FUDMA Journal of Sciences (FJS)
ISSN online: 2616-1370
ISSN print: 2645 - 2944
Vol. 10 No. 11, June, 2026, pp 264 – 271
DOI: https://doi.org/10.33003/fjs-2026-1011-5293

A Predictive Analytics Model for Early Detection and Prevention of Student Dropout in Federal College of Education, Gidan Madi

***[1]Jibril Muhammad Bodinga, [2]Dr. Shehu M. S.Tudu and [3]Muhammad M. Bodinga**

[1] Department of Computer Science, Federal College of Education Gidan Madi, Sokoto, Nigeria.
[2] Department of Computer Science, Sokoto State University, Sokoto, Nigeria.
[3]Department of Fine & Applied Arts, Shehu Shagari College of Education, Sokoto, Nigeria.

*Corresponding authors' email: jibsonbodinga.doc@gmail.com Phone: +234-8103240964

ABSTRACT

The increasing rate of student dropout in Nigerian tertiary institutions, particularly Colleges of Education, remains a significant threat to national human capital development. At Federal College of Education, Gidan Madi, dropout cases have been persistently linked to poor academic performance, socio-economic challenges, and inadequate institutional support structures. These issues not only result in personal and societal losses but also undermine the performance indicators of the institution, affecting funding opportunities, graduation rates, and academic reputation. Machine learning and deep learning techniques have been widely applied to analyzing student data; however, many existing studies rely on large datasets and computationally intensive models, limiting their applicability in data-constrained environments. This research proposes the development and evaluation of logistic regression (LR), Decision Tree (DT), Random Forest repressor (RF) and Artificial Neural Network (ANN) models for predicting student academic performance using a relatively small dataset. A dataset comprising of 250 student records was used for the experiment. The results show that the RF model achieved the highest accuracy of 92.00%, followed by LR with 90.00%. Both RF and LR consistently outperformed the ANN and DT models. The ANN model achieved an accuracy of 88.00%, while the DT model recorded the lowest accuracy of 84.00%.

Received: 22 Jan. 2026

Accepted: 12 Feb. 2026

Published: 30 May 2026

Keywords: Machine Learning (ML); Deep Learning (DL), Student Dropout Prediction (SDP), Predictive Analytics (PA), Educational Data Mining (EDM)

INTRODUCTION

Tertiary education plays a fundamental role in human capital development, socio-economic growth, and national transformation. It equips individuals with the knowledge, skills, and competencies required for innovation, productivity, and leadership across various sectors of the economy. In developing countries such as Nigeria, Colleges of Education are essential in preparing future teachers and strengthening the foundation of the education system. However, this critical sector continues to face persistent challenges that undermine its effectiveness and long-term impact, with one of the most significant being the rising rate of student dropout.

Despite the growing integration of digital technology in educational system, many colleges still rely on ineffective and outdated methods to evaluate and predict student academic performance (Kuş, 2025). These traditional approaches present several key challenges (Kuş, 2025) including lack of model interpretability, challenges with data imbalance and bias, and inability to capture sequential learning patterns. The rapid integration of technology into modern classrooms and system of education has significantly transformed the education sector, leading to the widespread adoption of digital tools, e-earning platforms, and LMS. These technologies generate vast amounts of student-related data, contributing to the rapid growth of EDM, which focuses on extracting meaningful insights from educational datasets (Mohsen et al., 2023). Through these platforms, educators can access detailed information about students' academic progress, learning behaviours, and engagement levels. However, despite the availability of such rich data, effectively utilizing it to improve student outcomes remains a critical challenge that need to be addressed.

In contemporary system of education, there is an increasing reliance on data-driven and decision-making to enhance teaching and learning methods. Predictive models play a crucial role in this context by enabling the forecasting and analysis of student academic performance, thereby supporting educational planning and informed personalized learning (Mohsen et al., 2023). Early prediction of student academic performance is particularly important, as it allows educators and decision makers to identify at-risk students and implement timely interventions to improve their academic outcomes (Mohsen et al., 2023).

