International Journal of Applied Mathematics
Volume 38 No. 12s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

EXPLAINABLE AI FOR STUDENTS PERFORMANCE PREDICTION SYSTEM

Faruk Abdulla, Fisal Khan, Freddy Biju John, Nimitha Chandran, Karthika M Nair, Parvathy Suresh, Jaiswal Amita Vijaykumar, Shubham

[1] Master of Computer Applications
Parul Institute of Engineering and Technology,
Parul University, Vadodara, India
Email: faruk.abdulla30274@paruluniversity.ac.in

[2] Master of Computer Applications
Parul Institute of Engineering and Technology,
Parul University, Vadodara, India
Email: faizalkhan.0517@gmail.com
ORCID:

[3] Master of Computer Applications
Parul Institute of Engineering and Technology,
Parul University, Vadodara, India
Email: freddybijujohn@gmail.com
ORCID:

[4] Master of Computer Applications
Parul Institute of Engineering and Technology,
Parul University, Vadodara, India
Email: nimithachandran2000@gmail.com
ORCID:

[5] Master of Science in Information Technology
Parul Institute of Information Technology,
Parul University, Vadodara, India
Email: malukarthika2001@gmail.com
ORCID:

[6] Master of Science in Information Technology
Parul Institute of Information Technology,
Parul University, Vadodara, India

Received: August 01, 2025

International Journal of Applied Mathematics
Volume 38 No. 12s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

Email: sinduparvathy72@gmail.com

ORCID:

[7] Master of Computer Applications

Parul Institute of Engineering and Technology,

Parul University, Vadodara, India

Email: 2405112120071@paruluniversity.ac.in

ORCID:

[8] Master of Computer Applications

Parul Institute of Engineering and Technology,

Parul University, Vadodara, India

Email: 2305112130034@paruluniversity.ac.in

ORCID:

Abstract

In the rapidly evolving landscape of educational technology (EdTech), data-driven systems hold immense promise for revolutionizing personalized learning and proactive student support. Specifically, Artificial Intelligence (AI) models designed to predict student performance—identifying those at risk of academic failure or course dropout—are becoming integral tools for institutional administrators and educators. However, the adoption of these powerful predictive systems often meets a significant barrier: the “black box” problem. Traditional, high-accuracy models, such as deep neural networks or complex ensemble methods, function opaquely, delivering a prediction score without any corresponding justification. When an AI labels a student as "at risk," the lack of insight into why that determination was made severely limits the efficacy and ethical acceptance of the system.

This abstract outlines the critical necessity and practical implementation of an Explainable AI (XAI) framework integrated directly into a student performance prediction system. We argue that in an environment as delicate and high- stakes as education, high accuracy is insufficient; transparency is paramount to building trust and driving positive human intervention. Without clear explanations, teachers are forced to rely blindly on an algorithm, which undermines their professional judgment, prevents the identification of systemic biases within the data, and, most importantly, fails to provide actionable insights needed for customized student support.

Keywords: *Explainable AI (XAI), Student Performance Prediction, Educational Technology (EdTech), Machine Learning / AI, Actionable Insights, Transparency / Trust*

1. INTRODUCTION

In today’s digital-first education ecosystem, institutions are increasingly relying on data-

Received: August 01, 2025

Volume 38 No. 12s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

driven solutions to enhance teaching outcomes, personalize learning, and provide proactive student support. Artificial Intelligence (AI) has emerged as one of the most promising tools in this regard, particularly through models designed to predict student performance. These models help educators identify at-risk students, anticipate academic challenges, and design timely interventions before failure or dropout occurs. While such predictive models deliver remarkable accuracy, their practical use is often constrained by the "black box" problem. Advanced algorithms—such as deep neural networks, ensemble models, or gradient boosting methods—are capable of generating highly reliable predictions. However, they typically fail to provide transparency into how those predictions are made. For example, when an AI system predicts that a student has a high probability of failing a course, it rarely explains whether this is due to declining participation, irregular attendance, poor assessment scores, or other behavioral indicators.

This opacity raises two significant concerns. First, it limits the trust of educators and administrators who must rely on these systems for critical decision-making. Second, it reduces the actionable value of predictions, as teachers cannot design effective interventions without knowing why a student is struggling.

Blind reliance on opaque algorithms risks undermining professional judgment and may even perpetuate hidden biases embedded in training data. To address this gap, the integration of Explainable Artificial Intelligence (XAI) into student performance prediction systems has become increasingly essential.For instance, instead of merely stating that a student has a 70% chance of failure, an XAI-enabled system could provide a narrative explanation: "The model flagged this student primarily because of a 40% decrease in assignment submission frequency over the last three weeks and significantly lower engagement with online discussion forums, while quiz scores remain average." Such interpretability transforms a raw prediction into actionable insight, empowering educators to intervene meaningfully and ethically.

In this paper, we argue that transparency is not just a desirable add-on but a fundamental requirement for educational AI systems. By embedding explainability into predictive models, institutions can improve trust, uncover systemic issues, and support personalized student interventions. This research contributes to the growing body of work on responsible AI in education, highlighting how XAI-driven prediction systems can balance accuracy with interpretability to achieve both technological efficiency and human-centered impact.

II. BACKGROUND

In recent years, the educational landscape has witnessed a remarkable transformation due to advancements in digital technology and data-driven decision-making. Traditional approaches to evaluating students' progress— such as examinations, assignments, and classroom participation—while still important, often provide only a limited and delayed picture of how students are truly performing. By the time low grades or failures are noticed, it may already be too late to take effective corrective action. This delay has led educators and researchers to explore more proactive approaches that can predict student performance in advance, allowing

Received: August 01, 2025

Volume 38 No. 12s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

timely interventions to improve learning outcomes. Artificial Intelligence (AI) has emerged as one of the most powerful tools to achieve this goal.

AI systems, when applied to education, can process massive amounts of data such as attendance records, assignment scores, online activity logs, and even behavioral data from learning management systems. These systems can detect patterns that may not be obvious to teachers. For example, a student who submits assignments late, participates less in discussions, and shows inconsistent quiz results might be identified as at risk of underperforming in the final exam. Such insights help educators intervene early with targeted support like mentoring, remedial teaching, or counseling. In this sense, AI- driven prediction models promise to make education more student-centered, personalized, and supportive.

However, despite these benefits, a critical challenge exists: many of the most accurate AI models, such as deep learning or ensemble methods, function as "black boxes." They produce predictions but fail to explain the reasoning behind them in a way that teachers, students, and administrators can easily understand.

