Cureus Journal of Computer Science
https://doi.org/10.7759/s44389-026-00036-8

Original Article

A Machine Learning and Explainable Artificial Intelligence Approach to Student Dropout Prediction Using Multidimensional Educational Data

Subhiksha S R [1]

1. *Information Technology, Sri Ramakrishna Institute of Technology, Coimbatore, IND*

Received: February 05, 2026 | Review began: February 10, 2026 | Review ended: February 24, 2026 | Published: February 27, 2026

Abstract

Student dropout prediction is a critical task for educational institutions seeking to enhance academic performance and reduce attrition. This study presents a machine learning-based framework that integrates comprehensive preprocessing, baseline model evaluation, and advanced ensemble learning for accurate dropout prediction. Six classical classifiers Logistic Regression, Decision Tree, Random Forest, Support Vector Machine, K-Nearest Neighbors, and Naïve Bayes are assessed to establish baseline performance. To improve predictive effectiveness, a Light Gradient-Boosting Machine model is employed and further enhanced through a hybrid stacked ensemble combining Random Forest, Light Gradient-Boosting Machine, and Support Vector Machine, with Logistic Regression as the meta-learner. The system extends beyond binary classification by introducing a three-level risk categorization (low, medium, high), enabling more focused interventions. Explainable AI techniques, specifically SHapley Additive exPlanations and Local Interpretable Model Agnostic Explanation, are incorporated to provide transparent global and local factor interpretations. The framework is supported by an interactive dashboard, demonstrating strong predictive accuracy, interpretability, and practical applicability for early identification of at-risk students.

Categories: AI/ML-based decision support systems, Predictive Analytics, Machine Learning (ML)

Keywords: student dropout prediction, machine learning, ensemble learning, explainable artificial intelligence, shap analysis, lime interpretation, educational data mining, risk classification

Introduction

Student dropout remains a major challenge for educational institutions worldwide, affecting academic quality, institutional performance, and long-term student success. Understanding and addressing dropout is therefore a critical concern in higher education systems.

Early foundational research by Tinto [1] conceptualized student dropout as a multifaceted phenomenon driven by academic and social integration within institutions, rather than as a predictive or computational problem. This theoretical perspective was later empirically validated by Cabrera et al. [2], who demonstrated the combined influence of academic, social, and environmental factors on student persistence using structural equation modeling. While these studies established a strong conceptual and statistical foundation, they did not address automated or early-stage dropout prediction.

With the increasing availability of large-scale educational data, educational data mining and machine learning techniques have been widely adopted to predict student performance and dropout risk. Early studies demonstrated the effectiveness of mining enrollment and academic data for student success prediction [3], while subsequent research explored various data mining and genetic programming approaches to identify at-risk students [4,5]. Several works further confirmed the suitability of classification-based models for improving student performance and retention outcomes [6,7]. More recent studies proposed machine learning-based dropout early warning systems and advanced predictive models that achieve high accuracy, precision, and recall across diverse educational settings [8-11].

Despite their predictive effectiveness, many machine learning models function as black-box systems, offering limited transparency into their decision-making processes. This lack of interpretability reduces trust among educators and limits practical adoption. To address this issue, explainable artificial intelligence techniques such as SHapley Additive exPlanations (SHAP) and Local Interpretable Model Agnostic Explanation (LIME) have been introduced to provide meaningful explanations of model predictions [12,13]. Recent research further emphasizes the importance of interpretable machine learning specifically within educational contexts to support informed decision-making and stakeholder trust [14-16].

Motivated by these challenges, this study proposes a student dropout prediction framework that combines classical machine learning models, boosting techniques, and a hybrid stacked ensemble to enhance predictive performance. The framework further integrates SHAP and LIME to deliver both global and instance-level explanations, thereby supporting transparent, accurate, and actionable dropout prediction for effective academic intervention.

