

An Explainable Machine Learning and Internet of Things (IoT) Framework for Academic Performance Prediction and Personalized Learning in Middle School

Deepali Sharma

techdeepsmba@gmail.com

Research Article

Keywords: Internet of Things, Machine Learning, Academic Performance Prediction, Personalized Learning, Explainable Artificial Intelligence, Random Forest, XGBoost, SHAP, LIME, Smart Education

Posted Date: August 25th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10490303/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

An Explainable Machine Learning and Internet of Things (IoT) Framework for Academic Performance Prediction and Personalized Learning in Middle School (July 2026)

Deepali Sharma Ph.D. Scholar, Department of Computer Applications, Graphic Era Deemed to be University,
Guide: **Dr. Manish Sharma**.

I. Abstract

The rapid advancement of the Internet of Things (IoT) and Machine Learning (ML) has created new opportunities for developing intelligent educational systems that support academic performance prediction and personalized learning. Traditional prediction models primarily rely on historical academic records, often overlooking real-time behavioural and environmental data collected within smart classrooms. IoT technologies, including RFID attendance systems, wearable devices, environmental sensors, and Learning Management Systems (LMS), continuously generate valuable educational data that can provide deeper insights into students' learning behaviour. However, many Machine Learning models function as black-box systems, limiting teachers' ability to interpret prediction outcomes and reducing trust in AI-assisted educational decision-making.

This research proposes an **Explainable Machine Learning and IoT Framework** for predicting middle school students' academic performance and generating personalized learning recommendations. The framework integrates IoT-generated educational data with ensemble Machine Learning algorithms, namely Random Forest and Extreme Gradient Boosting (XGBoost), to achieve accurate performance prediction. Explainable Artificial Intelligence (XAI) techniques, including SHAP and LIME, are incorporated to provide transparent explanations of prediction results and identify the factors influencing student performance. The proposed framework enables early identification of academically at-risk students while supporting teachers in designing personalized interventions based on real-time learning analytics. By combining IoT, Machine Learning, and Explainable AI, the framework aims to improve prediction accuracy, enhance

transparency, and facilitate evidence-based educational decision-making in smart learning environments.

Keywords: Internet of Things, Machine Learning, Academic Performance Prediction, Personalized Learning, Explainable Artificial Intelligence, Random Forest, XGBoost, SHAP, LIME, Smart Education

II. 1. Introduction

The integration of **Machine Learning (ML)** and the **Internet of Things (IoT)** has significantly transformed educational environments by enabling intelligent data collection, predictive analytics, and personalized learning. Smart classrooms equipped with IoT devices continuously collect information related to student attendance, classroom participation, environmental conditions, learning behaviour, and digital interactions. These technologies provide educators with real-time insights into students' academic progress and create opportunities for data-driven educational decision-making.

Traditional academic performance prediction systems mainly rely on examination scores, attendance records, and assignment marks. Although these variables provide useful information, they do not capture students' real-time engagement, classroom behaviour, or environmental factors that influence learning outcomes. IoT technologies address this limitation by collecting multimodal educational data from sensors, wearable devices, RFID attendance systems, and Learning Management Systems (LMS). When combined with Machine Learning algorithms, these data can significantly improve prediction accuracy and support early identification of academically at-risk students.

Despite the effectiveness of Machine Learning algorithms, many predictive models operate as **black-box systems**, making it difficult for teachers to understand how prediction outcomes are generated. Explainable Artificial Intelligence (XAI)

Manuscript submitted for review on July 25, 2026.

This research was carried out as part of the Ph.D. programme in the Department of Computer Applications, Graphic Era Deemed to be University, Dehradun, Uttarakhand, India. This work did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Deepali Sharma is a Ph.D. Scholar with the Department of Computer Applications, Graphic Era Deemed to be University, Dehradun, Uttarakhand 248002, India (e-mail: deepalisharma.phd@gehu.ac.in). *Corresponding author:* **Deepali Sharma** (e-mail: techdeepsmba@gmail.com).

addresses this challenge by providing transparent explanations that improve model interpretability and increase teacher confidence in AI-assisted educational systems.

This research proposes an **Explainable Machine Learning and IoT Framework** that integrates smart classroom data with ensemble Machine Learning models and Explainable AI techniques to predict academic performance and generate personalized learning recommendations for middle school students.

III. 1.1 Internet of Things in Education

The Internet of Things has become a key technology in smart education by enabling interconnected devices to collect, exchange, and analyze educational data. IoT devices such as RFID attendance systems, wearable sensors, environmental monitoring systems, and smart boards provide continuous information about student behaviour and classroom activities. These technologies enable educators to monitor learning environments in real time and support intelligent educational management.

