IBIMA Publishing
Communications of International Proceedings
https://ibimapublishing.com/p-articles/47ISM/2026/4726126/
Vol. 2026 (7), Article ID 4726126

AI-Driven Student Success: A Systematic Review of Retention and Academic Performance Prediction in Higher Education*

Rubiah MOHD YUNUS

Multimedia University, Selangor, Malaysia

Correspondence should be addressed to: Rubiah MOHD YUNUS, rubiah.yunus@mmu.edu.my

*Presented at the 47th IBIMA International Conference, 29-30 June 2026, Madrid, Spain

Abstract

Artificial Intelligence (AI) has emerged as a promising approach for addressing challenges related to student retention and academic performance in higher education. Despite the growing adoption of AI-based predictive systems, existing studies are fragmented across different methodologies, datasets, and educational contexts, making it difficult to obtain a comprehensive understanding of current developments and future research directions. This study aims to systematically review the application of AI techniques for student retention and performance prediction in universities.

A systematic literature review was conducted on studies published between 2020 and 2026. The review examined AI methodologies, predictive models, evaluation metrics, key predictive factors, application areas, implementation challenges, and emerging research opportunities. Relevant studies were identified, screened, and analysed to provide a comprehensive overview of the current state of research.

The findings indicate that AI-based predictive models consistently outperform traditional statistical approaches in identifying students at risk of academic failure or dropout. Machine learning and deep learning techniques, including Random Forest, Decision Trees, Support Vector Machines, Artificial Neural Networks, and Long Short-Term Memory models, demonstrated strong predictive performance across diverse educational datasets. The review further reveals that academic achievement, learning management system engagement, behavioural patterns, and socioeconomic factors are among the most influential predictors of student outcomes. However, challenges related to data privacy, algorithmic bias, model interpretability, and cross-institutional generalizability remain significant concerns.

The study concludes that future research should prioritize explainable and ethical AI frameworks while promoting human-centred implementation strategies to enhance decision-making and student success in higher education.

Keywords: Artificial Intelligence, Student Retention, Student Performance Prediction, Higher Education, Machine Learning, Learning Analytics.

I. Introduction

Student retention and academic performance are among the most significant concerns in higher education institutions worldwide. Universities continuously face challenges associated with student dropout, poor academic achievement, and low graduation rates. Student attrition negatively affects institutional reputation, financial sustainability, and workforce development goals [1]. Consequently, universities seek effective strategies to identify at-risk students and improve educational outcomes.

Cite this Article as: Rubiah MOHD YUNUS Vol. 2026 (7) " AI-Driven Student Success: A Systematic Review of Retention and Academic Performance Prediction in Higher Education " Communications of International Proceedings, Vol. 2026 (7), Article ID 4726126, https://doi.org/10.5171/2026.4726126

Traditional approaches to academic monitoring often rely on manual evaluation, historical records, and retrospective analysis. These methods are limited in scalability and predictive capability. The rapid digital transformation of higher education, combined with the increasing adoption of Learning Management Systems (LMS), online learning platforms, and student information systems, has generated vast amounts of educational data that can be analyzed using AI technologies [2].

Artificial Intelligence refers to computational systems capable of performing tasks requiring human intelligence, including reasoning, learning, prediction, and decision-making. In higher education, AI technologies such as Machine Learning, Deep Learning, Educational Data Mining, and Learning Analytics are increasingly used to predict student performance and retention outcomes [3]. These systems enable universities to analyze academic, behavioral, and demographic data to identify students at risk of failure or dropout.

Machine Learning algorithms such as Decision Trees, Random Forest, Support Vector Machines, and Neural Networks can identify hidden patterns within educational datasets and support predictive analytics [4]. AI-driven systems further enable personalized learning, adaptive educational environments, intelligent tutoring systems, and AI-assisted academic advising.