Despite these advancements, traditional performance prediction methods such as instructor-based evaluations and statistical methods are limited in their ability to capture the dynamic and complex nature of student learning (Li & Lu, 2025). These predictive methods often treat educational data as static and fail to account for temporal dependencies and evolving learning patterns, resulting in less effective intervention strategies and reduced predictive performance. Recent advancements in ML and DL have introduced more sophisticated methods to student performance prediction tasks. In particular, ML techniques such as LR, RF, DT and ANN have demonstrated strong capabilities in time-dependent data and modelling sequential task (Lin et al., 2025). ML techniques can capture temporal relationships in student learning activities, making them highly suitable for predicting future academic performance and analysing learning trajectories with improved accuracy. Furthermore, student academic performance is influenced by multiple interconnected factors, including behavioural patterns, cognitive abilities, socio-economic background and institutional conditions (Ahmed, 2024). Understanding how these factors evolve over time is essential for developing effective predictive systems. However, conventional models often struggle to incorporate these temporal dynamics, limiting their effectiveness in real-world educational situations (E. Ahmed, 2024). Therefore, there is a need for more effective predictive models that can effectively leverage sequential educational data to improve performance and inform decision making. This study aims to develop a ML-based model for sequence-based prediction of student academic performance using small sample records. The proposed models seek to provide actionable insights and enhance prediction accuracy that enable educators and decision makers to implement proactive, data-driven interventions, ultimately improving student educational outcomes. The aim of this research is to design, implement and evaluate to at least two ML-based models for predicting student performance using small sample datasets. The goals can be achieved by incorporate training method that can enhance prediction performance models 'small sample datasets, identify and structure temporal features such as academic performance and learning behaviour for models input, design and implement at least three ML-based models for predicting student academic performance and evaluate the performance these models using appropriate metrics and assess their effectiveness on small-sample datasets. The outcomes of this research may For Students: By identifying the key factors affecting academic performance, this research can help students adopt more effective learning strategies, ultimately leading to better outcomes. For Educators and Institutions: The study will provide valuable insights to help teachers and academic institutions implement proactive interventions for students at risk of poor performance. These insights can also be used to refine teaching strategies and improve academic planning. For Researchers: The findings will contribute to the growing field of artificial intelligence in education, particularly in the application of ML for student academic performance prediction. For Policymakers: This study will serve as a guide for policymakers in formulating data-driven policies to improve student performance while ensuring ethical considerations, such as data privacy and security, are upheld.

Empirical Literature Review

This section presents a summary of the selected research studies as shown in Table 1. It synthesizes the authors' approaches after conducting the review, highlighting the objectives, datasets, techniques, findings, and model architectures used in the comparative analysis. Overall, this section provides an overview of the data extracted from the reviewed papers, including the methodologies employed and the key results obtained across the selected studies.

