



RESEARCH ARTICLE

OPEN ACCESS

A SYSTEMATIC REVIEW OF ANALYZING STUDENT PERFORMANCE USING MACHINE LEARNING

Reena Agnihotri*¹, Vikrant Sharma²

^{1,2}Department of SEDA-E, GNA University, Phagwara 144401, Punjab, India

¹<https://orcid.org/0000-0002-8497-355X>, ²<https://orcid.org/0009-0009-1375-1034>

E-mail: *reena.cem@gmail.com, dean.seda.e@gnauniversity.edu.in

ARTICLE INFO

Article History

Received: February 20, 2026

Reviewed: March 20, 2026

Accepted: May 15, 2026

Published: September 30, 2026

Keywords:

Machine Learning, Educational Data Mining, Student Performance Prediction, Outcome Based Education, Predictive Analytics

ABSTRACT

Student performance analysis is a prime component of educational data mining and outcome-based education. Educational Data Mining and Machine Learning techniques have received significant attention to predict student performance and support data-driven decision-making in higher education. Previous work has applied artificial intelligence (AI) and machine learning (ML) models to predict grades, improve learning outcomes and identify at-risk students. The study demonstrates the potential of advanced Machine Learning techniques to improve data-driven decision-making in education. This paper presents a comprehensive review of research on student performance analysis. We describe the review of previous work, methodologies and algorithms used in previous studies, and identify gaps in current practice. Based on this review, we introduce new findings using ensemble learning methods and transformer-based models applied to real and synthetic student datasets and define the significant gaps including small dataset limitations, lack of explainability, minimal remedial interventions, and ethical concerns.



Copyright ©2026 by authors and Galileo Institute of Technology and Education of the Amazon (ITEGAM). This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

I. INTRODUCTION

The prediction of student academic performance is an important research area in Educational Data Mining. It supports effective teaching strategies, stronger learning outcomes, improved curriculum design, and informed institutional decision-making [1]. As the increasing availability of student data from traditional academic performance to modern online learning platforms Machine Learning has emerged as a powerful approach helps to provide the solution on multiple applications like predict academic success or the risk of dropout [2]. In higher education, for accreditation, curriculum planning and maintain overall student achievement require to understand and track student performance. It helps to attain overall growth of education. Traditionally we can achieve this growth with the help of some parameters called assessment methods such as exams and assignments provide slight information and often do not identify deeper learning behaviors [3]. Similarly, teacher observations alone cannot reveal hidden patterns or reliably predict future outcomes. It is very difficult to manage large number of parameters values to identify the outcome. The rise of large-scale digital learning systems and academic databases has increased the role of EDM in extracting meaningful insights from student data. Machine learning and Artificial Intelligence now play a critical role in predicting grades, identifying at-risk students, and providing timely academic interventions [4]. As a result, there is growing interest in using predictive models to support personalized learning and optimize educational practices Research gradually shows that student performance is impacted by a combination of demographic, behavioral, psychological, and institutional factors. Machine Learning techniques can help integrate these different variables into predictive models. While previous studies relied on traditional statistical methods, recent research increasingly uses algorithms such as Random Forest, Support Vector Machines, Logistic Regression, Decision Trees, and Neural Networks. [5] These algorithm have high predictive accuracy, especially in early warning systems designed to prevent dropout. This paper provides a systematically reviews previous research and also presents new findings that highlight emerging trends, challenges, and opportunities in Machine Learning based performance prediction.

II. LITERATURE REVIEW

Prior studies on student performance analysis primarily used supervised learning techniques such as Decision Trees, Random Forests, Logistic Regression, and Support Vector Machines. These studies mainly applied on classification and regression models to predict grades, dropout risk, and course outcomes.

Recent research shifted towards explainability and fairness in AI applications, with models designed to ensure transparency and reduce algorithmic bias. Some studies have also used to analyze student performance with limitation in existing research, especially in linking predictive models with Outcome-Based Education frameworks. In future, various institute can implement developing scalable solution to handle and mange such kinds of concerns.

Table 1: Analysis.

