dataset is often utilized for predicting diabetes diseases using machine learning models [35]. It is accessible publicly in the Kaggle repository at <https://www.kaggle.com/datasets/uciml/pima-indians-diabetes-database>.

This dataset included 268 patients who have diabetes and 500 who have no diabetes, with 8 characteristics listed in Table II. [35].

The Pima dataset serves as a common benchmark for exploring a range of machine learning models, such as random

forest and XGboost. However, it's important to recognize that while the PIMA dataset is valuable for instructional uses and testing machine learning models, it does have limits. These limitations originate from its small size, values missed, and possible biases. As a result, it is prudent to take care when making assumptions or creating prediction models based exclusively on this dataset for applications in real life.

TABLE II. PIMA DATASET METADATA

Dataset columns			
#	Column	Count	Datatype
0	Pregnancies	768	int64
1	Glucose	768	int64
2	Blood Pressure	768	int64
3	Skin Thickness	768	int64
4	Insulin	768	int64
5	BMI	768	float64
6	Diabetes Pedigree Function	768	float64
7	Age	768	int64
8	Outcome	768	int64

L. Validation Dataset

To assess the efficacy of our introduced stacking model, we conducted validation using a newly acquired diabetes dataset. This dataset is known as the Diabetes Prediction Dataset (DPD). It is a publicly available compilation of electronic health records. These records encompass digital archives of patients' medical histories, diagnoses, treatments, and outcomes. EHRs are frequently collected and stored by healthcare institutions such as medical facilities as a component of their usual clinical protocols. With a large dataset comprising about 100,000 records, the DPD dataset can contribute perfectly to evaluating the performance of the proposed model.

It can be accessed on the Kaggle through the following link: <https://www.kaggle.com/datasets/iammustafatz/diabetes-prediction-dataset>. Table III outlines the features present in the DPD dataset.

TABLE III. DPD DATASET METADATA

Dataset Columns			
#	Column	Null values	Datatype
0	Gender	Not null	object
1	Age	Not null	float64
2	hypertension	Not null	int64
3	Heart disease	Not null	int64
4	Smoking history	Not null	object
5	BMI	Not null	float64
6	HbA1c_level	Not null	float64
7	Blood_glucose_level	Not null	int64
8	Diabetes	Not null	int64

IV. EXPERIMENTAL SETUP

In this study, we employed Jupyter Notebook to construct our stacking model, utilizing a Microsoft Intel (R) Core i5-1035G7 CPU operating at 1.20 GHz and 8 GB of RAM. The Pima dataset was chosen to examine our proposed model. Due to some defects found in the dataset, we decided to preprocess the dataset and clean up these defects, such as zero values in some feature columns. We handle this issue by imputing the mean or median for each feature column based on his data distribution. The base learners' models were initialized with a Random Forest model, and we utilized the Grid-searchCV Hyperparameters Tuner to optimize their performance. The hyperparameters in the tuning process were such as bootstrap training, min_samples_split, and n_estimators. Second and third-base learners, CNN-LSTM, and SDL models optimized via the Optuna optimizer have the following parameters: number of convolutional layers, filter and kernel sizes, activation functions, dropout rates, early shopping, and pooling strategies. Also, Optuna was used for the third base learner, Sequential Dense Layers, with three dense layers.

We implemented the XGBoost model as a meta-learner to tackle the problem of imbalanced datasets, along with the ADASYN (Adaptive Synthetic Sampling) technique for oversampling our study dataset, which frequently results in overfitting and inconsistencies in the results. XGBoost uses an ensemble learning approach, enabling us to effectively handle unbalanced dataset classes. Additionally, we used a cross-validation method with a 5-fold validation for the training of Random Forest (RF), the base learner, and the stacking model in general. We used the GridSearchCV hyperparameter method to improve the RF performance. The Recursive Feature Elimination with Cross-Validation technique was applied to enhance model performance by systematically eliminating less informative features during the training process. Lastly, we tested our suggested stacking model using a DPD dataset with 100,000 records.