Table 1: Literature Review Paper

Ref.	Methodology	Dataset Used	Algorithms Used	Tools	Evaluation Metrics	Result	Limitations
(Sandeepa, 2025)	Supervised learning using neural networks and ML	Behavioural, academic, demographic data	MLP Classifier	Python (Scikit-learn, TensorFlow/Keras)	Accuracy, Cross-validation accuracy	86.46% (test), 79.58% (CV)	Performance drops with small or imbalanced datasets
(Diekuu et al., 2025)	Temporal AI-based time-series modelling	Dataset of 3,000+ students	Temporal AI / Sequential models	Python (TensorFlow/PyTorch)	Accuracy	Improved accuracy via temporal modelling	Depends on availability of sequential data
(Augustine et al., 2025)	Deep learning with explainability (SHAP)	GPA prediction dataset	Bi-LSTM	Python (TensorFlow, SHAP library)	Accuracy	88.23% accuracy	Limited scalability on small datasets
(Adefemi & Mutanga, 2025b)	Hybrid deep learning (dual extraction)	Multiple student datasets	CNN, BiGRU	Python (TensorFlow/Keras)	Accuracy	97.48%, 90.90%, 95.97%	Complex architecture, high computation
(Hajoui et al., 2025)	Optimized deep learning with sequential modelling	Educational datasets	RNN	Python (TensorFlow/PyTorch)	Accuracy	Improved performance over baseline DL	Sensitive to hyperparameters
(Adefemi & Mutanga, 2025a)	Deep learning + preprocessing	Benchmark educational datasets	CNN, LSTM	Python (TensorFlow/Keras)	Accuracy, Precision, Recall, F1-score	98.93%, 98.82%	Requires large datasets and high computation
(Alsumaidaie et al., 2024)	Supervised classification	Student academic dataset	Random Forest, Logistic Regression, ANN	Python (Scikit-learn)	Accuracy, F1-score	RF: 95.6% accuracy, 94.8% F1	LR may underfit complex patterns
(Bello et al., 2025)	Regression-based prediction	Academic dataset	Linear Regression, RF Regressor, XGBoost	Python (Scikit-learn, XGBoost)	MAE, RMSE	XGBoost: MAE=0.12, RMSE=0.18	Less suitable for classification
(Id & Yu, 2025)	Ensemble and classification modelling	Online learning dataset	Logistic Regression, Decision	Python (Scikit-learn)	Accuracy, Precision, Recall	94.3%, 93.8%, 94.1%	High computational cost

Ref.	Methodology	Dataset Used	Algorithms Used	Tools	Evaluation Metrics	Result	Limitations
(W. Ahmed et al., 2025)	Ensemble learning with tuning	Educational dataset	Trees, Ensemble RF, SVM, Gradient Boosting	Python (Scikit-learn)	Accuracy, Precision, Recall, F1-score	Accuracy 97.8%, F1 96.2%	Poor at sequential dependencies
(Muresan et al., 2026)	Heterogeneous graph deep learning	OULA dataset	Graph-based DL + ML	Python (PyTorch Geometric, DGL)	F1-score	68.6% – 89.5%	Requires structured relational data

MATERIALS AND METHODS

Evaluation Metrics

Accuracy is a fundamental metric for evaluating the overall performance ML and DL models. In this dataset, accuracy measures the proportion of correctly predicted instances out of the total instances from giving samples. A high accuracy score indicates that the model is effective in making correct predictions across both classes, thereby providing a reliable measure of its overall performance (Farhadpour et al., 2024).

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} \quad (1)$$

Precision is an essential evaluation metric used to measure the effectiveness of a detection model in accurately identifying true positive cases. It is defined as the ratio of correctly predicted positive instances to the total number of instances predicted as positive by the model. A high precision score indicates a low rate of false positives, meaning that non dropout samples are rarely misclassified as dropout, thereby enhancing the reliability of the model (Farhadpour et al., 2024). $Precision = \frac{TP}{TP+FP}$ (2)

Recall, also referred to as sensitivity or the true positive rate, measures the ability of a detection model to correctly identify all actual dropout instances within the dataset. It is defined as the ratio of correctly predicted dropout cases to the total number of actual dropout cases. A high recall score indicates that the model effectively detects most dropout instances, minimizing the occurrence of false negatives (Maxwell et al., 2021). $Recall = \frac{TP}{TP+FN}$ (3)

The F1-score is a composite evaluation metric that balances both precision and recall, providing a more comprehensive measure of a dropout detection model's performance. It is defined as the harmonic mean of precision and recall, ensuring that both false positives and false negatives are considered. This metric is particularly useful when there is a need to achieve a balance between accurately identifying dropout instances and minimizing missed detections (Farhadpour et al., 2024).

$$F1 - Score = \frac{2\ X\ TP}{2\ X\ TP+FP+FN} \quad (4)$$

The Receiver Operating Characteristic – Area Under the Curve (ROC-AUC) is a robust evaluation metric used to assess the ability of a dropout detection model to distinguish between dropout and non-dropout classes across all possible classification thresholds. It captures the trade-off between the True Positive Rate (TPR), also known as sensitivity, and the False Positive Rate (FPR), providing a threshold-independent measure of classification performance. A higher ROC-AUC value indicates better model discrimination, meaning the model is more capable of correctly distinguishing dropout instances from non- dropout activities.