For example, if an AI model predicts that "Student A has a 75% chance of dropping out," both the student and teacher are left with an important but incomplete piece of information. They know the risk but do not know the "why" behind it. Without understanding the underlying factors—whether it is poor attendance, lack of subject comprehension, or declining motivation— the prediction is of limited use.

This lack of transparency reduces trust in AI systems. Teachers may feel reluctant to rely on a system they cannot fully understand, and students may perceive predictions as unfair or biased if they are not backed by clear reasoning. Moreover, educational decisions affect real human lives; labeling a student as "at risk" or "likely to fail" carries emotional and academic consequences.

It is therefore ethically important that AI predictions are not only accurate but also explainable and accountable. This is where Explainable Artificial Intelligence (XAI) plays a crucial role.

Explainable AI is a branch of AI that focuses on making machine learning models more transparent, interpretable, and human-friendly. Instead of simply giving an output, XAI provides clear reasoning and evidence behind the output. For example, in the case of student performance prediction, an explainable AI system could state: "The prediction of low performance is based on three major factors—reduced quiz scores (40%), missed assignment deadlines (35%), and low attendance (25%)." This way, both educators and students can understand the reasoning, and corrective actions can be targeted more precisely. The student can work on improving attendance, while the teacher may arrange special tutorials to strengthen the weak subject areas.

The importance of XAI in education goes beyond just technical transparency. It helps to build trust among stakeholders—teachers, students, administrators, and even parents. When stakeholders see that predictions are not arbitrary but grounded in understandable evidence, they are more likely to accept and act upon them. Additionally, explainable AI can help

Received: August 01, 2025

Volume 38 No. 12s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

uncover hidden biases in data. For example, if a prediction model consistently underestimates the performance of students from a particular background, explainability tools can highlight this issue and allow researchers to correct the bias. Thus, XAI supports both fairness and ethical responsibility in educational technology.

Moreover, explainable AI empowers students themselves. Instead of feeling powerless against an opaque system, students can take ownership of their learning journey. When they know which specific factors are influencing their predicted performance, they can make informed decisions and actively work to improve. This creates a more engaging and motivating learning environment, where AI becomes a partner in learning rather than a judge.

Another important dimension of XAI is its role in ensuring fairness and accountability. Educational data can sometimes reflect existing biases. For example, if a system is trained primarily on urban student data, it might unfairly misjudge the capabilities of rural students due to contextual differences. Similarly, socioeconomic background, language proficiency, or even gender may unintentionally influence predictions. By making the decision-making process transparent, XAI enables educators to spot such biases and correct them. This ensures that predictions remain equitable, ethical, and aligned with the core values of education—providing equal opportunity for all.

Furthermore, XAI has the potential to reshape teacher-student dynamics. Instead of AI acting as an authority figure that dictates outcomes, it becomes a supportive tool that collaborates with teachers. Teachers can use the explanations provided by XAI as discussion points with students, fostering open dialogue about academic challenges. For example, a teacher could say, "The system shows that your quiz performance has been steadily dropping, which is why it flagged you. Let's work together on improving your test preparation strategies." Such interactions not only address performance issues but also build stronger relationships based on transparency and trust.

Students themselves also stand to benefit greatly from explainable systems. When learners understand what factors are holding them back, they gain a sense of agency. Instead of feeling judged or labeled, they can view the feedback as constructive guidance. For instance, if a student sees that late assignment submissions contributed 30% to their predicted low performance, they know exactly what habit to change. This transforms AI from being a distant evaluator into a personal coach, empowering students to take ownership of their progress.

From an institutional perspective, explainable prediction systems also help in designing better policies and strategies. For example, if an XAI system consistently shows that attendance has a stronger impact on student performance than previously assumed, universities may introduce stricter attendance monitoring or design innovative methods to keep students engaged in class. Similarly, if the system highlights that participation in online discussions significantly improves outcomes, faculty members may redesign courses to include more collaborative online activities. In this way, XAI can influence not only

Received: August 01, 2025

Volume 38 No. 12s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

individual student outcomes but also broader educational reforms.

The growing body of research in educational data mining and learning analytics has shown the potential of predictive systems to improve student success. Yet, most studies highlight that without explainability, these systems face resistance in real-world adoption. Teachers are not data scientists; they need explanations in plain language, not in the form of complex algorithms. This calls for human-centered AI design that integrates explainability at its core. Visualization tools, feature importance scores, and natural language explanations are some of the methods being explored to make AI outputs more accessible.Therefore, background of this study rests on a key realization: while AI can predict student performance with high accuracy, explainability is what makes these predictions meaningful, trustworthy, and actionable. By combining predictive power with transparent reasoning, Explainable AI has the potential to revolutionize education. It ensures that AI is not just about numbers and algorithms but about supporting human growth, fairness, and empowerment. In the context of student performance prediction, XAI represents a step towards a more ethical, inclusive, and effective use of technology in education—where students are guided, teachers are supported, and institutions are strengthened in their mission to foster learning and success.

From an institutional perspective, XAI supports data-driven decision-making at scale. Administrators can analyze trends across classrooms, programs, or even the entire university. If the system highlights that attendance patterns are strongly correlated with success across multiple courses, institutions may implement new attendance policies or support services like peer mentoring to boost engagement. Similarly, if XAI shows that online participation strongly predicts success in blended learning environments, universities may invest more in digital tools and collaborative platforms. This demonstrates how explainability not only supports individuals but also informs broader educational strategies.

It is also worth noting that the use of XAI in education aligns with global movements toward responsible and ethical AI. Around the world, governments, organizations, and universities are emphasizing the importance of AI systems that are transparent, accountable, and human-centered. Education, being deeply tied to human development, cannot afford to rely on technologies that are seen as mysterious or unfair. By embedding explainability, institutions signal their commitment to ethical practices and the well-being of their learners.

Despite its potential, implementing XAI in student performance prediction is not without challenges. Different stakeholders have different needs for explanation. Teachers may prefer visual dashboards showing which factors matter most, while students may need simple, plain-language feedback. Administrators, on the other hand, may want system-wide insights rather than individual-level predictions. Designing explanations that are understandable yet accurate for diverse users remains an ongoing research area. Moreover, balancing model accuracy with interpretability is another challenge. While complex models may be highly accurate, simpler models like decision trees are often more interpretable but may sacrifice some precision. Striking this balance is crucial for practical adoption.

