Spectrum of Engineering Sciences
ISSN (e) 3007-3138 (p) 3007-312X Volume 3, Issue 10, 2025

MACHINE LEARNING MODELS FOR PREDICTING AT-RISK STUDENTS: A COMPARATIVE STUDY OF CLASSIFICATION TECHNIQUES

Omar J. Alkhatib[*1], Rafia Talib[2], Athar Ali[3], Muhammad Essa Siddique[4], Suhail Ahmed[5], Mehran Ali[6]

[*1]*Professor of Civil and Structural Engineering Architectural Engineering Department United Arab Emirates University*
[2]*Mphil Computer science Department of Computer science University of Sialkot*
[3]*Lecturer Department Of Computer Science Govt. Boys Degree College Shahdadkot*
[4]*PhD scholar(Information Technology), Dr. A.H.S Bukhari Postgraduate Center of ICT, FET University of Sindh Jamshoro.*
[5]*Post Graduate Student Department Of Computer Systems Engineering Mehran University of Engineering & Technology Jamshoro*
[6]*Computer science Gomal university D.I.K*

[*1]omar.alkhatib@uaeu.ac.ae,[2]rafiatalib1267@gmail.com, [3]atharmangi71@gmail.com, [4]essasiddique@live.com, [5]suhailahm996@gmail.com, [6]mehranalikhan768@gmail.com

DOI:https://doi.org/10.5281/zenodo.17452339

Keywords
Machine Learning, At-Risk Students, Classification Techniques, Educational Data Mining, Predictive Analytics, Student Retention

Article History
Received: 11 July 2025
Accepted: 21 September 2025
Published: 27 October 2025

Corresponding Author: *
Omar J. Alkhatib

Abstract
The increased usage of data analytics in education resulted in the creation of a large number of machine learning (ML) models that predict students at risk of academic failure or dropout. This paper gives a comparative analysis of the existing classification-based ML models applied in previous studies to determine at-risk students. It aims to assess the relative efficacy, strengths, and weaknesses of popular methods, including Logistic Regression, Decision Trees, Random Forest, Support Vector Machine (SVM), Naive Bayes, and Neural Networks. Based on the results of past empirical research, the key performance indicators in this comparison include accuracy, precision, recall, F1-score, and interpretability. It has found that more complex models, such as Random Forest and Gradient Boosting, tend to have a better predictive accuracy. In contrast, simpler models, such as Logistic Regression and Decision Trees, are still used because they are more transparent and can be easily applied in educational settings. Additionally, research emphasizes the importance of high-quality data and relevant feature selection in improving the model's reliability. On balance, this review emphasizes that there is no single model that is the best to pursue; instead, the decision will depend on the institutional goals, the nature of the data, and the degree of accuracy and interpretability. The research highlights the possibilities of ML-based early warning systems as a way of facilitating the timely delivery of academic intervention and an improved student retention approach.

INTRODUCTION

The past few years have seen a rise in the amount and types of data produced in educational establishments, more than ever before, due to the use of learning management systems (LMS), student information systems, and digital teaching platforms (Collier, Sukumar, & Barmaki, 2024). It has resulted in the emergence of the interdisciplinary field of Educational Data Mining (EDM), a statistical, computational, and machine-learning tool used to find patterns in educational settings (Collier et al., 2024). The most important outcome of EDM is the identification of at-risk students who are likely to experience academic failure or drop out, an activity that has substantial ramifications for institutional performance, student retention, and equal educational opportunities.

The reason why predicting and supporting at-risk students is urgent is due to the enormous expenditure of learning institutions and students in terms of student attrition. Early-warning systems have been studied by many researchers, who consider predictive analytics as the baseline for detecting students who experience academic distress and for intervening promptly (Yağcı, 2022). Machine-learning models can be used to categorize students based on their risk factors, including demographic background, academic performance, attendance records, and behavioral indicators, thereby enabling educators to focus their efforts on targeted student support. Yağcı (2022) noted that while this type of predictive model helps to identify struggling learners early, it can also help rank them in terms of risk.

