ISSN 2277-2685
IJESR/April-June. 2024/ Vol-14/Issue-2/1799-1808
Nikita Deshmukh *et. al.,* / International Journal of Engineering & Science Research

Ai-Based Student Performance Prediction: A Systematic Review And Meta-Analysis

Nikita Deshmukh[1], Dr. Diwakar Tripathi[2]
[1]Research Scholar, Department of Computer Science, ISBM University, Nawapara (Kosmi), Chhattisgarh, India.
[2]Assistant Professor, Department of Computer Science, ISBM University, Nawapara (Kosmi), Chhattisgarh, India.

Abstract

Student performance prediction has emerged as a critical challenge in contemporary educational systems, necessitating the development of intelligent, data-driven solutions to identify at-risk learners and optimize academic outcomes. This paper presents a comprehensive review and meta-analysis of artificial intelligence (AI)-based approaches applied to student performance prediction, synthesizing findings from seminal and recent studies published between 2010 and 2024. The review systematically examines machine learning algorithms including decision trees, random forests, support vector machines, neural networks, and deep learning architectures deployed across diverse educational datasets encompassing demographic, behavioral, academic, and socio-economic variables. Through rigorous analysis of prediction accuracy, model interpretability, feature importance, and deployment contexts, this paper identifies prevailing methodological paradigms, critical research gaps, and emergent directions in the field. The meta-analysis reveals that ensemble methods and deep learning models consistently achieve superior predictive performance, with accuracy rates ranging from 75% to 98% across reviewed studies. Furthermore, the study highlights the growing significance of learning management system (LMS) data, student engagement metrics, and multimodal data fusion in enhancing model robustness. Ethical considerations, including algorithmic bias, data privacy, and model transparency, are also critically examined. The findings underscore the transformative potential of AI-driven educational analytics while advocating for standardized benchmarking and inclusive dataset curation to ensure equitable and generalizable prediction systems.

Keywords: *Artificial Intelligence, Student Performance Prediction, Machine Learning, Educational Data Mining, Deep Learning, Learning Analytics, Academic Outcome Forecasting*

1. Introduction

The global education sector is undergoing a profound transformation driven by the proliferation of digital learning platforms, institutional data repositories, and advanced computational techniques. Among the most consequential applications of this transformation is the prediction of student academic performance a domain that intersects educational psychology, data science, and institutional policy. Early identification of students likely to underperform or disengage from their academic programs empowers educators and administrators to deploy timely, targeted interventions, thereby reducing dropout rates, improving retention, and fostering equitable learning environments. Traditional approaches to performance monitoring, which relied predominantly on periodic assessments and faculty intuition, have proven insufficient in the face of increasingly large and heterogeneous student populations, particularly within online and blended learning environments. Artificial intelligence (AI) and its subfields machine learning (ML), deep learning (DL), and natural language processing (NLP) have emerged as powerful enablers of predictive educational analytics. These technologies leverage vast and varied datasets, including academic records, attendance logs, psychological assessments, socio-economic indicators, and digital engagement metrics, to construct models capable of forecasting student outcomes with remarkable precision. The integration of AI into student performance prediction systems (SPPS) has catalyzed a new era in educational data mining (EDM) and learning analytics (LA), disciplines explicitly dedicated to extracting actionable insights from academic data ecosystems. Despite rapid advancements in the field, significant heterogeneity persists across published studies in terms of algorithmic choices, feature engineering strategies, dataset characteristics, evaluation metrics, and deployment contexts. This diversity complicates cross-study comparisons and hampers the establishment of universal benchmarks or best practices. A systematic review and meta-analysis that critically synthesizes the existing body of literature is therefore essential to consolidate scientific knowledge, identify methodological strengths and weaknesses, and chart productive directions for future research. The present paper undertakes this task, offering a structured examination of AI-based student performance prediction systems developed over the past decade.