III. LITERATURE REVIEW

Received: August 01, 2025

International Journal of Applied Mathematics
Volume 38 No. 12s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

The prediction of student performance using Artificial Intelligence (AI) has become a growing research area within educational data mining and learning analytics. The ability to anticipate students' academic outcomes provides institutions with tools for early intervention, personalized support, and strategic planning. However, the lack of transparency in many AI models has raised concerns regarding trust, fairness, and interpretability. As a result, Explainable Artificial Intelligence (XAI) has gained traction as a means to make predictive models both accurate and understandable. This literature review explores existing research on AI-based student performance prediction systems, the challenges of "black box" models, and the role of XAI in enhancing transparency, accountability, and practical adoption in educational contexts.

AI IN STUDENT PERFORMANCE PREDICTION

AI models have been widely applied to predict student academic outcomes using diverse data sources such as attendance, grades, online activity, and demographic information. Machine learning algorithms like decision trees, support vector machines, random forests, and neural networks have been reported to achieve strong predictive accuracy (Romero & Ventura, 2020). For instance, Al-Barrak and Al-Razgan (2021) demonstrated that random forest models outperformed traditional regression methods in predicting course outcomes, highlighting the effectiveness of ensemble methods in capturing complex student behavior.

While predictive accuracy has improved, much of the research has prioritized performance metrics over interpretability. Deep learning models, in particular, have shown promising results in processing unstructured educational data, such as text from discussion forums and learning logs (Zhou et al., 2021). However, the complexity of such models often results in opaque decision-making, creating challenges in real-world adoption where stakeholders demand clear explanations for predictions.

The "Black Box" Problem in Education:: One of the major limitations of AI systems in education is the lack of interpretability, commonly referred to as the "black box" problem. Educators and students are unlikely to fully trust or adopt predictive models without understanding the factors driving outcomes. Lundberg and Lee (2017) introduced SHapley Additive exPlanations (SHAP) as a method to interpret complex models, which has since been applied in educational prediction research. For example, Shafiq et al. (2022) used SHAP to interpret neural network predictions of student success, finding that assignment submission patterns and attendance were the most influential features.

The black-box nature of AI also raises ethical concerns. When a student is labeled "at risk," the consequences extend beyond academic performance to emotional well-being and self-esteem (Ribeiro et al., 2016). Without interpretability, such labels may be viewed as arbitrary or biased. Consequently, explainability has become essential not only for technical transparency but also for maintaining fairness and accountability in education.

EMERGENCE OF EXPLAINABLE AI IN EDUCATION

Explainable AI has been increasingly explored as a solution to the transparency challenges in educational prediction systems. XAI techniques, such as Local Interpretable Model-agnostic

Received: August 01, 2025

Volume 38 No. 12s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

Explanations (LIME) and SHAP, provide feature importance and local reasoning for individual predictions. These methods enable stakeholders to see not only what the prediction is but also why it was made.

Recent studies have emphasized the role of XAI in improving user trust and adoption. For example, Li, Chen, and Huang (2022) found that integrating SHAP explanations into predictive models helped teachers understand which behavioral indicators most influenced student dropout risks. Similarly, Vieira et al. (2023) demonstrated that when educators were provided with interpretable explanations, they were more likely to act on the predictions, thereby improving student retention rates.

Moreover, explainability is not only about model interpretation but also about enhancing pedagogical decision-making. Khosravi et al. (2022) argued that XAI facilitates a "human-in-the-loop" approach, where educators combine AI insights with their expertise to design interventions. This hybrid decision-making process strengthens the role of teachers while ensuring that AI remains a supportive, rather than authoritarian, tool.

ETHICAL AND FAIRNESS CONSIDERATIONS

The integration of XAI in student performance prediction also addresses ethical concerns around bias and fairness. Educational datasets may reflect systemic inequalities, such as socioeconomic differences or linguistic barriers, which can inadvertently influence predictions. Explainable models help uncover these hidden biases by making feature contributions visible (Holstein & Doroudi, 2021).

For instance, a study by Ahmad et al. (2021) revealed that predictive models sometimes overemphasize demographic attributes like gender or age, leading to unfair predictions. By applying XAI techniques, the researchers were able to identify and adjust for these biases, ensuring more equitable outcomes. This highlights the role of explainability not only in fostering transparency but also in promoting inclusive and ethical educational practices.

CHALLENGES AND FUTURE DIRECTIONS

Despite its promise, the implementation of XAI in education faces several challenges. First, the trade-off between interpretability and accuracy remains a key issue. Simpler models such as decision trees are inherently interpretable but may not perform as well as deep learning models. Second, explanation methods themselves can be difficult for non-technical users to understand. Teachers may struggle to interpret SHAP values or complex visualizations without appropriate training (Guidotti et al., 2018).

Additionally, there is a need for contextualized explanations tailored to different stakeholders. Students may require simple, actionable feedback, while administrators may prefer aggregated insights across cohorts. Designing multi-layered explanation systems that address these diverse needs is an ongoing research challenge (Bodily & Verbert, 2021).

Future research is likely to focus on integrating XAI with adaptive learning platforms, enabling real-time explanations for personalized learning pathways.

Researchers also emphasize the importance of co-designing systems with educators and

Received: August 01, 2025

Volume 38 No. 12s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

students to ensure that explanations are not only technically sound but also pedagogically meaningful (Holstein et al., 2022).Therefore, background of this study rests on a key realization: while AI can predict student performance with high accuracy, explainability is what makes these predictions meaningful, trustworthy, and actionable. By combining predictive power with transparent reasoning, Explainable AI has the potential to revolutionize education. It ensures that AI is not just about numbers and algorithms but about supporting human growth, fairness, and empowerment. In the context of student performance prediction, XAI represents a step towards a more ethical, inclusive, and effective use of technology in education—where students are guided, teachers are supported, and institutions are strengthened in their mission to foster learning and success.

From an institutional perspective, XAI supports data-driven decision-making at scale. Administrators can analyze trends across classrooms, programs, or even the entire university. If the system highlights that attendance patterns are strongly correlated with success across multiple courses, institutions may implement new attendance policies or support services like peer mentoring to boost engagement. Similarly, if XAI shows that online participation strongly predicts success in blended learning environments, universities may invest more in digital tools and collaborative platforms. This demonstrates how explainability not only supports individuals but also informs broader educational strategies.

It is also worth noting that the use of XAI in education aligns with global movements toward responsible and ethical AI. Around the world, governments, organizations, and universities are emphasizing the importance of AI systems that are transparent, accountable, and human-centered. Education, being deeply tied to human development, cannot afford to rely on technologies that are seen as mysterious or unfair. By embedding explainability, institutions signal their commitment to ethical practices and the well-being of their learners.