V. PERFORMANCE MEASURE

We evaluated the effectiveness of our ensemble learners using the next metrics:

A. Accuracy

It measures the percentage of the predication data that were accurately predicted from all given data. It assesses how well the algorithm can distinguish between positive and negative instances [35]. The definition of accuracy is captured by Eq. (1).

$$Acc = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

Assuming the outcome for patient P indicates that there is a presence of diabetes (+DM) or NOT (-DM).

- True positive (TP): if the outcome is +DM, then P is diabetic.
- True negative (TN): if the outcome is -DM, then P is non-diabetic.
- False positive (FP): if the outcome is +DM, then P is non-diabetic.

- False negative (FN): if the outcome is -DM, then P is diabetic.

B. Precision

It evaluates how well the positive values are predicted. When the model accurately categorizes predictions as positive when it asserts, they are, so we have a strong precision score [35]. Eq. (2) can be utilized to represent precision.

$$Precision = \frac{TP}{TP+FP} \quad (2)$$

Assuming the outcome for patient P indicates that there is a presence of diabetes (+DM) or NOT (-DM).

- True positive (TP): if the outcome is +DM, then P is diabetic.
- False positive (FP): if the outcome is +DM, then P is non-diabetic.

C. Recall

Expressed as the sensitivity or the actual positive ratio. The remarkable recall rating means that most positive predicted values are correctly true according to the model data [35]. Recall can be expressed using Eq. (3).

$$Recall = \frac{TP}{TP+FN} \quad (3)$$

Assuming the outcome for patient P indicates that there is a presence of diabetes (+DM) or NOT (-DM).

- True positive (TP): if the outcome is +DM, then P is diabetic.
- False negative (FN): if the outcome is -DM, then P is diabetic.

D. Cohen's Kappa Metric

It assesses the agreement between two class targets and evaluates the concordance between the model predictions and the real results. It is valuable when confronted with unbalanced classes [35]. Eq. (4) can be employed to denote Cohen's Kappa Score.

$$CKS = \frac{P_o - P_e}{1 - P_e} \quad (4)$$

Here, P₀ shows the models' accuracy, whereas P_e reflects the correlation between the expected and real values [35].

E. F1-Score

The F1-score, a standard measure used in binary classification problems, is determined as a harmonic average of precision and recall, considering both precision in properly identifying positive cases and recall in capturing all positive cases [53]. Provided by the Eq. (5):

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (5)$$

F. Receiver Operating Characteristic: Area Under the Curve (ROC-AUC)

It evaluates the effectiveness of a classification model by assessing its ability to distinguish between positive and

negative instances across various decision thresholds. It is particularly useful for imbalanced classes or when the model needs to be evaluated at different levels of sensitivity and specificity [23].

Utilizing a diverse range of metrics, including accuracy, precision, recall, Cohen's Kappa, and ROC-AUC, is paramount for gaining deep insights into the performance of classification models. These metrics offer nuanced perspectives on the model's strengths and weaknesses, illuminating its ability to accurately classify instances, detect true positives while minimizing false positives and false negatives, and maintain consistency in predictions. By comprehensively assessing these metrics, practitioners can decipher the intricacies of model behavior and make informed decisions regarding model selection, optimization, and deployment strategies.

VI. RESULTS

In this study, we constructed a stacking ensemble model for detecting diabetes mellitus disease utilizing an ensemble of ML and DL techniques. Our proposed model comprised base learners' models including random forest, CNN-LSTM, and SDLs, with the extreme gradient boosting model working as the meta-learner. Additionally, several methods like ADASYN, Optuna, RFECV, cross-validation, and GridSearchCV were employed to refine the model's performance. Furthermore, we addressed the issue of zeros in certain feature columns of the Pima dataset by replacing them with values derived from the median or median of data distribution based on whether the type of distribution is normal or skewed. The proposed model achieved a remarkable 99% accuracy in detecting diabetes mellitus disease using the Pima dataset. Moreover, we assessed the model's effectiveness on a sizable dataset comprising approximately 100,000 records, referred to as the DPD dataset, achieving an accuracy of 97%. More details are focused on the discussion section.