Materials And Methods

The proposed framework follows a multi-stage methodological pipeline designed to achieve high predictive accuracy, model robustness, and strong interpretability. The overall methodology consists of four major components: (i) data preprocessing and cleaning, (ii) baseline model evaluation, (iii) advanced ensemble and hybrid modeling, and (iv) explainable AI integration. The overall workflow of the proposed student dropout prediction framework, including data preprocessing, baseline modeling, ensemble learning, and explainable AI integration, is illustrated in Figure *1*.

How to cite this article:
S R S (February 27, 2026) A Machine Learning and Explainable Artificial Intelligence Approach to Student Dropout Prediction Using Multidimensional Educational Data. Cureus J Comput Sci 3 : es44389-026-00036-8. DOI https://doi.org/10.7759/s44389-026-00036-8

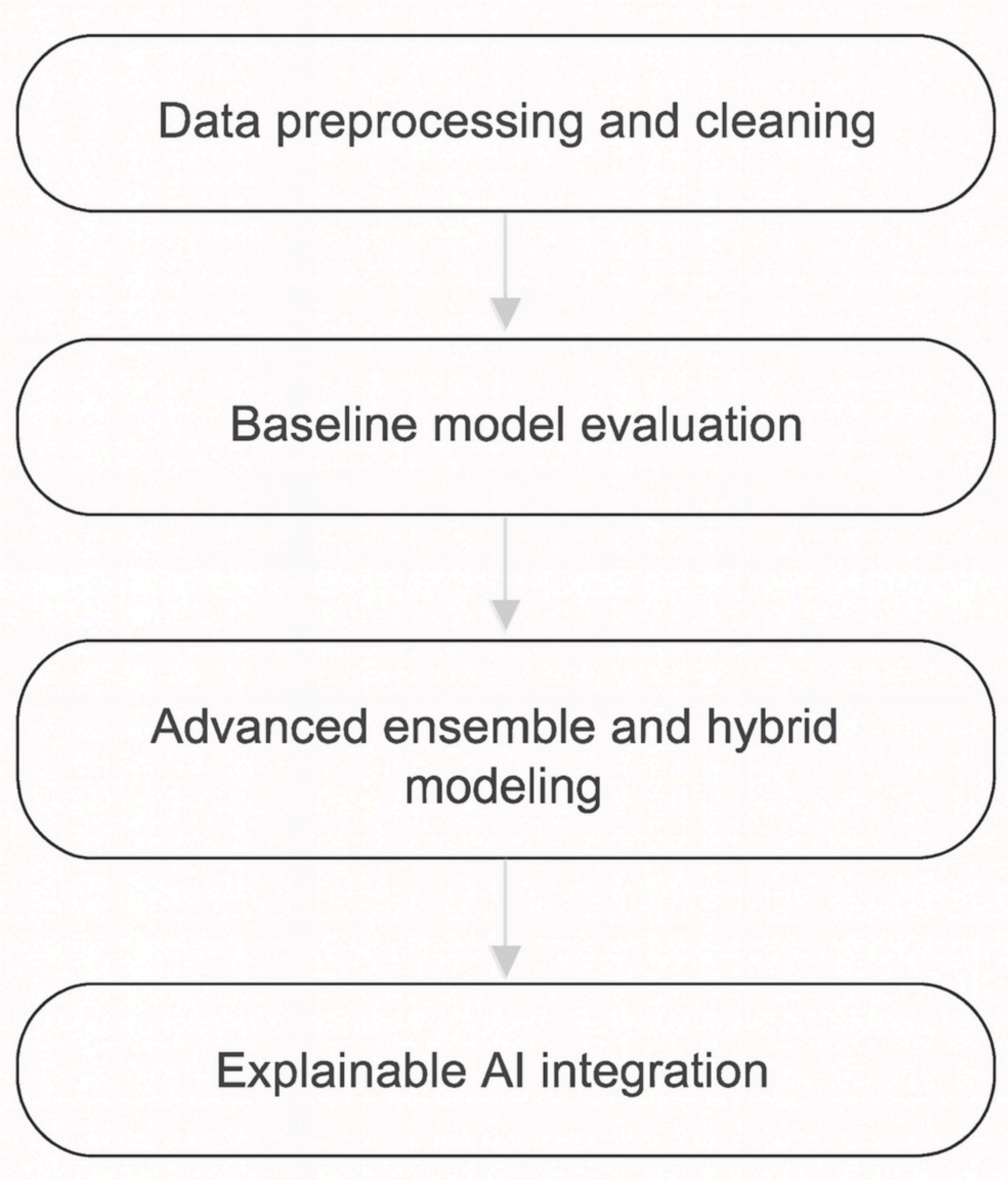

FIGURE 1: Workflow of the proposed framework

Dataset description

The dataset consists of 3,000 student records and 31 attributes, covering demographic details, academic performance, behavioral indicators, engagement metrics, and historical patterns. Key numerical features include GPA scores, attendance rates, assignment submission percentage, study hours, family income, and psychological stress levels.

How to cite this article:
S R S (February 27, 2026) A Machine Learning and Explainable Artificial Intelligence Approach to Student Dropout Prediction Using Multidimensional Educational Data. Cureus J Comput Sci 3 : es44389-026-00036-8. DOI https://doi.org/10.7759/s44389-026-00036-8

Categorical features include gender, study mode, region, parental education, and access to learning devices. The target variable is a binary label, indicating whether a student dropped out (1) or continued (0). The class distribution is moderately imbalanced, with approximately 20% dropout instances.

Data preprocessing and feature engineering

Data preprocessing is performed as a foundational step to ensure data quality and reliability. Missing numerical values are imputed using median statistics, while categorical attributes are filled using mode-based imputation. Numerical features are standardized using z-score normalization, and categorical variables are encoded using one-hot encoding.