The adoption of AI in universities is driven by several factors:

1. Availability of educational big data.
2. Growth of online and blended learning.
3. Demand for evidence-based educational decision-making.
4. Need for personalized learning experiences.
5. Institutional goals to improve retention and graduation rates.

Despite the advantages of AI-driven predictive systems, major concerns remain unresolved. Challenges include privacy issues, algorithmic bias, lack of explainability, and limited transferability of predictive models across institutions [5]. Ethical considerations regarding student data usage and automated decision-making further influence AI implementation in higher education.

This paper aims to provide a systematic review of AI applications for student retention and performance prediction in universities. The objectives of this study are:

1. To identify AI techniques commonly used in educational prediction systems.
2. To examine factors influencing retention and academic performance.
3. To analyse AI applications in higher education.
4. To discuss challenges and ethical concerns.
5. To identify future research opportunities.

II. Research Methodology

Systematic Review Approach

This study adopted a systematic literature review (SLR) methodology based on the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) framework. PRISMA provides a structured and transparent process for identifying, screening, assessing, and selecting relevant studies for inclusion in a review[6]. The methodology ensures rigor, reproducibility, and minimizes selection bias during the literature review process.

The review process consisted of four stages:

1. **Identification** – Relevant studies were retrieved from selected academic databases using predefined search keywords related to artificial intelligence, student retention, academic performance prediction, machine learning, and higher education.

2. **Screening** – Duplicate records and studies that did not meet the scope of the review were removed based on title and abstract screening.

3. **Eligibility Assessment** – The full texts of the remaining studies were assessed against the established inclusion and exclusion criteria to determine their relevance to the research objectives.

4. **Final Inclusion** – Studies that satisfied all eligibility criteria were included for detailed analysis and synthesis.

5. The study selection process is illustrated in **Figure 1**, which presents the PRISMA flow diagram showing the number of records identified, screened, excluded, assessed for eligibility, and finally included in the review.

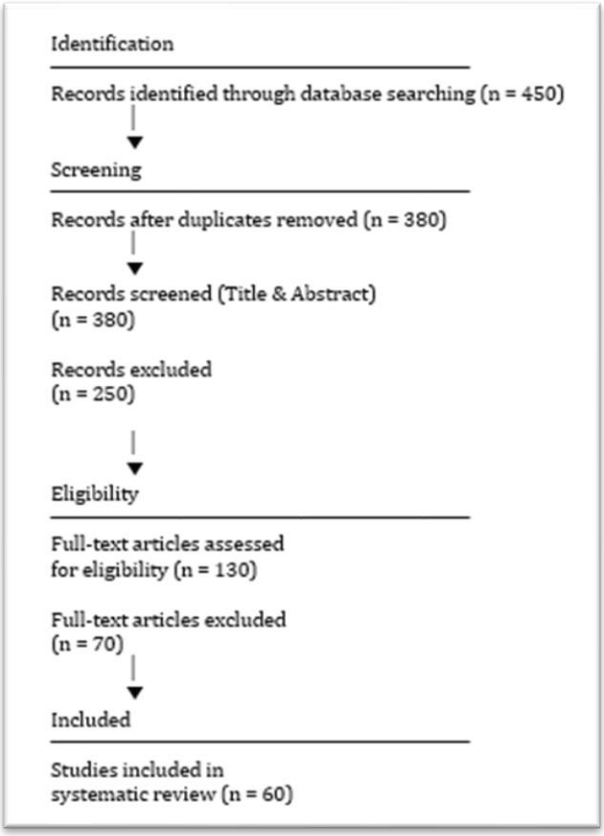

Figure 1: PRISMA Flow Diagram of the Study Selection Process.

Database Selection

Relevant studies for this systematic review were collected from several academic databases, including IEEE Xplore, ScienceDirect, SpringerLink, Scopus, ACM Digital Library, and Google Scholar. These databases were selected because they provide access to high-quality peer-reviewed publications related to Artificial Intelligence (AI), educational technology, learning analytics, and predictive modelling.

