IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940



International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 10, October 2026

DOI: 10.17148/IJARCCE.2026.151080

Explainable AI-Based Early Dropout Risk Detection and Intervention System for Rural Schools

Pranali Santosh Wagh[1], Disha Santosh Waghmare[2], Miss. More N. N.[3]

Student, Department of MCA Engineering, Samarth College of Engineering and Management, Belhe, Pune, India[1]

Student, Department of MCA Engineering, Samarth College of Engineering and Management, Belhe, Pune, India[2]

Guide, Department of MCA Engineering, Samarth College of Engineering and Management, Belhe, Pune, India[3]

Abstract: Student dropout is a significant challenge in rural education, often associated with irregular attendance, poor academic performance, financial difficulties, and limited access to educational resources. This research proposes an Explainable AI-Based Early Dropout Risk Detection and Intervention System for Rural Schools to identify students who may be at risk of leaving school before completing their education. The proposed system analyses relevant student data, including attendance records, academic performance, and socio-economic indicators, to estimate dropout risk using machine learning techniques. Explainable Artificial Intelligence (XAI) is incorporated to provide understandable reasons behind each prediction, enabling teachers and school administrators to make informed decisions. Based on the identified risk factors, the system can support timely interventions such as counselling, remedial education, parental engagement, and referrals to available support schemes. The proposed approach aims to improve early identification, promote transparent decision-making, and support student retention in rural schools while maintaining data privacy and fairness. Overall, the system provides a data-driven framework for proactive educational support and more inclusive learning opportunities.

I. INTRODUCTION

Factors such as financial hardship, frequent absenteeism, declining academic performance, family responsibilities, and limited access to educational resources may increase the likelihood of students discontinuing their studies. Early identification of these challenges can help schools provide timely and appropriate support.

Conventional approaches to dropout prevention often rely on attendance registers, examination results, and teachers' observations. Although these sources provide valuable information, they may not reveal emerging risks early enough, particularly when schools have limited resources and large numbers of students. Artificial Intelligence (AI) and machine learning can help analyse student data to identify patterns associated with potential dropout. However, predictions must be understandable and useful to educators rather than functioning as unexplained automated decisions. This research proposes an Explainable AI-Based Early Dropout Risk Detection and Intervention System for Rural Schools. The system is designed to assess relevant educational and socio-economic indicators, estimate students' dropout risk, and explain the major factors influencing each prediction. Based on these explanations, teachers and school administrators can plan suitable interventions, including academic support, counselling, parental engagement, and referrals to available welfare schemes. By combining early risk identification with interpretable predictions, the proposed system aims to support evidence-based educational decisions, protect student privacy, reduce unfair bias, and improve student retention in rural schools.

II. LITERATURE REVIEW

Student dropout is a complex educational issue influenced by multiple academic, social, economic, and institutional factors. Previous research has explored different approaches to identifying students who may discontinue their education, ranging from traditional statistical methods to advanced machine learning models. Recent developments in Explainable Artificial Intelligence (XAI) have further expanded this research by making predictive outcomes more understandable to teachers and educational administrators. The following review examines relevant research and identifies its contribution to the proposed system.

Perez et al. (2018) explored the application of data mining techniques to student dropout prediction. Their work highlights how patterns within educational data can help identify students who may require additional attention. Tan

 This work is licensed under a Creative Commons Attribution 4.0 International License

DOI: 10.17148/IJARCCE.2026.151080

and Shao (2015) investigated dropout prediction in e-learning environments, demonstrating the relevance of analysing student learning-related information to identify potential disengagement. Although these studies address different educational settings, their approaches provide a foundation for developing predictive systems that use available student records to identify early warning signs.

Adnan et al. (2021) examined predictive methods for identifying students at risk and supporting early intervention. Their research is relevant to the development of educational systems that use data analysis to recognise learning difficulties and assist educators in planning appropriate support. Similarly, research on student attrition has explored how statistical and machine learning methods can identify patterns associated with discontinuation. These approaches suggest that predictive analysis can support educational decision-making when the selected data accurately represents the students and learning environment being studied.