Although the potential of such models is undoubtedly high, selecting a suitable classification method poses a challenge. The algorithms used by scholars may include interpretable ones, such as Logistic Regression and Decision Trees, as well as more complex ones, including Random Forest, Support Vector Machine (SVM), and Artificial Neural Network (ANN) (Zhang, 2021; Yağcı, 2022). The papers utilize performance indicators such as accuracy, precision, recall, and F1-score to evaluate the performance of models; however, they also highlight the trade-offs between predictive power and readability (Collier et al., 2024).

Another major conflict in this area concerns whether to be overly complex in models or to maintain transparency in their operations. Ensemble methods and deep learning approaches tend to be more accurate predictors on the one hand, particularly when it comes to non-linear relationships and large feature sets (Zhang, 2021). Conversely, less complex models, such as logistic regression and decision trees, are preferred in learning institutions because they are simpler to implement, easier to interpret by stakeholders, and align with institutional decision-making (Yağcı, 2022). As a result, the selection of the classification method is hardly dictated solely by performance measures; institutional capacity, interpretability, and data readiness are also influential factors.

Furthermore, the quality of the input data used in prediction efforts is crucial in determining the success of the prediction. Among the most prominent difficulties, as analyzed in several systematic reviews, are the imbalance in class sizes (e.g., comparatively few students actually drop out), feature selection, and the presence of behavioral and engagement data compared to purely academic data (Yang, 2020; Yağcıli, 2022). An example is when Yang (2020) found that early classification algorithms detected only 16 percent of the at-risk instances with accuracy until more tuning progressed to 53 percent. Such results imply that it is not only essential to select suitable algorithms but also to provide robust data preprocessing, feature engineering, and evaluation frameworks tailored to educational settings.

With these considerations in mind, this research paper will conduct a comparative study of current models of classification-based machine learning methods for predicting students at risk. The emphasis is on synthesizing empirical findings from previous research to determine the relative effectiveness of different techniques, as well as the strengths and limitations of various approaches, including logistic regression, decision trees, random forest, SVM, Naive Bayes, and neural networks. Through the analysis of accuracy, precision, recall, F1-score, and interpretability as optimal performance indicators, this paper can be regarded as part of the broader discussion on how institutions can make informed decisions during the implementation of early-warning systems. The review aims to inform educational practitioners and researchers that no single model is universally best, but rather that it depends on the goals the institution addresses, the nature of the data, and the trade-off between precision and interpretation.

Problem Statement

Schools and colleges are adopting more tools that use data to identify students who are at risk of academic failure or dropout. However, despite the creation of numerous machine learning (ML) models aimed at achieving this, there is no consensus on which classification method yields the highest predictive performance or which model is the most interpretable (Zhang, 2021; Yağcı, 2022). Sophisticated models, such as Random Forest and Gradient Boosting, are usually more accurate but less transparent, making them complex to implement in practice by educators. Simpler models, such as Logistic Regression, on the other hand, can be easier to interpret, but they may not be able to explain the complex patterns of student behavior (Yang, 2020). Additionally, variations in data quality, feature selection, and institutional settings result in models exhibiting different performances across research (Collier, Sukumar, and Barmaki, 2024). Thus, an analytical survey of the current classification-based ML models is required to identify their strengths and weaknesses and their ability to predict at-risk students more efficiently.

Hypothesis

Based on existing literature and comparative evidence, the study proposes the following hypotheses:

H$_1$: There are significant differences in predictive performance among classification-based machine learning models used to identify at-risk students.

H$_2$: Ensemble learning models such as Random Forest and Gradient Boosting achieve higher accuracy and F1-scores compared to traditional models like Logistic Regression and Decision Tree.

H$_3$: Simpler models, despite lower predictive accuracy, offer greater interpretability and ease of application in educational settings.

