International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2026.0194
e ISSN: 2584-2854
Volume: 04 Issue: 05
May 2026
Page No: 1294 - 1298

An AI-Based Early Warning System for Identifying at- Risk Students in Higher Education

Bhavana R K[1], Karmjit SG[2], Bhavana S[3], Saniya K[4], Amritha Kallettumthara Reji[5], Daniya[6]
[1, 2, 3, 4, 5, 6]B.tech – Computer Science Engineering (AI & ML) Yenepoya University, Bangalore, India.
***Email Id:** rkbhavana31@gmail.com[1]*

Abstract
Student dropout and academic underperformance remain persistent challenges in higher education, often caused by factors such as academic difficulty, lack of engagement, and socio-economic conditions. Traditional monitoring approaches are mostly reactive and fail to identify at-risk students at an early stage, limiting the effectiveness of timely intervention. With the growing availability of educational data, Artificial Intelligence (AI) and machine learning techniques have been increasingly used to develop Early Warning Systems (EWS) for predicting student risk. This paper reviews and analyses various AI-based approaches that utilize data such as academic performance, attendance, and behavioural patterns to identify students who may require support. Different machine learning models, including decision trees, logistic regression, support vector machines, and ensemble methods such as Random Forest and XGBoost, are examined in terms of their effectiveness in prediction. The study also highlights the importance of integrating predictive systems with timely and personalized intervention strategies to improve student outcomes. In addition, key challenges such as data privacy, model interpretability, and ethical concerns are discussed. Overall, the paper emphasizes that AI-driven Early Warning Systems can significantly enhance student retention and academic success when implemented responsibly and supported by effective institutional practices.

Keywords: *Artificial Intelligence, Early Warning Systems, Machine Learning, At-Risk Students, Learning Analytics, Higher Education.*

1. Introduction

Student retention and academic success have become major concerns for higher education institutions worldwide. Many universities report significant dropout rates due to factors such as academic difficulties, lack of engagement, financial constraints, and challenges in adapting to the learning environment. These issues not only affect individual students but also impact institutional performance and overall educational quality. Traditionally, institutions have relied on manual monitoring methods such as grade evaluation and attendance tracking to identify struggling students. However, these approaches are largely reactive and often detect problems only after a student's performance has already declined. With the rapid growth of digital learning platforms and Learning Management Systems (LMS), large volumes of student data are now generated, including interaction patterns, assignment submissions, and engagement levels. This creates an opportunity to shift from reactive systems to proactive and data-driven approaches. Recent advancements in Artificial Intelligence (AI) and machine learning have enabled the development of predictive models that can analyse student data and identify at-risk students at an early stage. Studies have demonstrated that models such as Random Forest, XGBoost, and deep learning techniques can achieve high accuracy in predicting student outcomes. In addition, research highlights that combining predictive systems with timely intervention strategies significantly improves student retention and academic performance. Despite these advancements, several challenges remain, including data quality issues, model interpretability, and ethical concerns such as privacy and bias. Furthermore, many existing systems focus primarily on prediction

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2026.0194
e ISSN: 2584-2854
Volume: 04 Issue: 05
May 2026
Page No: 1294 - 1298

without effectively integrating intervention mechanisms. In this study, we aim to understand how AI-based Early Warning Systems can help identify at-risk students and support timely intervention strategies in higher education. It focuses on understanding different machine learning approaches, data sources, and practical challenges, whileemphasizing the importance of combining prediction with timely and effective intervention strategies.

2. Related Work

Several studies have explored the use of Artificial Intelligence and machine learning techniques for identifying at-risk students in higher education. Hussain et al. (2025) proposed an AI-driven framework integrating models such as Random Forest, XGBoost, TabNet, and BERT, demonstrating improved prediction accuracy using both structured and unstructured data. However, the study also highlighted challenges related to model interpretability and computational complexity. Elbouknify et al. (2025) developed a machine learning- based predictive system that achieved high accuracy in identifying at-risk students using real-world educational data. Their approach also incorporated explainable AI techniques to understand key contributing factors influencing predictions. Gathani et al. (2026) conducted a scoping review and found that machine learning-based Early Warning Systems can achieve high prediction accuracy, especially when multiple data sources are used. However, their effectiveness depends heavily on timely intervention and institutional support. Chang et al. (2025) emphasized the importance of combining predictive models with personalized intervention strategies. Their study demonstrated that integrating prediction with intervention significantly improves student performance. Carballo-Mendívil et al. (2025) proposed an early warning system using XGBoost based on pre-enrolment data, showing that early prediction before academic engagement can help institutions take proactive measures. Overall, existing research shows that AI-based Early Warning Systems are effective in predicting student risk, but challenges remain in implementation, interpretability, and integration with intervention strategies[1].