To identify influential predictors, correlation analysis and mutual information scores are computed. Redundant or weakly associated features are removed, resulting in a refined feature subset used for modeling. This reduces noise, prevents multicollinearity, and enhances downstream model performance.

Class Imbalance Handling

To mitigate potential bias arising from class imbalance in the dropout dataset, stratified sampling was employed during the train-test split to preserve the original class distribution across training and testing subsets. This ensures that both subsets maintain proportional representation of dropout and non-dropout instances.

Furthermore, class-weight adjustments were incorporated in model training for imbalance-sensitive algorithms. Specifically, for the LightGBM classifier, the scale_pos_weight parameter was computed as the ratio of negative to positive samples in the training set. This weighting mechanism penalizes misclassification of minority-class instances more heavily, thereby reducing bias toward the majority class and improving recall performance for dropout prediction.

No synthetic resampling techniques (e.g., SMOTE) were applied in this study in order to avoid introducing artificial distributional artifacts into the dataset.

Baseline model development

Six classical machine learning algorithms - Logistic Regression, Decision Tree, Random Forest, Support Vector Machine (SVM), K-Nearest Neighbors (KNN), and Naïve Bayes - are implemented to establish baseline performance. These models are selected to represent a diverse range of linear, tree-based, probabilistic, and distance-based learning techniques. Each model is trained and evaluated using accuracy, precision, recall, F1-score, and area under the ROC curve (AUC).

The baseline comparison provides insight into the strengths and limitations of traditional classifiers and serves as a benchmark for more advanced models.

Boosting and hybrid stacked ensemble

To enhance predictive performance, a high-performance LightGBM model is employed due to its fast training speed, ability to handle large feature spaces, and suitability for imbalanced datasets.

Building on this, a hybrid stacked ensemble is developed, combining three high-performing Level-1 models - Random Forest, LightGBM, and SVM. Logistic Regression is used as the Level-2 meta-learner, aggregating outputs from the base learners. This stacked architecture reduces variance, improves generalization, and consistently outperforms individual models.