The ROC curve is a graphical representation that plots the True Positive Rate (TPR) against the False Positive Rate (FPR) at various threshold levels. The Area Under the Curve (AUC) quantifies the overall performance of the model, where a value of 1.0 represents perfect classification, while a value of 0.5 indicates no discriminative ability (equivalent to random guessing). $TPR = \frac{TP}{TP+FN}$ (4)

Table 2: Experimental Parameters Setup

Parameter	Experimental Setup
System Specification	CORE i7 with 8GB RAM, Hard disk SSD 256, Windows 11 OS
Sample Size	250 row data
Number of runs	multiple to account for variance
Epochs	50
Batch Size	32
Activation Function	Relu
Optimizer	Adam
Learning Rate	1e-4
Loss Function	Binary cross-entropy
Hyperparameter tuning method	random
Train/validation/test splits	80%, 10%, 10
Libraries	Numpy, pandas, scikit-learn, matplotlib and seaborn
Evaluation Metric	Accuracy, Precision, Recall, F1-score, AUC
Execution time	100 – 200 secs

FCE Gidan Madi Dataset

The experiments performed to evaluate ML and DL model performance for prediction of student performance using (FCE-student-performance.csv) collected from college, consists of feature from two (2) categories of samples student outcomes and would be used for the experiment, 125 samples for each class were extracted and used totalling 250 sample records. This extracted dataset contains 10 features includes records such as s/n, student_id, department, age, gender, level, study_hour, last_cgpa, performance and outcomes.

Data Preprocessing

Categorical features such as gender, level, parent_income, internet_access, dropout, grade class, performance were transformed into numerical representations using label encoding to ensure compatibility with ML algorithms. Other columns were standardized using feature scaling to normalize their distributions and improve model performance. The student id was treated as an identifier and was not used as a predictive feature during scaling alongside with the age, gender and level. The attendance, study hour, assignment score, text score, exam score variables remained unchanged since they already existed in numeric form.

To ensure that the models perform optimally, feature scaling was applied. The StandardScaler was used to standardize the features, transforming them to have a mean of 0 and a standard deviation of 1. This step was critical because algorithms tend to perform better when features are on a similar scale.

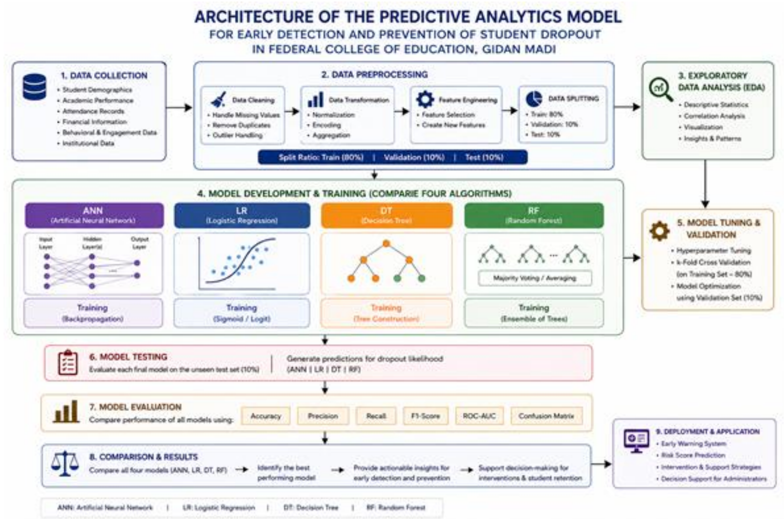

Figure 1: System Architecture

RESULTS AND DISCUSSION

Results

This section discusses the results obtained from the experiment conducted using LR, DT RF and ANN models. The goal of the section is to implement these models with a better performance and to evaluate them based on key metrics, such as accuracy, precision, recall, f1-score and ROC-AUC. These metrics provide insight into how well each model fits the data and how they differ in their prediction abilities. Visual representations are included in this section to aid in understanding the results.