A. Stacking Model Performance

In this study, several performance metrics were applied to evaluate the effectiveness of our proposed model, which integrates both ML and DL techniques. The outcomes of multiple metrics are presented in Table IV.

TABLE IV. THE STACKING MODEL PERFORMANCE RESULTS

Accuracy Score	ROC-AUC Score	Cohen's Kappa Score	Precision Score	Recall Score	Accuracy Score
0.9887	0.99	0.98970	0.9934	0.995	0.9887

VII. DISCUSSION

A. Results using Oversampling ADASYN

In this experiment, an ensemble of Machine Learning and Deep Learning techniques was used to detect the occurrence of diabetes mellitus disease. Fig. 1 elucidates the methodology of our proposed model, where each Random Forest (RF), CNN-LSTM, and SDLs was initialized as initial learners, and an XGBoost model was designated as the meta-learner. Additionally, the GridSearchCV Hyperparameters optimizer was utilized to determine best findings for the Random Forest model, also Optuna optimizer for CNN-LSTM and SDLs

models. To address the challenge of imbalanced datasets and mitigate overfitting, we applied the ADASYN (Adaptive Synthetic Sampling) technique for oversampling the study data. This increased the Pima dataset from 768 records to 4870 records in an adaptive manner to avoid overfitting and increase the size of the data set. Additionally, the XGBoost model served as a meta-learner, utilizing ensemble learning to handle class imbalances effectively.

For robust evaluation, a CV method was used in our proposed model, employing 5-fold cross-validation. Table V presents the comprehensive outcomes of our proposed model. Furthermore, Table VI displays the results of both the base models and the final model.

TABLE V. THE STACKING MODEL FINDINGS

Targets	Precision	Recall	F1-score support	Support
0	1.0	0.99	1.0	515
1	0.99	1.0	0.99	459
Model Accuracy			0.99	974
Macro average	0.99	0.99	0.99	974
Weighted average	0.99	0.99	0.99	974

TABLE VI. THE META AND BASE MODELS FINDINGS

#	Model	Score
0	Sequential Dense Layers	0.98
1	CNN-LSTM	0.94
2	Random Forest	0.98
3	Stacking Model	0.98

B. Results using ADASYN & RFECV

In this experiment, we employed the ADASYN technique that mentioned in point 7.1 plus the RFECV technique, which is a pivotal step in enhancing the performance of a stacking ensemble model. The aim was to optimize feature selection, thereby improving the overall effectiveness of the ensemble. Our investigation successfully identified a set of six optimal features. These features are glucose, blood pressure, insulin, BMI, diabetes pedigree function, and age. After that, we examined the proposed model on the selected features, and the results were as follows in Table VII:

TABLE VII. THE STACKING MODEL FINDINGS USING RFECV

Targets	Precision	Recall	F1-score support	Support
0	0.99	0.96	0.97	515
1	0.95	0.99	0.97	459
Model Accuracy			0.97	974
Macro average	0.97	0.97	0.97	974
Weighted average	0.97	0.97	0.97	974

C. Results on the Validation Dataset

In this latest experiment, the proposed stacking model underwent rigorous testing on a substantial test dataset comprising approximately 100,000 electronic patient records. Notably, the experiment deliberately abstained from employing the RFECV technique. The results obtained from this comprehensive evaluation revealed significant new perspectives into our model performance and its ability to handle the complexity inherent in the diverse patient data. These findings contribute valuable information to the ongoing discourse on the effectiveness of stacking models in healthcare analytics, shedding light on their potential without the aid of feature selection techniques. Table VIII shows the results after applying the proposed model.