Inclusion Criteria

Studies were included in this systematic review if they focused on higher education institutions and applied Artificial Intelligence (AI), Machine Learning (ML), Deep Learning (DL), Educational Data Mining (EDM), or Learning Analytics techniques. The selected studies also investigated student retention or academic performance prediction and were published between 2020 and 2026. In addition, only peer-reviewed journal articles and conference papers written in English were considered for inclusion.

Exclusion Criteria

Studies were excluded if they focused on primary or secondary education, did not involve predictive AI techniques, were editorials, reports, or theses, lacked empirical evidence, were duplicate studies.

Data Extraction and Analysis

Selected studies were analysed based on their research objectives, AI techniques used, datasets and variables, evaluation metrics, main findings, as well as the challenges and limitations identified in each study. A thematic analysis approach was applied to identify recurring themes and emerging research trends.

III. Artificial Intelligence in Higher Education

Artificial Intelligence has transformed various sectors, including healthcare, finance, transportation, and education. In higher education, AI technologies are increasingly applied to support teaching, learning, academic advising, and institutional decision-making [7].

Educational institutions generate substantial data through:

- Learning Management Systems (LMS)
- Student Information Systems (SIS)
- Online learning platforms
- MOOCs
- Digital assessment systems

These educational datasets create opportunities for predictive analytics and intelligent educational systems.

A. Educational Data Mining (EDM)

Educational Data Mining (EDM) focuses on extracting meaningful information, patterns, and insights from large-scale educational datasets such as student records, learning management system logs, assessment results, and attendance data. By applying data mining and machine learning techniques, EDM helps educators and institutions better understand how students learn, perform, and interact within academic environments.

EDM techniques support several key educational applications. These include student performance prediction, where algorithms are used to forecast academic outcomes such as grades or exam results based on historical and behavioural data. It also includes dropout analysis, which identifies students who are at risk of withdrawing from their studies by detecting early warning signs such as poor attendance, low engagement, or declining performance.

In addition, EDM is widely used for learning behaviour analysis, which examines how students interact with digital learning platforms, including study patterns, time spent on tasks, and participation in online activities. This helps educators understand learning habits and improve instructional design. EDM also supports adaptive learning systems, where learning content is dynamically adjusted according to individual student needs, ensuring a more personalized learning experience. Furthermore, recommendation systems are developed to suggest relevant courses, learning materials, or activities based on students' interests, performance levels, and learning history [8].

Overall, EDM enables universities and educational institutions to identify academic risks at an early stage, allowing timely intervention to support struggling students. It also enhances decision-making processes, improves teaching strategies, and contributes to more effective and data-driven educational interventions.

B. Learning Analytics (LA)

Learning Analytics refers to the systematic collection, measurement, and analysis of data related to students' learning activities in order to understand and optimize educational outcomes [9]. It involves processing data from various sources such as learning management systems, online quizzes, discussion forums, and digital interactions to gain insights into student learning patterns, engagement levels, and academic progress.

Learning Analytics supports several important educational functions. One key application is monitoring student engagement, where data is used to track participation in online learning platforms, attendance in virtual classes, and interaction with learning materials. This helps educators identify how actively students are involved in the learning process. It also plays a crucial role in identifying at-risk students by detecting early warning indicators such as low login frequency, poor assignment submission rates, or declining performance, enabling timely academic support.

In addition, Learning Analytics enables personalized learning by tailoring educational content, pace, and learning pathways according to individual student needs and performance levels. This ensures a more student-centered learning experience that can improve comprehension and achievement. It is also used for curriculum improvement, where aggregated learning data helps institutions evaluate the effectiveness of teaching materials, course structure, and instructional methods, leading to continuous enhancement of academic programs.

Furthermore, Learning Analytics supports academic advising by providing advisors with data-driven insights into student progress, strengths, and weaknesses. This allows for more informed guidance on course selection, academic planning, and career pathways.