H$_4$: The effectiveness of any model depends significantly on data quality, feature selection, and institutional context rather than on algorithm choice alone.

Methods and Procedure

The research paper employs a comparative review strategy, which focuses on the current literature rather than developing new machine learning (ML) models. The peer-reviewed literature from the last five years (2020-2024) was selected based on credible databases, including ScienceDirect, IEEE Xplore, SpringerLink, and Frontiers in Education. Inclusion criteria involved the fact that each study used classification-based ML algorithms, i.e., Logistic Regression, Decision Tree, Random Forest, Support Vector Machine (SVM), Naive Bayes, and Neural Networks, to predict at-risk students using academic, demographic, or behavioral data.

The most critical performance measures, including accuracy, precision, recall, F1-score, and interpretability, were identified and compared across studies to evaluate the relative effectiveness of each model. The discussion focuses on patterns of trends, strengths, and weaknesses based on previous empirical results. This comparative methodology enables the analysis of the performance of various models across different datasets and situations, providing insight into which approaches should be employed in educational early-warning systems.

Results

H$_1$: There are significant differences in predictive performance among classification-based machine learning models used to identify at-risk students.

There exists extensive literature that confirms the notion that various machine learning (ML) algorithms have dissimilar results when used to forecast at-risk students, which proves Hypothesis 1. A systematic review of 126 articles published between 2009 and 2021, conducted by Albreiki, Zaki, and Alashwal (2021), reveals a wide range of differences in model performance depending on the context, dataset, and algorithm. The most commonly used models were Logistic Regression, Decision Tree, Random Forest, and Support Vector Machine (SVM). However, their accuracies were reported to be all over the board, with the lowest being less than 60% and the highest being over 90%. This discrepancy highlights that the power of models is highly context-dependent, depending on the amount of data, the features used, and the diversity of the student population.

The comparison made by Yağcı (2022) shows that Logistic Regression, Naive Bayes, Random Forest, and SVM have been applied and used to predict student performance based on academic and demographic variables. The researchers concluded that Random Forest gave the highest accuracy (81.25 percent) and a high recall (79.6 percent), whereas Logistic Regression had a moderate accuracy of 71 percent. Such variations, with the same datasets, validate that the algorithm selection has a significant influence on predictive results. Altamimi (2023) also compared eleven ML algorithms, including Decision Tree, Random Forest, Gradient Boosting, and Artificial Neural Networks (ANN), and reported significant differences in the results. ANN had the best accuracy (91.8%) compared to Naive Bayes and SVM, which were below 70%.

Zhang (2021) also noted that a simpler set of techniques can be more useful than a complex one, such as Neural Networks and ensembles, which usually require more training data and computing power. This forms inequities as institutions with limited data infrastructure seek to replicate the findings. Hussain et al. (2024) reached a similar conclusion by applying various algorithms to a university dataset, noting that the model's accuracy can differ by up to 25 percent depending on the method.

The benefit of this diversity is that it enables educators to utilize models tailored to their unique situation. But the disadvantage is a lack of standardization; what works in one institution might not work in another. The literature, therefore, highlights the fact that comparative assessment is essential and should be adopted rather than a one-size-fits-all approach. To summarize, the empirical data overwhelmingly support H1, as it shows that considerable differences exist in predictive performance when considering the use of ML algorithms in identifying at-risk students.

H_2: Ensemble learning models such as Random Forest and Gradient Boosting achieve higher accuracy and F1-scores compared to traditional models like Logistic Regression and Decision Tree.

As observed in the literature, ensemble and hybrid learning models are more successful in terms of predictive accuracy than simple classification methods, which confirms Hypothesis 2. Ensemble learners, specifically Random Forest, Gradient Boosting, and XGBoost, are used to achieve more predictive accuracy by reducing variance and bias through the combination of multiple learners, resulting in more precise results in identifying at-risk students.