IoT-based educational systems improve classroom efficiency, enhance student engagement, and facilitate early detection of learning difficulties. By providing continuous data collection, IoT enables educational institutions to move from traditional record-based evaluation to real-time learning analytics.

IV. 1.2 Machine Learning for Academic Performance Prediction

Machine Learning enables computers to identify patterns within educational datasets and predict student academic outcomes without explicit programming. Educational institutions generate large amounts of data related to attendance, assessment scores, homework completion, classroom participation, and LMS activities. Machine Learning algorithms analyze these data to identify students who may require additional academic support.

Among various Machine Learning techniques, **Random Forest** and **Extreme Gradient Boosting (XGBoost)** have demonstrated excellent predictive performance due to their ability to handle complex educational datasets and reduce overfitting. These ensemble learning algorithms provide reliable predictions and support evidence-based educational decision-making.

V. 1.2.1 Explainable Artificial Intelligence

Although Machine Learning models provide accurate predictions, they often lack transparency. Explainable Artificial Intelligence (XAI) improves model interpretability by explaining how different input features contribute to prediction outcomes. **SHAP** identifies the importance of individual features, while **LIME** explains predictions for specific students. These techniques enable teachers to understand the reasons behind prediction results, improve trust in AI systems, and support personalized learning interventions.

VI. 2. Literature Review

The integration of Machine Learning (ML) and the Internet of Things (IoT) has attracted considerable attention in educational research due to its potential to enhance learning analytics and academic performance prediction. Educational institutions are increasingly adopting smart classroom technologies that

generate large volumes of real-time data through connected devices, enabling continuous monitoring of student learning behaviour. These developments have created opportunities for developing intelligent educational systems that support data-driven teaching and personalized learning.

Machine Learning techniques have been widely applied for predicting students' academic performance using structured educational datasets. Algorithms such as Decision Trees, Naïve Bayes, Support Vector Machines (SVM), Artificial Neural Networks (ANN), Random Forest, and Extreme Gradient Boosting (XGBoost) have demonstrated promising prediction accuracy. Among these methods, ensemble learning algorithms such as Random Forest and XGBoost consistently outperform traditional classifiers because they effectively manage high-dimensional educational data, reduce overfitting, and improve model generalization.

The Internet of Things has further expanded the scope of educational analytics by enabling continuous collection of real-time data from smart classroom environments. IoT devices, including RFID attendance systems, wearable sensors, smart boards, environmental monitoring systems, and Learning Management Systems (LMS), provide valuable information regarding student attendance, classroom participation, environmental conditions, and learning engagement. These real-time behavioural indicators complement traditional academic records and improve the understanding of factors influencing student performance.

Recent advances in Explainable Artificial Intelligence (XAI) have addressed one of the major limitations of Machine Learning models—their lack of interpretability. Techniques such as **SHAP (SHapley Additive exPlanations)** and **LIME (Local Interpretable Model-Agnostic Explanations)** provide transparent explanations by identifying the contribution of individual features to prediction outcomes. Explainability enables teachers to understand the reasoning behind AI-generated predictions, increasing confidence in educational decision-making and facilitating personalized interventions.

Although previous studies have independently explored Machine Learning, IoT, and Explainable AI, only a limited number of studies have integrated these technologies into a unified framework for academic performance prediction and personalized learning in middle school education. This research addresses this gap by proposing an explainable smart learning framework that combines IoT-generated educational data with Machine Learning and Explainable AI techniques.

VII. 3. Research Gap

The review of existing literature identifies several research gaps that motivate the proposed study:

- Most academic performance prediction models rely primarily on historical academic records and rarely utilize real-time IoT-generated educational data.
- Existing IoT-based educational systems mainly focus on attendance monitoring and classroom automation rather than predictive learning analytics.
- High-performing Machine Learning models often function as black-box systems, making it difficult for teachers to interpret prediction outcomes.

- Limited research integrates Machine Learning, IoT, and Explainable Artificial Intelligence within a single educational framework.
- Existing studies provide limited support for generating personalized learning recommendations based on real-time student behaviour and academic performance.

These limitations highlight the need for an integrated and explainable framework capable of combining IoT technologies, Machine Learning, and Explainable AI to improve academic performance prediction and support personalized learning.

VIII. 4. Problem Statement

Educational institutions increasingly collect large amounts of data from smart classroom environments; however, much of this information remains underutilized for academic performance prediction. Conventional prediction systems mainly depend on examination scores, attendance records, and assignment marks, while ignoring valuable real-time information generated by IoT devices such as classroom sensors, wearable devices, RFID systems, and Learning Management Systems. As a result, existing models often fail to provide a comprehensive understanding of student learning behaviour.

Although Machine Learning algorithms can accurately predict academic performance, their lack of transparency limits their practical adoption in educational settings. Teachers require interpretable prediction models that clearly explain the factors influencing student performance and provide meaningful recommendations for improving learning outcomes.