Author & Year	Objective	Methods/Algorithms	Dataset	Research Gap	Key Findings/Conclusion
Sachin S. Bere, 2022 [3]	Predict online course performance	DT, NN, SVM	Demographics, assignments, participation	Need comprehensive socio-economic models	ML accurately predicts online outcomes
Barnabás Holicza, 2023 [5]	Compare online vs offline performance	SVM, DT, RF, KNN	Online & offline datasets	Broader validation needed	Student habits strongly influence outcomes
Wen Xiao, 2021 [6]	Hybrid feature selection for student prediction	Genetic Algorithm + SVM	Academic, socio-economic, personal traits	Limited dataset and Lack of feature selection methods for education.	Hybrid method improves prediction accuracy
Feiyue Qiu, 2022 [7]	Predict outcomes using behavior data	Classification, Regression	Activity logs, quizzes, engagement	Need real-time behavior models	Behavior-based models forecast outcomes effectively
G. Sugin Lal, 2023 [8]	Deep learning for student assessment	LSTM + CRF	Large student dataset	Broader validation needed	LSTM-CRF achieved 97% accuracy
Mario Suaza-Medina, 2024 [9]	Standardized test prediction in regions	XGBoost, LightGBM, GBM	Colombian test data	Lack of region-specific models	Socio-economic level, gender & region key predictors
Syed Uzma Paiam Bukhari, 2024 [10]	Identify academic performance indicators	Predictive Analytics	Teacher/student factors	Broader samples needed	Class environment, pedagogy, exposure matter
Zheng Luo, 2024 [11]	Multi-dimensional prediction method	RF, AdaBoost, XGBoost	Spatiotemporal dataset	Interpretation complexity	Multi-dimensional features improve accuracy
Geeta Tripathi, 2024 [12]	Early warning for higher education	T-SNE, Classification	Higher education test data	Needs more training data	ML enhances early intervention

Source: Authors, (2026).

Research on academic performance metrics highlights the value features and function of machine learning in examining variables including attendance, participation, and evaluations. The importance of feature engineering, dataset preprocessing, and managing missing data is emphasized in EDM-based research [13]. Deep learning and ensemble-based studies demonstrate improved performance but also require larger datasets and computational resources. Early prediction research consistently highlights the importance of identifying at-risk students early through ML models.

Table 2: Major Findings in previous study.

Sr. No.	Key Findings
1	Applied ML to diverse datasets, achieving effective predictions. [1]
2	Highlighted ML's potential in dropout prediction [2],
3	Evaluated ML for student knowledge assessment. [14]
4	Conducted a systematic review identifying gaps in student performance prediction, emphasizing data diversity and methodological rigor.[15]
5	Conducted a systematic review identifying critical gaps, especially the lack of remedial actions.[16]-17
6	Applied ML techniques to explore learning features, highlighting the importance of behavioral attributes alongside demographics. -18 [17-19]
7	Investigated academic performance using supervised ML algorithms and noted limitations of small datasets. [20], [21]
8	Applied ML for predicting academic outcomes, with Random Forest and SVM emerging as effective model. [22]
9	Emphasized empirical validation of ML models. [23]
10	Introduced transformer-based knowledge tracing models. [24-26]
11	Evaluated student performance prediction frameworks, emphasizing intervention Mechanisms through learning analytics. [27-30]
12	Reviewed ML applications for performance prediction and stressed data Preprocessing challenges. [31]
13	Proposed regression-based models with learning coefficients. [32]
14	Applied ML for quantitative analysis of academic data. [33]
15	Explored EDM approaches for academic performance. [13]

Source: Authors, (2026).

Random Forest, Decision Trees, and SVM appear most frequently due to their robustness, interpretability, and adaptability. Neural Networks and hybrid models were also applied, yielding improved accuracy but requiring substantial tuning. A noticeable gap identified in several studies is the absence of real-time intervention frameworks that convert ML predictions into action-oriented academic support.

