A Deployment-Oriented Framework for Machine Learning-Based Learning Style Identification: A Systematic Computational Analysis

Sarafa Olasunkanmi Adeyemo[1], Mohd Shahizan Othman[2], Chan Weng Howe[3],
Muteb Sinhat Almarshadi[4], Siti Zaiton Mohd Hashim[5], Taofik Olasunkanmi Tafa[6], Abdulaziz Saidu Yalwa[7]
Faculty of Computing, Universiti Teknologi Malaysia, 81310, Johor Bahru, Johor, Malaysia[1, 2, 3, 5, 6, 7]
Department of Computer Science, College of Sciences and Human Studies, Shaqra University, Shaqra, Kingdom of Saudi Arabia[4]

Abstract—**This study presents a systematic and deployment-oriented analysis of machine learning (ML) techniques for learning style identification in adaptive digital environments. A total of 57 peer-reviewed studies published between 2020 and 2025 were analysed using a PRISMA-guided methodology. Beyond descriptive synthesis, the review systematically examines algorithmic paradigms, multimodal data integration strategies, evaluation protocols, and deployment readiness characteristics. The findings reveal that classical supervised models remain prevalent in small-scale applications, while deep learning and ensemble methods demonstrate improved performance in high-dimensional behavioural datasets. However, significant heterogeneity exists in validation strategies, fusion architectures, and system scalability. To address these limitations, this study proposes a deployment-oriented architectural framework that integrates: 1) context-aware model selection, 2) structured multimodal fusion design, 3) layered explainability mechanisms, and 4) a four-level deployment maturity evaluation model. The framework provides a unified system-level perspective that shifts emphasis from isolated performance optimization toward scalable, interpretable, and integration-ready ML system design. This work contributes a structured computational blueprint for developing robust and deployment-aware learning style identification systems in intelligent educational platforms.**

Keywords*—*Machine learning; learning style identification; multimodal data fusion; deep learning; ensemble learning; explainable artificial intelligence; deployment maturity; adaptive learning systems

I. Introduction

The rapid expansion of digital learning platforms has generated large-scale behavioural and interaction data, accelerating the development of machine learning (ML)-driven intelligent systems for educational analytics. Within this computational context, learning style identification (LSI) has evolved into a classification and modelling problem requiring robust algorithms, structured feature representation, and scalable system integration. Traditional LSI approaches, typically based on static self-report instruments such as the Felder-Silverman Learning Style Model (FSLSM), VARK, or Kolb's framework, are constrained by limited adaptability and delayed feedback mechanisms. Consequently, ML techniques have increasingly been employed to infer learner characteristics from behavioural logs, textual interactions, and multimodal data streams [1], [2].

Recent studies demonstrate the integration of ML models into intelligent tutoring systems and large-scale learning management systems (LMS), reflecting a shift toward data-driven learner modelling [3], [4]. A wide range of computational paradigms have been applied, including decision trees, support vector machines, deep neural networks, and ensemble learning architectures operating on heterogeneous data sources such as clickstream records and academic performance metrics [5], [6]. While classification performance has improved across reported studies, gains in accuracy alone do not ensure reproducibility, generalization, or scalable deployment.

Despite these technical developments, several computational limitations remain evident. Algorithmic comparisons are frequently conducted on isolated datasets without systematic cross-context validation, constraining generalization claims [7]. Multimodal data integration is often implemented through direct feature concatenation rather than architecturally principled fusion mechanisms, limiting robustness under heterogeneous or incomplete data conditions [8]. In addition, evaluation practices vary substantially, with heavy reliance on single-metric reporting and limited incorporation of uncertainty estimation or external validation protocols [9]. Although fairness, transparency, and privacy are increasingly discussed within AI systems research, they are seldom operationalized within LS modelling architectures in a structured manner [10]. These gaps suggest that while algorithmic experimentation is advancing, the field lacks a unified system-level design perspective for scalable and deployment-ready ML-based LS systems.

Existing review studies have largely catalogued algorithms or summarized reported performance metrics, providing descriptive overviews but limited architectural guidance. Unlike prior reviews that primarily enumerate algorithms, the present study develops a deployment-oriented computational framework that integrates modelling, evaluation, explainability, and system readiness within a unified design perspective. As machine learning systems transition from experimental prototypes toward integrated adaptive platforms, there is a growing need for structured computational frameworks that connect model selection strategies, multimodal fusion design, validation rigor, and deployment readiness.

To address this gap, this study conducts a systematic review of 57 peer-reviewed publications published between 2020 and 2025, following PRISMA 2020 guidelines [11]. Rather than providing a purely descriptive synthesis, the review reorganizes the literature across three primary analytical dimensions: 1) algorithmic paradigm evolution, 2) multimodal data integration strategies, and 3) evaluation and validation practices. Building upon this synthesis, the study proposes a deployment-oriented framework for machine learning-based learning style identification systems that integrates a context-sensitive model selection matrix, structured multimodal fusion architectures, layered explainability mechanisms, embedded ethical design checkpoints, and a four-level deployment maturity model.

The contributions of this work are threefold. First, it provides a structured computational synthesis of contemporary ML approaches for LSI, emphasizing scalability, robustness, and generalization characteristics. Second, it introduces a deployment-oriented architectural framework translating empirical trends into system-level design guidance. Third, it proposes a maturity-based evaluation model to assess deployment readiness within adaptive intelligent systems.

By reframing LSI as a system design and deployment challenge rather than solely a predictive task, this study contributes a structured computational blueprint for scalable, interpretable, and integration-ready ML-based learning style systems.

In this study, deployment maturity is defined as the extent to which a machine learning system progresses from experimental validation to scalable, real-time, and context-adaptive integration within operational educational environments.

This study is guided by the following overarching research question: How can machine learning-based learning style identification systems be systematically designed to achieve scalability, robustness, interpretability, and deployment readiness in real-world educational environments?

The remainder of this study is structured as follows: Section II reviews related work on machine learning-based learning style identification. Section III presents the systematic review methodology. Section IV provides the analytical synthesis of findings. Section V introduces the proposed deployment-oriented framework. Section VI discusses practical implications and best practices, while Section VII details the limitations of the study and outlines future research directions. Section VIII concludes the study.

II. Related Work

Research on learning style (LS) identification has progressively shifted from traditional rule-based and questionnaire-driven approaches toward automated machine learning (ML) and deep learning (DL) systems. Conventional LS models such as the Felder-Silverman Learning Style Model (FSLSM), VARK, and Kolb's experiential learning framework have been widely adopted as conceptual foundations for personalization. However, their operationalization through static instruments has raised concerns regarding scalability and real-time adaptability [12]. To address these limitations, data-driven ML techniques have increasingly been employed to infer learning styles from interaction data [13].

Recent studies have explored a range of supervised, semi-supervised, and clustering-based approaches. Classical classifiers, including decision trees, SVM, logistic regression, and k-nearest neighbours remain widely applied due to interpretability and suitability for small institutional datasets [14], [15]. Ensemble methods and feature selection techniques have been introduced to improve robustness, particularly in imbalanced datasets [16], [17]. At the same time, hybrid and deep learning architectures combining LSTM, transformer-based models, and clustering strategies have emerged in MOOC and large-scale LMS contexts, enabling automated feature extraction and sequential pattern modelling [18], [19].

Despite these advances, several structural limitations persist. Many studies rely on limited or course-specific datasets, restricting generalizability. Validation is often conducted within a single institutional context, and systematic cross-domain evaluation remains rare. Furthermore, multimodal integration strategies are frequently implemented using straightforward concatenation or aggregation methods, without principled fusion architectures. These patterns indicate that while predictive performance has improved, architectural and validation standards remain inconsistent.

Beyond learning style identification, ML has been widely applied to adjacent educational domains, including student dropout prediction, academic performance forecasting, adaptive content recommendation, and support for learners with special needs [20], [21], [22]. These applications demonstrate the versatility of ML techniques, ranging from classical classifiers to deep learning and explainable AI (XAI) tools such as SHAP and LIME, but also highlight recurring challenges related to dataset bias, interpretability, and deployment scalability.

Existing review studies in LS modelling have primarily catalogued algorithms or compared reported performance metrics. Few have systematically examined how algorithm choice, data modality integration, evaluation practices, and deployment considerations interact within adaptive educational systems. Consequently, there remains a need for structured synthesis that bridges technical modelling advances with system-level design guidance.

This study addresses that need by providing a systematic and analytically structured review of ML approaches for learning style identification, positioning algorithmic evolution, validation practices, and multimodal integration within a unified deployment-oriented perspective.

III. Methodology

This systematic review followed a structured and transparent process designed to address the study objectives and research questions. The methodology comprised research question formulation, database search, eligibility screening, quality assessment, data extraction, and analytical synthesis. Peer-reviewed studies published between 2020 and 2025 were retrieved from selected academic databases using predefined inclusion and exclusion criteria. Relevant variables, including machine learning techniques, dataset characteristics, evaluation metrics, and key findings, were extracted from the selected studies. Fig. 1 presents the overall systematic review workflow adopted in this study. The process began with planning and the formulation of specific research questions, followed by a structured search across selected databases. Identified records were screened using predefined inclusion and exclusion criteria, after which eligible studies underwent quality assessment. Only studies meeting the specified threshold were retained for detailed data collection and analysis. This structured approach ensured that the review remained focused, transparent, and aligned with the study objectives.

A qualitative analytical synthesis was conducted to identify recurring themes, research trends, and evidence gaps. The review process adhered to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines, as illustrated in Fig. 2, to ensure transparency and completeness [11], and the PRISMA checklist is contained in the Appendix. Additional research support tools were used to enhance review quality. Bibliometric analysis was performed using RStudio to explore keyword co-occurrence, Harzing's Publish or Perish was used to retrieve and analyse relevant articles, SciSpace assisted with literature screening and insight extraction [23], and Zotero was used to organize references. A total of 57 studies were ultimately included from an initial pool of 2,816 records after duplicate removal, date and language filtering, abstract screening, and full-text assessment. This rigorous process ensured that only relevant and recent studies on machine learning-based learning style identification were retained.

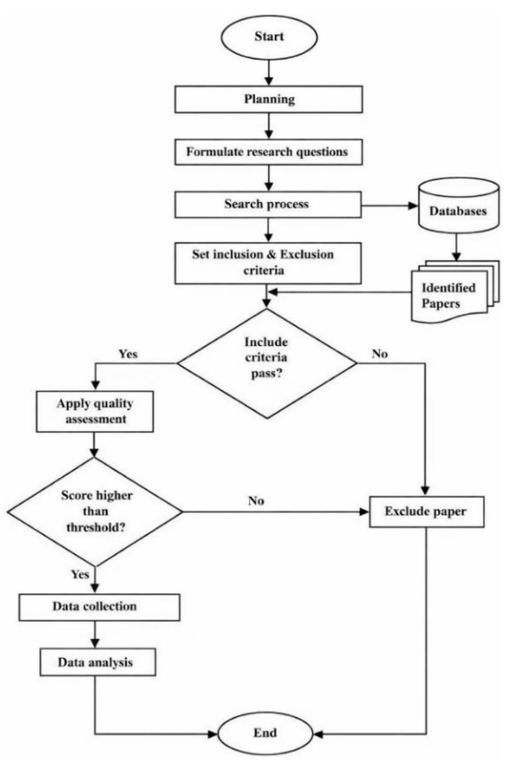

Fig. 1. Systematic review workflow.

A. Planning Phase

In this phase, the necessary steps to achieve the research objectives were identified. The study focuses on identifying LS using improved ML techniques. Also, both strategic and technical strategies to ensure that the rest of the proposed techniques can be executed systematically and consistently were identified. This planning phase served as the foundation for the successful implementation of the SLR method.