3. Dataset Description

This study is based on a review of existing research rather than a single experimental dataset. The datasets used in the referenced studies include a combination of academic records, attendance data, demographic information, and learning management system (LMS) activity. These datasets typically consist of features such as student grades, assignment submissions, engagement levels, socio- economic background, and behavioural patterns. Some studies also incorporate unstructured data such as text- based inputs and interaction logs. The use of diverse and multidimensional datasets allows machine learning models to capture complex patterns related to student performance and dropout risk. However, variations in dataset quality, size, and structure across different studies can influence the performance and generalizability of predictive models.

4. Methodology

In this study, we followed a structured approach to better understand how AI-based Early Warning Systems (EWS) can be used to identify at-risk students in higher education. The methodology is based on a review of existing research and focuses on the key stages involved in predictive modelling and intervention.

4.1 Data Collection

The first step involves collecting student-related data from multiple sources. These include academic records, attendance data, demographic information, and learning management system (LMS) activity. Previous studies have shown that combining multiple data sources improves prediction accuracy and provides a more comprehensive understanding of student behaviour patterns[3].

4.2 Data Preprocessing

The collected data is pre-processed to ensure quality and consistency. This step includes handling missing values, removing noise, normalizing data, and transforming it into a suitable format for analysis. Proper preprocessing is essential for improving

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2026.0194
e ISSN: 2584-2854
Volume: 04 Issue: 05
May 2026
Page No: 1294 - 1298

model performance and reducing bias in predictions.

4.3 Feature Selection

In this stage, the most relevant features influencing student performance are identified.[4] These may include grades, attendance, assignment submissions, engagement levels, and socio-economic factors. Selecting meaningful features helps in improving prediction accuracy and reducing computational complexity.

4.4 Model Development

Various machine learning algorithms are used to build predictive models. Traditional models such as Decision Trees, Logistic Regression, and Support Vector Machines are used for baseline prediction. In addition, advanced models such as Random Forest and XGBoost are applied to improve performance and handle complex patterns in data. Studies have shown that ensemble and deep learning models often provide better accuracy compared to basic models.

4.5 Prediction and Risk Classification

The trained models are used to classify students into categories such as "at-risk" and "not at-risk." These predictions are based on patterns identified from historical data. Machine learning models have demonstrated strong performance in predicting student risk, achieving high accuracy in multiple studies.

4.6 Early Warning and Intervention

Once at-risk students are identified, the system generates early warning alerts. These alerts enable educators and institutions to take timely action, such as providing academic support, counselling, or personalized guidance. Research indicates that combining prediction with intervention significantly improves student outcomes.

4.7 System Workflow

The overall process can be summarized as follows: Data Collection → Data Preprocessing → Feature Selection

→ Model Training → Prediction → Early Warning →InterventionThis workflow ensures a systematic approach to identifying and supporting at-risk students using AI-based techniques[2].

5. Experimental Setup

This study follows a comparative analysis approach based on previously published research. Different machine learning models are evaluated based on their performance in predicting at-risk students. The experimental setup in the referenced studies typically involves splitting datasets into training and testing sets, followed by model training and validation Figure 1. Performance metrics such as accuracy, precision, recall, F1-score, and AUC are commonly used to evaluate model effectiveness. Some studies also apply techniques such as cross-validation and data balancing to improve model performance and handle imbalanced datasets. The comparison of multiple models allows for a better understanding of their strengths and limitations in real-world educational scenarios.

6. Results and Analysis

From the studies we reviewed, it was observed that machine learning models are effective in predicting at-risk students using educational data[5]. Various models have been applied across different datasets, including academic performance, attendance, demographic information, and behavioural data. Traditional models such as Logistic Regression, Decision Trees, and Support Vector Machines provide moderate prediction accuracy and are widely used due to their simplicity and ease of implementation. These models perform well when the dataset is relatively small and less complex. In contrast, advanced machine learning models such as Random Forest, XGBoost, and deep learning techniques demonstrate higher accuracy and better performance when handling large and complex datasets[6]. Studies indicate that these models can achieve prediction accuracies ranging from approximately 85% to 90% in real-world scenarios. In some cases, especially when multidimensional data is used, prediction accuracy can reach up to 99%. The use of diverse input features, including academic records, attendance, engagement metrics, and socio-economic factors, contributes significantly to

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2026.0194
e ISSN: 2584-2854
Volume: 04 Issue: 05
May 2026
Page No: 1294 - 1298

improving model performance. Models that incorporate multiple data sources tend to produce more reliable and consistent predictions. In addition, studies highlight that preprocessing techniques such as data cleaning, normalization, and handling imbalanced datasets play an important role in enhancing model accuracy. The application of validation techniques such as cross-validation further improves the reliability of the results. Overall, the results indicate that AI-based predictive models are capable of accurately identifying at-risk students, with performance varying depending on the type of model used, the quality of data, and the range of features considered.