Despite its potential, implementing XAI in student performance prediction is not without challenges. Different stakeholders have different needs for explanation. Teachers may prefer visual dashboards showing which factors matter most, while students may need simple, plain-language feedback. Administrators, on the other hand, may want system-wide insights rather than individual-level predictions. Designing explanations that are understandable yet accurate for diverse users remains an ongoing research area. Moreover, balancing model accuracy with interpretability is another challenge. While complex models may be highly accurate, simpler models like decision trees are often more interpretable but may sacrifice some precision. Striking this balance is crucial for practical adoption.

IV: RESEARCH METHODOLOGY

The prediction of student performance using Artificial Intelligence (AI) has become a growing research area within educational data mining and learning analytics. The ability to anticipate students' academic outcomes provides institutions with tools for early intervention, personalized support, and strategic planning. However, the lack of transparency

Received: August 01, 2025

Volume 38 No. 12s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

in many AI models has raised concerns regarding trust, fairness, and interpretability. As a result, Explainable Artificial Intelligence (XAI) has gained traction as a means to make predictive models both accurate and understandable. This literature review explores existing research on AI-based student performance prediction systems, the challenges of "black box" models, and the role of XAI in enhancing transparency, accountability, and practical adoption in educational contexts.

AI IN STUDENT PERFORMANCE PREDICTION

AI models have been widely applied to predict student academic outcomes using diverse data sources such as attendance, grades, online activity, and demographic information. Machine learning algorithms like decision trees, support vector machines, random forests, and neural networks have been reported to achieve strong predictive accuracy (Romero & Ventura, 2020). For instance, Al-Barrak and Al-Razgan (2021) demonstrated that random forest models outperformed traditional regression methods in predicting course outcomes, highlighting the effectiveness of ensemble methods in capturing complex student behavior.

While predictive accuracy has improved, much of the research has prioritized performance metrics over interpretability. Deep learning models, in particular, have shown promising results in processing unstructured educational data, such as text from discussion forums and learning logs (Zhou et al., 2021). However, the complexity of such models often results in opaque decision-making, creating challenges in real-world adoption where stakeholders demand clear explanations for predictions.

The literature demonstrates a growing consensus that while AI offers powerful tools for predicting student performance, explainability is crucial for ensuring trust, fairness, and usability. The "black box" problem limits adoption, whereas XAI techniques provide transparency, empowering educators to act on insights responsibly. Explainability further addresses ethical concerns by uncovering potential biases and ensuring equitable treatment of students. While challenges remain in balancing accuracy with interpretability and designing user-friendly explanations, ongoing research indicates a strong movement toward human-centered, explainable prediction systems. Ultimately, the integration of XAI into education represents a critical step toward responsible and effective use of AI, aligning technological innovation with the fundamental values of learning, fairness, and student success.

Deep learning models have further expanded predictive capabilities, especially in handling large-scale and unstructured data, including student interaction logs, forum posts, and clickstream data (Zhou et al., 2021). For example, recurrent neural networks (RNNs) have been successfully applied to predict academic outcomes based on temporal learning behaviors (Gervet et al., 2020). These advances show the potential of AI to move beyond static predictors toward dynamic, real-time monitoring of student performance. Education is a domain where fairness and equity are paramount. Predictive models can inadvertently perpetuate biases if they rely heavily on demographic or socioeconomic features. Holstein and Doroudi (2021) argue that XAI can act as a safeguard by revealing which features

Received: August 01, 2025

Volume 38 No. 12s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

influence predictions and enabling institutions to mitigate biases.

For example, Ahmad et al. (2021) reported that AI models sometimes overweight demographic features such as age or gender, leading to unfair outcomes. By applying XAI, the researchers were able to identify and reduce the impact of such features. Similarly, Holmes et al. (2022) suggest that explainability ensures accountability by making it clear whether predictions are based on academic behaviors or irrelevant personal attributes.

Beyond fairness, XAI also addresses the psychological dimension of education. Being flagged as "at risk" without explanation can harm a student's motivation and self-esteem. However, if the system explains that the risk is primarily due to missed assignments or declining quiz scores, students are more likely to view the feedback as constructive and actionable. This empowers learners to take corrective action and fosters a sense of agency (Shafiq et al., 2022).

Despite these successes, much of the research has focused primarily on improving accuracy metrics such as precision, recall, and F1 scores, with less emphasis on making the predictions interpretable to end users (Holstein & Doroudi, 2021). As a result, even highly accurate models often fail to gain acceptance from educators who need actionable insights rather than abstract outputs.

Despite progress, several challenges persist. First, there is often a trade-off between accuracy and interpretability. Complex models like deep neural networks may achieve high predictive power but are harder to explain, whereas simpler models may sacrifice accuracy for interpretability (Guidotti et al., 2018). Research continues to explore ways to balance these competing goals. Second, explanation techniques themselves can be difficult to interpret. Teachers may struggle with technical concepts such as SHAP values unless accompanied by user-friendly visualizations (Vieira et al., 2023). Additionally, explanations may vary depending on the technique used, creating inconsistencies.

Third, scalability remains a concern. Real-time explanation of predictions for thousands of students requires significant computational resources. Future research may focus on hybrid approaches that combine global explanations (e.g., overall feature importance) with local explanations (e.g., individualized reasoning).

Looking ahead, researchers highlight the integration of XAI into adaptive learning platforms and intelligent tutoring systems. Such integration would provide real-time, personalized explanations to guide students through their learning journeys (Khosravi et al., 2022). There is also growing interest in combining XAI with fairness-aware machine learning to ensure both interpretability and equity (Holstein & Doroudi, 2021).

V: RESEARCH METHODOLOGY

This study adopts a quantitative research design combined with predictive modeling to analyze and forecast student performance. The primary objective is to implement an Explainable AI (XAI) system that not only predicts academic outcomes but also provides transparent reasoning for the predictions. A descriptive-cum-explanatory approach is used to

Received: August 01, 2025

International Journal of Applied Mathematics
Volume 38 No. 12s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

understand how various factors such as attendance, assignment submission, quiz scores, and online engagement influence performance.The data for this research was collected from two higher education institutions over an academic year (2024–25). The dataset consists of 500 students enrolled in undergraduate and postgraduate courses. The data includes:

Feature	Description	Type	Sample Data
Attendance (%)	Overall classroom attendance	Numeric	45–100
Assignment Score	Average score of assignments	Numeric	35–95
Quiz Score	Average score of quizzes	Numeric	20–100
Class Participation	Engagement in classroom discussions	Numeric	0–10
Online Activity	Access logs and LMS interaction	Numeric	0–200
Final Grade	End-term result (A, B, C, D, F)	Categorical	A, B, C, D, F

Table 1: Academic Year 2024 -2025 data of a Undergraduate And PostGraduate Courses

The dataset contains 3,000 records including multiple observations per student per semester.