B. Research Questions and Motivations

This section presents the research questions investigated in the current SLR study and the motivations behind them. The motivations come from the noteworthy achievements of the work using ML techniques to identify LS. For instance, most studies have indicated that ML is essential to accurately improve LSI. Table I provides a clear outline of the research questions along with the motivations for exploring each question in the context of identifying LSI using enhanced ML techniques.

TABLE I. RESEARCH QUESTIONS AND MOTIVATIONS

Research Question	Motivation
RQ1: What are the state-of-the-art ML techniques currently being employed to identify LS in educational contexts, and how do they compare in terms of accuracy and effectiveness?	To understand the current landscape of ML methods used in LS identification and evaluate their performance.
RQ2: How do enhanced ML techniques improve the identification of LS over traditional methods, and what are the key metrics used to evaluate these improvements?	To assess the advancements made over traditional methods and determine the effectiveness of these improvements using standard metrics.
RQ3: How do different data sources and feature selection methods impact the performance and accuracy of LS prediction models?	To understand the importance of different types of data and features in predicting LS accurately.

C. Search Strategy and Term Justification

A comprehensive literature search was conducted across IEEE, Web of Science, Scopus, Google Scholar, and ScienceDirect. The core search expression was ("Machine learning" OR "Enhanced Machine Learning") AND ("Learning styles" OR "Learning style"). To improve search sensitivity and capture a broader spectrum of relevant studies, the expanded Boolean strategy also considered related terms such as "classification", "deep learning", "ensemble learning", "personalized learning", "explainability", "student modeling", and "educational data mining". The review focused on studies published between 2020 and 2025 to reflect post-COVID-19 developments in online and adaptive learning, including the recent growth of deep learning, explainable AI, and hybrid modelling approaches. The databases and corresponding search strings are presented in Table II.

D. Inclusion and Exclusion Criteria

The selection of studies for this review was guided by clearly defined inclusion and exclusion criteria. To be considered, studies had to be peer-reviewed journal or conference papers published between 2020 and 2025, written in English, and explicitly focused on the application of machine learning techniques to learning style identification. Additionally, only studies that employed empirical data or included experimental analysis were included. On the other hand, studies were excluded if they did not align with the objectives of the review for instance, those centred solely on theoretical frameworks without practical implementation. Non-peer-reviewed materials such as editorials and opinion pieces were also excluded. Duplicate publications and studies that lacked sufficient detail regarding the machine learning methods used or their evaluation metrics were removed from consideration. A summary of these criteria is presented in Table III.

TABLE II. DATABASE SEARCH STRATEGY AND RETRIEVED STUDY DISTRIBUTION ACROSS DIGITAL LIBRARIES

Digital Database	Field	Search Strings	No. of result
IEEE	All	("Machine learning" OR "Enhanced Machine Learning") AND ("Learning styles" OR "Learning style")	204
Web of Science	All	("Machine learning" OR "Enhanced Machine Learning") AND ("Learning styles" OR "Learning style")	178
Scopus	TITLE-ABS-KEY	("Machine learning" OR "Enhanced Machine Learning") AND ("Learning styles" OR "Learning style")	325
Science Direct	All	("Machine learning" OR "Enhanced Machine Learning") AND ("Learning styles" OR "Learning style")	1336
Google Scholar	All	("Machine learning" OR "Enhanced Machine Learning") AND ("Learning styles" OR "Learning style")	773
Overall			2816

TABLE III. INCLUSION AND EXCLUSION CRITERIA

Inclusion Criteria for the paper to be included	Exclusion Criteria for the paper to be excluded
Papers published between 2020 and 2025 only.	Papers published before 2020.
Paper related to Research question 1 to Research question 3 (RQ1 to RQ3)	Paper not related to the research question RQ1 to RQ3.
Published conferences and journal paper.	Paper not published in conferences and journal.
Peer review Papers/article only	Not peer review Papers/article
Papers written in English language only.	Not written in English language.
Open access paper.	Not open access papers.
	Duplicate.

E. Quality Assessment

The quality assessment process was carried out to limit the scope of data collection and to ensure targeted and reliable analysis. The main objective was to assess the extent to which each selected study answered the research questions outlined in this review [24]. To maintain objectivity, two independent reviewers assessed the studies against a set of predefined questions aligned with the research objectives. Inter-rater reliability was measured using Cohen's Kappa. The coefficient of k = 0.75 indicates a high level of agreement [24]. This procedure helped to accurately identify relevant studies and filter out those that did not fulfil the criteria. Table IV shows the three quality assessment questions that were used to guide this assessment, while Table V shows the scoring matrix for each study. These questions were specifically selected based on their alignment with the objectives of the study and their effectiveness in determining relevance. Each paper could receive a maximum score of 3, and studies that achieved a score of 2 or higher were included because they either demonstrated a general, comprehensive analysis with sufficient detail or provided a strong, well-supported examination of the topic. The quality scores assigned to each of the articles reviewed are summarized in Table VI.

TABLE IV. QUALITY ASSESSMENT AND QUESTIONNAIRE

Quality Assessment Question	Relevant to the Research Question
QA1: Is the research objective clearly stated and aligned with the study's focus on identifying learning styles using machine learning techniques?	RQ1
QA2: Are the machine learning techniques and evaluation metrics used in the study clearly described and justified?	RQ2
QA3: Are the types of data and features used in the study clearly described and appropriate for the research questions?	RQ3

TABLE V. QUALITY ASSESSMENT SCORING MATRIX

Quality Assessment Scoring Criteria	Score
The author(s) have provided a detailed, explicit, and clear understanding of the answers of the specific RQ.	High=H=1
The author(s) have provided some general explanation, but not detailed, explicit, and clear understanding on the answers to the specific RQ.	Medium=M=0.5
The author(s) have provided no and very few technical details to the specific RQ.	Low=L=0

F. Data Collection

Following quality assessment, data extraction was conducted using a structured coding protocol. Each study was systematically analysed across predefined dimensions, including algorithmic paradigm, data modality, feature engineering strategy, evaluation metrics, and validation protocol. To ensure consistency, extracted variables were normalized into unified categorical and numerical representations. Studies were subsequently grouped into analytical clusters to support cross-study comparison and synthesis. This structured approach enabled the identification of recurring methodological patterns and facilitated cross-dimensional analysis aligned with the research questions. The overall review process followed the PRISMA framework, ensuring that the methodology was systematic, transparent, and rigorous, as illustrated in Fig. 2.

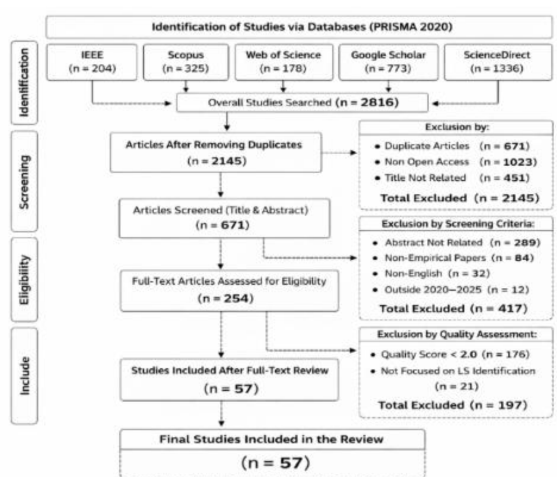

Fig. 2. PRISMA flow diagram.

G. Data Analysis

The data analysis stage synthesized the extracted evidence to answer the research questions defined for this study. Analytical grouping, comparative interpretation, and quality assessment scores were used to organize the reviewed literature and identify dominant patterns. The article list and corresponding quality assessment scores are presented in Table VI.

TABLE VI. ARTICLES STUDIED AND QUALITY ASSESSMENT REPORT

Article Studied	References	Publisher	Type	Q1	Q2	Q3	Score
AS01	[25]	Education and Information Technologies	Journal	1.0	1.0	0.5	2.5
AS02	[26]	IEEE	Journal	1.0	1.0	0.5	2.5
AS03	[27]	Applied Sciences	Journal	1.0	1.0	1.0	3.0
AS04	[28]	International Journal of Advanced Computer Science and Applications	Journal	1.0	1.0	1.0	3.0
AS05	[29]	Cluster Computing	Journal	1.0	1.0	1.0	3.0
AS06	[30]	Intelligent Systems with Applications	Conference	1.0	1.0	0.5	2.5
AS07	[31]	International Conference on Electrical Engineering	conference	1.0	1.0	1.0	3.0
AS08	[32]	IEEE Access	Journal	1.0	1.0	0.5	2.5
AS09	[33]	Journal of Computer Science	Journal	1.0	0.5	0.5	2.0
AS10	[34]	International Journal of Advanced Computer Science and Applications	Journal	1.0	1.0	1.0	3.0
AS11	[35]	International Journal	Journal	1.0	1.0	0.5	2.5
AS12	[36]	Journal of Cloud Computing	Journal	1.0	1.0	0.5	2.5
AS13	[37]	IEEE Access	Journal	1.0	1.0	1.0	3.0
AS14	[38]	Applied Sciences	Journal	1.0	1.0	1.0	3.0
AS15	[39]	Expert Systems with Applications	Journal	1.0	0.5	0.5	2.0
AS16	[40]	IEEE Access	Journal	1.0	0.5	0.5	2.5
AS17	[41]	IEEE Access	Journal	1.0	1.0	0.5	2.5
AS18	[42]	The journal of contemporary issues in business and government	Journal	1.0	1.0	1.0	3.0
AS19	[43]	Computational Intelligence and Neuroscience	Journal	1.0	1.0	0.5	2.5
AS20	[44]	IEEE ACCESS	Journal	1.0	0.5	1.0	2.0
AS21	[45]	Data Mining and Big Data	Conference	1.0	1.0	0.5	2.0
AS22	[15]	International Conference on Education Technology and Computers	Conference	1.0	1.0	1.0	3.0
AS23	[46]	IEEE ACCESS	Journal	1.0	1.0	0.5	2.5
AS24	[47]	Journal of Universal Computer Science	Journal	1.0	1.0	0.5	2.0
AS25	[48]	International Journal of Electrical and Computer Engineering	Journal	1.0	1.0	1.0	3.0
AS26	[49]	International Conference on Innovative Research in Applied Science, Engineering and Technology	Conference	1.0	0.5	0.5	2.0
AS27	[50]	Journal of Networking and Network Applications	Journal	1.0	1.0	1.0	3.0
AS28	[51]	Expert systems with applications	Journal	1.0	1.0	1.0	3.0
AS29	[52]	Information	Journal	1.0	1.0	0.5	2.5
AS30	[53]	IEEE Access	Journal	1.0	0.5	0.5	2.0
AS31	[54]	Cluster Computing	Journal	1.0	1.0	1.0	3.0
AS32	[55]	Indonesian Journal of Electrical Engineering and Computer Science	Journal	1.0	1.0	0.5	2.5
AS33	[56]	International Conference on Electrical, Computer and Energy Technologies	Conference	1.0	1.0	1.0	3.0
AS34	[57]	Applied Sciences	Journal	1.0	1.0	0.5	2.5
AS35	[58]	JURNAL INFOTEL	Journal	1.0	1.0	0.5	2.5
AS36	[59]	International Journal of Emerging Technologies in Learning	Journal	1.0	1.0	0.5	2.5
AS37	[60]	International Journal of Advanced and Applied Sciences	Journal	1.0	1.0	0.5	2.5
AS38	[12]	International Journal of Emerging Technologies in Learning	Journal	1.0	0.5	0.5	2.0
AS39	[13]	International Conference on Innovative Research in Applied Science, Engineering and Technology	Conference	1.0	1.0	1.0	3.0
AS40	[61]	IEEE Access	Journal	1.0	0.5	0.5	2.0
AS41	[62]	Cogent Education	Journal	1.0	0.5	0.5	2.0
AS42	[63]	High Tech and Innovation Journal	Journal	1.0	1.0	1.0	3.0
AS43	[64]	Education Sciences	Journal	1.0	1.0	0.5	2.5