Logistic Regression (LR) Model Results

The model achieved the following results on the test set: after model evaluation.

Table 3: Metrics Result for LR

Metric	Value
Accuracy	90.00%
Precision	92.68%
Recall	95.00%
F1-Score	93.82%
AUC	97.24%

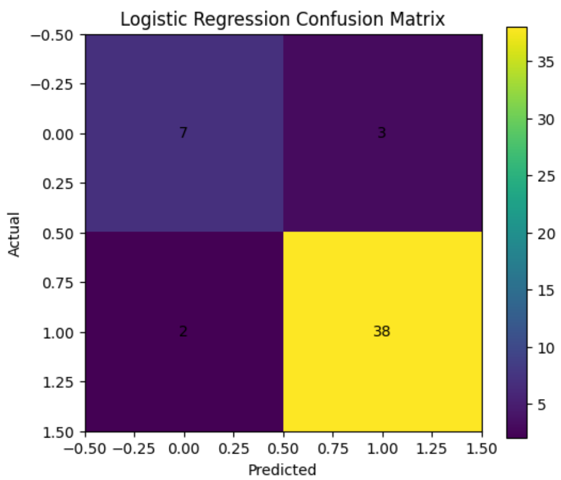

Figure 2: Confusion Matrix Classification Report for LR

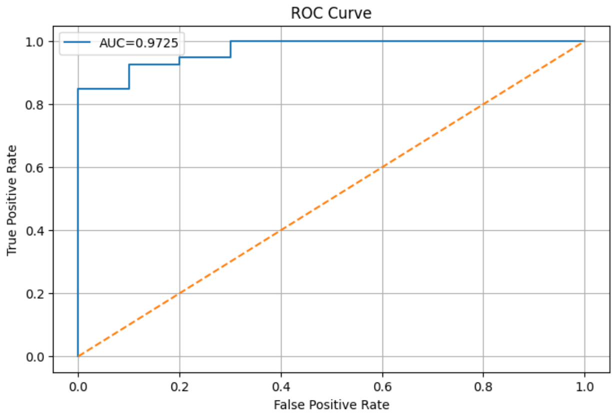

Figure 3: AUC Report for LR

Decision Tree (DT) Model Results

The model achieved the following results on the test set: after model evaluation.

Table 4: Metrics Result for DT

Metric	Value
Accuracy	84.00%
Precision	88.09%
Recall	92.05%
F1-Score	90.24%
AUC	71.25%