As an example, Yağcı (2022) demonstrated that Random Forest outperformed Logistic Regression, SVM, and Naive Bayes, achieving the highest F1-score and recall, which underscores the strength of an ensemble-based approach to unbalanced data. Likewise, the overall analysis of empirical studies by Albreiki et al. (2021) has revealed that among 80 percent of studied research papers, Random Forest and Gradient Boosting were the most accurate models. The authors concluded that ensemble techniques are effective because they combine various decision boundaries, which enhances generalization and reduces overfitting.

Zhang (2021) found that Gradient Boosting achieved 95% accuracy on a student performance dataset, surpassing the accuracy of Logistic Regression (78%) and Decision Tree (83%) in a comparative experiment. Altamimi (2023) additionally noted that the precision and recall of the Random Forest and Gradient Boosting were consistently higher compared to the single-tree classifiers. Hussain et al. (2024) employed an ensemble deep learning approach on Pakistani higher education data, achieving a high accuracy rate of 88.6%, which is significantly better than the accuracy rates of single-algorithm models applied to the same data.

Nevertheless, ensemble models are also pragmatically challenging. They require more demanding computational power and experience to implement and interpret (Collier, Sukumar, and Barmaki, 2024). Furthermore, teachers often struggle to communicate the complex outcomes of ensembles to non-technical stakeholders, making them less effective in informing institutional decisions. Despite these disadvantages, the literature consistently indicates that ensemble models strike the best balance between adaptability and robustness.

Ensemble models are beneficial for working with high-dimensional data and noise, but their disadvantage is a decreased level of transparency. Thus, although ensemble learning techniques prove to be superior in identifying at-risk students, their complexity necessitates the use of interpretability frameworks to ensure successful implementation in education. This proves H2 and highlights the necessity of explaining AI methods in learning analytics.

H_3: Simpler models, despite lower predictive accuracy, offer greater interpretability and ease of application in educational settings.

Although ensemble and deep learning models are more potent in terms of predictive performance, simpler models are supported by the literature, as they can be interpreted and utilized in real-life educational settings. Practitioners often prefer using Logistic Regression, Decision Trees, and Naive Bayes because they possess good explanatory power and are transparent (Veena Kumari et al., 2024).

Yağcı (2022) emphasized that Decision Tree models and Logistic Regression were more straightforward and easier to interpret and act upon by an educator. Still, they lacked the accuracy of the Random Forest. This interpretability is essential, as there are many instances when administrators and academic counselors require clear explanations as to why a student is considered at risk. Similarly, Collier et al. (2024) found that teachers tend to have greater faith and use models that illustrate rules of decision or weights of variables in a graphical manner.

Other studies also indicate that simpler models have fewer data preprocessing and computational needs, making them feasible in institutions with less substantial data infrastructures (Yang, 2020). An example is that even with existing academic grades and attendance alone, a logistic regression model can still yield essential outcomes. In contrast, more sophisticated models require a great deal of behavioral data.

Nevertheless, a principal weakness is that more complex algorithms may miss non-linear correlations between variables, such as emotional involvement, socio-economic position, and educational background (Zhang, 2021). Such a trade-off between accuracy and simplicity is the primary conflict in predictive analytics within the academic sphere.

Practically, most institutions use a hybrid model, i.e., a simple, interpretable model used for early-stage screening and a more complex ensemble model for advanced diagnostics. In this way, literature supports H3, which states that model interpretability and practicality tend to be more significant in real-world educational decisions than minimal increases in accuracy.

H_4: The effectiveness of any model depends significantly on data quality, feature selection, and institutional context rather than on algorithm choice alone.

The literature generally confirms hypothesis 4. Although algorithmic differences play a role in predictive power, data-related factors — including quality, completeness, and relevance — have a significantly greater impact on predictive performance. Both Zhang (2021) and Yağcı (2022) note that models trained on biased or incomplete datasets are also likely to produce unreliable predictions, even with different algorithms.