1.1 Background and Motivation

The prediction of student academic performance has been a subject of scholarly inquiry for several decades, with early studies relying on statistical regression techniques and psychometric assessments. The introduction of educational data mining as a formal discipline in the early 2000s marked a turning point, as researchers began applying computational methods such as clustering, classification, and association rule mining to institutional datasets to uncover latent patterns in student learning behavior. Foundational work by Baker and Yacef (2009) and Romero and Ventura (2010) established the theoretical and methodological scaffolding upon which subsequent AI-driven approaches were constructed. The explosive growth of learning management systems (LMS) such as Moodle, Blackboard, and Canvas generated unprecedented volumes of timestamped behavioral data, including login frequency, assignment submission patterns, discussion forum participation, and resource access logs. These data streams, combined with demographic and prior academic records, provided fertile ground for the application of supervised machine learning algorithms to performance prediction tasks. Simultaneously, advances in neural network architectures particularly recurrent neural networks (RNNs), long short-term memory (LSTM) networks, and transformer models opened new avenues for modeling temporal dependencies in student learning trajectories. The motivation for developing accurate SPPS extends beyond academic curiosity: institutions that deploy such systems can achieve significant improvements in retention, graduation rates, and student satisfaction, translating into both social and economic benefits. Furthermore, in resource-constrained educational settings prevalent across developing nations, AI-powered early warning systems can maximize the impact of limited counseling and support resources by directing them toward students most in need.

1.2 Scope and Objectives of the Review

This review is scoped to encompass peer-reviewed journal articles, conference proceedings, and book chapters published between 2010 and 2024 that address the application of AI and machine learning techniques to the prediction of student academic performance at the primary, secondary, and tertiary educational levels. Studies employing purely statistical methods without any machine learning component are excluded, as are papers focused exclusively on administrative or financial aspects of student success. The primary objectives of this review are threefold. First, to systematically catalog and classify AI methodologies applied in student performance prediction, including supervised learning algorithms, ensemble methods, deep learning architectures, and hybrid models. Second, to critically evaluate the datasets, feature sets, and evaluation criteria employed across studies, assessing their comprehensiveness, generalizability, and alignment with real-world educational contexts. Third, to perform a meta-analysis that quantifies predictive performance trends, identifies moderating factors such as dataset size, educational level, and prediction target and synthesizes overarching conclusions regarding the state of the art. By fulfilling these objectives, the review aims to serve as a definitive reference for researchers, educational technologists, and policy-makers seeking to design, evaluate, or deploy AI-based student performance prediction systems.

1.3 Organization of the Paper

The remainder of this paper is organized as follows. Section 2 presents a comprehensive literature survey covering key studies on AI-based student performance prediction, organized thematically by algorithmic paradigm. Section 3 describes the research methodology adopted for the systematic review and meta-analysis, including search strategies, inclusion/exclusion criteria, and analytical frameworks. Section 4 provides a critical analysis of past work, examining methodological limitations, inconsistencies, and biases identified across reviewed studies. Section 5 discusses overarching findings, emerging trends, and implications for educational practice and policy. Section 6 concludes the paper with a synthesis of key insights and recommendations for future research.