AS44	[65]	Journal of Engineering Education Transformations	Journal	0.5	0.5	1.0	2.0
AS45	[66]	International Conference on Computer Science and Education	Conference	1.0	0.5	1.0	2.5
AS46	[14]	IEEE Transactions on Learning Technologies	Journal	1.0	0.5	0.5	2.0
AS47	[2]	Education and Information Technologies	Journal	1.0	1.0	0.5	2.5
AS48	[17]	IEEE Transactions on Learning Technologies	Journal	1.0	0.5	0.5	2.0
AS49	[16]	International Journal of Advanced Computer Science and Application	Journal	1.0	1.0	1.0	3.0
AS50	[67]	Asian Journal of Research in Computer Science	Journal	1.0	1.0	0.5	2.5
AS51	[18]	Information	Journal	1.0	1.0	1.0	3.0
AS52	[19]	Applied Soft Computing	Journal	1.0	1.0	1.0	3.0
AS53	[68]	IEEE Access	Journal	1.0	1.0	0.5	2.5
AS54	[1]	Journal of Computer in Education	Journal	1.0	1.0	0.5	2.5
AS55	[69]	IEEE Access	Journal	1.0	0.5	1.0	2.5
AS56	[70]	Intelligent Systems with Applications	Journal	1.0	1.0	1.0	3.0
AS57	[71]	IEEE Access	Journal	1.0	1.0	1.0	3.0

IV. Results and Analytical Synthesis

This section presents a structured analytical synthesis of 57 studies published between 2020 and 2025. Rather than merely enumerating algorithms, the literature is reorganized across three primary analytical dimensions: algorithmic paradigm evolution, evaluation and validation practices, and multimodal data integration, followed by a cross-dimensional synthesis emphasizing deployment readiness. This analytical structure provides a clearer understanding of how learning style identification research is evolving toward scalable adaptive systems.

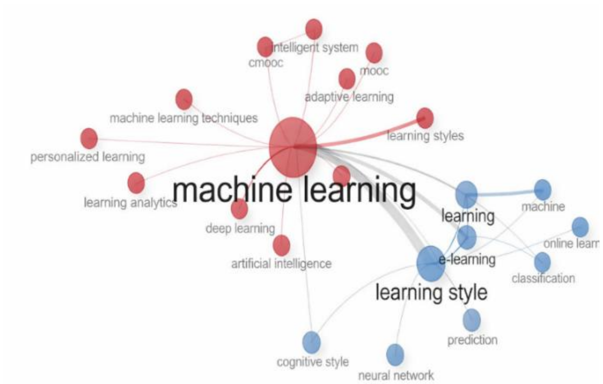

Fig. 3. Keyword co-occurrence network of ML-based learning style research.

Fig. 3 presents the keyword co-occurrence network derived from the reviewed corpus, illustrating the conceptual structure of ML-based learning style research. Two dominant thematic clusters are observable. The first cluster centres on technological constructs; including machine learning, deep learning, artificial intelligence, and learning analytics indicating a strong emphasis on computational modelling and predictive techniques. The second cluster emphasizes educational constructs such as learning style, e-learning, classification, and adaptive learning, reflecting application-driven personalization within digital environments. The interconnection between these clusters highlights the interdisciplinary nature of the field, where algorithmic innovation intersects with pedagogical objectives. The presence of terms such as MOOCs and intelligent systems suggests increasing attention toward scalable and adaptive deployment contexts. Collectively, the network structure reinforces the need for integrative frameworks that coordinate technological advancement with educational system design.

Unless otherwise stated, percentage values reported in this section represent proportional counts of reviewed studies categorized according to their primary methodological orientation. Some studies exhibit methodological overlap across categories; therefore, percentages are intended to indicate dominant patterns rather than mutually exclusive classification totals.

A. RQ1: Algorithmic Landscape and Paradigm Evolution

This section addresses RQ1 by examining the dominant machine learning paradigms employed for learning style identification and analysing their comparative performance characteristics, contextual suitability, and evolutionary trends across the reviewed studies.

1) Classical supervised models: Classical supervised learning approaches, including Decision Trees, Support Vector Machines (SVM), Logistic Regression, and k-Nearest Neighbours (k-NN), dominated early studies between 2020 and 2022 [15], [29], [39], [47], [55]. These models were frequently adopted due to their interpretability, relatively low computational requirements, and suitability for small or institution-level datasets. In contexts where dataset sizes were limited and feature sets were manually engineered, such models provided stable and explainable classification performance. However, their predictive effectiveness often depended heavily on handcrafted feature extraction and preprocessing strategies. Several studies reported performance degradation when datasets were imbalanced or insufficiently representative, with overfitting risks particularly evident in small-sample institutional deployments [39], [57], [70]. While these models remain attractive for classroom-scale adaptive tools, their robustness across heterogeneous learning environments appears constrained.

Analytical observation: Classical supervised approaches continue to provide accessible and interpretable solutions but demonstrate limited generalizability across diverse educational contexts.

2) Deep representation learning: From 2022 onward, there has been a noticeable increase in the adoption of deep learning architectures, particularly Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN), and autoencoder-based models [14], [46], [54], [68]. These approaches were predominantly applied in MOOC-scale and large LMS datasets characterized by high-dimensional interaction logs and sequential behavioural data. Deep architectures enabled automatic extraction of latent behavioural patterns, reduced dependence on manual feature engineering, and improved classification stability in data-rich environments. In several cases, deep models outperformed traditional classifiers when sufficient training data were available. However, most implementations were validated within single datasets, and cross-institution or cross-context validation was rarely performed. Additionally, limited integration of interpretability mechanisms was observed, raising concerns regarding transparency and stakeholder trust in educational settings.

Analytical observation: Deep representation learning enhances predictive capacity in large-scale digital environments but introduces transparency and generalization challenges that may constrain institutional adoption.

3) Ensemble optimization approaches: Ensemble-based methods, including Random Forest, Gradient Boosting, and Extreme Gradient Boosting (XGBoost), demonstrated consistently strong reported performance across multiple datasets [15], [33], [43], [52]. By aggregating predictions from multiple base learners, these approaches improved robustness under noisy feature conditions and mitigated overfitting risks associated with single-model classifiers. Ensemble strategies were particularly effective when heterogeneous feature types were combined, including behavioural metrics, academic records, and demographic indicators. Nevertheless, while ensembles often provided improved predictive stability, interpretability trade-offs persisted, especially in boosting-based implementations.

Key pattern identified: Ensemble methods appear to offer a practical balance between predictive robustness and interpretability, positioning them as strong candidates for intermediate-scale adaptive learning systems.

4) Hybrid and multimodal architectures: A growing subset of studies explored hybrid and multimodal architectures that integrate multiple data sources, including behavioural logs, textual content, physiological signals, and demographic information [17], [18], [19]. These approaches often combined deep feature extraction modules with classical classifiers or ensemble layers, thereby leveraging complementary strengths across modelling paradigms. Hybrid configurations demonstrated improved cross-feature robustness and adaptability, particularly in contexts where behavioural signals alone were insufficient for accurate classification. However, structured multimodal fusion strategies such as attention mechanisms, modality weighting, or dropout-based robustness techniques were rarely implemented. Most studies relied on simple feature concatenation or late-stage aggregation.

Emerging trend: The field is progressively shifting from single-modality classification toward multimodal adaptive modelling, although systematic architectural principles for fusion remain underdeveloped.

Across the 57 reviewed studies, 51% employed classical supervised models as their primary modelling paradigm, 32% utilized deep learning architectures, and 16% implemented ensemble techniques. Although deep learning adoption is increasing, classical approaches remain dominant in small-scale institutional contexts, while multimodal and hybrid systems represent an emerging but still developing direction.

Table VII provides a consolidated distribution of machine learning techniques identified across the reviewed studies, supporting the aggregated paradigm-level analysis presented above.

TABLE VII. DISTRIBUTION OF MACHINE LEARNING TECHNIQUES ACROSS REVIEWED STUDIES (2020-2025)

ML Category	Technique	Frequency	Studied IDs
Clustering	K-Means,	12	AS47, AS21, AS22, AS38, AS42, AS45, AS25, AS27, AS16, AS01, AS04, AS06
	Hierarchical Clustering,	1	AS14
	Gaussian Mixture Models (GMM)	1	AS28
Classification	Logistic Regression,	4	AS49, AS24, AS55, AS57
	Decision Trees,	11	AS40, AS46, AS42, AS19, AS13, AS14, AS17, AS05, AS55, AS56. AS57
	Support Vector Machines (SVM),	7	AS47, AS15, AS32, AS22, AS35, AS55, AS56
	k-Nearest Neighbours (k-NN),	5	AS32, AS22, AS29, AS56, AS57
	Naive Bayes	2	AS24, AS55
NLP	TF-IDF,	7	AS37, AS27, AS16, AS08, AS09, AS10, AS18
Bayesian Network	Naive Bayes,	3	AS41, AS19, AS08
	Bayesian Belief Networks,	1	AS42
	Gaussian Naive Bayes,	1	AS35
	Bayesian Network Classifiers	3	AS33, AS04, AS07
Deep learning	Convolutional Neural Networks (CNN),	10	AS50, AS23, AS31, AS34, AS36, AS17, AS03, AS09, AS12, AS57
	Recurrent Neural Networks (RNN),	6	AS53, AS54, AS48, AS51, AS10, AS18
	Autoencoders,	1	AS46

	Generative Adversarial Networks (GANs),	1	AS52
Ensemble	Random Forests,	5	AS50, AS43, AS19, AS34, AS09
	Gradient Boosting Machines (GBM),	2	AS54, AS29
	XGBoost,	2	AS15, AS22
Others		8	AS43, AS44, AS20, AS26, AS30, AS33, AS34, AS11

The distribution reinforces the continued dominance of classical supervised approaches, while highlighting the increasing presence of deep learning and ensemble techniques in recent years.

B. *RQ2: Evaluation Practices and Performance Patterns*

This section addresses RQ2 by examining how machine learning models for learning style identification are evaluated, which performance metrics are reported, and how validation strategies influence the interpretation of effectiveness across educational contexts.

1) Performance metrics and reporting standards: Across the reviewed literature, accuracy was the dominant evaluation metric, reported in 89% of studies. F1-score was reported in 54% of cases, while precision and recall were separately documented in 37% of studies. However, fewer than 10% of publications reported confidence intervals, and only 18% provided measures of variance such as standard deviation. This heavy reliance on accuracy as a primary metric limits meaningful cross-study comparison, particularly in contexts where class imbalance is present. In learning style identification tasks, where distributional skew may occur across preference categories, exclusive dependence on accuracy can obscure misclassification patterns. The relatively limited reporting of uncertainty measures further constrains the reproducibility and statistical interpretability of results.

Analytical observation: Evaluation practices remain metric-centric rather than reliability-centric, with insufficient emphasis on uncertainty quantification and robustness analysis.

2) Validation protocols and generalization: Validation strategies varied considerably across the reviewed studies. Approximately 63% employed k-fold cross-validation, while 28% relied on simple hold-out validation. Only 7% of studies conducted cross-dataset or cross-institution validation. Longitudinal validation, in which models are tested across different academic cohorts or time periods, was rarely observed. Although k-fold cross-validation enhances internal reliability, it does not guarantee cross-context generalization. The limited adoption of external validation protocols suggests that many models remain optimized for dataset-specific conditions rather than scalable educational deployment. Furthermore, few studies reported hyperparameter tuning strategies or ablation analyses, making replication challenging.