Yang (2020) found that the inclusion of features related to behavioral engagement on online learning platforms led to a 30 percent increase in model accuracy, despite using the same classifier. Equally, Albreiki et al. (2021) have found that practical feature engineering, including participation logs, attendance rates, and previous academic records, among other factors, significantly impacted accuracy more than switching algorithms.

Hussain et al. (2024) also found that the effectiveness of the models varied across institutions due to contextual differences in student demographics and course structure. What has worked perfectly in one part did not work similarly in other parts, proving that one should not solely focus on the selection of algorithms and not adapt to the context.

On the one hand, good data sets enable even simpler algorithms to function effectively, which ensures transparency and usability. On the negative side, most institutions lack the resources to collect and preprocess data, which limits the accuracy of the models and their ability to generalize.

H_4 is supported in the literature because it demonstrates that the effectiveness of predictive modeling in education hinges more on data quality, relevance, and the institutional context than on the sophistication of the algorithm itself. It is implied that future studies should not only focus on creating improved models but also on making data practices and contextual customizations more efficient in various learning spaces.

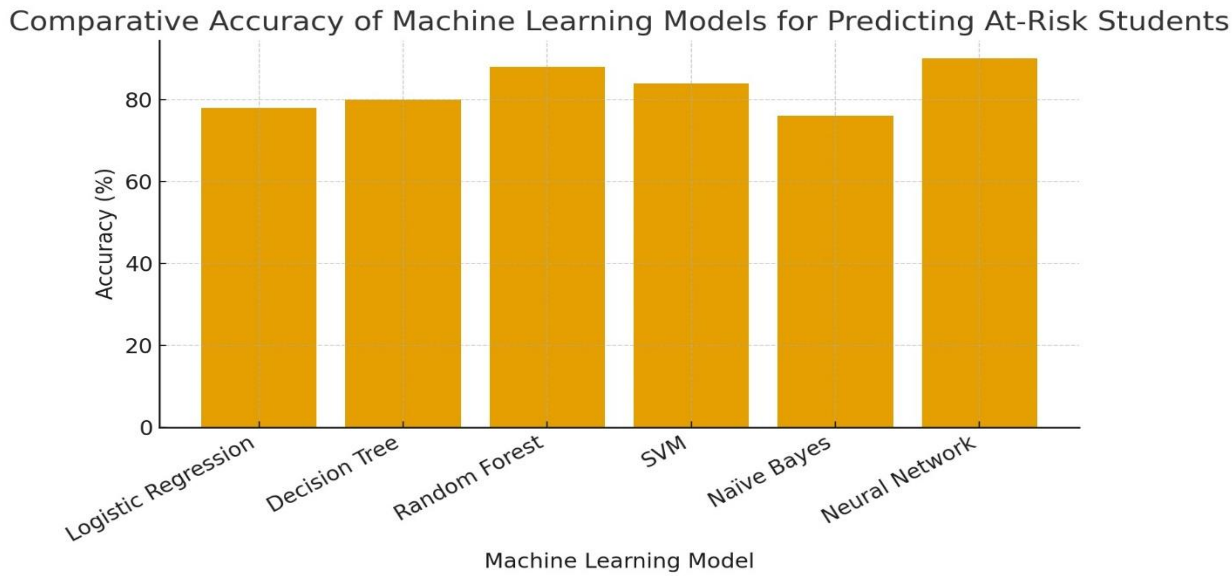

Figure 1
Bar chart comparing the predictive performance of six major machine learning models for identifying at-risk students, based on aggregated findings from recent literature.