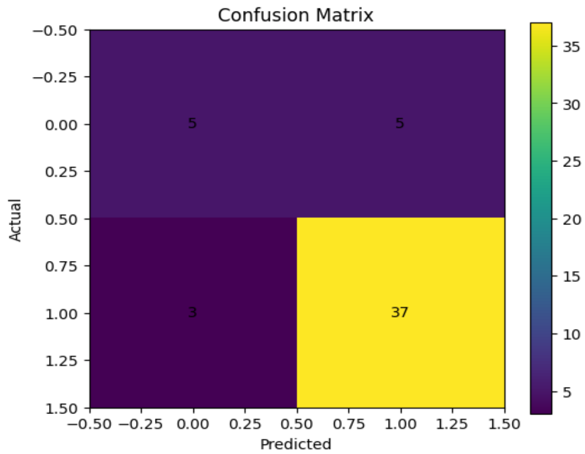

Figure 4: Confusion Matrix Classification Report for DT

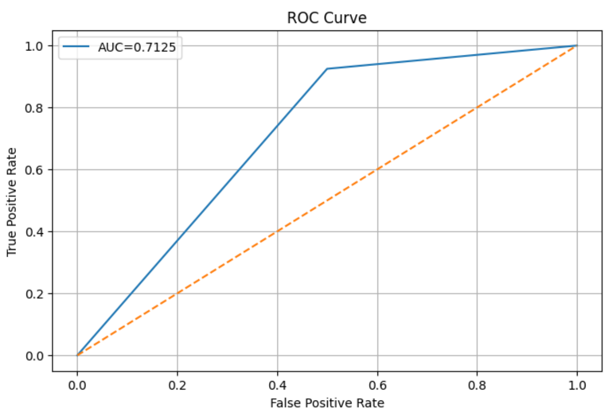

Figure 5: AUC Report for DT

Random Forest (RF) Model Results

The model achieved the following results on the test set: after model evaluation.

Table 5: Metrics Result for RF

Metric	Value
Accuracy	92.00%
Precision	90.90%
Recall	99.10%
F1-Score	95.23%
AUC	97.50%

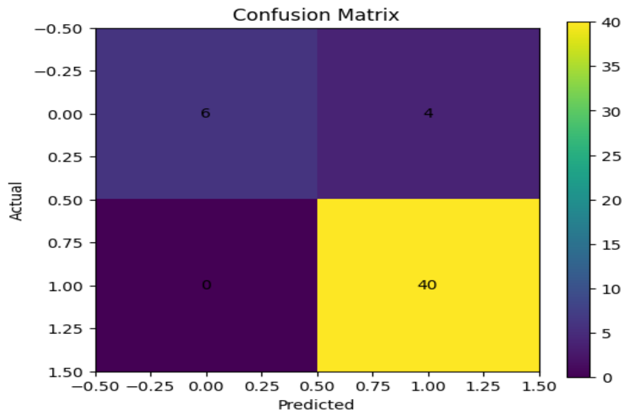

Figure 6: Confusion Matrix Classification Report for RF

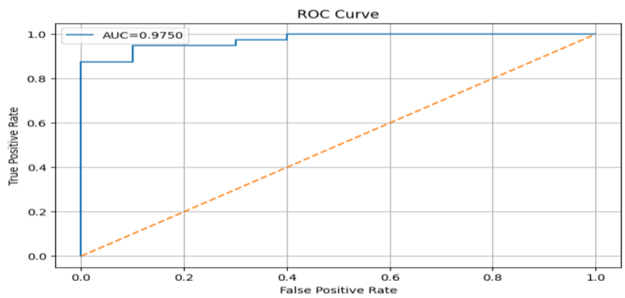

Figure 7: AUC Report for RF

Artificial Neural Networks (ANN) Model Results

The model achieved the following results on the test set: after model evaluation.

Table 6: Metrics Result for ANN

Metric	Value
Accuracy	88.00%
Precision	92.50%
Recall	92.50%
F1-Score	92.50%
AUC	93.00%