TABLE VIII. THE STACKING MODEL FINDINGS ON DPD DATASET

Targets	Precision	Recall	F1-score support	Support
0	0.97	1.0	0.98	18292
1	0.97	0.69	0.81	1708
Model Accuracy			0.97	20000
Macro average	0.97	0.85	0.90	20000
Weighted average	0.97	0.97	0.97	20000

D. Comparative Evaluation with Existing Work

1) *First study:* The ensemble stacking model for diabetes detection proposed by S. Härner and D. Ekman in 2022 [13] integrates various ML models, such as DT and Naive Bayes (NB) models, and leverages Pima Dataset. The study's findings demonstrate a 75.56% accuracy in predicting diabetes using this stacking approach. However, it is important to note certain limitations highlighted in the study, like the absence of a hyperparameter tuner to systematically search for optimal hyperparameters for the base learners within the stacking model.

2) *Second study:* Patil et al. (2023) [29] introduced a stacking model for diabetes detection, utilizing ML techniques including decision trees, NB, multilayer perceptron, support vector machines SVMs, and K-Nearest Neighbor KNN. This study utilized the Pima dataset and reported an 82% accuracy in diabetes prediction using the stacking model. Also, the absence of any mention of the CV technique in their suggested model is noteworthy, as this technique is crucial for robustly constructing the stacking model. Additionally, the study did not employ an optimizer to systematically search for optimal hyperparameters during the training of base learners, which could potentially enhance the overall model performance.

3) *Third study:* Lei Qin (2022) [49] presented a stacking model for diabetes detection that combined diverse ML techniques such as LR, KNN, DT, Gaussian Naive Bayes, and SVMs. Employing the Pima dataset, findings showed an accuracy of 81.6% in predicting diabetes. However, there are lack of a hyperparameter tuner for the hyperparameters in initial learner model training may restrict the pursuit of improved findings. Furthermore, the study acknowledged the

challenge posed by the dataset's restricted size, which could influence the achievement of ideal outcomes.

4) *Forth study*: Kumari et al. (2021) [50] suggested a soft voting approach for diabetes prediction, incorporating ML techniques like RF, LR and NB. Utilizing the same Pima dataset, their findings demonstrated a 79.04% accuracy in predicting diabetes. Notably, the proposed methodology excluded CV method, which is assuring reliability by evaluating the efficacy of the ML models among diverse of data samples, hence increasing the overall effectiveness of the predictions to the proposed model.

5) *Fifth study*: Bhopte and Rai (2022) [51] employed the CNN-LSTM model to detect DM disease using Pima dataset, achieving an accuracy of 89.30%. However, the study acknowledged limitations, specifically the absence of an optimizer for systematically searching hyperparameters to optimize results during the model development process.

6) *Sixth study*: Niharika et al. (2022) [52] utilized the MLP model to predict diabetes, employing the Pima dataset and achieving a 77% accuracy. The study highlighted limitations, including concerns about the sample size used and

the necessity for employing diverse DL techniques for disease prediction, specifically regarding various types of diabetes.

So, Compared to all six studies [13-29-49-50-51-52], our proposed stacking model exhibits higher accuracy in diabetes mellitus detection. In our methodology, we harnessed the power of GridsearchCV and Optuna optimizers to find the optimal hyperparameters in the base learners that are supported with dataset using the oversampling ADASYN method. Notably, these optimization techniques were absent in the First, Second, Third, Fourth, and Fifth Studies.

Additionally, Optuna was not utilized in the Sixth Study. The integration of these optimization techniques significantly improved the learning process of our base learners, resulting in the extraction of optimal results. It is crucial to highlight that the [29–50–51] research studies did not apply cross-validation, which is a critical method for evaluating a prediction model's generalization ability. Whereas our approach used this method for performing k-fold cross-validation iterations, this method successfully analyzed and prevented overfitting, significantly improving the quality of our predictive model. Table IX emphasizes the differences and benefits of our study compared to others, underscoring the major improvements in our proposed model.