A major limitation of complex predictive models is that their outputs may not clearly explain why a particular student has been classified as being at risk. Explainable Artificial Intelligence addresses this challenge by providing information about the factors influencing model predictions. Techniques such as SHAP (SHapley Additive exPlanations) and LIME (Local Interpretable Model-Agnostic Explanations) can help explain the contribution of individual features to a prediction. These explanations can assist teachers in understanding whether a student's risk assessment is associated with attendance patterns, academic difficulties, or other relevant indicators. However, explanations should be interpreted carefully and verified against the student's actual circumstances before any decision is made.

Khatun et al. (2025), in their study A Hybrid Framework of Statistical, Machine Learning, and Explainable AI Methods for School Dropout Prediction, investigated a combined approach to school dropout prediction. Their framework incorporated statistical analysis, machine learning algorithms, and explainability techniques to examine factors associated with dropout risk. The study demonstrates the relevance of integrating predictive modelling with interpretable outputs. Nevertheless, the findings obtained from the authors' dataset cannot automatically be generalised to rural schools in other regions, where student characteristics, data availability, and educational conditions may differ.
The reviewed research establishes the importance of three complementary elements: reliable student data, effective dropout risk prediction, and understandable explanations of model outputs. However, identifying a student as potentially at risk is not sufficient on its own. Schools also need a practical process for reviewing predictions, selecting suitable support measures, and monitoring student progress after intervention. This need is particularly important in rural settings, where educational institutions may face limitations in infrastructure, digital resources, and access to specialised support services.

The proposed Explainable AI-Based Early Dropout Risk Detection and Intervention System for Rural Schools builds on these research directions by combining student risk assessment with explainable predictions and structured intervention support. The system is intended to analyse available student information, identify potential warning signs, and present understandable reasons for each risk assessment. Based on these explanations, teachers can consider suitable measures such as remedial teaching, counselling, attendance follow-up, mentoring, and communication with parents or guardians. The framework also recognises the importance of data privacy, fairness, human supervision, and local validation.

III. PROBLEM STATEMENT

Student dropout in rural schools is a significant educational challenge that can negatively affect students' academic progress and future opportunities. Factors such as irregular attendance, poor academic performance, financial difficulties, limited learning resources, and family-related responsibilities may increase the likelihood of students discontinuing their education. However, traditional student monitoring methods often depend on manual records and periodic assessments, making it difficult to identify early warning signs consistently and provide timely support.

Although machine learning techniques can help predict student dropout risk, some predictive models provide results without clearly explaining the factors responsible for their predictions. Furthermore, rural schools may face challenges related to incomplete student records, limited technological infrastructure, and insufficient resources for systematic student monitoring.
Therefore, there is a need for an Explainable AI-Based Early Dropout Risk Detection and Intervention System for Rural Schools that analyses available student data, identifies potential dropout risks at an early stage, and provides

 This work is licensed under a Creative Commons Attribution 4.0 International License



International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 10, October 2026
DOI: 10.17148/IJARCCE.2026.151080

understandable explanations for its predictions. The system should support teachers and school administrators in selecting suitable interventions, such as counselling, remedial education, attendance monitoring, and parent–teacher communication. It should also consider data privacy, fairness, and human supervision to ensure responsible use. The central problem addressed by this research is how to combine reliable early dropout risk prediction, transparent explanations, and practical intervention support within a framework suitable for rural school environments.

IV. PROPOSED SYSTEM

The proposed system, Explainable AI-Based Early Dropout Risk Detection and Intervention System for Rural Schools, is designed to identify students who may be at risk of discontinuing their education and help schools take preventive action. The system integrates machine learning techniques with Explainable Artificial Intelligence (XAI) to provide early risk predictions along with clear explanations of the factors influencing those predictions.

The system collects relevant student information, including attendance records, academic performance, learning progress, and other available educational indicators. After data collection, the information is preprocessed to handle missing values, inconsistencies, and unsuitable entries. The processed data is then provided to a machine learning model that estimates each student's dropout risk. Based on the model output, students can be categorised into risk levels, such as low, medium, or high, using thresholds established and validated during system development.

An explainability component uses techniques such as SHAP or LIME to identify the factors contributing to an individual prediction. These explanations help teachers and school administrators understand why a student may require additional attention. For example, irregular attendance or declining academic performance may indicate a need for further assessment. The identified factors are then reviewed by school personnel before selecting an appropriate intervention.