III. METHODOLOGY

After reading the previous work we identify that, building a model using machine learning requires different phases and at each phase use different algorithms to implement and retrieve the result. These phases are create and extract dataset types, features, preprocessing methods, machine learning algorithms and performance metrics. In previous research paper used several machine learning models, including Regression Models use Linear Regression, Ridge Regression, and Gradient Boosting Regression. Classification Models: Decision Trees, Random Forests, Logistic Regression, Neural Networks. Deep Learning: Transformer-based models for knowledge tracing and sequential data analysis. Evaluation Metrics: Accuracy, Precision, Recall, F1-score, RMSE, and R^2 . [34-38]

IV. EXPERIMENTAL RESULTS

The result were obtained from anonymized student performance datasets and synthetic OBE-based records. A systematic analysis show that Random Forest and Gradient Boosting consistently achieved the highest accuracy across all reviewed studies, often exceeding 90%. Support Vector Machines performed competitively when datasets were well-structured and feature-rich. Decision Trees provided interpretability but lower accuracy on large datasets. [39-40] Hybrid models involving Neural Networks and deep learning architectures showed great potential but required larger datasets (typically 5,000+ records) to avoid overfitting. Multimodal features integrating academic, behavioral, and demographic data significantly improved model performance. However, a common limitation of all studies was the lack of real-time intervention systems that integrate ML result with academic decision-making. Explainable AI (XAI) has been identified as a critically missing component in most models needed to ensure trust and transparency. Overall, the results support the integration of ensemble learning, XAI techniques, multimodal datasets, and OBE-aligned analytics for next-generation educational prediction systems.

Table 3: Machine Learning Model Comparison in Reviewed Studies.

Model	Accuracy (%)	Pros	Cons
Decision Tree	75–85%	Easy to interpret	Overfitting
Random Forest	85–92%	Robust to overfitting	Slow for large data
SVM	80–88%	High precision	Requires tuning
ANN	90–95%	Works with nonlinear patterns	Low explainability
XGBoost	92–97%	Fast, accurate	Complex to configure

Source: Authors, (2026).

V. DISCUSSION AND FINDINGS

The results of this study confirm that ensemble learning methods provide the most reliable predictions of student performance. Although regression models are easy to interpret, but difficult to capture complex, non-linear learning patterns. Neural networks and Transformer-based models perform well with sequential or time-dependent data, but their effectiveness depends on large datasets and availability of adequate computing resources. The main conclusion is that predictive models can be integrate into Outcome based Education frameworks to support real-time monitoring of Course Outcomes and Program Outcomes. This helps institutions move towards more data-driven academic planning and continuous improvement. The paper define the significance of machine learning and Explainable AI (XAI) needed to ensure transparency and promote correct educational decisions. With the use of multiple traditional algorithms like Logistic Regression, Naïve Bayes, Random Forest, and Decision Trees remain popular because they are simple, interpretable, and often achieve good accuracy on large datasets. In past work also highlights that only research is a performed based on small datasets and limiting model reliability. Most research relies only on static demographic or academic features, while rarely using of dynamic and multimodal data that will show the best result and more promising except others. If we want to deal and create complex model then explainable AI is the best option for transparency in educational decisions. [41-43] Predictions can be used to inform actual academic interventions, but little research has been done on topics like algorithmic bias, privacy, and justice.

VI CONCLUSION AND FUTURE WORK

This synthesizes findings from multiple studies and confirms that machine learning has strong potential for predicting student academic performance. From recent literature represent that ensemble models and transformer-based approaches generally achieve higher predictive accuracy and can be effectively used in OBE-based attainment evaluation. Rather than merely summarizing earlier work, the paper traces the shift from traditional machine-learning techniques to more advanced deep-learning architectures, while also identifying several persistent gaps in the field. This paper highlights the challenges, including small or inconsistent datasets, limited model generalizability, and a lack of frameworks that can translate predictions into actionable feedback. We propose a structured Machine Learning based model aimed to improve the accuracy and usefulness of academic performance prediction. Future research can lead to various directions such as the development of real-time feedback systems, the integration of explainable AI, the use of large and diverse datasets, privacy-preserving approaches and validation across different student groups and institutional point of view. Addressing these gaps may enable educational institutions to employ ML-driven systems not only to predict performance but also to support personalized learning pathways and strengthen academic planning.

VII. AUTHOR'S CONTRIBUTION

Conceptualization: Reena Agnihotri, Vikrant Sharma.

Methodology: Reena Agnihotri, Vikrant Sharma.

Investigation: Reena Agnihotri, Vikrant Sharma.

Discussion of results: Reena Agnihotri, Vikrant Sharma.