Analytical Observation: While internal validation practices are common, cross-context generalization remains underexplored, constraining claims of scalability.

3) Comparative performance patterns: When comparing algorithmic categories, ensemble and deep learning approaches frequently reported higher predictive performance relative to classical supervised models, particularly in data-rich environments [1], [17], [64], [67], [68]. However, performance improvements were often incremental rather than transformative and were highly dependent on dataset characteristics, feature engineering quality, and validation protocol. Notably, studies incorporating multimodal data sources tended to report improved classification stability compared to unimodal implementations. Nevertheless, direct cross-study performance comparison remains difficult due to inconsistent reporting standards and heterogeneous dataset characteristics. A cross-study comparative analysis further indicates that ensemble and deep learning approaches generally outperform classical supervised models in data-rich environments, particularly when multimodal inputs are incorporated. However, performance improvements remain context-dependent and strongly influenced by dataset scale, feature engineering quality, and validation strategy. This indicates that algorithmic superiority cannot be assessed in isolation but must be interpreted within the broader system design context.

Key pattern identified: Reported performance gains are strongly influenced by dataset scale, feature design, and validation strategy, suggesting that algorithmic superiority cannot be assessed independently of contextual factors.

Fig. 4 visualizes comparative performance trends across modelling paradigms using a qualitative aggregation scale (1-5) to reflect relative consistency and strength of reported outcomes. Owing to methodological heterogeneity across studies, the scale represents synthesized performance patterns rather than standardized benchmark comparisons. The heatmap, therefore, highlights cross-metric trends and variability without implying direct statistical equivalence between algorithms.

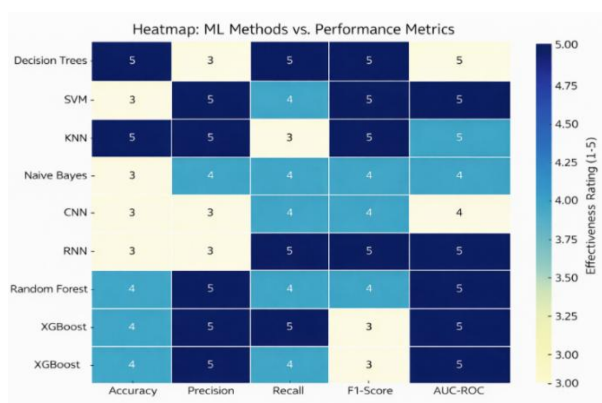

Fig. 4. Comparative performance heatmap of major ML paradigms across evaluation metrics.

This heatmap illustrates the relative performance patterns of classical, deep learning, and ensemble methods across five commonly reported metrics: accuracy, precision, recall, F1-score, and AUC-ROC based on aggregated findings from the reviewed studies.

4) Reliability and reproducibility gaps: A recurring limitation across the reviewed literature is the absence of standardized benchmarking datasets and shared evaluation protocols. Few studies made datasets publicly available, and replication studies were virtually non-existent. The lack of consistent reporting on preprocessing steps, feature normalization, and model hyperparameters further impedes reproducibility. These limitations highlight a structural gap between experimental model development and robust, transferable educational implementation. Without standardized validation frameworks, reported improvements may reflect dataset-specific optimization rather than generalizable advancement.

Analytical observation: The field demonstrates methodological diversity, but lacks consolidated benchmarking standards necessary for cumulative progress.

C. *RQ3: Data Modalities and Feature Engineering Strategies*

This section addresses RQ3 by analysing how different data modalities and feature engineering strategies influence predictive robustness, model scalability, and deployment feasibility in learning style identification systems.

1) Dominant data modalities: The reviewed studies employed a range of data sources, with behavioural interaction logs and academic performance records being the most frequently utilized modalities. Behavioural data including clickstream activity, time-on-task, navigation sequences, and assessment attempts were particularly prevalent in LMS-based environments [1], [18]. Academic performance metrics such as grades, quiz scores, and assignment outcomes were commonly integrated as structured input variables. Textual data, including forum posts and reflective writing submissions, appeared less frequently but were leveraged in studies employing natural language processing techniques. Physiological signals, such as EEG or eye-tracking data, were present in a limited subset of experimental settings due to infrastructural and ethical constraints. Demographic features were occasionally incorporated as supplementary variables but rarely served as primary predictive drivers. Across the 57 reviewed studies, only 21% implemented true multimodal integration involving two or more distinct data types. The majority relied on single-modality datasets, predominantly behavioural or academic records.

Analytical observation: While multimodal approaches are emerging, unimodal implementations remain dominant, limiting the depth of learner representation and adaptive personalization potential.

2) Feature engineering and representation strategies: Feature engineering practices varied significantly across studies. In classical supervised models, manual feature extraction was the norm, involving statistical summaries, interaction frequencies, session durations, and categorical encodings. Feature selection methods such as correlation analysis, principal component analysis (PCA), and information gain were occasionally employed to reduce dimensionality. In contrast, deep learning models relied more heavily on automated representation learning. Sequential modelling techniques, particularly RNN-based architectures, captured temporal behavioural patterns without extensive manual preprocessing. However, even in deep learning implementations, preprocessing decisions such as normalization, tokenization (for textual data), and imbalance handling remained critical determinants of performance. Notably, few studies reported systematic ablation analyses to quantify the impact of individual feature groups. Additionally, handling of missing data and modality imbalance was inconsistently addressed.

Analytical observation: Although representation learning reduces dependence on manual feature design, feature engineering practices remain heterogeneous and insufficiently standardized.

3) Multimodal fusion strategies: Among studies employing multimodal data, fusion strategies were generally implemented using simple early or late fusion mechanisms. Early fusion typically involved concatenating feature vectors before model training, whereas late fusion aggregated predictions from modality-specific classifiers. Intermediate fusion approaches where embeddings from separate subnetworks are integrated at latent layers were comparatively rare. Advanced fusion strategies such as attention-based weighting, adaptive modality selection, or modality dropout regularization were seldom observed. Consequently, most multimodal implementations lacked dynamic weighting mechanisms capable of handling modality-specific noise or missing inputs.

These findings suggest that while multimodal interest is growing, architectural sophistication in cross-modal integration remains limited.

Key pattern identified: Multimodal integration is conceptually recognized as beneficial, but is often operationalized through simplistic fusion techniques rather than structured representation learning.

4) Scalability and practical constraints: Data modality selection is closely linked to scalability and deployment feasibility. Behavioural logs and academic records are readily available in LMS platforms and therefore support scalable implementation. In contrast, physiological data collection requires specialized hardware and raises privacy concerns, restricting broader adoption.

Textual and multimodal approaches introduce increased computational complexity and data preprocessing demands. Few studies discussed infrastructure requirements or real-time processing constraints, further reinforcing the gap between experimental modelling and deployable system design.

Analytical observation: Data modality choices directly influence system scalability, cost, and ethical compliance, underscoring the need for context-sensitive model design.

D. Cross-Dimensional Synthesis

The analytical synthesis of RQ1-RQ3 reveals several interconnected structural patterns shaping the current landscape of machine learning-based learning style identification. First, algorithmic evolution (RQ1) demonstrates a gradual shift from interpretable classical classifiers toward ensemble and deep representation learning models, particularly in data-rich LMS and MOOC environments. However, this algorithmic sophistication has not been matched by corresponding advances in evaluation rigor (RQ2), as validation practices remain predominantly internal and dataset specific.

Second, the dominance of unimodal behavioural datasets (RQ3) partly explains the persistence of classical and ensemble approaches, as limited data diversity constrains the benefits of advanced multimodal architectures. Although multimodal modelling is increasing, most implementations rely on simplistic fusion strategies without systematic treatment of modality imbalance or missing data. Consequently, architectural complexity does not consistently translate into demonstrable generalization gains.

Third, evaluation and feature engineering practices appear strongly interdependent. Models employing automated representation learning frequently lack transparent reporting of preprocessing steps, while manually engineered feature pipelines often omit robustness testing. The absence of standardized benchmarking datasets and cross-context validation further complicates meaningful comparison across algorithmic paradigms.

Taken together, these observations indicate that the field has achieved notable progress in predictive modelling but remains structurally fragmented. Algorithm selection, data modality integration, and validation design are often treated as isolated decisions rather than coordinated components of a unified adaptive system architecture. This fragmentation limits scalability, reproducibility, and institutional deployment readiness. To further consolidate the structural patterns identified across algorithmic paradigms and design strategies, Table VIII classifies the reviewed studies according to their primary developmental focus.

TABLE VIII. THEMATIC CLASSIFICATION OF REVIEWED STUDIES BY DEVELOPMENTAL FOCUS

Category	Frequency	Study IDs
Feature Engineering	20	AS21, AS45, AS29, AS15, AS13, AS08, AS07, AS52, AS54, AS46, AS44, AS31, AS26, AS25, AS05, AS35, AS42, AS22, AS55, AS12
Algorithmic Improvements	28	AS20, AS39, AS40, AS21, AS45, AS29, AS30, AS15, AS17, AS24, AS43, AS32, AS23, AS33, AS27, AS14, AS16, AS36, AS08, AS47, AS01, AS54, AS53, AS48, AS49, AS03, AS56, AS57
Hybrid Models	21	AS40, AS41, AS19, AS28, AS23, AS20, AS02, AS07, AS50, AS01, AS53, AS37, AS48, AS51, AS44, AS34, AS31, AS45, AS38, AS55, AS57

An important observation emerging from the thematic classification is the presence of cross-category overlap among several studies. Several works simultaneously incorporate feature engineering refinement, algorithmic optimization, and hybrid architectural strategies. This overlap indicates that methodological advancement in learning style identification is increasingly multi-layered rather than confined to isolated improvements. However, while combinatorial approaches are emerging, integration is often partial and lacks a coordinated system-level design perspective. In many cases, feature refinement, model optimization, and multimodal fusion are implemented independently rather than within a unified deployment framework.

These cross-dimensional findings motivate the need for a structured, deployment-oriented framework that integrates model selection, multimodal design, evaluation standards, explainability mechanisms, and maturity-based deployment assessment. The following section builds on this synthesis to propose such a framework, translating empirical trends into system-level design guidance for scalable educational AI applications.

E. Evolution of Research Focus and System Maturity

To further contextualize the cross-dimensional findings, the reviewed studies were mapped across four thematic orientations: performance improvement, enhanced personalization, data quality and stability, and scalable/adaptable model development. Fig. 5 presents the temporal distribution of these focus areas from 2020 to 2025.

Fig. 5. Temporal evolution of research focus in ML-based learning style systems (2020-2025).

The mapping reveals that early research (2020-2021) predominantly emphasized performance optimization through algorithm comparison and accuracy improvement. From 2022 onward, a gradual expansion toward personalization refinement and data robustness became evident. More recent studies increasingly addressed scalability and system adaptability, indicating a transition from isolated classification experiments toward implementation-aware design considerations.

Notably, the emergence of scalability-focused research aligns with the broader maturation of educational AI systems. However, while thematic diversification is observable, integration across these dimensions remains limited. Most studies concentrate on a single improvement objective rather than coordinating performance, robustness, personalization, and scalability within a unified architectural framework.

These findings reinforce the need for structured system-level guidance that consolidates these thematic strands into a coherent deployment strategy. The following section translates this thematic evolution into a deployment-oriented design framework for ML-based learning style systems.

This figure categorizes reviewed studies into four thematic orientations: performance improvement, personalization enhancement, data stability, and scalable/adaptable model development, illustrating the progressive diversification of research priorities across the review period.