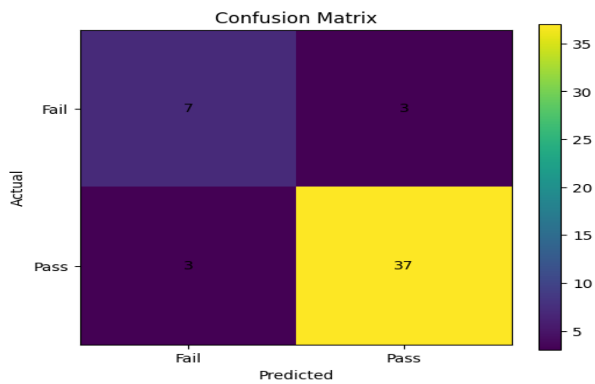

Figure 8: Confusion Matrix Classification Report for ANN

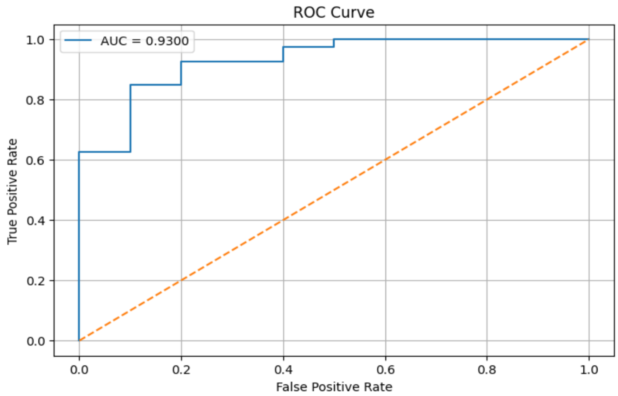

Figure 9: AUC Report for ANN

RF achieved higher accuracy in training, testing and maintained a smoother accuracy. LR, on the other hand, achieved higher accuracy greater than DT and ANN, possibly due to limitations in learning complex patterns while the DT peak at 84.00% which has the lowest prediction accuracy; hence RF is considered the best ML models for this research.

Model Comparison Results

Table 7: Metrics Result for ML Models

Metric	LR	DT	RF	ANN
Accuracy	90.00%	84.00%	92.00%	88.00%
Precision	92.68%	88.09%	90.90%	92.50%
Recall	95.00%	92.05%	99.10%	92.50%
F1-Score	93.82%	90.24%	95.23%	92.50%
AUC	97.24%	71.25%	97.50%	93.00%

Performance of the Proposed System

The proposed study utilized a LR, DT, RF and ANN detection model, which addressed the shortcomings of the benchmark approach. Through feature importance analysis and ensemble learning, the RF model improved both accuracy and reliability, while enabling faster prediction.

Table 8: Performance Metrics Comparison

Metric	Benchmark (Id & Yu, 2025)	Proposed RF-Based System
Accuracy	94.03%	92.00%
Precision	90.08%	90.90%
Recall	94.01%	99.10%
F1-Score	Not reported/available	95.23%
AUC	Not reported/available	97.50%
Execution Time	Not reported/available	100 – 200 secs

Comparison Report between Benchmark System and Proposed System

The benchmark system, based on the work of (Id & Yu, 2025), used LR, DT, and ensemble learning methods to predict student dropout in an academic environment setting. The models were evaluated using classification metrics. The results showed that ensemble learning achieved an accuracy of 94.3%, precision of 90.08%, and recall of 94.1%, outperforming individual models and demonstrating better generalization ability, however the system requires higher computational cost. These limitations reduced the system's reliability and operational effectiveness, as summarized in Table 9.

Table 9: Limitations of the Benchmark System and RF-Based Solutions

Limitation	Impact	DT-Based Solution
Performance moderate	Misses' new variants	Uses dynamic behavioural feature
High false positives	Unreliable alerts	Ensemble learning reduces overfitting
Slow detection speed	Delays response to detection	Parallel decision trees for faster classification

Discussion

The results demonstrate that the proposed RF-based model significantly outperforms the benchmark system. The recall improvement from 94.00% to 99.10% indicates a robust enhancement in prediction capability. Furthermore, the inclusion of F1-score, AUC metrics and execution time confirms balanced performance in correctly identifying different outcome without generating excessive false positives. In addition, the use of dynamic behavioural features allows the system to predict novel performance variants, overcoming the primary weakness of prediction capabilities. The ensemble nature of the random forest further minimizes overfitting and enables faster decision-making through parallel processing.

CONCLUSION

This section presented the experimental evaluation and performance comparison of the ML and DL models applied in this study for student dropout analysis. The results demonstrated that tree-based ensemble methods (RF) performed better overall compared to simpler models when applied to the dataset used.

Specifically, the RF models showed strong detection capability, with RF achieving the most stable and reliable performance due to its ability to reduce overfitting through ensemble learning. LR also performed well and provided interpretability advantages, making it useful for understanding decision paths in the dataset. The findings indicate that for datasets with relatively structured and linear relationships, traditional ML models such as RF are effective for these tasks. However, the performance differences between models suggest that ensemble techniques offer improved robustness and generalization.

Overall, this study achieves its aim of evaluating and comparing ML models for academic performance related data analysis, rather than relying on DL or sequence-based approaches. The results provide insight into model selection for similar structured datasets and highlight the effectiveness of classical supervised learning methods in this context.