With the integration of AI-driven learning analytics dashboards, educators can now monitor student progression in real time. These dashboards provide visual representations of learning data, enabling quick interpretation of student performance trends and facilitating timely interventions to improve academic success and overall learning outcomes.

C. AI-Based Educational Decision Making

AI systems enable universities to transition from reactive management to proactive intervention systems. Predictive models can identify students requiring academic support before failure occurs [10].

IV. AI Techniques for Student Retention and Performance Prediction

A. Machine Learning Techniques

Machine Learning is the dominant AI approach in student prediction systems.

1) Decision Trees

Decision Trees are used to classify students based on various academic and behavioural variables such as GPA, attendance, assignment completion, and class participation. By splitting data into branches according to these attributes, decision trees help identify patterns that distinguish different student performance categories.

One of the main advantages of decision trees is their ease of interpretation, as the resulting model is simple to understand and can be visualized clearly. They also provide transparent decision rules, making it easy for educators and researchers to understand how classifications are made. In addition, decision trees are fast to implement and require relatively low computational resources compared to more complex machine learning models.

However, decision trees also have some limitations. They are susceptible to overfitting, especially when the tree becomes too complex and closely fits the training data, which can reduce generalization to new data. They may also be less effective when dealing with highly complex or large-scale datasets, as their performance can degrade compared to more advanced ensemble or deep learning methods.

2) Random Forest

Random Forest combines multiple decision trees to improve predictive performance.

Advantages:

- High predictive accuracy
- Reduced overfitting
- Effective for large datasets

Several studies identify Random Forest as one of the best-performing algorithms for student retention prediction [4].

3) Support Vector Machine (SVM)

Support Vector Machines (SVM) classify students into different risk categories by identifying the optimal hyperplanes that best separate data points belonging to different classes. This method works by finding the most appropriate boundary that maximizes the margin between groups such as high-performing and at-risk students.

One of the main advantages of SVM is its effectiveness in handling high-dimensional datasets, where the number of features is large and complex relationships exist among variables. It also has strong classification capability, making it highly accurate in distinguishing between different student performance levels.

However, SVM also has some limitations. It is computationally intensive, especially when applied to large datasets, as it requires significant processing power and time. In addition, the model is often difficult to interpret, making it less suitable when explainability and transparency are important in educational decision-making.

4) Logistic Regression

Logistic Regression is widely used in educational data analysis due to its simplicity and strong interpretability. It is commonly applied to predict outcomes such as whether a student is at risk or not based on input variables.

One of its main advantages is that it is easy to implement and does not require complex computational resources. It also provides transparent output in the form of probabilities, allowing educators to clearly understand the likelihood of different outcomes. In addition, Logistic Regression is particularly suitable for binary classification problems, such as classifying students into pass/fail or at-risk/not at-risk categories.

However, Logistic Regression has limitations. It has a limited ability to model nonlinear relationships between variables, which means it may not perform well when student data contains complex or highly interactive patterns that require more advanced modeling techniques.

5) K-Nearest Neighbor (KNN)

K-Nearest Neighbors (KNN) predicts student outcomes by comparing a student's data with historical student records that are most similar to them. It classifies or predicts results based on the majority outcome of the nearest data points.

One of the main advantages of KNN is its simple implementation, as it does not require complex training processes or model assumptions. It is also effective for smaller datasets, where the computation of distances between data points remains manageable and efficient.

However, KNN has limitations. It can become computationally expensive when applied to large datasets because it must calculate the distance between the new instance and all existing data points, which increases processing time and reduces efficiency.

B. Deep Learning Techniques

Deep Learning models are increasingly used because of their ability to process complex educational data.

1) Artificial Neural Networks (ANN)

Artificial Neural Networks (ANN) simulate how the human brain processes information using interconnected nodes called neurons. They are used to learn patterns from data and make predictions.