V. A Deployment-Oriented Framework for ML-Based Learning Style Systems

While the preceding sections examined algorithmic and methodological trends, the present section translates those findings into a structured framework for the design, evaluation, and deployment of machine learning (ML) systems for learning style (LS) identification in adaptive educational environments. Recent research in educational artificial intelligence increasingly emphasizes the need to bridge experimental model development with scalable, accountable system integration [72], [73], [74]. However, structured deployment guidance remains limited. To address this gap, the proposed framework integrates four interdependent dimensions:

- Context-sensitive model selection,
- Multimodal fusion architecture design,
- Layered explainability integration, and
- Deployment maturity assessment.

These dimensions are derived from the cross-dimensional patterns identified in Section IV and are intended to operationalize evidence-based design principles for scalable adaptive systems.

A. Model Selection Matrix for Educational Contexts

A recurring limitation in the reviewed literature is the absence of explicit criteria for selecting machine learning models according to educational constraints. Model choice is frequently performance-driven rather than context-driven, despite evidence that interpretability, data volume, and deployment scale substantially influence practical feasibility [4], [75]. To address this issue, this study proposes the context-sensitive model selection matrix, as shown in Table IX.

Analytical rationale: Educational decision-making often involves pedagogical accountability; therefore, interpretable models are preferable in high-stakes contexts. Tree-based models and SVM classifiers offer transparency and lower computational demands. Conversely, large-scale LMS and MOOC environments generate high-dimensional interaction logs that benefit from representation learning through deep neural architectures [76]. In multimodal adaptive systems, hybrid architectures combining ensemble learning with structured fusion mechanisms provide improved robustness and adaptability [77]. This matrix reframes model selection from isolated performance optimization toward context-aware system design.

TABLE IX. Model Selection Guidelines for Learning Style Systems

Educational Context	Data Availability	Interpretability Requirement	Recommended Model Type
Small classroom tools	Limited	High	Decision Trees/SVM
Institutional dashboards	Moderate	Medium	Random Forest/XGBoost
MOOC-scale platforms	Large behavioral logs	Low-Medium	Deep Learning (RNN/CNN)
Adaptive intelligent tutoring systems	Multimodal	Medium	Hybrid/Ensemble + Fusion

B. Multimodal Fusion Architecture Patterns

Section IV revealed increasing adoption of multimodal inputs, yet most implementations rely on relatively shallow fusion strategies. Similar observations have been reported in broader educational AI research, where multimodal integration often lacks principled architectural design [78], [79].

Fusion strategies can be categorized into three structural levels as follows:

1) Early fusion: In early fusion architectures, raw features from different modalities are concatenated prior to model training. While computationally straightforward, this approach is sensitive to modality imbalance and noise dominance, particularly when behavioural and textual data differ substantially in scale.

2) Intermediate fusion: Intermediate fusion employs separate subnetworks for each modality, merging learned embeddings at latent layers. This design supports representation learning while preserving modality-specific structures. Recent multimodal educational systems increasingly adopt this approach to enhance robustness and generalization [80].

3) Late fusion: Late fusion involves independent modality-specific classifiers whose outputs are aggregated through ensemble voting or stacking mechanisms. This approach improves modularity and resilience to missing modalities, making it suitable for deployment environments with incomplete learner data.

Identified architectural gap: Across the reviewed corpus, advanced mechanisms such as modality dropout regularization, cross-modal attention, and adaptive fusion weighting were rarely implemented. Yet these techniques are increasingly recognized as critical for robust multimodal AI systems [81]. Their limited adoption indicates architectural immaturity relative to system-level deployment demands.

C. Explainability Integration Framework

Interpretability remains a central barrier to educational ML deployment. Regulatory, ethical, and pedagogical accountability require transparency in automated learner modelling [82], [83]. A three-layer explainability integration structure is proposed, as shown in Table X.

TABLE X. EXPLAINABILITY INTEGRATION LAYERS

Layer	Purpose	Representative Tools
Model-Level	Feature contribution analysis	SHAP, LIME
Decision-Level	Rule extraction / local explanation	Tree-based rules, surrogate models
System-Level	Transparent reporting dashboards	Visual analytics panels

Embedding explainability during model design rather than post hoc enhances educator trust, supports regulatory compliance, and strengthens institutional accountability. Recent educational AI research emphasizes that explainability must be structurally integrated into adaptive systems rather than appended as an auxiliary feature [84].

D. Ethical and Fairness Design Integration

Although ethical issues are now widely discussed in adaptive learning research, they are still not sufficiently built into the actual design of most systems [85]. In many cases, ethics is treated as a general principle rather than a practical requirement that shapes how the system is developed and used. To make machine learning-based learning style systems more responsible and trustworthy, ethical and fairness considerations should be integrated throughout the full system lifecycle. At the data collection stage, this includes obtaining informed consent from learners, protecting personal information through anonymization, and checking the data for possible sources of bias. During model training, attention should be given to reducing class imbalance, applying fairness-aware learning strategies, and examining whether model performance differs across learner groups. At the deployment stage, ethical integration should continue through human oversight, regular monitoring for emerging bias, and clear reporting mechanisms that make system decisions more transparent. Embedding these safeguards across data collection, model development, and deployment helps move ethical AI from abstract discussion into practical system design and supports current expectations for responsible governance in educational AI [86].

E. Deployment Maturity Model for ML-Based Learning Style Systems

To systematically evaluate readiness for real-world implementation, this study introduces a four-level deployment maturity model, as shown in Table XI.

TABLE XI. DEPLOYMENT MATURITY LEVELS

Level	Description	Characteristics	Example Scenario
Level 1: Experimental	Offline prototype	Single dataset; no real-time integration	Research laboratory model
Level 2: Dataset-Validated	Cross-validation reported	Moderate robustness; limited scope	Institutional pilot
Level 3: Platform-Integrated	LMS-connected system	Real-time classification; baseline explainability	University dashboard
Level 4: Cross-Context Adaptive	Multi-institution validation	Multimodal, explainable, fairness-audited	Scalable adaptive platform

Observations from reviewed studies: Most studies remain within Level 1 or Level 2, focusing on offline evaluation. A smaller subset demonstrates characteristics of Level 3 integration. None fully satisfy Level 4 criteria, which require cross-context validation, structured multimodality, embedded explainability, and operational fairness monitoring.

Strategic implication: These findings indicate that although predictive modelling techniques are advancing, the field remains structurally immature regarding scalable and context-robust deployment. The maturity model provides a structured benchmark for evaluating progression toward adaptive system integration.

F. Integrated Framework Overview

The proposed framework brings together five closely connected elements: context-aware algorithm selection, structured multimodal architecture design, layered explainability integration, embedded ethical safeguards, and deployment maturity assessment. Rather than treating these as separate concerns, the framework combines them into a unified design structure for next-generation adaptive learning systems. This integrated view shifts attention away from focusing only on algorithmic accuracy and instead emphasizes how different components work together to support real-world educational use. In doing so, the framework provides a practical foundation for building machine learning-driven educational systems that are scalable, transparent, and accountable. Fig. 6 illustrates this integrated deployment-oriented architecture by showing how the five layers interact with one another and with the broader system properties that influence successful implementation.

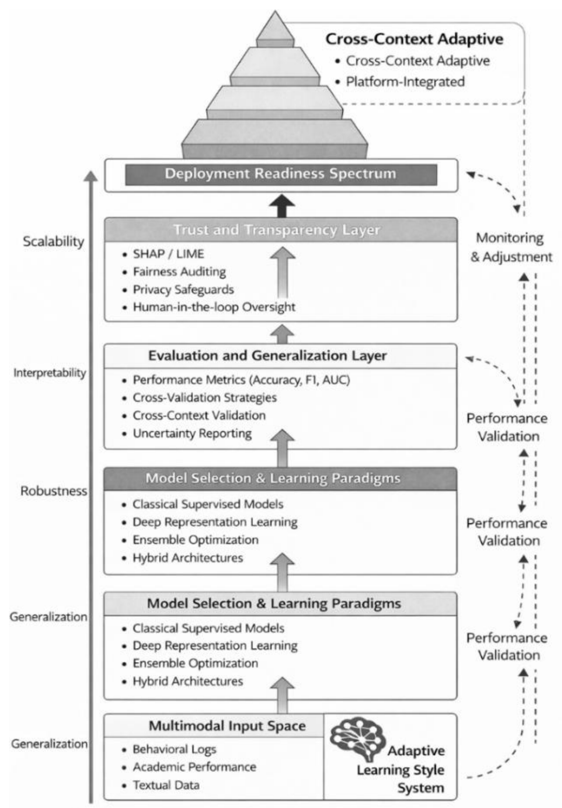

Fig. 6. Deployment-oriented framework for ML-based learning style systems.

The architecture illustrates a bottom-to-top progression from multimodal input space through model selection, evaluation and generalization, trust and transparency integration, and finally deployment readiness, with cross-cutting properties (scalability, interpretability, robustness, and generalization) guiding system maturation.

VI. Discussion

This review reveals an increasingly sophisticated yet structurally fragmented landscape in the application of machine learning (ML) to learning style identification (LSI). Although predictive performance has improved across modelling paradigms, methodological consistency, cross-context generalization, and deployment readiness remain uneven.

A. Comparative Evaluation of Modelling Paradigms

Deep learning and ensemble approaches consistently report high performance across accuracy, precision, recall, and F1-score metrics. Their strength lies in handling high-dimensional behavioural logs and complex learner interaction patterns. However, these models frequently lack interpretability and are rarely validated beyond single datasets, limiting institutional trust and scalability.

Classical supervised methods, including decision trees and support vector machines, remain prevalent in small-scale contexts due to their transparency and computational efficiency. Nevertheless, they demonstrate reduced robustness when applied to multimodal or unstructured data environments.

Clustering and unsupervised methods are effective in exploratory settings and where labelled data are limited. Yet their outputs often require post hoc interpretation and provide limited operational guidance without additional modelling layers.

Hybrid architectures represent a promising middle ground. By combining representation learning with interpretable classifiers or ensemble aggregation, they offer improved performance stability while partially addressing transparency concerns. This suggests that balanced, context-aware modelling strategies may be more sustainable than performance-driven model selection alone.

B. Structural Challenges and Persistent Gaps

Despite technical advances, several structural limitations constrain progress.

First, generalizability remains weak. Many high-performing models are trained and evaluated on proprietary or context-specific datasets, restricting claims of cross-institution robustness. Second, deployment validation is limited. Most systems are evaluated offline rather than within live LMS or classroom workflows, reducing confidence in real-world applicability. Third, interpretability is inconsistently embedded. Complex architectures such as CNNs and autoencoders often function as black boxes, posing barriers to educator adoption and accountability [87], [88]. Fourth, data quality issues, including sparsity, class imbalance, and contextual bias continue to affect model reliability and fairness [48]. These findings reinforce the need to shift from algorithm-centric experimentation toward deployment-aware system design.

C. Best Practices for Robust LS Modelling

Several methodological practices emerged as consistently associated with stronger contributions.

From a system development perspective, several practical guidelines emerge. First, model selection should be aligned with data availability, interpretability requirements, and deployment constraints rather than solely predictive performance. Second, multimodal fusion strategies should prioritize intermediate or adaptive architectures to improve robustness under heterogeneous data conditions. Third, evaluation protocols should incorporate multi-metric assessment and cross-context validation to enhance reliability. Finally, explainability and fairness mechanisms should be embedded during model design to support transparency, accountability, and institutional adoption.

These practices collectively support the development of scalable and context-aware learning analytics systems.

D. Practical and Ethical Implementation Considerations

For responsible educational deployment, implementation must prioritize interpretability, privacy, and fairness alongside predictive strength.