Therefore, there is a need for an explainable educational framework that integrates IoT-generated educational data with Machine Learning and Explainable Artificial Intelligence. Such a framework can improve prediction accuracy, enhance model transparency, support early identification of academically at-risk students, and enable personalized learning interventions for middle school education.

IX. 5. Research Objectives

The primary objective of this research is to develop an **Explainable Machine Learning and Internet of Things (IoT) Framework** for predicting academic performance and generating personalized learning recommendations for middle school students.

The specific objectives are:

- To develop an integrated IoT and Machine Learning framework for academic performance prediction.
- To collect real-time educational data using IoT-enabled smart classroom technologies.
- To identify significant academic and behavioural factors affecting student performance.
- To compare the performance of Random Forest and XGBoost algorithms for academic prediction.
- To integrate Explainable Artificial Intelligence techniques (SHAP and LIME) for improving model transparency.
- To generate personalized learning recommendations based on prediction outcomes.
- To support teachers in evidence-based educational decision-making.

X. 6. Research Methodology

The proposed study adopts a **quantitative research methodology** based on the **Design Science Research (DSR)** approach. The objective is to design, develop, and evaluate an Explainable Machine Learning and IoT framework for predicting middle school students' academic performance.

XI. 6.1 Data Collection

Educational data are collected from multiple smart classroom sources, including:

- RFID Attendance System
- Learning Management System (LMS)
- Assignment and Homework Records
- Internal Assessment Scores
- Previous Examination Results
- Smart Classroom Sensors
- Wearable Devices
- Classroom Participation Logs
- Environmental Sensors (Temperature, Light, Noise)
- Teacher Feedback

These data represent students' academic progress, behavioural engagement, and classroom environment.

XII. 6.2 Data Preprocessing

The collected data undergo preprocessing to improve consistency and quality.

The preprocessing process includes:

- Missing value handling
- Duplicate record removal
- Data normalization
- Feature scaling
- Categorical encoding
- Outlier detection
- IoT data synchronization

These steps improve dataset quality before Machine Learning model development.

XIII. 6.3 Feature Engineering

Feature engineering identifies the most influential variables affecting academic performance.

Important features include:

- Attendance Percentage
- Assignment Completion Rate
- Internal Assessment Scores
- LMS Usage Frequency
- Classroom Participation
- Smart Device Interaction
- Environmental Conditions
- Homework Submission
- Student Engagement Index

The selected features are combined into a unified educational dataset.

XIV. 6.4 Machine Learning Models

The proposed framework employs two ensemble Machine Learning algorithms.

XV. Random Forest

Random Forest constructs multiple decision trees using randomly selected subsets of training data. The final prediction

is obtained through majority voting, improving robustness and reducing overfitting. It also estimates feature importance for educational variables.

XVI. Extreme Gradient Boosting (XGBoost)

XGBoost is a gradient boosting algorithm that sequentially builds decision trees by correcting previous prediction errors. It provides high prediction accuracy, computational efficiency, and excellent performance on heterogeneous educational datasets.

XVII. 6.5 Explainable Artificial Intelligence

To improve model transparency, the framework integrates two Explainable AI techniques.

XVIII. SHAP (SHapley Additive exPlanations)

SHAP explains the contribution of each input feature toward prediction outcomes by providing both global and local explanations. It enables teachers to identify the factors that most strongly influence student performance.

XIX. LIME (Local Interpretable Model-Agnostic Explanations)

LIME explains individual student predictions by approximating the Machine Learning model with a simple local model. This enables teachers to understand why a particular student has been classified as academically at risk or high performing.

XX. 6.6 Performance Evaluation

The developed models are evaluated using the following performance metrics:

- Accuracy
- Precision
- Recall
- F1-Score
- ROC-AUC
- Confusion Matrix

These metrics enable objective comparison of Random Forest and XGBoost performance.