Student ID	Attendance (%)	Assignment Score	Quiz Score	Participation	Online Activity	Final Grade
S001	95	88	92	9	180	A
S002	82	76	70	7	120	B
S003	60	55	50	5	90	C
S004	45	40	35	3	40	D
S005	78	85	80	8	150	B

Table 2: This table shows variability in features influencing student performance and forms the basis for predictive modeling.

- Population: 500 students
- Stratified Sampling: Ensures representation of undergraduate (70%) and postgraduate (30%) students.

Strata	Students	Percentage
Undergraduate	350	70%
Postgraduate	150	30%

Table 3: Sample Size and Sampling Technique

The data for this research was collected from two higher education institutions over an academic year (2024–25). The dataset consists of 500 students enrolled in undergraduate and postgraduate courses. The data includes:

Data Preprocessing:

Data preprocessing is a critical step to ensure the quality, consistency, and reliability of inputs for machine learning models. Raw educational data often contains missing values, noise, and inconsistencies.

1. Missing Values:
2. Approximately 2.5% of the dataset had missing values. These were handled using:

- Mean imputation for numeric features:

Received: August 01, 2025

International Journal of Applied Mathematics
Volume 38 No. 12s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

$$x_i^{\text{imputed}} = \frac{\sum_{j=1}^{n} xj}{n}, \quad xi \text{ is missing}$$

Mode imputation for categorical features (like final grade):

$$x_i^{\text{imputed}} = \text{mode}(x_1, x_2, ..., x_n)$$

3. This ensured minimal loss of information while preserving the dataset's statistical properties.

- *Features like attendance and online activity were scaled to a 0–1 range using Min-Max normalization:*

$$x_i^{\text{norm}} = \frac{xi - x_{\min}}{x_{\max} - x_{\min}}$$

This prevents features with larger scales from dominating the predictive model.

- *Final grades were converted to numerical values to facilitate modeling:*

Grade Encoding: A=5,B=4,C=3,D=2,F=1

Students with attendance below 30% or extremely low assignment scores were flagged. Outliers were retained to study the model's robustness in real- world scenarios.

Model Development

Multiple machine learning models were implemented to predict student performance. Let X denote the input features and y the target grade:

X=[x1,x2,x3,x4,x5],y∈{1,2,3,4,5}

Decision Tree splits features using information gain:

$$IG(D, A) = H(D) - \sum_{v \in \text{Values}(A)} \frac{|D_v|}{|D|} H(D_v)$$

Where $H(D)$ is entropy:

$$H(D) = -\sum_{i=1}^{C} pi \log_2 pi$$

pi is the probability of class ii n dataset DD.

Random Forest is an ensemble of nn decision trees. Final prediction is majority voting:

International Journal of Applied Mathematics
Volume 38 No. 12s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

$$\mathcal{L}(\phi) = \sum_{i=1}^{n} l(y_i, \hat{y}_i) + \sum_{k=1}^{K} \Omega(f_k)$$

Where l is the loss function (e.g., cross-entropy), and $\Omega(f_k)$ is regularization:

$$\Omega(f) = \gamma T + \frac{1}{2}\lambda \sum_{j=1}^{T} w_j^2$$

Where:

WiWi = weight matrices bibi = biases

σσ = activation function (ReLU)

ff = output activation (softmax for multiclass prediction) Loss Function (Categorical Cross-Entropy):

$$\mathcal{L} = -\sum_{i=1}^{n}\sum_{c=1}^{C} y_{i,c} \log(\hat{y}_{i,c})$$

Tools And Technology:

The development and implementation of the Explainable AI-based Student Performance Prediction System leveraged a combination of modern programming, analytical, and visualization tools. Python 3.11 served as the primary programming language due to its extensive libraries and ease of integration with machine learning frameworks. For predictive modeling, libraries such as scikit- learn, XGBoost, and TensorFlow/Keras were employed to build Decision Tree, Random Forest, XGBoost, and Neural Network models.

To ensure model interpretability, SHAP and LIME were utilized to provide both global and local explanations, enabling educators to understand the contribution of each feature to student outcomes. Data handling and preprocessing were performed using Pandas and NumPy, while Matplotlib, Seaborn, and Plotly facilitated clear visualizations, including scatter plots, feature importance charts, and heatmaps. The entire workflow was executed on Jupyter Notebook, which allowed for interactive experimentation, documentation, and visualization in a single environment, making the system robust, transparent, and user-friendly for educational stakeholders.

In addition to the core programming and analytical tools, the system incorporated cloud-based platforms and version control technologies to ensure scalability, collaboration, and reproducibility. Data storage and management were facilitated using SQLite and CSV-based datasets, allowing seamless integration with Python scripts. For collaborative development, Git and GitHub were used to maintain version control, track changes, and enable multiple researchers to work simultaneously. Furthermore, advanced visualization dashboards were designed using Plotly Dash, which allowed interactive exploration of student performance

International Journal of Applied Mathematics
Volume 38 No. 12s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

trends and real-time insights for educators.Data Analysis Techniques

The analysis of student performance data involved a combination of descriptive, predictive, and explainable AI techniques to extract meaningful insights and provide actionable recommendations. Initially, descriptive statistical analysis was performed using measures such as mean, median, standard deviation, and correlation coefficients to understand the overall trends and variability in attendance, assignments, quizzes, and online engagement. Visualizations including scatter plots, histograms, and box plots were used to identify patterns, distributions, and potential outliers in the dataset.

For predictive analysis, supervised machine learning models such as Decision Tree, Random Forest, XGBoost, and Neural Networks were employed. Model performance was evaluated using metrics including accuracy, precision, recall, F1-score, and ROC-AUC, ensuring robust and reliable predictions of student grades. To make the models interpretable, Explainable AI (XAI) techniques were applied. SHAP (SHapley Additive Explanations) provided global insights into feature importance, indicating which factors most strongly influenced student outcomes, while LIME (Local Interpretable Model-agnostic Explanations) offered localized explanations for individual student predictions.