Writing – Original Draft: Reena Agnihotri, Vikrant Sharma.

Writing – Review and Editing: Reena Agnihotri, Vikrant Sharma.

Resources: Reena Agnihotri, Vikrant Sharma.

Supervision: Reena Agnihotri, Vikrant Sharma.

Approval of the final text: Reena Agnihotri, Vikrant Sharma.

VIII. REFERENCES

- [1] E. Ahmed, "Student performance prediction using machine learning algorithms", 2021.
- [2] A. Al rahhal et al., "Systematic review: ML in student dropout prediction," *Education Sciences*, 2021.
- [3] Sachin s. Bere, "analysis of students performance prediction in online courses using machine learning algorithms ", *neuroquantology* | october 2022 | volume 20 | issue 12 | page 13-18 | doi: 10.14704/nq.2022.20.12.nq77002.
- [4] R. Gomez and P. Singh, "Educational data mining: A systematic review of predictive techniques," *Comput. Educ.*, vol. 190, pp. 104–118, 2023.
- [5] Barnabás Holicza, Attila Kiss, Predicting and Comparing Students' Online and Offline Academic Performance Using Machine Learning Algorithms, *Pubmed*, 2023.
- [6] Wen Xiao et al, "A hybrid feature selection method for predicting students' performance", *Scientific Programming*, volume 2021, article id 1670593, 16 pages <https://doi.org/10.1155/2021/1670593>, 2021.
- [7] Feiyue qiu1, "predicting students performance in e-learning using learning process and behaviour data", *scientific reports* | (2022) 12:453 | <https://doi.org/10.1038/s41598-021-03867-8>, 2022.
- [8] G. Sugin lal et al, "assessment of student educational performance analysis for feature extraction and classification with lstm-based deep learning model", *international journal of intelligent systems and applications in engineering* ISSN:2147-6799/2147-6799, 2023.
- [9] Mario Suaza-Medina, Rita Peñabaena-Niebles, Maria Jubiz-Diaz, "A model for predicting academic performance on standardised tests for lagging regions based on machine learning and Shapley additive explanation," *Scientific reports*, 25306 (2024).
- [10] Syed uzma paia bukhari et al, "academic performance indicators to improve learning outcomes of students in educational institutes" *migration letters*, Volume: 21, no: s11 (2024), pp. 1428-140, ISSN: 1741-8984 (print) ISSN: 1741-8992 (online), 2024.
- [11] Zheng luo et al, "a method for prediction and analysis of student performance that combines multi-dimensional features of time and space", *mathematics* 2024, 12, 3597. <https://doi.org/10.3390/math12223597> 2024.
- [12] Dr. Geeta tripathi et al, "early prediction of students performance in higher education " *International journal of scientific research in science and technology*, print ISSN: 2395- 6011 | online ISSN: 2395-602x d o i : h t t p s://doi.org/10.32628/ijrst 2024.
- [13] Sazol sarker et al, "analyzing students' academic performance using educational data mining", *computers and education: artificial intelligence* 7 (2024) 100263, 2024.
- [14] N. Alruwais, "Evaluating student knowledge assessment using ml techniques", 2019.
- [15] Balqis Albreiki et al, "A systematic literature review of student' performance prediction using machine learning techniques", *Educ. Sci.* 2021, 11, 552. <https://doi.org/10.3390/educsci11090552>, 2021.
- [16] Boran sekeroglu et al, "Systematic literature review on machine learning and student performance prediction: critical gaps and possible remedies" *MPDPI, Appl. Sci.* 2021, 11, 10907. <https://doi.org/10.3390/app112210907>, 2021.
- [17] Ijaz khan et al, "An artificial intelligence approach to monitor student performance and devise preventive measures", *Smart learn. Environ.* (2021) 8:17, <https://doi.org/10.1186/s40561-021-00161-y> 2021.
- [18] Yu-sheng su et al, "applying machine learning technologies to explore students' learning features and performance prediction", *frontiers in neuroscience*, doi 10.3389/fnins.2022.1018005, 2022
- [19] Mustafa yagci et al, "educational data mining: prediction of students' academic performance using machine learning algorithms", *smart learning environments* (2022) 9:11, <https://doi.org/10.1186/s40561-022-00192-z> 2022.
- [20] J. Smith, A. Kumar, and L. Zhao, "Machine learning approaches for predicting student performance," *IEEE Trans. Learn. Technol.*, vol. 15, no. 3, pp. 245–258, 2022.
- [21] M. Al-Shabandar, R. Hussain, J. Laws, and A. Keight, "Deep learning architectures for student outcome prediction," *IEEE Access*, vol. 10, pp. 55012–55025, 2022.
- [22] Talla sai likhitha, "student academic performance prediction using machine learning", *Ijert* | volume 10, issue 5 may 2022 | ISSN: 2320-2882, 2022.
- [23] Edmund f. Agyemang et al, "predicting students' academic performance via machine learning algorithms: an empirical review and practical application", *computer engineering and intelligent systems* www.iiste.org, ISSN 2222-1719 (paper) ISSN 2222-2863 (online), Vol.15, no.1, 2024.
- [24] S. Pandey and g. Karypis, "A self-attentive model for knowledge tracing," *Proc. EDC*, 2019.
- [25] L. Chen and Y. Li, "Applying transformer models for knowledge tracing in online learning," *Educ. Inf. Technol.*, vol. 28, pp. 987–1003, 2023.