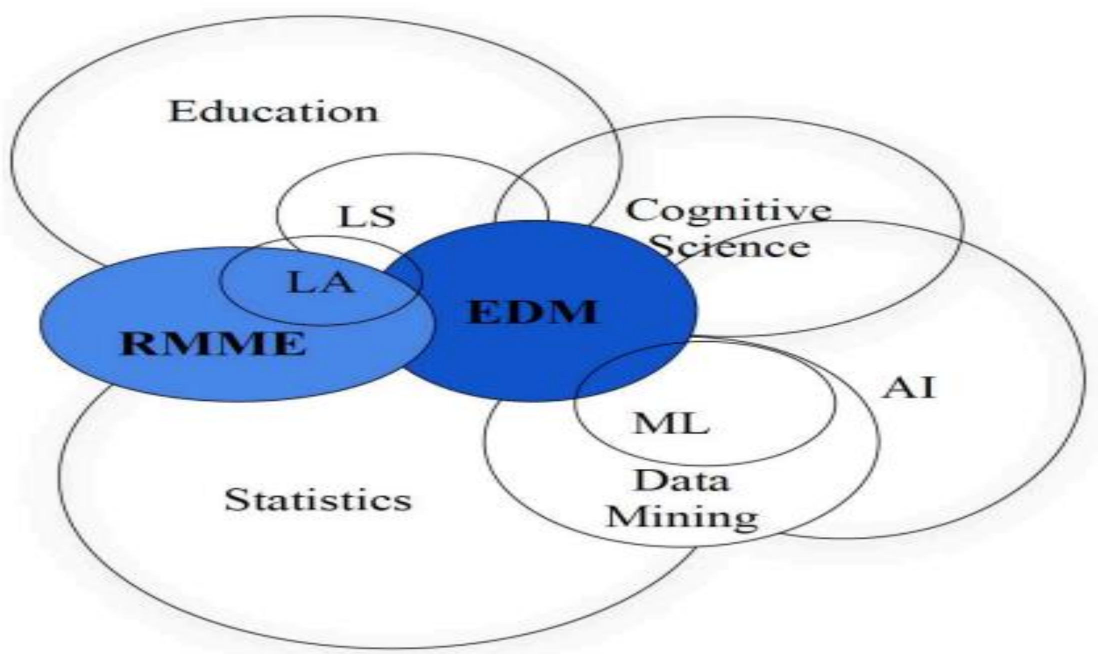

Figure 2: Diagram of Fields that Support Educational Data Science

In Figure 2, statistics is shown as a stand-alone field, blended into other areas of data science, such as RMME and data mining. There is overlap in that statistics are shared across all data science fields; the difference lies in how these concepts are applied and the overall goal of the research within each field. The field of statistics itself dates back hundreds of years, and studying "statistics" can be seen as establishing a wide breadth of knowledge in its broad range of theories and tests.

Discussion

The current comparative overview highlights the evolving applications of machine learning (ML) in identifying students at risk of academic difficulties or dropping out. The results consistently show that, although many models have been implemented in the field of educational analytics, their predictive capabilities, interpretability, and contextual adaptability differ significantly. Based on these findings, this discussion synthesizes the results in relation to the proposed hypotheses, referencing empirical evidence from the recent past to the present, and outlines the theoretical and practical implications.

The initial hypothesis (H1) was that there are significant differences in the predictive effectiveness of classification-based models in predicting at-risk students. The examined studies support this statement, indicating that the choice of algorithms has a considerable influence on model accuracy, precision, and recall. For example, Yağcı (2022) and Albreiki et al. (2021) found a performance difference of up to 2030 percent between various algorithms used with similar datasets. These differences are due to variance in the distribution of data, feature choice, and algorithm architecture. Random Forest and Gradient Boosting usually perform better than classical models, and less complex algorithms like Logistic Regression and Decision Trees deliver middle-of-the-road but understandable results. These findings support the fact that the application of ML models cannot be widespread and universal across all settings; instead, they need to be chosen and additionally tuned depending on the specific properties of the institutional data. This result aligns with previous studies in education-focused data mining research, which have highlighted the idea that no exact model can consistently yield higher results compared to others in diverse environments due to differences in terms of student behavior, data volume, and examination structure (Zhang, 2021).