One of the main advantages of ANN is its high predictive performance, as it can produce very accurate results when trained properly. It is also able to model nonlinear relationships, making it suitable for complex educational data.

However, ANN also has limitations. It requires large datasets to train effectively, which may not always be available in educational settings. In addition, its decision-making process has limited explainability, making it difficult to interpret how results are generated.

2) Long Short-Term Memory (LSTM)

Long Short-Term Memory (LSTM) networks are a type of deep learning model that analyze sequential educational data, such as weekly learning behaviors and interactions on Learning Management Systems (LMS). They are designed to understand patterns over time by remembering important information from previous data points.

LSTM networks are commonly used in online learning monitoring to track how students progress over time. They are also applied in student engagement analysis to observe changes in participation and activity levels across different learning periods. In addition, they support temporal prediction, where future student performance or behavior is predicted based on past learning trends.

3) Convolutional Neural Networks (CNN)

Convolutional Neural Network (CNN) models are increasingly applied to multimodal educational datasets that combine different types of data such as text, images, audio, and video from learning environments. In educational settings, CNNs are particularly useful for analyzing visual data like handwritten assignments, facial expressions during online learning, and recorded classroom interactions. By extracting spatial features from these inputs, CNNs help identify patterns related to student engagement, attention, and learning behavior. When integrated with other data sources such as LMS logs and assessment records, CNN-based multimodal systems can provide a more comprehensive understanding of student performance and learning experiences. This approach has been shown to improve prediction accuracy and support more advanced learning analytics applications in higher education contexts [10].

C. Explainable Artificial Intelligence (XAI)

Explainable Artificial Intelligence (XAI) improves the transparency and understanding of AI prediction systems [11]. It helps users see how and why a model makes certain decisions.

In universities, interpretable models are important because educators and administrators need to understand why students are classified as at-risk, which factors influence the predictions, and how they can design suitable interventions to help students improve.

Common XAI methods include SHAP, LIME, and feature importance analysis. These methods show which features are most important in making predictions. Overall, XAI is important because it makes AI systems more ethical, transparent, and easier to trust in education.

V. Factors Influencing Student Retention and Academic Performance

A. Academic Factors

Academic variables are consistently identified as the strongest predictors of student success. These variables provide important information about a student's learning progress and overall academic performance.

Common indicators include GPA, midterm scores, final examination results, assignment completion rates, attendance records, and course participation levels. Together, these factors help educators assess how well a student is performing throughout a course or program.

Poor academic performance is strongly associated with an increased risk of student dropout, as students with consistently low grades or low engagement are more likely to disengage from their studies [3].

B. Behavioral Factors

Behavioral analytics derived from Learning Management System (LMS) platforms significantly improve the accuracy of student performance predictions. These analytics focus on how students interact with online learning systems and provide valuable insights into their learning behaviors.

Common examples of behavioral indicators include login frequency, time spent on learning materials, discussion forum participation, number of quiz attempts, and video viewing duration. These data points help educators understand student engagement levels and learning habits in detail.

Overall, behavioral data supports real-time monitoring of student activity and enables early intervention systems. This allows educators to identify at-risk students early and provide timely academic support to improve learning outcomes [9].

C. Socioeconomic Factors

Socioeconomic conditions play an important role in influencing students' educational persistence and overall academic success. These factors affect a student's ability to continue their studies and perform well in academic environments.

Examples of socioeconomic factors include financial status, access to the internet, employment commitments, and the level of family support. Students with limited financial resources may struggle to afford educational materials or tuition fees, while lack of stable internet access can affect participation in online learning activities. In addition, students who are working while studying may have limited time for academic tasks, and insufficient family support can reduce motivation and academic stability.

Overall, students from disadvantaged socioeconomic backgrounds often face higher academic risks, including lower performance and an increased likelihood of dropout due to these external challenges.

D. Psychological Factors

Recent studies increasingly include psychological indicators to better understand student learning behaviour and academic performance. These factors provide insight into how students feel, think, and manage their learning process.