7. Discussion

Based on our analysis, machine learning models appear to be highly effective in identifying at-risk students when compared to traditional monitoring approaches. Advanced models such as Random Forest and XGBoost demonstrate higher accuracy due to their ability to handle large datasets and capture complex relationships between multiple factors influencing student performance. However, increased accuracy often comes at the cost of reduced interpretability. Simpler models such as logistic regression and decision trees are easier to understand but may not perform as well when dealing with complex and multidimensional data. This creates a trade-off between model performance and explainability, which is an important consideration in educational environments. Another key observation is that prediction alone is not sufficient to improve student outcomes. The effectiveness of Early Warning Systems largely depends on how well institutions respond to predictions. Systems that integrate timely and personalized interventions, such as academic support and counselling, show significantly better results in improving student retention. Additionally, the use of diverse data sources, including academic, behavioural, and socio-economic factors, enhances prediction accuracy. However, this also raises concerns related to data privacy and ethical use of student information. Ensuring transparency and fairness in AI- based systems is essential for their successful adoption. Overall, the findings suggest that AI-based Early Warning Systems should not be viewed purely as predictive tools, but as decision-support systems that combine technology with human intervention to effectively support students.

8. Proposed System Architecture

This study proposes an AI-based Early Warning System designed to identify at-risk students using machine learning techniques[7]. The system integrates multiple stages, including data collection, preprocessing, feature selection, and predictive modelling. Student data such as academic performance, attendance, and behavioural patterns are processed and analysed using machine learning models like Random Forest and XGBoost. These models classify students into risk categories, enabling early identification of those who may require support. The system generates early warning alerts, allowing institutions to implement timely intervention strategies such as academic assistance, counselling, and personalized guidance. This approach ensures a proactive and data- driven method for improving student outcomes and retention.

Figure 1 Proposed Early Warning System Architecture

Conclusion

In this study, we explored the role of Artificial Intelligence- based Early Warning Systems in identifying at-risk students in higher education. The analysis shows that machine learning models can effectively use student data such as academic performance, attendance, and behavioural patterns to predict potential risks at an early stage. The study highlights that advanced models like Random Forest and XGBoost generally provide higher prediction accuracy compared to traditional methods. However, accuracy alone is not sufficient. The effectiveness of these systems

e ISSN: 2584-2854
Volume: 04 Issue: 05
May 2026
Page No: 1294 - 1298

depends on how well predictions are combined with timely and appropriate interventions. In addition, important challenges such as data quality, model interpretability, privacy, and ethical concerns must be carefully addressed to ensure responsible implementation. Educational institutions need to balance technological advancements with human-centered approaches to support students effectively. In the future, Early Warning Systems can be improved by incorporating real-time data, more personalized intervention strategies, and explainable AI techniques to enhance transparency and trust. Overall, AI-based systems have strong potential to make education more proactive and student-focused by enabling early support and improving retention rates.

References

[1]. Hussain, F., Hammad, M., & Al Qahtani, H. I. (2025). AI- driven predictive analytics for student success and institutional decision-making in higher education. International Journal of Information Technology. https://doi.org/10.1007/s41870-025-03076-w
[2]. Permana, A. E., & Santoso, E. B. (2025). Student retention early warning procedures: A case study of higher education institution management. International Journal of Economics, Business and Accounting Research, 9(3).
[3]. Elbouknify, I., Berrada, I., Mekouar, L., Iraqi, Y., Bergou, E. H., Belhabib, H., Nail, Y., & Wardi, S. (2025). AI-based identification and support of at-risk students: A case study of the Moroccan education system. arXiv preprint arXiv:2504.07160.
[4]. Janahi, Y., & Obeidat, M. (2025). Predictive analytics and AI in education: A systematic literature review on identifying and supporting at-risk students.
[5]. Gathani, S., Faiz, H., Hazilan, W. F., Khandhadia, R., & Fenn, C. J. (2026). Exploring the effectiveness of machine learning-based early warning systems in education: A scoping review. Global Journal of Educational Thoughts, 3(1).
[6]. Chang, Y.-H., Chen, F.-C., & Lee, C.-I. (2025). Developing an early warning system with personalized interventions to enhance academic outcomes for at-risk students in Taiwanese higher education. Education Sciences, 15(10), 1321. https://doi.org/10.3390/educsci15101321
[7]. Carballo-Mendívil, B., Arellano-González, A., Ríos- Vázquez, N. J., & Lizardi-Duarte, M. P. (2025). Predicting student dropout from day one: XGBoost-based early warning system using pre-enrolment data. Applied Sciences, 15(16), 9202. https://doi.org/10.3390/app15169202