2. Literature Survey

The literature on AI-based student performance prediction is vast and multidisciplinary, drawing upon contributions from computer science, educational psychology, statistics, and information systems. The following survey is organized into thematic categories corresponding to the primary algorithmic paradigms and methodological innovations documented in the reviewed literature. Machine learning techniques, particularly decision trees and their ensemble derivatives, have constituted the most extensively studied class of algorithms for student performance prediction. Romero et al. (2008) pioneered the use of decision tree classifiers applied to Moodle LMS data, demonstrating that student interaction patterns such as quiz attempts and resource downloads were strong predictors of final course grades. Subsequent work by Kotsiantis et al. (2010) employed a diverse battery of classifiers, including naive Bayes, k-nearest neighbors (kNN), and support vector machines (SVM), on a dataset of distance-learning students, reporting that SVMs achieved the highest accuracy (84.6%) in binary pass/fail classification tasks. The prevalence of decision tree-based methods in early studies was attributed to their interpretability, which facilitated actionable insights for educators. Random forests and gradient boosting algorithms, introduced to the SPPS domain in the mid-2010s, significantly elevated predictive performance. Huang and Fang (2013) applied random forest classifiers to engineering student data, achieving 91.2% accuracy in GPA prediction, while also identifying cumulative GPA, study hours, and part-time employment as the most influential predictive features. Gradient boosting methods, notably XGBoost, gained prominence following the work of Chen and He (2015), with subsequent adaptations to educational datasets by Hussain et al. (2019) demonstrating accuracy improvements of 6–12% over conventional tree models in multi-class grade prediction tasks. Support vector machines (SVMs) have been extensively applied in the context of binary and multi-class performance prediction. Kabakchieva (2013) utilized SVMs alongside neural networks to classify university students into performance categories based on pre-enrollment data, reporting that SVMs outperformed neural networks on smaller datasets but were surpassed by ensemble methods as dataset size increased. The study underscored the sensitivity of SVM performance to kernel selection and hyperparameter tuning. Logistic regression, while less computationally sophisticated, has remained a competitive baseline in many benchmark studies, particularly when interpretability and computational efficiency are prioritized. Delen et al. (2013) applied logistic regression, decision trees, and artificial neural networks to a large-scale undergraduate dataset, finding that artificial neural networks (ANNs) achieved marginally superior accuracy (80.5%) in predicting student attrition, though logistic regression offered comparable performance with substantially lower computational overhead. The advent of deep learning introduced qualitatively new capabilities for student performance prediction, particularly in modeling non-linear relationships and temporal dynamics in learning trajectories. Recurrent neural networks (RNNs) and their LSTM variants have been applied to sequential learning data extracted from LMS platforms. Piech et al. (2015) developed the Knowledge Tracing model using LSTMs to predict student performance on knowledge components in intelligent tutoring systems, achieving a significant improvement over Bayesian Knowledge Tracing baselines. Zhang et al. (2017) extended deep learning approaches by incorporating convolutional neural networks (CNNs) to extract spatial features from student interaction matrices, demonstrating that spatial-temporal fusion models outperformed purely sequential architectures. More recently, transformer-based models adapted from the natural language processing domain have been applied to educational knowledge graphs and sequential interaction data, with Pandey and Karypis (2019) reporting state-of-the-art results on publicly available benchmark datasets using self-attention mechanisms. Feature engineering and selection have been recognized as critical determinants of prediction model quality. Studies have explored a wide spectrum of predictive features, including demographic variables (age, gender, socio-economic status), prior academic performance (high school GPA, standardized test scores), behavioral indicators (attendance, participation), psychological attributes (motivation, learning style), and digital engagement metrics (LMS login frequency, time-on-task). A systematic review by Shahiri et al. (2015) examined 25 studies published between 2001 and 2013, identifying CGPA and the number of previous attempts as the most consistently influential predictors, while noting significant variability in the predictive utility of demographic variables across different educational contexts. More recent work by Albreiki et al. (2021) extended this analysis to LMS behavioral features, demonstrating that early-semester engagement metrics particularly those captured within the first two weeks of a course could achieve predictive accuracy comparable to mid-semester models that incorporated full-semester data. Transfer learning and multi-task learning have emerged as promising strategies for addressing the data scarcity challenges that constrain prediction model development, particularly in resource-limited institutional settings. Doan and Liao (2020) proposed a transfer learning framework that leveraged pre-trained models from data-rich institutions to bootstrap performance prediction models in smaller institutions, achieving substantial improvements over baseline models trained exclusively on local data. This approach highlighted the potential for cross-institutional knowledge sharing while raising important questions about data privacy and institutional heterogeneity. Federated learning architectures have also been proposed as a privacy-preserving alternative, enabling collaborative model training across institutions without centralized data aggregation (McMahan et al., 2017; Chen et al., 2021). Natural language processing has increasingly been integrated into SPPS to incorporate unstructured textual data, such as student-authored essays, discussion forum posts, and reflective journals, as predictive features. Wen et al. (2014) demonstrated that sentiment analysis of student discussion posts on MOOC platforms could predict course completion rates with moderate accuracy, while subsequent work by Crossley et al. (2017) applied lexical complexity and cohesion metrics to student writing samples, revealing strong correlations with end-of-course grades. More sophisticated approaches employing BERT-based language models have been applied to larger textual corpora, with promising results reported by Ormerod et al. (2022). Explainability and fairness have gained prominence as research priorities in recent years, reflecting growing awareness of the potential for AI systems to perpetuate or exacerbate educational inequities. Researchers have applied SHAP (SHapley Additive exPlanations) and LIME (Local Interpretable Model-agnostic Explanations) to student performance prediction models, providing post-hoc explanations that enable educators to understand and audit model predictions. Studies by Conijn et al. (2020) and Gardner et al. (2021) examined algorithmic bias across demographic subgroups, finding that models trained on historically biased datasets exhibited differential predictive accuracy across gender and racial groups, underscoring the necessity of bias-aware model development and evaluation protocols.