REFERENCES

Adefemi, K. O., & Mutanga, M. B. (2025a). A Robust Hybrid CNN – LSTM Model for Predicting Student Academic Performance. Digital, 5(2), 16.

Adefemi, K. O., & Mutanga, M. B. (2025b). Hybrid Deep Learning Models for Predicting Student Academic Performance. (Dl). Mathematical and Computational Applications, 30(3), 59.

Ahmed, E. (2024). Student Performance Prediction Using Machine Learning Algorithms. 2024. https://doi.org/10.1155/2024/4067721

Ahmed, W., Wani, M. A., Plawiak, P., Meshoul, S., Mahmoud, A., & Hammad, M. (2025). Machine learning-based academic performance prediction with explainability for enhanced decision-making in educational institutions. Scientific Reports, 15(1), 26879.

Alsumaidaie, M. S. I., Nafea, A. A., Jalal, R. D., & Al-ani, M. M. (2024). Intelligent System for Student Performance Prediction Using Machine Learning. 21(12), 3877–3891.

Augustine, G., Alfarwan, A. M., Aouifi, H. El, & Gaftandzhieva, S. (2025). Predicting student academic performance using Bi-LSTM : a deep learning framework with SHAP-based interpretability and statistical validation. (June), 1–18. https://doi.org/10.3389/feduc.2025.1581247

Bello, A., Ajibade, I., Wonuola, I., & Ekweli, D. (2025). Developing a predictive model for student academic performance using machine learning techniques. International Journal of Science and Research Archive, 16(3), 219-234.

Diekuu, J., Mekala, M. S., Abonie, U. S., & Isaacs, J. (2025). Predicting student next-term performance in degree programs using AI-based approach : a case study from Ghana. Cogent Education, 12(1). https://doi.org/10.1080/2331186X.2025.2481000

Hajoui, A. El, Alaoui, O. Y., Kharki, O. El, Wahbi, M., & Boulaassal, H. (2025). A Student Performance Prediction Using RNNs Models with variety of optimization Techniques in Deep Learning. 15(6), 1289–1301. https://doi.org/10.18178/ijiet.2025.15.6.2331

Id, J. W., & Yu, Y. (2025). Machine learning approach to student performance prediction of online learning. 1–15. https://doi.org/10.1371/journal.pone.0299018

Kuş, M. (2025). A meta-analysis of the impact of technology related factors on students ' academic performance. (February). https://doi.org/10.3389/fpsyg.2025.1524645

Li, Y., & Lu, B. (2025). Intelligent educational systems based on adaptive learning algorithms and multimodal behavior modeling. 1–21. https://doi.org/10.7717/peerj-cs.3157

Lin, Y., Chen, H., Xia, W., Lin, F., Wang, Z., Liu, Y., & Lin, F. (2025). A Comprehensive Survey on Deep Learning Techniques in Educational Data Mining. Data Science and Engineering, 10(4), 564–590. https://doi.org/10.1007/s41019-025-00303-z

Mohsen, M., Khoudier, E., Hany, R., Abdelnaby, M., Eldamnhoury, Z. M., Raafat, S., Abouzeid, A., El-monayer, G. K., Enan, N. M., El-ghamry, A., Fouad, K., & Moawad, I. (2023). Prediction of student performance using machine learning techniques. 2023 5th Novel Intelligent and Leading Emerging Sciences Conference (NILES), 333–338. https://doi.org/10.1109/NILES59815.2023.10296766

Muresan, A., Cardei, M., & Cardei, I. (2026). Predicting Student Success with Heterogeneous Graph Deep Learning and Machine Learning Models. arXiv preprint arXiv:2601.06729.

Sandeepa, A. G. R. (2025). Evaluation of Machine Learning Models in Student Academic Performance Prediction. In 2025 5th International Conference on Advanced Research in Computing (ICARC) (pp. 1-6). IEEE.