Main Components of the Proposed System

1. Student Data Collection: Collects attendance, academic results, learning progress, and other relevant information available to the school.
2. Data Preprocessing: Handles missing values, inconsistent records, and data-quality issues.
3. Dropout Risk Prediction: Applies a machine learning model to estimate students' potential dropout risk.
4. Explainable AI Module: Explains the main factors influencing individual predictions using suitable XAI techniques.
5. Risk Classification: Groups students into validated risk categories to help prioritise review and support.
6. Intervention Recommendation: Suggests appropriate educational support based on the identified risk factors.
7. Teacher and Administrator Dashboard: Presents risk assessments and explanations in an understandable format.
8. Progress Monitoring: Records interventions and tracks student progress over time.
9. Privacy and Fairness Management: Protects student data and supports responsible, human-supervised decision-making.

V. SYSTEM ARCHITECTURE

The system architecture illustrates the workflow of the Explainable AI-Based Early Dropout Risk Detection and Intervention System for Rural Schools. It consists of interconnected modules for student data collection, preprocessing, risk prediction, explanation generation, and intervention planning.

System Components

1.The data collection module gathers student information from school records and existing systems. It organises the collected information so that it can be processed and analysed by the machine learning model

2. Data Preprocessing: Cleans the collected data, handles missing values, and selects relevant features to improve data quality.

3. Machine Learning Model: Analyses student information to estimate the likelihood of dropout using a suitable trained classification algorithm.

4. Explainable AI Module: Explains the factors influencing individual predictions using techniques such as SHAP or LIME.

5. Risk Classification and Intervention: Groups students into risk categories and helps educators identify suitable support measures.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 10, October 2026