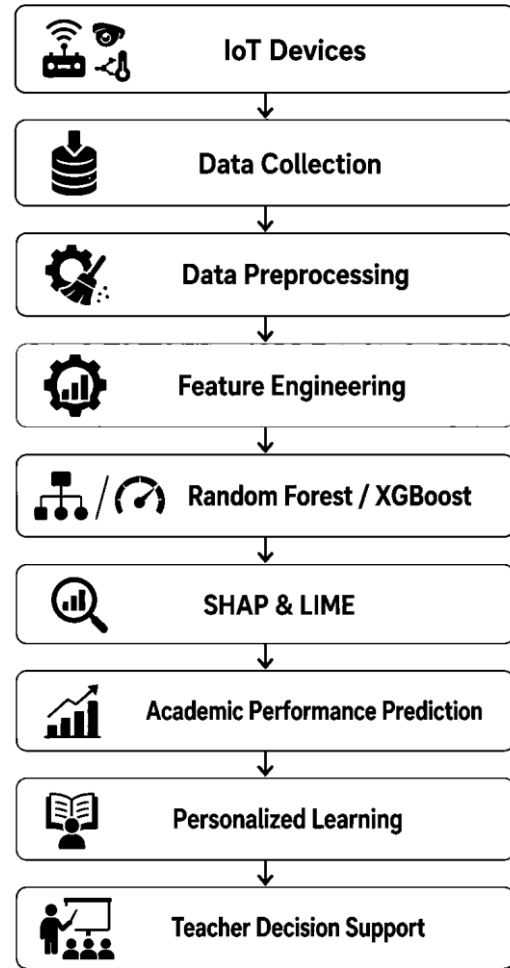
XXI. 7. Proposed Framework

The proposed framework integrates Internet of Things devices, Machine Learning algorithms, and Explainable Artificial Intelligence to create an intelligent educational decision-support system.

Initially, educational data are collected through IoT-enabled devices such as RFID attendance systems, wearable sensors, Learning Management Systems, classroom sensors, and academic records. The collected data undergo preprocessing and feature engineering to improve quality and identify the most significant variables influencing student performance.

The processed dataset is used to train Random Forest and XGBoost models for predicting academic performance. To enhance transparency, SHAP and LIME generate explanations for both global model behaviour and individual student predictions. Based on these explanations, the system provides personalized learning recommendations that support teachers in designing appropriate interventions for academically at-risk students.

XXII. Workflow



XXIII. 8. Expected Results

The proposed Explainable Machine Learning and IoT framework is expected to improve academic performance prediction by combining traditional educational records with real-time IoT-generated behavioural and environmental data.

The expected outcomes include:

- Improved prediction accuracy using ensemble Machine Learning models.
- Early identification of academically at-risk students.
- Increased transparency through SHAP and LIME explanations.
- Better understanding of factors affecting academic performance.
- Personalized learning recommendations based on real-time educational data.
- Enhanced teacher confidence in AI-assisted educational decision-making.

The proposed framework is also expected to contribute to the development of intelligent smart classroom systems capable of supporting evidence-based teaching practices.

XXIV. 9. Conclusion

This study proposed an **Explainable Machine Learning and Internet of Things (IoT) Framework** for academic

performance prediction and personalized learning in middle school. The framework integrates IoT-enabled smart classroom technologies with Machine Learning algorithms and Explainable Artificial Intelligence to improve educational analytics.

Unlike traditional prediction systems that rely solely on historical academic records, the proposed framework utilizes real-time behavioural and environmental information collected through IoT devices. The incorporation of SHAP and LIME enhances transparency by explaining prediction outcomes and supporting teacher trust in AI-assisted educational systems.

Future research may validate the proposed framework using real classroom datasets, integrate deep learning models for multimodal educational analytics, and deploy the system within smart classroom environments to support real-time personalized learning.

XXV. References (IEEE Style)

- [1] L. Breiman, "Random Forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.
 - [2] T. Chen and C. Guestrin, "XGBoost: A Scalable Tree Boosting System," *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2016.
 - [3] S. M. Lundberg and S. I. Lee, "A Unified Approach to Interpreting Model Predictions," *Advances in Neural Information Processing Systems*, 2017.
 - [4] M. T. Ribeiro, S. Singh, and C. Guestrin, "Why Should I Trust You? Explaining the Predictions of Any Classifier," *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2016.
 - [5] L. Atzori, A. Iera, and G. Morabito, "The Internet of Things: A Survey," *Computer Networks*, vol. 54, no. 15, pp. 2787–2805, 2010.
 - [6] D. Evans, "The Internet of Things: How the Next Evolution of the Internet Is Changing Everything," Cisco IBSG White Paper, 2011.
 - [7] UNESCO, *Artificial Intelligence and Education: Guidance for Policy Makers*, UNESCO Publishing, Paris, 2021.
 - [8] OECD, *Digital Education Outlook 2023*, Organisation for Economic Co-operation and Development, Paris, 2023.
-

Declarations

Clinical Trial Registration

Not applicable. This study did not involve a clinical trial.

Ethics Declaration

Not applicable.

Consent to Participate

Not applicable.

Consent to Publish

Not applicable.

Conflict of Interest

The authors declare that they have no competing interests.

Funding

No funding was received for conducting this study.

Author Contributions

All authors contributed to the study conception and design. Material preparation, data analysis, and manuscript writing were performed by the authors. All authors read and approved the final manuscript.

Data Availability

Data sharing is not applicable to this article as no datasets were generated or analyzed during the current study.