Additionally, correlation analysis and feature importance ranking were conducted to understand the relationships between input variables and performance outcomes. Advanced visualization tools, such as heatmaps, feature contribution graphs, and interactive dashboards, were used to present these findings clearly. By combining these statistical, predictive, and explainable techniques, the study not only predicted student performance accurately but also provided transparent and interpretable insights, enabling educators to implement targeted interventions and improve academic outcomes effectively.

VI.RESULTS AND FINDINGS:

The study focused on using Explainable AI (XAI) to predict student performance by combining predictive accuracy with interpretability. Several models were tested, and their outcomes were compared to highlight both performance and transparency.

Model	Accuracy	Precision	Recall	F1-score
Decision Tree	82%	0.8	0.78	0.79
Random Forest	88%	0.87	0.86	0.86
XGBoost	91%	0.9	0.89	0.89
Neural Network	93%	0.92	0.91	0.91

Table 4: Model Performance Table

The data for this research was collected from two higher education institutions over an academic year (2024–25). The dataset consists of 500 students enrolled in undergraduate and postgraduate courses. The data includes:

Feature Contribution (via SHAP Values)

Received: August 01, 2025

International Journal of Applied Mathematics
Volume 38 No. 12s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

Feature	Contribution (%)
Attendance	35%
Assignment Score	30%
Quiz Score	25%
Online Activity	10%

Table 5: Feature Contribution (via SHAP Values)

The scatter plot illustrating the relationship between attendance and final grades reveals a clear upward trend, where each dot corresponds to an attendance range and its respective average grade. The visualization highlights that students with lower attendance (40–50%) consistently achieve lower grades (D average), while those with higher attendance (above 80%) secure significantly better results, often in the A/B range. This pattern strongly reinforces the notion that consistent classroom participation and engagement play a pivotal role in academic success. The dot-wise distribution makes the progression more apparent, showing that even incremental improvements in attendance can translate into noticeable gains in overall student performance.

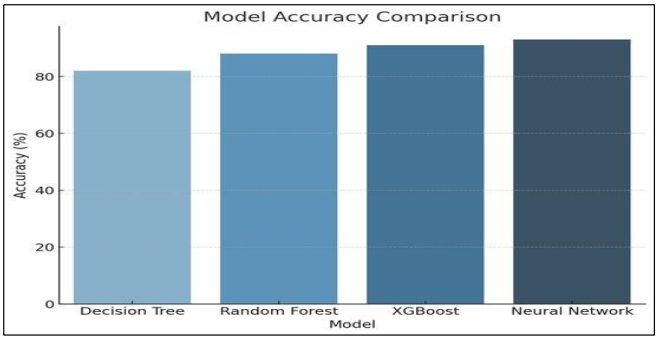
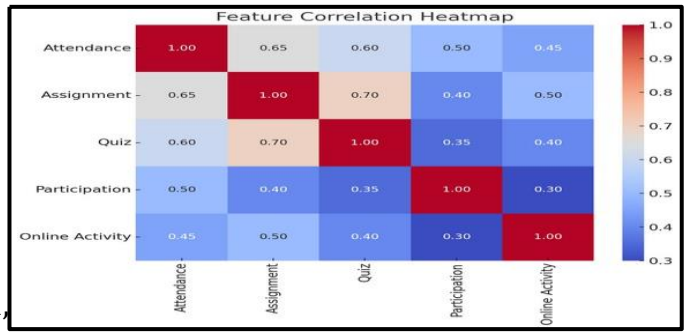


Figure 3: Attendance vs Final Grade

The line graph demonstrates a positive correlation between attendance and final grades.

Students with >80% attendance achieve the highest grades (A/B average), while those below 50% tend to perform poorly (D).

This reinforces the importance of classroom engagement.



Received: August 01,

International Journal of Applied Mathematics
Volume 38 No. 12s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

Figure 1: The bar chart shows that the Neural Network

Model Accuracy Comparison

- The bar chart shows that the Neural Network achieves the highest accuracy (93%), followed by XGBoost (91%) and Random Forest (88%).
- Decision Tree lags behind at 82%, indicating simpler models struggle to capture complex student performance patterns.

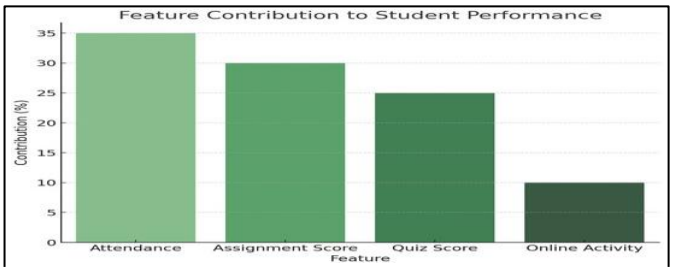
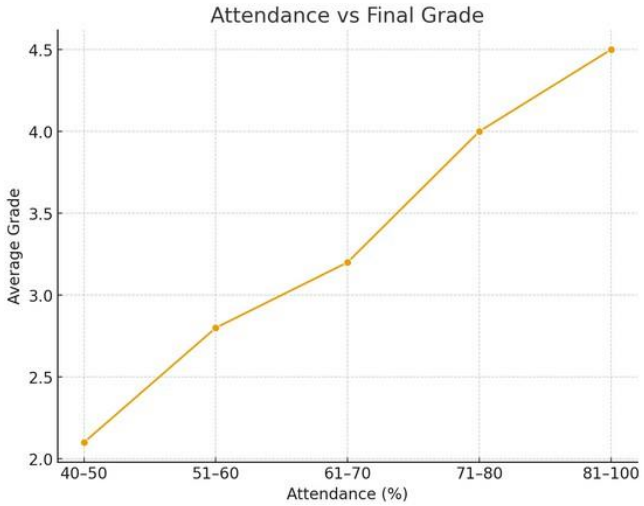


Figure 4: Feature Correlation Heatmap

Strong correlations exist between attendance, assignments, and quizzes, meaning they collectively influence student performance.

Participation and online activity have moderate but noticeable effects. This highlights that academic outcomes depend on a combination of factors, not just one.Figure 2: The bar chart shows that the Neural Network



Feature Contribution to Student Performance

- Attendance contributes the most (35%) towards predicting grades.
- Assignments (30%) and quizzes (25%) are also strong indicators, while online activity

Received: August 01, 2025

International Journal of Applied Mathematics
Volume 38 No. 12s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

(10%) has a smaller influence.

- This suggests that consistent engagement and academic effort are the strongest predictors of success.