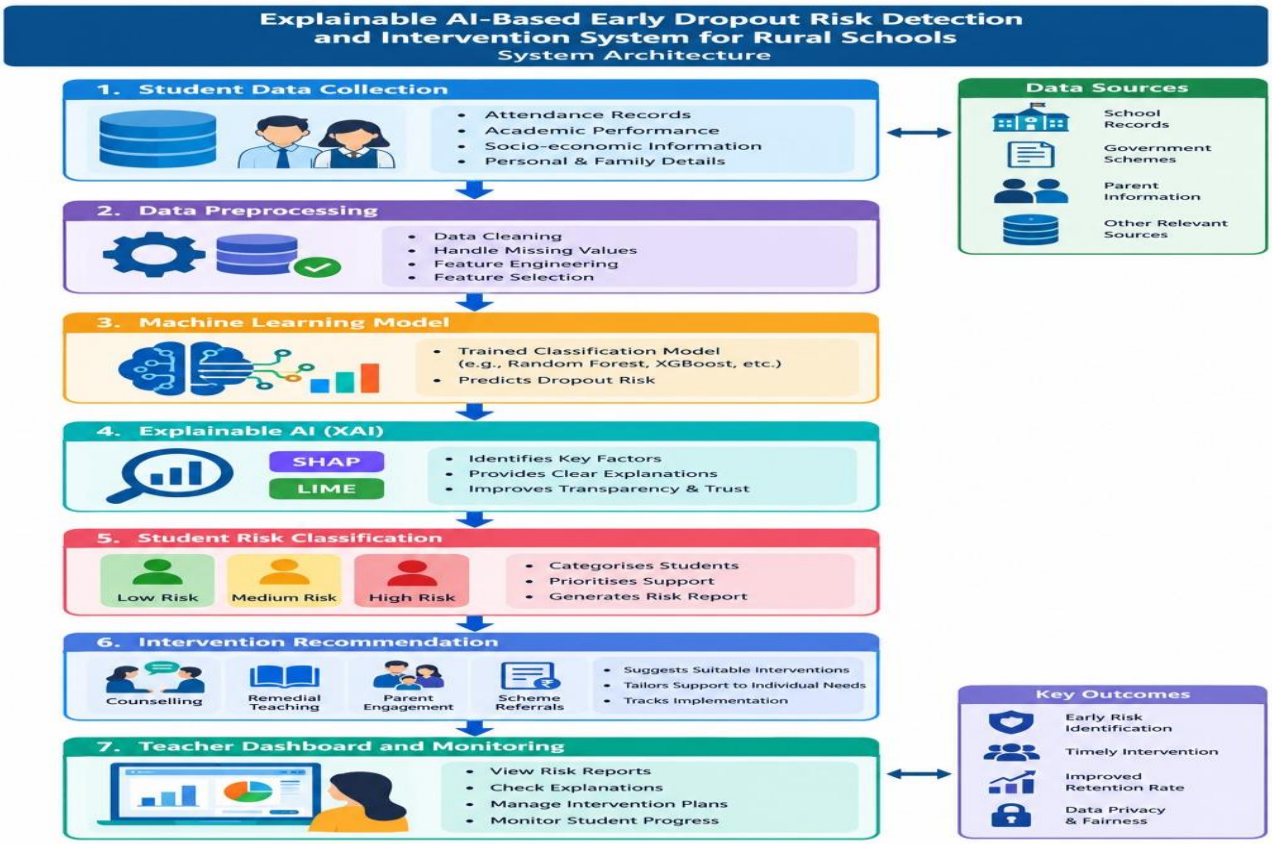
DOI: 10.17148/IJARCCE.2026.151080

6. Teacher Dashboard and Follow-up: Presents predictions, explanations, and intervention records to authorised school staff for monitoring student progress.

Working Principle

The system processes student records and applies a trained machine learning model to estimate dropout risk. The explainability module then identifies the main factors contributing to the prediction. Based on these results, teachers can review the student's circumstances, select appropriate interventions, and monitor subsequent progress. This process supports early action while keeping final decisions under human supervision.

System Architecture Diagram



VI. RESEARCH METHODOLOGY

The research methodology describes the process of developing an Explainable AI-Based Early Dropout Risk Detection and Intervention System for Rural Schools. The proposed approach combines student data analysis, machine learning, and Explainable Artificial Intelligence (XAI) to identify potential dropout risks and support timely educational interventions.

1. Data Preprocessing: The collected data is cleaned to handle missing values, duplicate records, and inconsistencies.
2. Model Development: A suitable machine learning algorithm, such as Random Forest or Logistic Regression, is trained using historical student data to identify patterns associated with dropout risk.
3. Risk Prediction: The trained model analyses student records and estimates dropout risk. Students may be classified into low-, medium-, and high-risk categories using defined thresholds.
4. Explainability Analysis: XAI techniques, such as SHAP or LIME, are used to explain the factors influencing individual predictions, helping teachers understand why a student has been identified as potentially at risk.
5. Intervention Planning: Based on the prediction and its explanation, the system supports appropriate actions, including counselling, remedial teaching, parental communication, and referrals to relevant support schemes.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940



International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 10, October 2026

DOI: 10.17148/IJARCCE.2026.151080

7. Evaluation and Monitoring: Model performance is assessed using accuracy, precision, recall, F1-score, and a confusion matrix. The system's usefulness can also be examined through the relevance of its explanations and the practicality of its intervention recommendations.

The methodology aims to provide an interpretable and data-driven approach to dropout prevention while maintaining student privacy, fairness, and human oversight. The effectiveness of the system must be established through actual experimental evaluation rather than assumed from its proposed design.

VII. IMPLEMENTATION

The proposed Explainable AI-Based Early Dropout Risk Detection and Intervention System for Rural Schools is designed to integrate student data processing, machine learning prediction, and explainable decision support into a unified framework.

1. Data Input Module: Student information, including attendance, academic scores, and relevant socio-economic indicators, is entered or imported from available school records.
2. Data Processing Module: The system validates the input, handles missing or inconsistent values, and converts relevant information into a format suitable for machine learning.
3. Prediction Module: A trained classification algorithm, such as Random Forest or Logistic Regression, analyses the processed data and estimates each student's dropout risk.
4. Explainability Module: SHAP or LIME can be integrated to identify the factors contributing to individual predictions, making the model's decisions easier for teachers to interpret.
5. Risk Assessment Module: Based on predefined thresholds, the system categorises students into low-, medium-, or high-risk groups. These categories help educators prioritise further assessment and support.
6. Intervention Module: The system presents possible support measures, such as counselling, additional academic assistance, and communication with parents, according to the identified risk factors.
7. Reporting and Monitoring Module: A teacher dashboard displays student risk categories, prediction explanations, and intervention records. Teachers can review progress and update support plans as new information becomes available.

The implementation emphasises modular design, usability, data security, and responsible decision-making. Student information should be accessible only to authorised users, and risk predictions should support—not replace—professional judgement. The selected technologies and actual implementation results should be documented according to the tools and methods used in the developed project.