Common psychological indicators include motivation, stress levels, emotional engagement, and self-regulated learning abilities. Motivation reflects a student's drive to succeed, while stress can affect concentration and performance. Emotional engagement shows how connected students feel to their learning activities, and self-regulated learning refers to how well students manage their own study habits and learning strategies.

Combining these psychological factors with AI-based predictive models can improve the accuracy of student performance predictions. It allows a more holistic understanding of student success by integrating both academic data and emotional or behavioral states, leading to better identification of at-risk students and more effective intervention strategies.

VI. Applications of AI in Universities

AI in education is widely applied in several key areas to improve student outcomes and institutional efficiency. Early Warning Systems help identify at-risk students before academic failure by using predictive dashboards, automated alerts, real-time monitoring, and risk classification, leading to improved retention rates and faster interventions [12]. Personalized Learning Systems create adaptive learning environments by offering personalized learning paths, adaptive assessments, intelligent tutoring systems, and customized feedback, which enhance student engagement and learning outcomes.

In addition, AI supports academic advising by helping advisors identify struggling students, recommend suitable courses, design intervention strategies, and monitor academic progress more effectively. At the institutional level, AI is used for decision-making processes such as enrolment forecasting, resource allocation, curriculum planning, and optimizing student support services. Overall, AI contributes to more efficient, data-driven, and evidence-based educational management and governance.

VII. Evaluation Metrics

Predictive systems are assessed using several performance metrics to measure their effectiveness. These include accuracy, which indicates the overall percentage of correct predictions, and precision, which measures how many of the predicted positive cases are actually correct. Recall evaluates the system's ability to correctly identify at-risk students, while the F1-score provides a balanced measure between precision and recall. ROC-AUC is also used to assess classification performance across different decision thresholds. Recent studies highlight the importance of using multiple evaluation metrics, as educational datasets are often imbalanced and may not be fully represented by a single metric [13].

VIII. Challenges and Limitations

A. Data Privacy and Security

Educational datasets contain sensitive student information, making data privacy and security critical concerns in educational environments. These datasets often include academic records, personal details, and behavioral data that must be carefully protected to prevent misuse or exposure.

Key challenges include the risk of data breaches, unauthorized access to confidential information, ethical issues in data handling, and compliance with privacy regulations. If these issues are not properly addressed, they can compromise student confidentiality and institutional trust.

Therefore, universities must implement responsible data governance frameworks to ensure that student data is securely stored, ethically managed, and compliant with relevant privacy laws and institutional policies [14].

B. Algorithmic Bias

AI systems may unintentionally reinforce historical inequalities that exist within training datasets. As a result, the predictions or decisions made by these systems can reflect and amplify existing social biases.

This bias can affect different groups of students, including those based on gender, socioeconomic background, and minority status. For example, certain groups may be incorrectly predicted as at-risk or may not receive equal academic support due to biased data patterns.

Therefore, bias mitigation strategies are essential to ensure fair and ethical AI implementation in education, helping to promote equal opportunities for all students [14].

C. Lack of Explainability

Many Deep Learning models operate as “black boxes,” meaning their internal decision-making processes are not easily understandable. This lack of transparency makes it difficult for users to see how predictions are generated.

As a result, the reduced interpretability can lower trust among educators and administrators, especially when AI outputs are used for important academic decisions such as identifying at-risk students or recommending interventions.

Therefore, Explainable AI is increasingly necessary to ensure responsible and trustworthy educational AI systems, as it helps make model decisions more transparent, interpretable, and suitable for real-world academic use [11].

D. Generalizability Issues

Models developed using datasets from a single university may not perform effectively when applied to different institutional contexts. This is because each institution may have unique characteristics that influence student data patterns and learning behaviors.

Several factors affect the generalizability of these models, including differences in curriculum structure, institutional policies, student demographics, and cultural contexts. These variations can cause predictive models to perform well in one setting but less accurately in another.