TABLE IX. ASSESSMENT WITH THE EXISTING INVESTIGATIONS

Authors	Approaches	Studies Dataset	Accuracy
S. Härner and D. Ekman (2022)	Ensemble Stacking approach. (DT, NB, Cross-validation).	Pima dataset	75.6%
Patil et al (2023)	Ensemble Stacking approach. (DT, NB, multilayer perceptron, SVM, and KNN).	Pima dataset	81.9%
Bhopte and Rai (2022)	CNN-LSTM.	Pima dataset	89.30%
Niharika et al. (2022)	Multilayer perceptron (MLP), GridSearchCV.	Pima dataset	77%
Lei Qin (2022)	Ensemble Stacking approach. (LR, K-NN, DT, Gaussian Naive Bayes, and SVM).	Pima dataset	82%
Kumari et al (2021)	Ensemble Soft voting approach. (RF, LR, and NB).	Pima dataset	79.04%
Our proposed model	Ensemble Stacking approach. (RF, CNN-LSTM, SDLs, XGboost) ADASYN GridSearchCV, Optuna, Cross-validation, RFECV.	Pima Dataset	99%
Our proposed model on the validation dataset	Ensemble Stacking approach. (RF, CNN-LSTM, SDLs, XGboost) GridSearchCV, Cross-validation.	DPD Dataset	97%

VIII. CONCLUSION

Diabetes mellitus is a prevalent condition that poses a significant threat to public health, giving rise to various severe complications like renal failure, cardiovascular disorders, and sightlessness. In our study, we introduce an innovative stacking model designed for detecting diabetes mellitus diasease at early stage, utilizing the Pima dataset and integrating both ML and DL models using ADASYN oversampling method. The ensemble comprises Random Forest (RF), CNN-LTSM, and SDLs as base learner models, with XGBoost serving as the Meta-Learner model. Cross-validation techniques were applied for the meta learner. Moreover, incorporating Grid Search optimization for the RF model and adopting Optuna optimization for the CNN-LTSM and SDLs models to secure optimal results. To mitigate the challenges posed by an imbalanced dataset, which can lead to over-fitting and unexpected outcomes, the XGBoost model is employed as a

meta-learner. Additionally, our study dataset underwent preprocessing to address zero values, which could adversely impact prediction accuracy, particularly in columns like blood and glucose. To tackle this issue, zero values were replaced with the median or mean from the total values in each feature column, considering the feature data distribution type. REFCV technique was applied to our proposed model. The results highlight the efficacy of our proposed model in detecting DM, achieving an accuracy of 99% across the Pima dataset and 97% in the DPD dataset. As a recommendation, our stacking model holds potential for deployment in diagnostic applications for diabetes mellitus. Furthermore, its performance can be validated on larger and more diverse datasets to enhance precision. Additionally, exploring the use of deep-learning models to uncover new patterns for robust diabetes diagnosis, applicable across different diabetes types (T1DM, T2DM, and gestational diabetes), is recommended.