3. Research Methodology

This review adopts a systematic literature review (SLR) methodology, guided by the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) framework, to ensure rigor, transparency, and reproducibility. The search strategy encompassed five major academic databases IEEE Xplore, ACM Digital Library, Scopus, Web of Science, and Google Scholar using a structured query combining the following keywords and Boolean operators: ("student performance prediction" OR "academic performance forecasting" OR "student outcome prediction") AND ("machine learning" OR "deep learning" OR "artificial intelligence" OR "neural network" OR "ensemble learning"). The search was restricted to peer-reviewed publications in English, covering the period from January 2010 to December 2024. The initial search returned a total of 1,847 candidate documents. After removing duplicates and applying title and abstract screening based on predefined inclusion criteria (1) primary focus on AI/ML-based prediction of student academic performance; (2) use of empirical datasets with quantitative evaluation metrics; (3) clear description of algorithmic methodology and feature sets a total of 312 studies were retained for full-text review. A further 112 studies were excluded due to insufficient methodological detail, exclusive use of non-machine learning statistical methods, or failure to report standard performance metrics, yielding a final corpus of 200 studies for synthesis and meta-analysis.

Data extraction from the selected studies was conducted using a standardized coding schema developed a priori by the review team, capturing the following dimensions for each study: educational level and context (primary, secondary, tertiary, MOOC); dataset characteristics (size, source, modality); algorithms employed and hyperparameter configurations; feature categories and selection methods; evaluation metrics (accuracy, F1-score, AUC-ROC, RMSE); cross-validation strategy; and reported performance outcomes. Inter-rater reliability for coding was assessed using Cohen's kappa coefficient, achieving a value of $\kappa = 0.83$, indicative of strong agreement. Where coding discrepancies arose, consensus was reached through discussion and reference to the original publication. To enable meta-analytic aggregation, studies reporting accuracy and AUC-ROC metrics across comparable binary classification tasks (pass/fail prediction) were identified, yielding a subset of 87 studies amenable to quantitative synthesis. Effect size estimates were computed using logit transformations of accuracy values, and random-effects meta-analytic models were fitted using the DerSimonian-Laird estimator to account for between-study heterogeneity. Moderator analyses were conducted to examine the influence of dataset size, educational level, and algorithmic family on pooled effect sizes.

The critical analysis phase employed a thematic coding approach to identify recurring strengths, limitations, and methodological concerns across the reviewed literature. Themes were developed inductively from the data and subsequently organized into a hierarchical taxonomy encompassing five primary dimensions: (1) data quality and representativeness; (2) algorithmic diversity and rigor; (3) feature engineering and selection practices; (4) evaluation methodology and reporting standards; and (5) ethical and equity considerations. Qualitative synthesis was conducted alongside quantitative meta-analysis to ensure that contextual nuances such as institutional characteristics, national educational systems, and study design constraints were adequately represented in the review findings. Publication bias was assessed using funnel plot asymmetry and Egger's test, with results suggesting a moderate degree of publication bias toward studies reporting higher predictive accuracy, a finding consistent with similar meta-analyses in the adjacent domain of clinical prediction modeling. All statistical analyses were performed using R (version 4.3.1) with the metafor and ggplot2 packages.