Therefore, cross-institutional collaboration is necessary to improve model robustness, as sharing diverse datasets and insights can help develop more generalizable and reliable AI models for educational use.

IX. Ethical Considerations in AI for Education

Ethical considerations are a critical aspect of applying AI in education, as these systems directly influence student learning outcomes, academic opportunities, and institutional decision-making. One major concern is data privacy and confidentiality, since educational AI systems rely on sensitive student information such as academic records, behavioral logs, and personal details. Improper handling of this data can lead to privacy violations and unauthorized access, making strong data protection measures essential [14].

Another key ethical issue is fairness and bias in AI models. AI systems trained on historical data may unintentionally reinforce existing inequalities, leading to biased outcomes against certain groups of students, such as those from different genders, socioeconomic backgrounds, or minority populations. To address this, bias detection and mitigation strategies are necessary to ensure fairness in predictions and recommendations [14].

Transparency and explainability are also important ethical requirements. Many AI models, especially deep learning systems, operate as black boxes, making it difficult for educators and administrators to understand how decisions are made. This lack of interpretability can reduce trust and accountability in AI-based educational tools. Explainable AI approaches are therefore essential to improve transparency and support responsible decision-making [11].

In addition, ethical considerations include accountability and responsibility, where institutions must clearly define who is responsible for AI-driven decisions, especially when errors or harmful outcomes occur. There is also a need for compliance with legal and institutional regulations, ensuring that AI systems adhere to data protection laws and educational policies.

Overall, ethical considerations ensure that AI in education is used in a responsible, fair, and transparent manner that supports student well-being and maintains institutional trust.

X. Future Research Directions in AI for Education

Future research in AI-driven educational systems is expected to focus on improving model performance, fairness, and real-world applicability. One important direction is the development of more generalizable predictive models that can work effectively across different institutions. Current studies show that many models fail to perform well when applied to new educational settings due to differences in curriculum, student populations, and institutional policies [5][2].

Another key area is the improvement of model transparency and interpretability. Future work should focus on integrating Explainable AI (XAI) techniques to ensure that educators can clearly understand how predictions are made, especially in high-stakes decisions such as identifying at-risk students and recommending interventions [11].

Researchers are also encouraged to explore the integration of multimodal and non-traditional data sources, such as psychological, behavioral, and socioeconomic factors, to improve prediction accuracy and provide a more holistic understanding of student performance [2]. In addition, combining AI with real-time learning analytics systems is expected to enhance early warning mechanisms and adaptive learning interventions [7].

Ethical and responsible AI is another major research direction. Future studies should address issues such as data privacy, algorithmic bias, fairness, and accountability, ensuring that AI systems do not reinforce educational inequalities and remain aligned with ethical standards [14]. Human-centered design approaches are also increasingly recommended to involve educators and students in the development process, improving trust and usability [10].

Finally, future research should focus on cross-institutional collaboration and large-scale datasets, which can improve model robustness and reduce limitations in generalizability. This will help develop more reliable AI systems that can be applied across diverse educational contexts [5].

Overall, future research is moving toward building AI systems that are more accurate, transparent, ethical, and adaptable to real-world educational environments.

References

- A. Almalawi, B. Soh, A. Li, and H. Samra, "Predictive Models for Educational Purposes: A Systematic Review," *Big Data and Cognitive Computing*, vol. 8, no. 12, pp. 1–29, 2024.
- B. Albreiki, N. Zaki, and H. Alashwal, "A Systematic Literature Review of Student Performance Prediction Using Machine Learning Techniques," *Education Sciences*, vol. 11, no. 9, pp. 1–32, 2021.
- D. A. Shafiq, M. Marjani, R. A. A. Habeeb, and D. Asirvatham, "Student Retention Using Educational Data Mining and Predictive Analytics: A Systematic Literature Review," *IEEE Access*, vol. 10, pp. 78427–78444, 2022.
- T. Cardona, E. A. Cudney, R. Hoerl, and J. Snyder, "Data Mining and Machine Learning Retention Models in Higher Education," *Journal of Educational Technology Systems*, vol. 50, no. 1, pp. 1–25, 2020.