REFERENCES

- [1] H. Sone, "Diabetes Mellitus", in *Encyclopedia of Cardiovascular Research and Medicine*, R. S. Vasani and D. B. Sawyer, Eds., Oxford: Elsevier, 2018, pp. 9–16.
- [2] T. Andoh, "Subchapter 19A - Insulin", in *Handbook of Hormones*, Y. Takei, H. Ando, and K. Tsutsui, Eds., San Diego: Academic Press, 2016, pp. 157-e19A-3.
- [3] J. Hippisley-Cox and C. Coupland, "Diabetes treatments and risk of amputation, blindness, severe kidney failure, hyperglycaemia, and hypoglycaemia: open cohort study in primary care", *BMJ*, p. i1450, Mar. 2016.
- [4] A. N. Baanders and M. J. W. M. Heijmans, "The Impact of Chronic Diseases: The Partner's Perspective", *Family & Community Health*, vol. 30, no. 4, pp. 305–317, Oct. 2007.
- [5] P. Saeedi et al., "Global and regional diabetes prevalence estimates for 2019 and projections for 2030 and 2045: Results from the International Diabetes Federation Diabetes Atlas, 9th edition", *Diabetes Research and Clinical Practice*, vol. 157, p. 107843, Nov. 2019.
- [6] C. V. A. Collares et al., "Transcriptome meta-analysis of peripheral lymphomononuclear cells indicates that gestational diabetes is closer to type 1 diabetes than to type 2 diabetes mellitus", *Mol Biol Rep*, vol. 40, no. 9, pp. 5351–5358, Sep. 2013.
- [7] A. E. Butler and D. Misselbrook, "Distinguishing between type 1 and type 2 diabetes", *BMJ*, p. m2998, Aug. 2020.
- [8] G. Roglic, "WHO Global report on diabetes: A summary," in *International Journal of Noncommunicable Diseases*, vol. 1, no. 1, p. 3, 2016.
- [9] C.-C. Chang, Y.-Z. Li, H.-C. Wu, and M.-H. Tseng, "Melanoma Detection Using XGB Classifier Combined with Feature Extraction and K-Means SMOTE Techniques", *Diagnostics*, vol. 12, no. 7, p. 1747, Jul. 2022.
- [10] Z. Zhao, H. Peng, C. Lan, Y. Zheng, L. Fang, and J. Li, "Imbalance learning for the prediction of N6-Methylation sites in mRNAs", *BMC Genomics*, vol. 19, no. 1, p. 574, Dec. 2018.
- [11] N. H. N. B. M. Shahri, S. B. S. Lai, M. B. Mohamad, H. A. B. A. Rahman, and A. B. Ramli, "Comparing the Performance of AdaBoost, XGBoost, and Logistic Regression for Imbalanced Data", *ms*, vol. 9, no. 3, pp. 379–385, May 2021.
- [12] L. Breiman, "Random Forests", *Machine Learning*, vol. 45, no. 1, pp. 5–32, Oct. 2001.
- [13] D. Ekman, "Comparing Ensemble Methods with Individual Classifiers in Machine Learning for Diabetes Detection," *KTH Royal Institute of Technology in Stockholm, Sweden*, 2022.
- [14] I. E. Livieris, E. Pintelas, and P. Pintelas, "A CNN-LSTM model for gold price time-series forecasting", *Neural Comput & Applic*, vol. 32, no. 23, pp. 17351–17360, Dec. 2020.
- [15] I. H. Sarker, "Deep Cybersecurity: A Comprehensive Overview from Neural Network and Deep Learning Perspective", other, preprint, Feb. 2021.
- [16] D.-Y. Kim et al., "Intelligent Ensemble Deep Learning System for Blood Glucose Prediction Using Genetic Algorithms", *Complexity*, vol. 2022, pp. 1–10, Oct. 2022.
- [17] Department of Computer Science and Engineering, Khulna University of Engineering & Technology, Khulna-9203, Bangladesh, S. Islam Ayon, and Md. Milon Islam, "Diabetes Prediction: A Deep Learning Approach", *IJIEEB*, vol. 11, no. 2, pp. 21–27, Mar. 2019.
- [18] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning", *Nature*, vol. 521, no. 7553, pp. 436–444, May 2015.
- [19] A. Thammano and A. Meengen, "A New Evolutionary Neural Network Classifier", in *Advances in Knowledge Discovery and Data Mining*, T. B. Ho, D. Cheung, and H. Liu, Eds., in *Lecture Notes in Computer Science*, vol. 3518. Berlin, Heidelberg: Springer Berlin Heidelberg, 2005, pp. 249–255.
- [20] A. Ameh Joseph, M. Abdullahi, S. B. Junaidu, H. Hassan Ibrahim, and H. Chiroma, "Improved multi-classification of breast cancer histopathological images using handcrafted features and deep neural network (dense layer)", *Intelligent Systems with Applications*, vol. 14, p. 200066, May 2022.