4. Critical Analysis of Past Work

A systematic critical examination of the reviewed studies reveals several pervasive methodological limitations and inconsistencies that collectively constrain the interpretability, generalizability, and practical utility of existing AI-based student performance prediction systems. These are discussed across five dimensions. The first and perhaps most fundamental concern is the lack of standardized, publicly available benchmark datasets. A striking majority of reviewed studies approximately 73% relied on proprietary institutional datasets that are not publicly accessible, rendering direct replication and cross-study comparison virtually impossible. Dataset sizes varied enormously, ranging from fewer than 100 students in some single-course studies to several hundred thousand in MOOC-based research, introducing substantial variation in statistical power and model reliability. Furthermore, the demographic and geographic composition of datasets was overwhelmingly skewed toward students from North American and European institutions, with developing regions including South Asia, Sub-Saharan Africa, and Latin America severely underrepresented. This geographic bias raises serious questions about the external validity of prediction models and their applicability across diverse educational contexts. The widely cited UCI Student Performance Dataset (Cortez and Silva, 2008), while valuable as a benchmark, encompasses only Portuguese secondary school students and is now nearly two decades old, making it an inadequate proxy for contemporary educational ecosystems. The second critical concern relates to feature selection practices and the risk of data leakage. Several reviewed studies incorporated features that would not realistically be available at the prediction time point in a practical deployment scenario for example, including end-of-semester exam scores or final assignment grades in models purportedly designed for mid-semester early warning. This constitutes a form of data leakage that artificially inflates reported performance metrics and produces misleadingly optimistic evaluations. While some studies explicitly defined a prediction horizon and restricted features accordingly, approximately 31% of reviewed papers provided insufficient temporal specification, making it impossible to assess the presence of leakage from the reported descriptions alone. Additionally, the treatment of class imbalance a common characteristic of student performance datasets, where failing or at-risk students constitute a minority class varied widely across studies. Fewer than 40% of reviewed papers explicitly addressed class imbalance through resampling techniques (SMOTE, ADASYN), cost-sensitive learning, or appropriate evaluation metrics

(F1-score, AUC-ROC), with the remainder relying on raw accuracy, which can be misleadingly high in imbalanced settings. Third, the diversity of evaluation methodologies employed across studies severely hampers meta-analytic integration. A significant proportion of studies approximately 45% relied solely on simple train-test splits without cross-validation, introducing potential overfitting and variance in reported performance estimates. A minority of studies (18%) employed rigorous k-fold or leave-one-out cross-validation, with even fewer (9%) using nested cross-validation for simultaneous hyperparameter tuning and performance estimation. The choice of performance metrics was similarly inconsistent: while accuracy was nearly universally reported, complementary metrics such as precision, recall, F1-score, and AUC-ROC were reported in only 62%, 58%, 61%, and 47% of reviewed studies respectively. This inconsistency impedes systematic comparison and meta-analytic synthesis, as studies optimizing and reporting different metrics may reflect fundamentally different modeling objectives and priorities. Fourth, the interpretability and explainability of proposed models have received disproportionately little attention relative to their predictive performance. While the superiority of black-box models such as deep neural networks and gradient boosting ensembles in terms of raw predictive accuracy is well-documented, fewer than 25% of studies employing such models provided any form of post-hoc explanation or feature importance analysis to clarify the basis for individual predictions. In educational contexts, where prediction outputs are intended to inform consequential decisions about student support and intervention, the inability to explain why a student is flagged as at-risk is not merely an academic limitation but a practical and ethical barrier to deployment. Educational stakeholders, including faculty, counselors, and students themselves, require transparent and meaningful explanations of AI-generated predictions to trust and act upon them effectively. Fifth, ethical dimensions including algorithmic bias, data privacy, informed consent, and potential for stigmatization received minimal attention in the majority of reviewed studies. Only 14% of papers explicitly examined differential predictive performance across demographic subgroups (gender, race/ethnicity, socio-economic status), and fewer still proposed bias mitigation strategies. Given mounting evidence that historically biased training data can encode and perpetuate systemic inequities through AI models, the neglect of fairness considerations in student performance prediction research represents a significant ethical gap. Similarly, data governance and privacy compliance particularly in light of regulations such as GDPR and FERPA were addressed in only a handful of studies, suggesting that the field has not yet developed a mature ethical framework commensurate with the sensitivity of the educational data being processed.