- K. Alalawi, R. Athauda, and R. Chiong, “Contextualizing the Current State of Research on the Use of Machine Learning for Student Performance Prediction: A Systematic Literature Review,” *Engineering Reports*, vol. 5, no. 6, pp. 1–20, 2023.
- S. Helal, J. Li, L. Liu, E. Ebrahimie, and D. Dawson, “Predicting Academic Performance by Considering Student Heterogeneity,” *Knowledge-Based Systems*, vol. 161, pp. 134–146, 2018.
- H. Aldowah, H. Al-Samarraie, and W. M. Fauzy, “Educational Data Mining and Learning Analytics for 21st Century Higher Education: A Review and Synthesis,” *Telematics and Informatics*, vol. 37, pp. 13–49, 2019.
- M. Viberg, O. Hatakka, M. Balter, and A. Mavroudi, “The Current Landscape of Learning Analytics in Higher Education,” *Computers in Human Behavior*, vol. 89, pp. 98–110, 2018.
- “The Power of Deep Learning Techniques for Predicting Student Performance in Virtual Learning Environments: A Systematic Literature Review,” *Computers and Education: Artificial Intelligence*, vol. 6, pp. 100231, 2024.
- “Predicting Student Performance: A Comprehensive Review of Machine Learning, Deep Learning, and Explainable AI Approaches,” *Computers and Education: Artificial Intelligence*, vol. 10, pp. 100548, 2026.
- R.-M. Nechita, D.-C. Deselnicu, P.-C. Simion, and M. A. M. Ichimov, “Structuring a Factor-Based Framework for Student Retention: A Systematic Review and Clustering for MCDM Applications,” *Frontiers in Education*, vol. 11, pp. 1–22, 2026.
- S. T. K. and M. Chakkaravarthy, “Predictive Analytics in Education: Machine Learning Approaches and Performance Metrics for Student Success – A Systematic Literature Review,” *Data and Metadata*, vol. 4, pp. 1–18, 2025.
- J. Luan, “Data Mining and Its Applications in Higher Education,” *New Directions for Institutional Research*, vol. 2002, no. 113, pp. 17–36, 2002.
- R. S. Baker and K. Yacef, “The State of Educational Data Mining in 2009: A Review and Future Visions,” *Journal of Educational Data Mining*, vol. 1, no. 1, pp. 3–17, 2009.
- G. Siemens and R. S. J. d. Baker, “Learning Analytics and Educational Data Mining: Towards Communication and Collaboration,” in *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge*, 2012, pp. 252–254.
- E. Fernandes, M. Holanda, M. Victorino, V. Borges, R. Carvalho, and G. Van Erven, “Educational Data Mining: Predictive Analysis of Academic Performance of Public School Students in the Capital of Brazil,” *Journal of Business Research*, vol. 94, pp. 335–343, 2019.
- A. Pena-Ayala, “Educational Data Mining: A Survey and a Data Mining-Based Analysis of Recent Works,” *Expert Systems with Applications*, vol. 41, no. 4, pp. 1432–1462, 2014.
- C. Romero and S. Ventura, “Educational Data Mining: A Review of the State of the Art,” *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 40, no. 6, pp. 601–618, 2010.
- Y. Park and I.-H. Jo, “Development of the Learning Analytics Dashboard to Support Students’ Learning Performance,” *Journal of Universal Computer Science*, vol. 21, no. 1, pp. 110–133, 2015.
- M. Ifenthaler and D.-K. Mah, “Students’ Perceptions Toward Academic Competencies: The Case of Learning Analytics in Higher Education,” *Technology, Knowledge and Learning*, vol. 22, pp. 233–252, 2017.