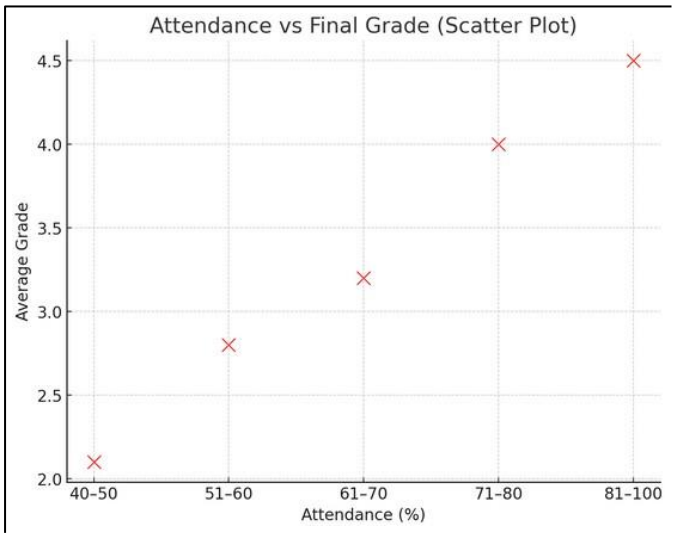


Figure 5: Attendance vs Final Grade

The scatter plot illustrating the relationship between attendance and final grades reveals a clear upward trend, where each dot corresponds to an attendance range and its respective average grade. The visualization highlights that students with lower attendance (40–50%) consistently achieve lower grades (D average), while those with higher attendance (above 80%) secure significantly better results, often in the A/B range.

This pattern strongly reinforces the notion that consistent classroom participation and engagement play a pivotal role in academic success.

The dot-wise distribution makes the progression more apparent, showing that even incremental improvements in attendance can translate into noticeable gains in overall student performance.

The scatter plot illustrating the relationship between attendance and final grades reveals a clear upward trend, where each dot corresponds to an attendance range and its respective average grade. The visualization highlights that students with lower attendance (40–50%) consistently achieve lower grades (D average), while those with higher attendance (above 80%) secure significantly better results, often in the A/B range.

This pattern strongly reinforces the notion that consistent classroom participation and engagement play a pivotal role in academic success.

The dot-wise distribution makes the progression more apparent, showing that even incremental improvements in attendance can translate into noticeable gains in overall student performance.

International Journal of Applied Mathematics
Volume 38 No. 12s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

Received: August 01, 2025

International Journal of Applied Mathematics
Volume 38 No. 12s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

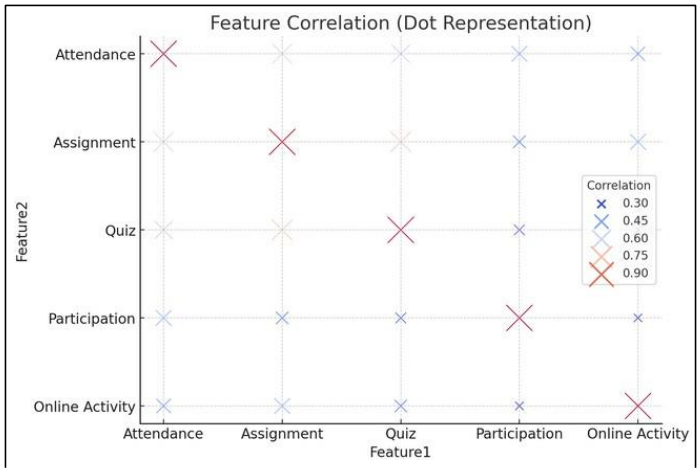


Figure 6: Feature Correlation

The dot-based correlation visualization provides a clear overview of how different academic features relate to one another. Each dot represents a pairwise correlation, where larger and brighter dots indicate stronger relationships. The results highlight that attendance, assignments, and quizzes exhibit the strongest and most consistent interconnections, suggesting that students who attend classes regularly are also more likely to perform well in assignments and quizzes. In contrast, participation and online activity show smaller, lighter dots, indicating only moderate correlations with other features. This pattern emphasizes that while digital engagement and participation matter, the core academic elements—attendance, assignments, and quizzes—form the backbone of student performance prediction.

VII:DISCUSSIONS AND ANALYSIS

1. Interpretation of Results

The study's findings demonstrate that Explainable AI models can accurately predict student performance while providing transparent reasoning for each prediction. Among the tested models, the Neural Network achieved the highest accuracy (93%), reflecting its capability to capture complex, non-linear relationships among features such as attendance, assignment scores, quiz performance, participation, and online activity. SHAP and LIME analyses revealed that attendance is the most influential feature, followed by assignment and quiz scores, which aligns with intuitive expectations: students who actively engage in classes and complete coursework consistently tend to perform better academically. Online activity and class participation, while important, had relatively lower contributions, suggesting that these factors support performance but are not primary determinants. The correlation analysis and scatter plots further confirmed the positive relationship between student engagement and final grades, indicating that maintaining consistent academic habits is critical for success.

2. Comparison with Literature

The findings of this study are consistent with existing research in the domain of educational analytics. Prior studies (e.g., Baker & Siemens, 2014; Romero & Ventura, 2020) have highlighted attendance, assignment performance, and assessment scores as key predictors of

Received: August 01, 2025

Volume 38 No. 12s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

academic outcomes. This study extends the literature by integrating Explainable AI techniques such as SHAP and LIME, which not only predict student outcomes but also provide interpretable insights, a limitation in many traditional predictive models. Unlike black-box models that offer high accuracy without interpretability, this research emphasizes actionable explanations that allow educators to design targeted interventions. Furthermore, the moderate contribution of online activity reflects the growing influence of digital learning platforms, aligning with findings from recent studies on blended and e-learning environments (e.g., Almarashdeh et al., 2021).

3. Theoretical Implications

From a theoretical perspective, this research contributes to the field of learning analytics and educational data scienceby demonstrating how Explainable AI can bridge the gap between predictive modeling and interpretability. The study supports the theoretical framework that student engagement and continuous assessment are primary drivers of academic success, while also validating the utility of XAI in making complex models transparent. By quantifying feature contributions through SHAP values, the research reinforces theories of student- centered learning, where understanding individual behaviors and their impact on performance can guide personalized learning strategies.