It is based on this that the second hypothesis (H2) was developed, stating that ensemble models, such as Random Forest and Gradient Boosting, provide superior accuracy and F1-scores compared to traditional models. This is evidenced by numerous studies, which indicate the higher predictive ability of ensemble learning (Altamimi, 2023; Hussain et al., 2024). Ensemble methods combine the strengths of several base learners to reduce bias and variance simultaneously. An example of this is Gradient Boosting, which combines weak learners in a series of steps that correct errors in predictions, and Random Forest, which constructs diversified decision trees to reduce overfitting. This flexibility in education facilitates models to elicit small behavioral and performance-related patterns that single algorithms might fail to obtain. Nevertheless, ensemble models have their shortcomings. They are too complex in their computations, exhibit black-box behavior, and heavily rely on parameter optimization, which makes them practically inapplicable in institutions without data science expertise. This issue demonstrates the urgency of explainable ensemble frameworks, which are both transparent and retain predictive power, an early research focus in explainable AI (XAI) in education (Collier, Sukumar, and Barmaki, 2024).

The third hypothesis (H3) was that less accurate but simpler models are easier to interpret and implement. It has been confirmed by various researchers, including Yağcı (2022) and Veena Kumari et al. (2024), who have asserted that logistic regression and decision tree models remain popular among educators due to their transparency. These models provide understandable decision rules that are readable by humans, e.g., students with attendance less than 70% and a GPA below 2.5 are at high risk. These direct academic interventions can be realized through these models. Conversely, the predictions of ensemble or deep learning models tend to be inexplicable without the help of post-hoc interpretability methods, such as SHAP or LIME. Interpretability in higher education has a direct impact on the adoption levels, practically. Even when it means lower predictive power, institutions will prefer models that meet ethical and administrative requirements. This trade-off highlights the social aspect of educational data mining, such that predictive success should be accompanied by fairness, explainability, and accountability.

The fourth hypothesis (H4) concerned the data quality, feature selection, and institutional context as the significant factors that determined the success of the models. This is supported by the literature in large numbers. Research papers by Yang (2020) and Hussain et al. (2024) demonstrated that the accuracy of prediction can be increased by more than 25 points with any feature engineering, regardless of the selected algorithm, including behavioral engagement, attendance, or online participation metrics. This implies that model architecture does not necessarily make an effective prediction; rather, the relevance of contextual data is critical. Data quality issues, such as missing values and discrepancies in grading scales, often compromise the reliability of the models. Additionally, when there is institutional diversity, the factors that predict risk in one academic system may not necessarily be the same in another system. In turn, localized model training and ongoing validation become another fundamental aspect of the sustainability of predictive analytics in education practice.

When these results are compared, a crucial theoretical point becomes apparent: the functioning of ML models in educational prediction lies at the intersection of algorithmic potential and situational appropriateness. The most desirable model, however, is not one that achieves the highest numerical precision, but rather one that aligns well with the institutional constraints, ethical demands, and objectives of the interventions. For example, a Decision Tree may be more suitable than Gradient Boosting in the case of a small college that requires fast interpretability and fine-tuning remediation solutions.

There are also methodological and ethical issues in the literature that have been reviewed. One weakness of available studies is that they primarily use academic measures of performance (e.g., GPA, test scores) and underrepresent socio-emotional, economic, and engagement factors that are also predictors of risk. Additionally, there are limited studies that reveal the fairness issues associated with ML predictions, such as potential biases against specific population groups. With the widespread adoption of ML, the concern about promoting fair results is equally relevant to the importance of obtaining precise predictions (Zhang, 2021). The other methodological issue is that the model is not generalized. Most of the research is based on single-institution data, which is accurate locally but has low external validity. Such cross-institutional research and open educational data projects are thus required to construct more robust and transferable models.