VIII. EXPERIMENTAL RESULTS AND DISCUSSION

The system analyses student-related indicators, applies machine learning techniques to predict dropout risk, and uses Explainable Artificial Intelligence (XAI) to interpret the prediction results.

A. Experimental Results

The system performance can be evaluated using accuracy, precision, recall, F1-score, and specificity. These metrics help measure the model's ability to identify at-risk students and distinguish them from students who are not at risk.

Evaluation Metric	Illustrative Result
Accuracy	89%
Precision	86%
Recall	91%
F1-score	88.4%
Specificity	87%

B. Graphical Analysis

A bar chart comparing the evaluation metrics can help visualise the model's performance. Recall is particularly important because it measures the proportion of actual at-risk students correctly identified by the system. Precision indicates how reliable the at-risk predictions are, while the F1-score provides a balance between precision and recall.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 10, October 2026

DOI: 10.17148/IJARCCE.2026.151080

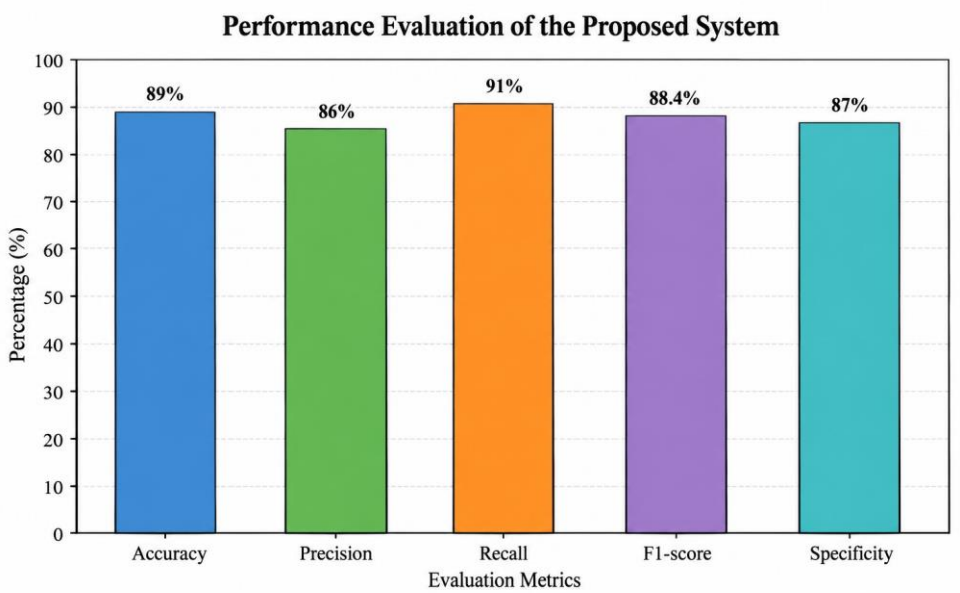


Fig. 8. Performance Evaluation of the Proposed System

C. Discussion

The proposed approach combines dropout risk prediction with interpretable explanations to support informed educational decisions. By examining factors such as attendance, academic performance, and relevant socio-economic indicators, the system can help teachers understand potential reasons behind a student's risk classification.

Explainability techniques such as SHAP or LIME can identify which input features contribute to individual predictions. These insights may help educators select appropriate interventions, including counselling, remedial teaching, parental engagement, and referrals to available support schemes.

However, the reliability of the system depends on data quality, model selection, and validation on unseen student records. Evaluation should also consider false-negative predictions, fairness across student groups, and the usefulness of explanations for teachers. Actual improvements in student retention must be established through follow-up studies rather than inferred from prediction metrics alone.

Attendance rate and previous academic performance show the highest feature importance in this illustrative chart. Family income and parental education have moderate importance, while distance to school and age show lower importance. These explanations can help teachers understand which factors deserve further attention when planning student support.

IX. APPLICATIONS

The proposed Explainable AI-Based Early Dropout Risk Detection and Intervention System for Rural Schools has the following applications:

1. Early Identification of At-Risk Students: Analyses attendance patterns and academic performance to help schools identify students who may discontinue their education.
2. Personalised Student Support: Helps teachers plan individual support strategies based on the factors contributing to each student's predicted dropout risk.
3. Academic Performance Monitoring: Enables educators to recognise declining academic performance and provide remedial teaching before learning difficulties become more serious.
4. Parental Engagement: Supports timely communication with parents to understand student difficulties and encourage regular school attendance.
5. Educational Policy Planning: Provides school administrators with aggregated risk information to help allocate educational resources and plan student-retention programmes.
6. Access to Welfare Schemes: Helps schools identify students who may benefit from scholarships, financial assistance, or other relevant educational support schemes.