Initial deployments should favour interpretable or hybrid models, particularly in high-stakes academic decision contexts. Embedding explainability mechanisms during model design rather than as post hoc add-ons enhances educator confidence and accountability. Data governance practices must include anonymization, secure storage, and informed consent procedures [46]. Additionally, fairness auditing should be systematically applied to detect demographic or contextual bias. Ultimately, ML-based LS systems must align with institutional policies and societal values. Technical accuracy alone is insufficient for sustainable adoption.

VII. Limitations and Future Research Directions

A. Study Limitations

This study is subject to several limitations. First, the inclusion period (2020-2025) prioritizes recent developments but excludes earlier foundational work. Second, heterogeneity in datasets, evaluation protocols, and reporting standards limit the feasibility of quantitative meta-analysis. Third, reliance on indexed digital libraries introduces potential database bias, while publication bias may favour studies reporting positive results. Finally, the limited availability of publicly accessible datasets restricts external validation and reproducibility. These constraints should be considered when interpreting the findings.

B. Future Research Directions

Future research should focus on developing robust multimodal learning style identification systems that integrate behavioural, textual, affective, and physiological data using adaptive fusion strategies. Greater attention is also needed for explainable and fairness-aware machine learning models to improve transparency, trust, and responsible educational deployment. In addition, standardized benchmark datasets, unified validation protocols, and multi-metric evaluation frameworks should be established to improve reproducibility and cross-study comparability. Finally, future studies should move beyond offline experimentation toward deployment-oriented architectures that support real-time inference, scalability, and continuous adaptation in practical educational environments.

VIII. Conclusion

This study provided a systematic and structured synthesis of machine learning approaches for learning style identification published between 2020 and 2025. Moving beyond descriptive cataloguing of algorithms, the review reorganized the literature across four analytical dimensions: algorithmic paradigm evolution, multimodal integration patterns, evaluation practices, and deployment readiness characteristics. This analytical reframing revealed that although predictive performance has improved, methodological consistency, cross-context validation, and system-level maturity remain limited.

The findings indicate that classical supervised models continue to dominate small-scale institutional settings, while deep and ensemble methods are increasingly adopted in large-scale platforms. However, multimodal architectures remain largely shallow, interpretability is inconsistently embedded, and ethical considerations are rarely operationalized within deployment workflows. Most systems remain confined to experimental or dataset-validated stages, with few demonstrating sustained real-world integration.

To address these structural limitations, this study introduced a layered deployment-oriented framework that integrates multimodal input design, context-sensitive model selection, evaluation rigor, explainability integration, and a four-level deployment maturity spectrum. By positioning learning style modelling within a system-level architectural perspective, the framework shifts emphasis from isolated accuracy improvements toward scalable, transparent, and ethically grounded adaptive learning infrastructures.

Ultimately, the advancement of ML-based learning style systems will depend not only on algorithmic innovation but on principled fusion strategies, cross-context generalization, embedded explainability, and deployment-aware validation. This study provides a structured foundation to guide that transition and to support the development of robust next-generation adaptive educational systems.

References

[1] A. Ezzaim, A. Dahbi, A. Aqqal, and A. Haidine, “AI-based learning style detection in adaptive learning systems: a systematic literature review,” *J. Comput. Educ.*, Jun. 2024, doi: 10.1007/s40692-024-00328-9.

[2] M. Jebbari, B. Cherradi, S. Hamida, and A. Raihani, “Identifying learning styles in MOOCs environment through machine learning predictive modeling,” *Educ Inf Technol*, vol. 29, no. 16, pp. 20977–21014, Nov. 2024, doi: 10.1007/s10639-024-12637-8.

[3] R. Isaeva, N. Karasartova, K. Dznunusnalieva, K. Mirzoeva, and M. Mokliuk, “ENHANCING LEARNING EFFECTIVENESS THROUGH ADAPTIVE LEARNING PLATFORMS AND EMERGING COMPUTER TECHNOLOGIES IN EDUCATION,” *Jur. Ilmh. Ilm. Ter. Un. Ja*, vol. 9, no. 1, pp. 144–160, Feb. 2025, doi: 10.22437/jiituj.v9i1.37967.

[4] M. Liu and D. Yu, “Towards intelligent E-learning systems,” *Educ Inf Technol*, vol. 28, no. 7, pp. 7845–7876, Jul. 2023, doi: 10.1007/s10639-022-11479-6.

[5] A. Rashad Sayed, M. Helmy Khafagy, M. Ali, and M. Hussien Mohamed, “Predict student learning styles and suitable assessment methods using click stream,” *Egyptian Informatics Journal*, vol. 26, p. 100469, Jun. 2024, doi: 10.1016/j.eij.2024.100469.

[6] S. Rizwan, C. K. Nee, and S. Garfan, “Identifying the Factors Affecting Student Academic Performance and Engagement Prediction in MOOC Using Deep Learning: A Systematic Literature Review,” *IEEE Access*, vol. 13, pp. 18952–18982, 2025, doi: 10.1109/ACCESS.2025.3533915.

[7] K. Alahmadi, S. Alharbi, J. Chen, and X. Wang, “Generalizing sentiment analysis: a review of progress, challenges, and emerging directions,” *Soc. Netw. Anal. Min.*, vol. 15, no. 1, p. 45, Apr. 2025, doi: 10.1007/s13278-025-01461-8.

[8] S. Chaabene, A. Boudaya, B. Bouaziz, and L. Chaari, “An overview of methods and techniques in multimodal data fusion with application to healthcare,” *Int J Data Sci Anal*, vol. 20, no. 4, pp. 3093–3117, Oct. 2025, doi: 10.1007/s41060-025-00715-0.

[9] A. Almalawi, B. Soh, A. Li, and H. Samra, “Predictive Models for Educational Purposes: A Systematic Review,” *BDCC*, vol. 8, no. 12, p. 187, Dec. 2024, doi: 10.3390/bdcc8120187.

[10] P. Radanliev, “AI Ethics: Integrating Transparency, Fairness, and Privacy in AI Development,” *Applied Artificial Intelligence*, vol. 39, no. 1, p. 2463722, Dec. 2025, doi: 10.1080/08839514.2025.2463722.

[11] M. J. Page *et al.*, “The PRISMA 2020 statement: an updated guideline for reporting systematic reviews,” *BMJ*, p. n71, Mar. 2021, doi: 10.1136/bmj.n71.

[12] C. Waladi, M. Khaldi, and M. Lamarti Sefian, “Machine Learning Approach for an Adaptive E-Learning System Based on Kolb Learning Styles,” *Int. J. Emerg. Technol. Learn.*, vol. 18, no. 12, pp. 4–15, Jun. 2023, doi: 10.3991/ijet.v18i12.39327.

[13] M. Jebbari, B. Cherradi, W. Moutaouakil, O. El Gannour, S. Hamida, and A. Raihani, “Prediction and Classification of Learning Styles using Machine Learning Approach,” in *2023 3rd International Conference on Innovative Research in Applied Science, Engineering and Technology (IRASET)*, Mohammedia, Morocco: IEEE, May 2023, pp. 1–7. doi: 10.1109/IRASET57153.2023.10152916.

[14] H. Y. Ayyoub and O. S. Al-Kadi, “Learning Style Identification Using Semisupervised Self-Taught Labeling,” *IEEE Trans. Learning Technol.*, vol. 17, pp. 1093–1106, 2024, doi: 10.1109/TLT.2024.3358864.

[15] H.-H. Nguyen *et al.*, “Applying machine learning techniques to detect student’s learning styles,” in *Proceedings of the 14th International Conference on Education Technology and Computers*, Barcelona Spain: ACM, Oct. 2022, pp. 456–462. doi: 10.1145/3572549.3572622.

[16] A. Lestari, A. Lawi, S. A. Thamrin, and N. Hidayat, “Automated Detection of Learning Styles using Online Activities and Model Indicators,” *ijacsa*, vol. 15, no. 6, 2024, doi: 10.14569/IJACSA.2024.0150661.

[17] Q. Ni, Y. Mi, Y. Wu, L. He, Y. Xu, and B. Zhang, “Design and Implementation of the Reliable Learning Style Recognition Mechanism Based on Fusion Labels and Ensemble Classification,” *IEEE Trans. Learning Technol.*, vol. 17, pp. 241–257, 2024, doi: 10.1109/TLT.2023.3263568.

[18] T. Hussain, L. Yu, M. Asim, A. Ahmed, and M. A. Wani, “Enhancing E-Learning Adaptability with Automated Learning Style Identification and Sentiment Analysis: A Hybrid Deep Learning Approach for Smart Education,” *Information*, vol. 15, no. 5, p. 277, May 2024, doi: 10.3390/info15050277.

[19] B. A. Muhammad, C. Qi, Z. Wu, and H. K. Ahmad, “An evolving learning style detection approach for online education using bipartite graph embedding,” *Applied Soft Computing*, vol. 152, p. 111230, Feb. 2024, doi: 10.1016/j.asoc.2024.111230.

[20] R. Goran *et al.*, “Identifying and Understanding Student Dropouts Using Metaheuristic Optimized Classifiers and Explainable Artificial Intelligence Techniques,” *IEEE Access*, vol. 12, pp. 122377–122400, 2024, doi: 10.1109/ACCESS.2024.3446653.

[21] L. R. Pelima, Y. Sukmana, and Y. Rosmansyah, “Predicting University Student Graduation Using Academic Performance and Machine Learning: A Systematic Literature Review,” *IEEE Access*, vol. 12, pp. 23451–23465, 2024, doi: 10.1109/ACCESS.2024.3361479.

[22] A. Zingoni, J. Taborri, and G. Calabrò, “A machine learning-based classification model to support university students with dyslexia with

personalized tools and strategies," *Sci Rep*, vol. 14, no. 1, p. 273, Jan. 2024, doi: 10.1038/s41598-023-50879-7.

[23] Y. A. Bena *et al.*, "Harnessing and Mitigating Big Data Governance Challenges Using Hybrid Approach: A Systematic Literature Review," *IEEE Access*, vol. 12, pp. 175151–175175, 2024, doi: 10.1109/ACCESS.2024.3498947.

[24] L. Yang *et al.*, "Quality Assessment in Systematic Literature Reviews: A Software Engineering Perspective," *Information and Software Technology*, vol. 130, p. 106397, Feb. 2021, doi: 10.1016/j.infsof.2020.106397.

[25] I. Azzi, A. Jeghal, A. Radouane, A. Yahyaouy, and H. Tairi, "A robust classification to predict learning styles in adaptive E-learning systems," *Educ Inf Technol*, vol. 25, no. 1, pp. 437–448, Jan. 2020, doi: 10.1007/s10639-019-09956-6.

[26] R. D. Da Costa, G. F. De Souza, T. B. De Castro, R. A. De Medeiros Valentim, and A. De Pinho Dias, "Identification of Learning Styles in Distance Education Through the Interaction of the Student With a Learning Management System," *IEEE R. Iberoamericana Tecnologias Aprendizaje*, vol. 15, no. 3, pp. 148–160, Aug. 2020, doi: 10.1109/RITA.2020.3008131.

[27] E. Gomede, R. Miranda De Barros, and L. De Souza Mendes, "Use of Deep Multi-Target Prediction to Identify Learning Styles," *Applied Sciences*, vol. 10, no. 5, p. 1756, Mar. 2020, doi: 10.3390/app10051756.

[28] N. Hidayat, R. Wardoyo, A. Sn, and H. Dwi, "Enhanced Performance of the Automatic Learning Style Detection Model using a Combination of Modified K-Means Algorithm and Naive Bayesian," *IJACSA*, vol. 11, no. 3, 2020, doi: 10.14569/IJACSA.2020.0110380.

[29] B. Hmedna, A. El Mezouary, and O. Baz, "A predictive model for the identification of learning styles in MOOC environments," *Cluster Comput*, vol. 23, no. 2, pp. 1303–1328, Jun. 2020, doi: 10.1007/s10586-019-02992-4.