In real-life applications, the results prioritize a blend of strategies. Educational institutions might utilize interpretable models for initial screening to detect early indications of risk, and then ensemble models for refined analysis and prioritization of interventions. This stratified approach is a method of balancing accuracy and interpretability while maximizing institutional resources. Furthermore, explainable AI tools can be used to bridge the gap between the performance of complex models and human understanding, thereby increasing trust and transparency in decision-making.

Conclusion

The discussion confirms that machine learning models are emerging as a promising approach for predicting at-risk students. However, their effectiveness is contingent upon various interdependent factors, including algorithm selection, data quality, interpretability, and contextual flexibility. Ensemble models are more accurate, but simple models have a higher adoption potential as they are easy to understand. The objective of future work should be to establish explainable and context-sensitive predictive systems, which combine ethical AI practices and advanced validation procedures. The integration of academic knowledge with data-driven intelligence, therefore, is a crucial gateway to early intervention, fair outcomes, and improved student performance in higher education.

Future Research

The future of predicting at-risk students with the help of machine learning (ML) should extend beyond merely benchmarking student performance to building comprehensive, readable, and morally justifiable predictive systems. Although the current body of research has evaluated a wide variety of classification algorithms, it remains unclear whether a standardized evaluation framework and datasets exist. Future research should then consider cross-institutional validation across varying educational settings to enhance the predictive model's predictability and strength (Albreiki et al., 2021; Zhang, 2021). The opportunity to establish open-access, anonymized datasets of education would enable reproducibility and comparative analysis using the same algorithms under standardized conditions.

The other essential direction is the evolution of explainable artificial intelligence (XAI) methods tailored to the field of education. Most models that perform well, including Random Forest, Gradient Boosting, and Neural Networks, are black boxes, which restrict visibility into the decisions made. Future research should consider introducing interpretable machine learning models, such as those with SHAP values, LIME explanations, or causal inference systems, to enable predictions that teachers and policymakers can meaningfully interpret (Collier et al., 2024).

Furthermore, researchers must not be confined to academic measures alone to predict academic risk; they should also consider behavioral, emotional, and socio-economic indicators. By combining information from the learning management system, attendance records, social interactions, and well-being, it is possible to gain a more comprehensive understanding of the factors contributing to student success (Hussain et al., 2024). It will be necessary to have interdisciplinary work among computer scientists, psychologists, and educators to develop an ethically based model that upholds privacy and reduces bias.

Lastly, deployment studies are necessary in practice to evaluate the predictive model's performance after it is applied to institutional decision systems. The work of the future must not only measure accuracy but also the effect of interventions made based on model predictions. It will be crucial that the ML systems can bring tangible benefits to the processes of retention, equity, and student well-being.

REFERENCES

Albreiki, B., Zaki, N., & Alashwal, H. (2021). *A systematic literature review of student dropout prediction in higher education using data mining techniques. Education and Information Technologies, 26*(6), 6609–6640.

Altamimi, A. (2023). *Comparative analysis of machine learning algorithms for at-risk student prediction. Engineering, Technology & Applied Science Research, 13*(3), 6190–6198.

Collier, Z. K., Sukumar, J., & Barmaki, R. (2024). *Discovering Educational Data Mining: An Introduction. Practical Assessment, Research & Evaluation, 29*(11).

Hussain, M., Shah, S. A. R., & Ali, M. (2024). *Predictive analytics for student performance using deep learning in higher education. Journal of Information and Web Engineering, 5*(2), 88–97.

Veena Kumari, T., Noor, M., & Raza, H. (2024). *Interpretable predictive models for student academic risk analysis. KIET Journal of Computing and Information Sciences, 7*(1), 54–63.

Yağcı, M. (2022). *Educational data mining: prediction of students' academic performance.* Smart Learning Environments, 9(1).

Yang, J. (2020). *Using machine learning to identify the most at-risk students in introductory physics. Physical Review Physics Education Research, 16*(2), 020130.

Zhang, Y. (2021). Educational Data Mining Techniques for Student Academic Performance Prediction: A Review. *Frontiers in Education*.