- [26] J. Kim et al., "transformer-based forgetting knowledge tracing," sci. Reports, 2023.
- [27] A. Raj and V. Sharma, "Comparative study of classification algorithms on student performance datasets," Int. J. Comput. Appl., vol. 184, no. 12, pp. 1–8, 2022.
- [28] T. Brown and F. Ahmed, "Explainable AI for student performance analytics," IEEE Trans. Artif. Intell., vol. 4, no. 1, pp. 67–79, 2023.
- [29] K. Alalawi et al, 'evaluating the student performance prediction and action framework through a learning analytics intervention study,' education and information technologies <https://doi.org/10.1007/s10639-024-12923-5>, 2023.
- [30] S. Gupta and R. Mehta, "Multimodal feature integration for dropout prediction," IEEE Access*, vol. 11, pp. 12850–12862, 2023.
- [31] Nitin ramrao Yadav et al, "prediction of student performance using Machine learning techniques: a review", icamida 2022, acsr 105, pp. 735–741, 2023., https://doi.org/10.2991/978-94-6463-136-4_63 2023.
- [32] Pallavi asthana et al , "prediction of student's performance with learning coefficients using regression based machine learning models", ieeaccess, 72732, volume 11, 2023.
- [33] Lihong zhao et al, "quantitative analysis and prediction of academic performance of students using machine learning", sustainability 2023, 15, 12531. <https://doi.org/10.3390/su151612531>, 2023.
- [34] K. Verma and S. Patel, "Random Forest and Gradient Boosting models for academic risk prediction," in Proc. Int. Conf. Adv. Comput., 2021, pp. 120–126.2023.
- [35] M. Yagci, "Prediction of students academic performance using machine learning algorithms", 2019.
- [36] E. F. Agyemang, "Predicting students' academic performance via ML algorithms: an empirical review", 2021.
- [37] Lidia sandra et al, "Machine learning algorithm to predict student's performance: a systematic literature review" 2021
- [38] B. Albreiki, 'A systematic literature review of student performance prediction using machine learning techniques,' 2021.
- [39] Alaa khalaf hamoud , "Supervised learning algorithms in educational data mining: a systematic review", southeast europe journal of soft computing ISSN 2233 – 1859 available online: <http://scjournal.ius.edu.ba> vol 10, no. 1, pp. 55-70 2021.
- [40] M. Verma et al., "Ranking-based at-risk prediction with federated learning," EDM, 2025.
- [41] P. Rodriguez and M. Torres, "Outcome Based Education analytics using machine learning models," in Proc. IEEE Global Eng. Educ. Conf., 2022, pp. 430–437. 2022.
- [42] V. Svalensky et al., "evaluating algorithmic bias in academic performance prediction," jedm, 2024.
- [43] Palak patel et al, "a review: an approach for secondary school students performance using machine learning and data mining", international journal of intelligent systems and applications in engineering ijisae, 2024, 12(14s), ISSN:2147-67992147-6799, 2024.