[30] S. M. Nafea, F. Siewe, and Y. He, "A Novel Algorithm for Dynamic Student Profile Adaptation Based on Learning Styles," in *Intelligent Systems and Applications*, vol. 1037, Y. Bi, R. Bhatia, and S. Kapoor, Eds., in Advances in Intelligent Systems and Computing, vol. 1037. , Cham: Springer International Publishing, 2020, pp. 29–51. doi: 10.1007/978-3-030-29516-5_4.

[31] G. L. Pritalia, S. Wibirama, T. B. Adji, and S. Kusrohmaniah, "Classification of Learning Styles in Multimedia Learning Using Eye-Tracking and Machine Learning," in *2020 FORTEI-International Conference on Electrical Engineering (FORTEI-ICEE)*, Bandung, Indonesia: IEEE, Sep. 2020, pp. 145–150. doi: 10.1109/FORTEI-ICEE50915.2020.9249875.

[32] R. Rajkumar and V. Ganapathy, "Bio-Inspiring Learning Style Chatbot Inventory Using Brain Computing Interface to Increase the Efficiency of E-Learning," *IEEE Access*, vol. 8, pp. 67377–67395, 2020, doi: 10.1109/ACCESS.2020.2984591.

[33] M. K. Sabariah, P. I. Santosa, and R. Ferdiana, "Identification of Children Learning Styles Using Elicitation Application," *Journal of Computer Science*, vol. 16, no. 6, pp. 856–862, Jun. 2020, doi: 10.3844/jcssp.2020.856.862.

[34] D. S. S. Sahid, R. Efendi, E. Hasri, and M. Wahyudi, "Categorizing Attributes in Identifying Learning Style using Rough Set Theory," *IJACSA*, vol. 11, no. 1, 2020, doi: 10.14569/IJACSA.2020.0110136.

[35] A. Al-Qerem, "Identifying Driver behaviors in learning style classification Using Machine Learning Technique," *IJETER*, vol. 8, no. 6, pp. 2668–2673, Jun. 2020, doi: 10.30534/ijeter/2020/73862020.

[36] H. Zhang *et al.*, "A learning style classification approach based on deep belief network for large-scale online education," *J Cloud Comp*, vol. 9, no. 1, p. 26, Dec. 2020, doi: 10.1186/s13677-020-00165-y.

[37] I. Uddin, A. S. Imran, K. Muhammad, N. Fayyaz, and M. Sajjad, "A Systematic Mapping Review on MOOC Recommender Systems," *IEEE Access*, vol. 9, pp. 118379–118405, 2021, doi: 10.1109/ACCESS.2021.3101039.

[38] P. Ramírez-Correa, J. Alfaro-Pérez, and M. Gallardo, "Identifying Engineering Undergraduates' Learning Style Profiles Using Machine Learning Techniques," *Applied Sciences*, vol. 11, no. 22, p. 10505, Nov. 2021, doi: 10.3390/app112210505.

[39] F. Rasheed and A. Wahid, "Learning style detection in E-learning systems using machine learning techniques," *Expert Systems with Applications*, vol. 174, p. 114774, Jul. 2021, doi: 10.1016/j.eswa.2021.114774.

[40] V. F. Mirama, L. E. Diez, A. Bahillo, and V. Quintero, "A Survey of Machine Learning in Pedestrian Localization Systems: Applications, Open Issues and Challenges," *IEEE Access*, vol. 9, pp. 120138–120157, 2021, doi: 10.1109/ACCESS.2021.3108073.

[41] S. M. Aslam, A. K. Jilani, J. Sultana, and L. Almutairi, "Feature Evaluation of Emerging E-Learning Systems Using Machine Learning: An Extensive Survey," *IEEE Access*, vol. 9, pp. 69573–69587, 2021, doi: 10.1109/ACCESS.2021.3077663.

[42] "Process of building a dataset and classification of vark learning styles with machine learning and predictive analytics models," *cibg*, vol. 26, no. 02, Feb. 2021, doi: 10.47750/cibg.2020.26.02.128.

[43] E. M. Onyema *et al.*, "Prospects and Challenges of Using Machine Learning for Academic Forecasting," *Computational Intelligence and Neuroscience*, vol. 2022, pp. 1–7, Jun. 2022, doi: 10.1155/2022/5624475.

[44] B. Pardamean, T. Suparyanto, T. W. Cenggoro, D. Sudigyo, and A. Anugrahana, "AI-Based Learning Style Prediction in Online Learning for Primary Education," *IEEE Access*, vol. 10, pp. 35725–35735, 2022, doi: 10.1109/ACCESS.2022.3160177.

[45] W. Song and Z. Wang, "Improved Clustering Strategies for Learning Style Identification in Massive Open Online Courses," in *Data Mining and Big Data*, vol. 1744, Y. Tan and Y. Shi, Eds., in Communications in Computer and Information Science, vol. 1744. , Singapore: Springer Nature Singapore, 2022, pp. 240–254. doi: 10.1007/978-981-19-9297-1_18.

[46] S. V. Mahadevkar *et al.*, "A Review on Machine Learning Styles in Computer Vision—Techniques and Future Directions," *IEEE Access*, vol. 10, pp. 107293–107329, 2022, doi: 10.1109/ACCESS.2022.3209825.

[47] Z. Mehenaoui, Y. Lafifi, and L. Zemmouri, "Learning Behavior Analysis to Identify Learner’s Learning Style based on Machine Learning Techniques," *jucs*, vol. 28, no. 11, pp. 1193–1220, Nov. 2022, doi: 10.3897/jucs.81518.

[48] K. Hosny and A. El-korany, "Applying adaptive learning by integrating semantic and machine learning in proposing student assessment model," *IJECE*, vol. 12, no. 2, p. 2014, Apr. 2022, doi: 10.11591/ijece.v12i2.pp2014-2025.

[49] M. Jebbari, B. Cherradi, O. El Gannour, S. Hamida, and A. Raihani, "Exploration Study on Learning Styles Identification and Prediction Techniques," in *2022 2nd International Conference on Innovative Research in Applied Science, Engineering and Technology (IRASET)*, Meknes, Morocco: IEEE, Mar. 2022, pp. 1–7. doi: 10.1109/IRASET52964.2022.9738030.

[50] B. A. Muhammad, B. Liu, H. K. Ahmad, M. Umar, and K. Qin, "GNN-LS: A Learning Style Prediction in Online Environments using Graph Neural Networks," *J-NaNA*, vol. 2, no. 4, pp. 172–182, 2022, doi: 10.33969/J-NaNA.2022.020405.

[51] B. Ahmad Muhammad, C. Qi, Z. Wu, and H. Kabir Ahmad, "GRL-LS: A learning style detection in online education using graph representation learning," *Expert Systems with Applications*, vol. 201, p. 117138, Sep. 2022, doi: 10.1016/j.eswa.2022.117138.

[52] H. Munir, B. Vogel, and A. Jacobsson, "Artificial Intelligence and Machine Learning Approaches in Digital Education: A Systematic Revision," *Information*, vol. 13, no. 4, p. 203, Apr. 2022, doi: 10.3390/info13040203.

[53] M. Murtaza, Y. Ahmed, J. A. Shamsi, F. Sherwani, and M. Usman, "AI-Based Personalized E-Learning Systems: Issues, Challenges, and Solutions," *IEEE Access*, vol. 10, pp. 81323–81342, 2022, doi: 10.1109/ACCESS.2022.3193938.

[54] S. Alshmrany, "Adaptive learning style prediction in e-learning environment using levy flight distribution based CNN model," *Cluster Comput*, vol. 25, no. 1, pp. 523–536, Feb. 2022, doi: 10.1007/s10586-021-03403-3.

[55] A. M. Altamimi, M. Azzeh, and M. Albashayreh, "Predicting students' learning styles using regression techniques," *IJEECS*, vol. 25, no. 2, p. 1177, Feb. 2022, doi: 10.11591/ijeecs.v25.i2.pp1177-1185.

[56] K. A. H. Assiry and A. Muniasamy, "Predicting Learning Styles Using Machine Learning Classifiers," in *2022 International Conference on*

Electrical, Computer and Energy Technologies (ICECET), Prague, Czech Republic: IEEE, Jul. 2022, pp. 1–7. doi: 10.1109/ICECET55527.2022.9872971.

[57] D. Goštautaitė and L. Sakalauskas, "Multi-Label Classification and Explanation Methods for Students' Learning Style Prediction and Interpretation," *Applied Sciences*, vol. 12, no. 11, p. 5396, May 2022, doi: 10.3390/app12115396.

[58] M. S. Hasibuan and R. A. Aziz, "Detection of learning styles with prior knowledge data using the SVM, K-NN and Naïve Bayes algorithms," *J.INFOTEL*, vol. 14, no. 3, pp. 209–213, Aug. 2022, doi: 10.20895/infotel.v14i3.788.

[59] S. Rami, S. Bennani, and M. Khalidi Idrissi, "Cognitive Learning Style Detection in e-Learning Environments using Artificial Neural Network," *Int. J. Emerg. Technol. Learn.*, vol. 17, no. 17, pp. 62–77, Sep. 2022, doi: 10.3991/ijet.v17i17.30243.

[60] L. K. Smirani, H. A. Yamani, L. J. Menzli, and J. A. Boulahia, "Using Ensemble Learning Algorithms to Predict Student Failure and Enabling Customized Educational Paths," *Scientific Programming*, vol. 2022, pp. 1–15, Apr. 2022, doi: 10.1155/2022/3805235.

[61] S. G. Essa, T. Celik, and N. E. Human-Hendricks, "Personalized Adaptive Learning Technologies Based on Machine Learning Techniques to Identify Learning Styles: A Systematic Literature Review," *IEEE Access*, vol. 11, pp. 48392–48409, 2023, doi: 10.1109/ACCESS.2023.3276439.

[62] A. Yaw Obeng, "Consequential effects of using competing perspectives to predict learning style in e-learning systems," *Cogent Education*, vol. 10, no. 1, p. 2218960, Dec. 2023, doi: 10.1080/2331186X.2023.2218960.

[63] M. S. Hasibuan, R. Z. Abdul Aziz, D. A. Dewi, T. B. Kurniawan, and N. A. Syafira, "Recommendation Model for Learning Material Using the Felder Silverman Learning Style Approach," *HighTech. Innov. J.*, vol. 4, no. 4, pp. 811–820, Dec. 2023, doi: 10.28991/HIJ-2023-04-04-010.

[64] R. Nazempour and H. Darabi, "Personalized Learning in Virtual Learning Environments Using Students' Behavior Analysis," *Education Sciences*, vol. 13, no. 5, p. 457, Apr. 2023, doi: 10.3390/educsci13050457.

[65] System Engineer, Tata Consultancy Services, Chennai, Tamil nadu *et al.*, "Machine Learning based Tutor Ward System (MLTWS) for Cognitive Learning style Management," *JEET*, vol. 36, no. 3, pp. 146–155, Jan. 2023, doi: 10.16920/jeet/2023/v36i3/23106.

[66] J. Huang and J. Liu, "Learning Styles Identification Model in a MOOC Learning Environment," in *Computer Science and Education*, W. Hong and Y. Weng, Eds., Singapore: Springer Nature Singapore, 2023, pp. 234–244.

[67] W. Waam and P. Hks, "Identifying the Learning Style of Students Using Machine Learning Techniques: An Approach of Felder Silverman Learning Style Model (FSLSM)," *AJRCoS*, vol. 17, no. 3, pp. 15–37, Jan. 2024, doi: 10.9734/ajrcos/2024/v17i3422.