- [21] T. G. Dietterich, "Ensemble Methods in Machine Learning", in *Multiple Classifier Systems*, in *Lecture Notes in Computer Science*, vol. 1857. Berlin, Heidelberg: Springer Berlin Heidelberg, 2000, pp. 1–15.
- [22] L. Kuncheva, "Combining Pattern Classifiers Methods and Algorithms, Second Edition." Hoboken, NJ: Wiley, 2014.
- [23] A. Dutta et al., "Early Prediction of Diabetes Using an Ensemble of Machine Learning Models", *IJERPH*, vol. 19, no. 19, p. 12378, Sep. 2022.
- [24] S. M. Ganie and M. B. Malik, "An ensemble Machine Learning approach for predicting Type-II diabetes mellitus based on lifestyle indicators", *Healthcare Analytics*, vol. 2, p. 100092, Nov. 2022.
- [25] U. e Laila, K. Mahboob, A. W. Khan, F. Khan, and W. Taekeun, "An Ensemble Approach to Predict Early-Stage Diabetes Risk Using Machine Learning: An Empirical Study", *Sensors*, vol. 22, no. 14, p. 5247, Jul. 2022.
- [26] G. Geetha and K. M. Prasad, "An Hybrid Ensemble Machine Learning Approach to Predict Type 2 Diabetes Mellitus", *WEB*, vol. 18, no. Special Issue 02, pp. 311–331, Apr. 2021.
- [27] R. D. Joshi and C. K. Dhakal, "Predicting Type 2 Diabetes Using Logistic Regression and Machine Learning Approaches", *IJERPH*, vol. 18, no. 14, p. 7346, Jul. 2021.
- [28] D. Pankaj Javale and S. Suhas Desai, "Machine learning ensemble approach for healthcare data analytics", *IJECS*, vol. 28, no. 2, p. 926, Nov. 2022.
- [29] R. N. Patil, S. Rawandale, N. Rawandale, U. Rawandale, and S. Patil, "An efficient stacking based NSGA-II approach for predicting type 2 diabetes", *IJECE*, vol. 13, no. 1, p. 1015, Feb. 2023.
- [30] H. Zhou, R. Myrzashova, and R. Zheng, "Diabetes prediction model based on an enhanced deep neural network", *J Wireless Com Network*, vol. 2020, no. 1, p. 148, Dec. 2020.
- [31] N. Gupta, B. Kaushik, M. Khalid Imam Rahmani, and S. Anwar Lashari, "Performance Evaluation of Deep Dense Layer Neural Network for Diabetes Prediction", *Computers, Materials & Continua*, vol. 76, no. 1, pp. 347–366, 2023.
- [32] F. Nazari and W. Yan, "Convolutional versus dense neural networks: comparing the two neural networks" performance in predicting building operational energy use based on the building shape", presented at the 2021 Building Simulation Conference, Sep. 2021.
- [33] T. Zhu, K. Li, P. Herrero, and P. Georgiou, "Deep Learning for Diabetes: A Systematic Review", *IEEE J. Biomed. Health Inform.*, vol. 25, no. 7, pp. 2744–2757, Jul. 2021.
- [34] Larabi-Marie-Sainte, Aburahmah, Almohaini, and Saba, "Current Techniques for Diabetes Prediction: Review and Case Study", *Applied Sciences*, vol. 9, no. 21, p. 4604, Oct. 2019.
- [35] M. Gollapalli et al., "A novel stacking ensemble for detecting three types of diabetes mellitus using a Saudi Arabian dataset: Pre-diabetes, T1DM, and T2DM", *Computers in Biology and Medicine*, vol. 147, p. 105757, 2022.
- [36] A. Singh, A. Dhillon, N. Kumar, M. S. Hossain, G. Muhammad, and M. Kumar, "eDiaPredict: An Ensemble-based Framework for Diabetes Prediction", *ACM Trans. Multimedia Comput. Commun. Appl.*, vol. 17, no. 2s, pp. 1–26, Jun. 2021.
- [37] A. H. Syed and T. Khan, "Machine Learning-Based Application for Predicting Risk of Type 2 Diabetes Mellitus (T2DM) in Saudi Arabia: A Retrospective Cross-Sectional Study", *IEEE Access*, vol. 8, pp. 199539–199561, 2020.
- [38] C.-Y. Chou, D.-Y. Hsu, and C.-H. Chou, "Predicting the Onset of Diabetes with Machine Learning Methods", *JPM*, vol. 13, no. 3, p. 406, Feb. 2023.
- [39] Haibo He, Yang Bai, E. A. Garcia, and Shutao Li, "ADASYN: Adaptive synthetic sampling approach for imbalanced learning", in *2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence)*, Hong Kong, China: IEEE, Jun. 2008, pp. 1322–1328.