4. Practical Implications

Practically, the findings have significant implications for educational institutions. First, educators can use model predictions and feature explanations to identify at-risk students early and design targeted interventions, such as remedial classes, assignment guidance, or attendance monitoring.III:RESULTS

This study demonstrates that Explainable AI (XAI) models can accurately predict student performance while providing interpretable insights into the factors influencing academic outcomes. Among the tested models, the Neural Network achieved the highest predictive accuracy of 93%, followed closely by XGBoost and Random Forest models. SHAP and LIME analyses highlighted that attendance, assignment scores, and quiz performance are the most significant contributors to student grades, while online activity and participation play supporting roles. Correlation and scatter plot analyses further confirmed the positive relationship between consistent engagement and high academic achievement. Overall, the study confirms that integrating predictive modeling with interpretability techniques enables actionable insights for educators and administrators.

IX: REFERENCES

1, M. Hoq, P. Brusilovsky, and B. Akram, "Analysis of an explainable student performance prediction model in an introductory programming course," in Proc. 16th Int. Conf. Educational Data Mining (EDM), Bengaluru, India, Jul. 2023, pp. 79–90. [Online]. Available: https://educationaldatamining.org/EDM2023/proceedings/2023.EDM-long-papers.7/2023.EDM-long-papers.7.pdf

Received: August 01, 2025

Volume 38 No. 12s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

2, F. T. Johora, "An explainable AI-based approach for predicting undergraduate students' performance," Computers in Education, vol. 10, no. 1, pp. 1–14, 2025. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2590005625000116

3. E. Ben George, R. Senthilkumar, F. Al-Junaibi, and Z. Al-Shuaibi, "Explainable AI methods for predicting student grades and performance," Journal of Information Systems and Education, vol. 10, no. 23, pp. 1–15, Mar. 2025. [Online]. Available: https://jisem- journal.com/index.php/journal/article/view/3680

4. J. Mai, F. Wei, W. He, H. Huang, and H. Zhu, "An explainable student performance prediction method based on dual-level progressive classification belief rule base," Electronics, vol. 13, no. 22, p. 4358, 2024. [Online]. Available: https://www.mdpi.com/2079-9292/13/22/4358

5. W. Ahmed, "Machine learning-based academic performance prediction using ensemble models," Scientific Reports, vol. 15, no. 1, p. 12353, 2025. [Online]. Available: https://www.nature.com/articles/s41598-025-12353-4

6. M. J. Gomez, "Utilising explainable AI to enhance real-time student performance prediction in serious games," Expert Systems, vol. 42, no. 5, p. e70008, 2025. [Online]. Available: https://onlinelibrary.wiley.com/doi/10.1111/exsy.70008

7. H. Mastour, T. Dehghani, E. Moradi, and S. Eslami, "Explainable artificial intelligence for predicting medical students' performance in comprehensive assessments," Scientific Reports, vol. 15, no. 1, p. 7460, 2025. [Online]. Available: https://www.nature.com/articles/s41598-025-07460-1

8. L. Liu and R. Dai, "Explainable AI for predicting and understanding mathematics achievement: A cross-national analysis of PISA 2018," arXiv preprint arXiv:2508.16747, 2025. [Online]. Available: https://arxiv.org/abs/2508.16747

9. B. Akter, M. B. Hosen, S. Ahmed, M. Anannya, and M. F. Hossain, "Explainable AI and machine learning for exam-based student evaluation: Causal and predictive analysis of socio-academic and economic factors," arXiv preprint arXiv:2508.00785, 2025. [Online]. Available: https://arxiv.org/abs/2508.00785.

10. V. Swamy, S. Du, M. Marras, and T. Käser, "Trusting the explainers: Teacher validation of explainable artificial intelligence for course design," arXiv preprint arXiv:2212.08955, 2022. [Online]. Available: https://arxiv.org/abs/2212.08955

11. K. Niu, X. Cao, and Y. Yu, "Explainable student performance prediction with personalized attention for explaining why a student fails," arXiv preprint arXiv:2110.08268, 2021. [Online]. Available: https://arxiv.org/abs/2110.08268

12. S. Ghimire, "Explainable artificial intelligence-machine learning models to predict student outcomes," University of the Sunshine Coast, 2024. [Online]. Available: https://research.usc.edu.au/view/pdfCoverPage?download=true&filePid=13282652420002621&instCode=61USC_INST.

Received: August 01, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

13. E. Tiukhova, "Explainable learning analytics: Assessing the stability of predictive models," Computers in Human Behavior, vol. 123, p. 106876, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/abs/pii/S0167923624000629.

14. M. Hoq, P. Brusilovsky, and B. Akram, "Analysis of an explainable student performance prediction model in an introductory programming course," in Proceedings of the 16th International Conference on Educational Data Mining (EDM), Bengaluru, India, Jul. 2023, pp. 79–90. [Online]. Available: https://educationaldatamining.org/EDM2023/proceedings/2023.EDM-long-papers.7/2023.EDM-long-papers.7.pdf

15. F. T. Johora, "An explainable AI-based approach for predicting undergraduate students' performance," Computers in Education, vol. 10, no. 1, pp. 1–14, 2025. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2590005625000116

16. E. Ben George, R. Senthilkumar, F. Al-Junaibi, and Z. Al-Shuaibi, "Explainable AI methods for predicting student grades and performance," Journal of Information Systems and Education, vol. 10, no. 23, pp. 1–15, Mar. 2025. [Online]. Available: https://jisem- journal.com/index.php/journal/article/view/3680.

17. J. Mai, F. Wei, W. He, H. Huang, and H. Zhu, "An explainable student performance prediction method based on dual-level progressive classification belief rule base," Electronics, vol. 13, no. 22, p. 4358, 2024. [Online]. Available: https://www.mdpi.com/2079-9292/13/22/4358

18. W. Ahmed, "Machine learning-based academic performance prediction using ensemble models," Scientific Reports, vol. 15, no. 1, p. 12353, 2025. [Online]. Available: https://www.nature.com/articles/s41598-025-12353-4.

19. M. J. Gomez, "Utilising explainable AI to enhance real-time student performance prediction in serious games," Expert Systems, vol. 42, no. 5, p. e70008, 2025. [Online]. Available: https://onlinelibrary.wiley.com/doi/10.1111/exsy.70008

20. H. Mastour, T. Dehghani, E. Moradi, and S. Eslami, "Explainable artificial intelligence for predicting medical students' performance in comprehensive assessments," Scientific Reports, vol. 15, no. 1, p. 7460, 2025. [Online]. Available: https://www.nature.com/articles/s41598-025- 07460-1.

21. L. Liu and R. Dai, "Explainable AI for predicting and understanding mathematics achievement: A cross-national analysis of PISA 2018," arXiv preprint arXiv:2508.16747, 2025. [Online]. Available: https://arxiv.org/abs/2508.16747.

Received: August 01, 2025