[68] W. Villegas-Ch, J. García-Ortiz, and S. Sánchez-Viteri, "Personalization of Learning: Machine Learning Models for Adapting Educational Content to Individual Learning Styles," *IEEE Access*, vol. 12, pp. 121114–121130, 2024, doi: 10.1109/ACCESS.2024.3452592.

[69] L. Genith Isaza Domínguez, A. Robles-Gómez, and R. Pastor-Vargas, "A Data-Driven Approach to Engineering Instruction: Exploring Learning Styles, Study Habits, and Machine Learning," *IEEE Access*, vol. 13, pp. 10978–11002, 2025, doi: 10.1109/ACCESS.2025.3528263.

[70] A. R. Sayed, M. H. Khafagy, M. Ali, and M. H. Mohamed, "Exploring the VAK model to predict student learning styles based on learning activity," *Intelligent Systems with Applications*, vol. 25, p. 200483, Mar. 2025, doi: 10.1016/j.iswa.2025.200483.

[71] S. Bousalem, F. Benchikha, and N. Marir, "Personalized Learning Through MBTI Prediction: A Deep Learning Approach Integrated With Learner Profile Ontology," *IEEE Access*, vol. 13, pp. 30861–30873, 2025, doi: 10.1109/ACCESS.2025.3542701.

[72] J. Govea, E. Ocampo Edye, S. Revelo-Tapia, and W. Villegas-Ch, "Optimization and Scalability of Educational Platforms: Integration of Artificial Intelligence and Cloud Computing," *Computers*, vol. 12, no. 11, p. 223, Nov. 2023, doi: 10.3390/computers12110223.

[73] K. T. Kotsis, "Integrating Artificial Intelligence into Higher Education Curricula: Challenges and Opportunities for Science-Based Programs," *IJAIET*, vol. 6, no. 1, pp. 04–12, 2025, doi: 10.54660/IJAIET.2025.6.1.04-12.

[74] R. I. E. Saragih, "AI-Powered Education: Transforming Learning through Personalized and Scalable Solutions," *Int. J. Inf. Syst. Innovative Technol.*, vol. 3, no. 2, pp. 1–9, Dec. 2024, doi: 10.63322/qm9dk118.

[75] D. Bakir, K. Moussaid, Z. Chiba, N. Abghour, and A. El Omri, "A SPAR-4-SLR Systematic Review of AI-Based Traffic Congestion Detection: Model Performance Across Diverse Data Types," *Smart Cities*, vol. 8, no. 5, p. 143, Aug. 2025, doi: 10.3390/smartcities8050143.

[76] S. Al Amoudi, A. Alhothali, R. Mirza, H. Assalahi, and T. Aldosemani, "Click-Based Representation Learning Framework of Student Navigational Behavior in MOOCs," *IEEE Access*, vol. 12, pp. 121480–121494, 2024, doi: 10.1109/ACCESS.2024.3450514.

[77] H. Chen and D. Tang, "Multimodal Data Fusion and Decision Algorithms in Deep Learning-Based Intelligent Systems: A Comprehensive Study," in *Proceedings of the 2025 International Conference on Artificial Intelligence and Smart Manufacturing*, Chongqing China: ACM, May 2025, pp. 802–811. doi: 10.1145/3756423.3756555.

[78] G. Lee *et al.*, "Multimodality of AI for Education: Toward Artificial General Intelligence," *IEEE Trans. Learning Technol.*, vol. 18, pp. 666–683, 2025, doi: 10.1109/TLT.2025.3574466.

[79] P. P. Liang, A. Zadeh, and L.-P. Morency, "Foundations & Trends in Multimodal Machine Learning: Principles, Challenges, and Open Questions," *ACM Comput. Surv.*, vol. 56, no. 10, pp. 1–42, Oct. 2024, doi: 10.1145/3656580.

[80] F. Zhao, C. Zhang, and B. Geng, "Deep Multimodal Data Fusion," *ACM Comput. Surv.*, vol. 56, no. 9, pp. 1–36, Oct. 2024, doi: 10.1145/3649447.

[81] S. Li and H. Tang, "Multimodal Alignment and Fusion: A Survey," Oct. 11, 2025, *arXiv*: arXiv:2411.17040. doi: 10.48550/arXiv.2411.17040.

[82] R. Meylani, "RETHINKING EXPLAINABILITY IN EDUCATIONAL ARTIFICIAL INTELLIGENCE: A CRITICAL SYSTEMATIC REVIEW OF MODELS, APPLICATIONS, AND ETHICAL DIMENSIONS," *Ver. Dir.*, vol. 22, no. 3, p. e223553, Oct. 2025, doi: 10.18623/rvd.v22.n3.3553.

[83] J. Niewint-Gori, "Explainable AI in education: A multi-stakeholder approach to transparency and ethical practice," *IUL Research*, vol. 6, no. 12, pp. 117–132, Dec. 2025, doi: 10.57568/iulresearch.v6i12.781.

[84] O. Embarak, "Bridging Theory and Application: A Review of Explainable AI with a Case Study in Adaptive Learning Systems," *Procedia Computer Science*, vol. 265, pp. 73–82, 2025, doi: 10.1016/j.procs.2025.07.158.

[85] ariyanto, F. X. D. Kristianingsih, and R. Maharani, "Artificial intelligence in adaptive education: a systematic review of techniques for personalized learning," *Discov Educ*, vol. 4, no. 1, p. 458, Oct. 2025, doi: 10.1007/s44217-025-00908-6.

[86] O. Al-Omari, A. Alyousef, S. Fati, F. Shannaq, and A. Omari, "Governance and Ethical Frameworks for AI Integration in Higher Education: Enhancing Personalized Learning and Legal Compliance," *JoE*, vol. 4, no. 2, Jan. 2025, doi: 10.62754/joe.v4i2.5781.

[87] V. Hassija *et al.*, "Interpreting Black-Box Models: A Review on Explainable Artificial Intelligence," *Cogn Comput*, vol. 16, no. 1, pp. 45–74, Jan. 2024, doi: 10.1007/s12559-023-10179-8.

[88] E. ŞAHiN, N. N. Arslan, and D. Özdemir, "Unlocking the black box: an in-depth review on interpretability, explainability, and reliability in deep learning," *Neural Comput & Applic*, vol.

APPENDIX: PRISMA 2020 CHECKLIST

Section and Topic	Item #	Checklist item	Location where item is reported
TITLE			
Title	1	Identify the report as a systematic review.	1
ABSTRACT			
Abstract	2	See the PRISMA 2020 for Abstracts checklist.	1
INTRODUCTION			
Rationale	3	Describe the rationale for the review in the context of existing knowledge.	2
Objectives	4	Provide an explicit statement of the objective(s) or question(s) the review addresses.	2
METHODS			
Eligibility criteria	5	Specify the inclusion and exclusion criteria for the review and how studies were grouped for the syntheses.	7
Information sources	6	Specify all databases, registers, websites, organisations, reference lists and other sources searched or consulted to identify studies. Specify the date when each source was last searched or consulted.	6
Search strategy	7	Present the full search strategies for all databases, registers and websites, including any filters and limits used.	7
Selection process	8	Specify the methods used to decide whether a study met the inclusion criteria of the review, including how many reviewers screened each record and each report retrieved, whether they worked independently, and if applicable, details of automation tools used in the process.	7
Data collection process	9	Specify the methods used to collect data from reports, including how many reviewers collected data from each report, whether they worked independently, any processes for obtaining or confirming data from study investigators, and if applicable, details of automation tools used in the process.	7-9
Data items	10a	List and define all outcomes for which data were sought. Specify whether all results that were compatible with each outcome domain in each study were sought (e.g. for all measures, time points, analyses), and if not, the methods used to decide which results to collect.	N/A
	10b	List and define all other variables for which data were sought (e.g. participant and intervention characteristics, funding sources). Describe any assumptions made about any missing or unclear information.	N/A
Study risk of bias assessment	11	Specify the methods used to assess risk of bias in the included studies, including details of the tool(s) used, how many reviewers assessed each study and whether they worked independently, and if applicable, details of automation tools used in the process.	N/A
Effect measures	12	Specify for each outcome the effect measure(s) (e.g. risk ratio, mean difference) used in the synthesis or presentation of results.	16
Synthesis methods	13a	Describe the processes used to decide which studies were eligible for each synthesis (e.g. tabulating the study intervention characteristics and comparing against the planned groups for each synthesis (item #5)).	7, 11
	13b	Describe any methods required to prepare the data for presentation or synthesis, such as handling of missing summary statistics, or data conversions.	3-4
	13c	Describe any methods used to tabulate or visually display results of individual studies and syntheses.	3-4, 12-15
	13d	Describe any methods used to synthesize results and provide a rationale for the choice(s). If meta-analysis was performed, describe the model(s), method(s) to identify the presence and extent of statistical heterogeneity, and software package(s) used.	N/A
	13e	Describe any methods used to explore possible causes of heterogeneity among study results (e.g. subgroup analysis, meta-regression).	N/A
	13f	Describe any sensitivity analyses conducted to assess robustness of the synthesized results.	N/A
Reporting bias assessment	14	Describe any methods used to assess risk of bias due to missing results in a synthesis (arising from reporting biases).	7
Certainty assessment	15	Describe any methods used to assess certainty (or confidence) in the body of evidence for an outcome.	7
RESULTS			
Study selection	16a	Describe the results of the search and selection process, from the number of records identified in the search to the number of studies included in the review, ideally using a flow diagram.	8-11
	16b	Cite studies that might appear to meet the inclusion criteria, but which were excluded, and explain why they were excluded.	N/A
Study characteristics	17	Cite each included study and present its characteristics.	10-11
Risk of bias in studies	18	Present assessments of risk of bias for each included study.	15-20
Results of individual studies	19	For all outcomes, present, for each study: (a) summary statistics for each group (where appropriate) and (b) an effect estimate and its precision (e.g. confidence/credible interval), ideally using structured tables or plots.	15-20

Section and Topic	Item #	Checklist item	Location where item is reported
Results of syntheses	20a	For each synthesis, briefly summarise the characteristics and risk of bias among contributing studies.	15-20
	20b	Present results of all statistical syntheses conducted. If meta-analysis was done, present for each the summary estimate and its precision (e.g. confidence/credible interval) and measures of statistical heterogeneity. If comparing groups, describe the direction of the effect.	15-20
	20c	Present results of all investigations of possible causes of heterogeneity among study results.	15-20
	20d	Present results of all sensitivity analyses conducted to assess the robustness of the synthesized results.	15-20
Reporting biases	21	Present assessments of risk of bias due to missing results (arising from reporting biases) for each synthesis assessed.	15-20
Certainty of evidence	22	Present assessments of certainty (or confidence) in the body of evidence for each outcome assessed.	15-20
DISCUSSION			
Discussion	23a	Provide a general interpretation of the results in the context of other evidence.	20, 22
	23b	Discuss any limitations of the evidence included in the review.	21-22
	23c	Discuss any limitations of the review processes used.	21-22
	23d	Discuss implications of the results for practice, policy, and future research.	21-22
OTHER INFORMATION			
Registration and protocol	24a	Provide registration information for the review, including register name and registration number, or state that the review was not registered.	N/A
	24b	Indicate where the review protocol can be accessed, or state that a protocol was not prepared.	N/A
	24c	Describe and explain any amendments to information provided at registration or in the protocol.	N/A
Support	25	Describe sources of financial or non-financial support for the review, and the role of the funders or sponsors in the review.	22
Competing interests	26	Declare any competing interests of review authors.	22
Availability of data, code and other materials	27	Report which of the following are publicly available and where they can be found: template data collection forms; data extracted from included studies; data used for all analyses; analytic code; any other materials used in the review.	22

From: Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021;372:n71. doi: 10.1136/bmj.n71.