- [40] Q. Zou, K. Qu, Y. Luo, D. Yin, Y. Ju, and H. Tang, "Predicting Diabetes Mellitus With Machine Learning Techniques", *Front. Genet.*, vol. 9, p. 515, Nov. 2018.
- [41] C. C. Aggarwal, "Neural Networks and Deep Learning: A Textbook", Cham: Springer International Publishing, 2018.
- [42] T. Akiba, S. Sano, T. Yanase, T. Ohta, and M. Koyama, "Optuna: A Next-generation Hyperparameter Optimization Framework". *arXiv*, Jul. 25, 2019.
- [43] M. Krishnamoorthy, M. S. A. Hameed, T. Kopinski, and A. Schwung, "Disease Prediction Based on Individuals Medical History Using CNN", in 2021 20th IEEE International Conference on Machine Learning and Applications (ICMLA), Pasadena, CA, USA: IEEE, Dec. 2021, pp. 89–94.
- [44] M. Maalouf, "Logistic regression in data analysis: an overview", *IJDATS*, vol. 3, no. 3, p. 281, 2011.
- [45] P. Misra and A. S. Yadav, "Improving the Classification Accuracy using Recursive Feature Elimination with Cross-Validation", 2020.
- [46] T. Chen and C. Guestrin, "XGBoost: A Scalable Tree Boosting System", in Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco California USA: ACM, Aug. 2016, pp. 785–794.
- [47] D. Berrar, "Cross-Validation", in *Encyclopedia of Bioinformatics and Computational Biology*, Elsevier, 2019, pp. 542–545.
- [48] F. Pedregosa et al., "Scikit-learn: Machine Learning in Python", *Journal of Machine Learning Research*, vol. 12, no. 85, pp. 2825–2830, 2011.
- [49] L. Qin, "A Prediction Model of Diabetes Based on Ensemble Learning", in Proceedings of the 2022 5th International Conference on Artificial Intelligence and Pattern Recognition, Xiamen China: ACM, Sep. 2022, pp. 45–51.
- [50] S. Kumari, D. Kumar, and M. Mittal, "An ensemble approach for classification and prediction of diabetes mellitus using soft voting classifier", *International Journal of Cognitive Computing in Engineering*, vol. 2, pp. 40–46, Jun. 2021.
- [51] M. Bhopte and M. Rai, "Hybrid Deep learning CNN-LSTM Model for Diabetes Prediction", *International Journal of Scientific Research*, vol. 8, no. 1, 2022.
- [52] M. Niharika, M. Gousia, M. G. Priya, O. P. Kalyani, R. Vasavi, and D. Sowjanya, "CLASSIFICATION OF DIABETES USING MLP CLASSIFIER", vol. 9, no. 5, 2022.
- [53] S. A. Hicks et al., "On evaluation metrics for medical applications of artificial intelligence", *Sci Rep*, vol. 12, no. 1, p. 5979, Apr. 2022, doi: 10.1038/s41598-022-09954-8.