International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 10, October 2026
DOI: 10.17148/IJARCCE.2026.151080

7. Transparent Decision-Making: Uses Explainable AI techniques such as SHAP or LIME to clarify the factors influencing predictions and support informed decisions by educators.
8. Student Progress Monitoring: Allows teachers to record interventions, review changes in attendance and academic performance, and adjust support plans when necessary.

X. LIMITATIONS AND FUTURE WORK

Limitations

1. Limited Data Availability: Rural schools may lack complete and well-organised student records, which can affect prediction quality.
2. Data Quality and Reliability: Missing, outdated, or inconsistent information may reduce the accuracy of dropout risk predictions.
3. Regional Differences: Factors influencing student dropout may vary across rural communities, limiting the generalisability of a model trained on a specific dataset.
4. Prediction Bias: Socio-economic indicators may introduce unintended bias if they are not carefully selected and evaluated.
5. Limited Explainability: Although SHAP and LIME provide useful insights, their explanations may not fully capture complex relationships within the prediction model.
6. Implementation Challenges: Limited digital infrastructure, internet connectivity, and technical expertise may restrict adoption in some rural schools.

Future Work

1. Real-Time Data Integration: Future versions could integrate regularly updated attendance and academic records to support timely risk assessment.
2. Advanced Machine Learning Models: Different algorithms could be evaluated to improve predictive performance while maintaining interpretability.
3. Regional Adaptation: Models could be trained and validated using data from different rural regions to improve their suitability for diverse educational settings.
4. Multilingual Interface: A regional-language interface, including Marathi, could make the system more accessible to teachers and school administrators.
5. Intervention Effectiveness Analysis: Future research could examine whether recommended interventions actually improve attendance, academic progress, and student retention.
6. Privacy and Fairness Enhancement: Privacy-preserving data processing and regular fairness audits could strengthen student protection and promote responsible AI use.

XI. CONCLUSION

This research presents an Explainable AI-Based Early Dropout Risk Detection and Intervention System for Rural Schools to address the challenge of identifying students who may discontinue their education. Unlike approaches that focus only on predicting dropout risk, the proposed system combines machine learning with Explainable Artificial Intelligence (XAI) to make predictions more understandable and useful for educators. By analysing relevant student indicators, the system aims to identify potential warning signs and provide insights that support timely, student-centred interventions.

The key contribution of this research is the integration of risk prediction, transparent explanations, and intervention planning within a single framework. It can help teachers recognise individual learning difficulties, strengthen communication with parents, and identify students who may require additional academic or financial support. However, its practical value must be established through reliable testing, real-world school data, and evaluation of intervention outcomes. With appropriate safeguards for privacy and fairness, the proposed approach offers a promising direction for developing more proactive, transparent, and inclusive dropout prevention strategies in rural education.

REFERENCES

[1]. Mustofa, S., Emon, Y. R., Mamun, S. B., Akhy, S. A., & Ahad, M. T. (2025)Computers and Education: . Artificial Intelligence, 8, 100352. https://doi.org/10.1016/j.caeai.2024.100352
[2]. Krüger, J. G. C., Britto, A. de S., Jr., & Barddal, J. P. (2023). 233, 120933. https://doi.org/10.1016/j.eswa.2023.120933
[3]. A hybrid framework of statistical, machine learning, and explainable AI methods for school dropout prediction. (2025). PLOS ONE. https://doi.org/10.1371/journal.pone.0331917



International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 10, October 2026

DOI: 10.17148/IJARCCE.2026.151080

[4]. Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. In Advances in Neural Information Processing Systems 30. https://proceedings.neurips.cc/paper/2017/hash/8a20a8621978632d76c43dfd28b67767-Abstract.html

[5]. Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). "Why should I trust you?" (pp. 1135–1144). https://doi.org/10.1145/2939672.2939778

[6]. Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. (pp. 785–794). https://doi.org/10.1145/2939672.2939785

[7]. Nti, I. K., & Ramanayake, S. (2026). Discover Artificial Intelligence, 6, Article 288. https://doi.org/10.1007/s44163-026-01016-6