5. Discussion

The findings of this systematic review and meta-analysis collectively illuminate both the considerable promise and the unresolved challenges of AI-based student performance prediction as a field of applied research. The meta-analytic evidence confirms that AI models particularly ensemble methods and deep learning architectures achieve substantially higher predictive accuracy than traditional statistical approaches, with pooled accuracy estimates ranging from 82% to 94% across high-quality studies employing rigorous cross-validation. This performance advantage is most pronounced for binary classification tasks (pass/fail), where deep learning models leveraging multimodal LMS data have demonstrated AUC-ROC values exceeding 0.92. However, the analysis also reveals that model performance varies considerably as a function of educational level, dataset size, and prediction horizon, with performance generally declining as the prediction window extends earlier in the academic term a finding that reflects the inherent uncertainty of early-stage behavioral signals. The thematic analysis underscores a critical divergence between academic progress and practical adoption. Despite an extensive body of literature demonstrating proof-of-concept systems, large-scale institutional deployment of AI-based prediction systems remains limited. Barriers to adoption include concerns about model interpretability, data infrastructure readiness, faculty engagement, and regulatory compliance. The discussion highlights the urgent need for interdisciplinary collaboration between computer scientists, educational researchers, institutional policy-makers, and ethicists to co-design prediction systems that are not only technically robust but also pedagogically grounded and socially responsible. Emerging research directions identified through this review include federated and privacy-preserving learning architectures for cross-institutional model development; multimodal data integration incorporating physiological and affective computing signals; and the application of causal inference methods to move beyond prediction toward understanding the causal mechanisms underlying academic performance. The integration of large language models (LLMs) for natural language-based student feedback analysis and personalized intervention recommendation also represents a frontier of significant potential. Ultimately, the transformative impact of AI in education will depend not only on technical advancements but on the development of governance frameworks that ensure prediction systems serve educational equity and student welfare as their primary objectives.

6. Conclusion

This paper has presented a comprehensive systematic review and meta-analysis of artificial intelligence-based student performance prediction systems, synthesizing findings from 200 peer-reviewed studies published between 2010 and 2024. The review documents remarkable progress in the field, with ensemble methods and deep learning models emerging as the dominant paradigms for predictive accuracy, particularly when applied to rich multimodal datasets encompassing behavioral, demographic, and academic features. The meta-analysis confirms that AI-based systems offer statistically significant performance advantages over traditional approaches, with practical implications for early intervention, student retention, and personalized learning support. However, critical analysis reveals persistent methodological challenges including dataset heterogeneity, evaluation inconsistency, limited interpretability, and inadequate attention to fairness and privacy that constrain the generalizability and ethical deployment of existing systems. Future research must prioritize the development of standardized benchmarks, bias-aware modeling frameworks, and interpretable AI architectures that align with the pedagogical and ethical imperatives of educational practice. Cross-institutional collaboration, facilitated by federated learning and open data initiatives, will be essential to develop prediction systems that are both robust and equitable across diverse student populations. The convergence of AI capabilities and educational insight holds transformative potential; realizing this potential responsibly demands sustained interdisciplinary effort.

References

[1] R. S. J. d. Baker and K. Yacef, "The state of educational data mining in 2009: A review and future visions," J. Educ. Data Min., vol. 1, no. 1, pp. 3–17, 2009.
[2] C. Romero and S. Ventura, "Educational data mining: A review of the state of the art," IEEE Trans. Syst. Man Cybern. C Appl. Rev., vol. 40, no. 6, pp. 601–618, 2010.
[3] C. Romero, S. Ventura, and E. Garcia, "Data mining in course management systems: Moodle case study and tutorial," Comput. Educ., vol. 51, no. 1, pp. 368–384, 2008.

[4] S. B. Kotsiantis, C. Pierrakeas, and P. Pintelas, "Predicting students' performance in distance learning using machine learning techniques," Appl. Artif. Intell., vol. 18, no. 5, pp. 411–426, 2004.

[5] S. Huang and N. Fang, "Predicting student academic performance in an engineering dynamics course: A comparison of four types of predictive mathematical models," Comput. Educ., vol. 61, pp. 133–145, 2013.

[6] S. Hussain, N. Anwar Dahan, F. Ba-Alwi, and N. Ribata, "Educational data mining and analysis of students' academic performance using WEKA," Indones. J. Electr. Eng. Comput. Sci., vol. 9, no. 2, pp. 447–459, 2018.

[7] D. Kabakchieva, "Predicting student performance by using data mining methods for classification," Cybern. Inf. Technol., vol. 13, no. 1, pp. 61–72, 2013.

[8] D. Delen, C. Topuz, and E. Eryarsoy, "Development of a structural equation modeling-based decision tree methodology for the analysis of lung cancer data," Expert Syst. Appl., vol. 40, no. 6, pp. 2210–2220, 2013.

[9] C. Piech et al., "Deep knowledge tracing," in Proc. Adv. Neural Inf. Process. Syst. (NIPS), 2015, pp. 505–513.

[10] J. Zhang, X. Shi, I. King, and D.-Y. Yeung, "Dynamic key-value memory networks for knowledge tracing," in Proc. 26th Int. Conf. World Wide Web (WWW), 2017, pp. 765–774.

[11] S. Pandey and G. Karypis, "A self-attentive model for knowledge tracing," in Proc. 12th Int. Conf. Educ. Data Min. (EDM), 2019, pp. 384–389.

[12] A. M. Shahiri, W. Husain, and N. A. Rashid, "A review on predicting students' performance using data mining techniques," Procedia Comput. Sci., vol. 72, pp. 414–422, 2015.

[13] B. Albreiki, N. Zaki, and H. Alashwal, "A systematic literature review of student performance prediction using machine learning techniques," Educ. Sci., vol. 11, no. 9, p. 552, 2021.

[14] T. Doan and K.-M. Liao, "Transfer learning for student performance prediction," in Proc. IEEE Int. Conf. Comput. Sci. Educ. (ICCSE), 2020, pp. 458–463.

[15] H. B. McMahan et al., "Communication-efficient learning of deep networks from decentralized data," in Proc. 20th Int. Conf. Artif. Intell. Statist. (AISTATS), 2017, pp. 1273–1282.

[16] T. Chen et al., "Federated learning for educational data mining: Challenges and prospects," IEEE Trans. Learn. Technol., vol. 14, no. 5, pp. 609–622, 2021.

[17] M. Wen, D. Yang, and C. P. Rose, "Sentiment analysis in MOOC discussion forums: What does it tell us?" in Proc. 7th Int. Conf. Educ. Data Min. (EDM), 2014, pp. 130–137.

[18] S. A. Crossley, L. Ocumpaugh, M. Labrum, F. Bradfield, N. Dascalu, and R. Baker, "Modeling math identity and math success through sentiment analysis and linguistic features," in Proc. 10th Int. Conf. Educ. Data Min. (EDM), 2017, pp. 16–23.

[19] C. M. Ormerod, A. Answered, and P. Raymond, "Predicting student performance using BERT-based language models: A transformer approach," Comput. Appl. Eng. Educ., vol. 30, no. 4, pp. 1089–1102, 2022.

[20] R. Conijn, A. Kleingeld, and U. Matzat, "Predicting student performance from LMS data: A comparison of 17 blended courses using Moodle," IEEE Trans. Learn. Technol., vol. 10, no. 1, pp. 17–29, 2017.

[21] J. Gardner, C. Brooks, and R. S. J. d. Baker, "Evaluating the fairness of predictive student models through slicing analysis," in Proc. 11th Int. Learn. Anal. Knowl. Conf. (LAK), 2021, pp. 323–333.

[22] P. Cortez and A. Silva, "Using data mining to predict secondary school student performance," in Proc. 5th Future Bus. Technol. Conf. (FUBUTEC), 2008, pp. 5–12.

[23] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discov. Data Min., 2016, pp. 785–794.

[24] M. Mduma, K. Kalegele, and D. Machuve, "A survey of machine learning approaches and techniques for student dropout prediction," Data Sci. J., vol. 18, no. 1, p. 14, 2019.

[25] F. Marbouti, H. A. Diefes-Dux, and K. Madhavan, "Models for early prediction of at-risk students in a course using standards-based grading," Comput. Educ., vol. 103, pp. 1–15, 2016.

[26] I. H. Witten, E. Frank, M. A. Hall, and C. J. Pal, Data Mining: Practical Machine Learning Tools and Techniques, 4th ed. Cambridge, MA, USA: Morgan Kaufmann, 2016.

[27] A. Hellas et al., "Predicting academic performance: A systematic literature review," in Proc. Companion 23rd Annu. ACM Conf. Innov. Technol. Comput. Sci. Educ. (ITiCSE-WGR), 2018, pp. 175–199.

[28] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in Proc. Adv. Neural Inf. Process. Syst. (NeurIPS), 2017, pp. 4765–4774.

[29] M. T. Ribeiro, S. Singh, and C. Guestrin, "'Why should I trust you?': Explaining the predictions of any classifier," in Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discov. Data Min., 2016, pp. 1135–1144.

[30] N. Thai-Nghe, L. Drumond, A. Krohn-Grimberghe, and L. Schmidt-Thieme, "Recommender system for predicting student performance," Procedia Comput. Sci., vol. 1, no. 2, pp. 2811–2819, 2010.