To mitigate potential overfitting in the stacked ensemble, out-of-fold cross-validation predictions were generated during Level-1 model training and used as inputs for the meta-learner. This prevents information leakage between training stages and improves generalization performance. Additionally, regularization parameters and early stopping mechanisms were applied within the LightGBM model to further enhance robustness.

Statistical validation and model stability assessment

Five-fold stratified cross-validation was employed to assess model robustness and reduce performance variance arising from random train-test splits. For each fold, accuracy, precision, recall, F1-score, and AUC were recorded. The final reported performance values correspond to the mean ± standard deviation across folds.

How to cite this article:
S R S (February 27, 2026) A Machine Learning and Explainable Artificial Intelligence Approach to Student Dropout Prediction Using Multidimensional Educational Data. Cureus J Comput Sci 3 : es44389-026-00036-8. DOI https://doi.org/10.7759/s44389-026-00036-8

To evaluate whether performance improvements of the proposed stacked ensemble were statistically significant compared to baseline models, a paired t-test was conducted on fold-wise F1-scores. Statistical significance was established at $p < 0.05$.

Risk categorization and explainable AI integration

The predicted dropout probabilities are converted into three risk categories Low, Medium, and High based on predefined threshold ranges, as shown in Table *1*. This facilitates actionable intervention strategies.

While traditional dropout prediction models often rely on binary classification, educational intervention strategies require more granular risk assessment. Therefore, predicted probabilities were converted into three actionable risk levels (Low, Medium, High) to enable differentiated institutional responses. This stratified approach allows preventive monitoring for low-risk students, targeted academic support for medium-risk students, and intensive intervention for high-risk cases.

Risk Level	Probability Range
Low	0.00–0.35
Medium	0.35–0.70
High	0.70–1.00

TABLE 1: Risk categorization thresholds

Explainability is incorporated through SHAP for global and local feature interpretation and LIME for instance-level explanations. SHAP provides summary, dependence, feature importance, force, and decision plots, revealing how each feature contributes to model output. LIME highlights key factors influencing individual student predictions. This ensures transparency and enhances trust among educators and decision-makers.

Use of artificial intelligence tools

The authors used a large language model-based tool to assist with language refinement, grammar correction, and improvement of clarity during manuscript preparation. The tool was not used for data analysis, result generation, or decision-making. All scientific content, interpretations, and conclusions were developed and verified by the authors, who take full responsibility for the integrity and accuracy of the work. External analytical resources and interactive enrichment platforms were consulted as supporting tools during the methodological design phase [17].

Results And Discussion

Experimental setup

The experiments were conducted on a dataset comprising 3,000 student records with 31 features. The data were split into training (80%) and testing (20%) sets using stratified sampling. All models were implemented in Python using scikit-learn and LightGBM libraries. Model performance was evaluated using accuracy, precision, recall, F1-score, and area under the ROC curve (AUC).

The comparative performance of all baseline machine learning models in terms of accuracy, precision, recall, F1-score, and AUC is presented in Table *2*. The ROC curves of the baseline classifiers, illustrating their discriminative ability, are shown in Figure *2*. However, tree-based models such as Decision Tree and Random Forest show inconsistent recall for minority dropout cases, indicating potential overfitting or sensitivity to imbalanced data. The comparative accuracy performance of all baseline machine learning models is presented in Figure 3.

Model	Accuracy	Precision	Recall	F1-Score	AUC
Logistic Regression	0.972	0.948	0.91	0.929	0.994
Support Vector Machine	0.96	0.938	0.861	0.897	0.988
Random Forest	0.92	0.963	0.631	0.762	0.978
Naïve Bayes	0.9	0.787	0.697	0.739	0.959
K-Nearest Neighbors	0.9	0.908	0.566	0.697	0.923
Decision Tree	0.873	0.68	0.713	0.696	0.814

TABLE 2: Baseline comparsion performance metrics

AUC, Area Under the ROC Curve

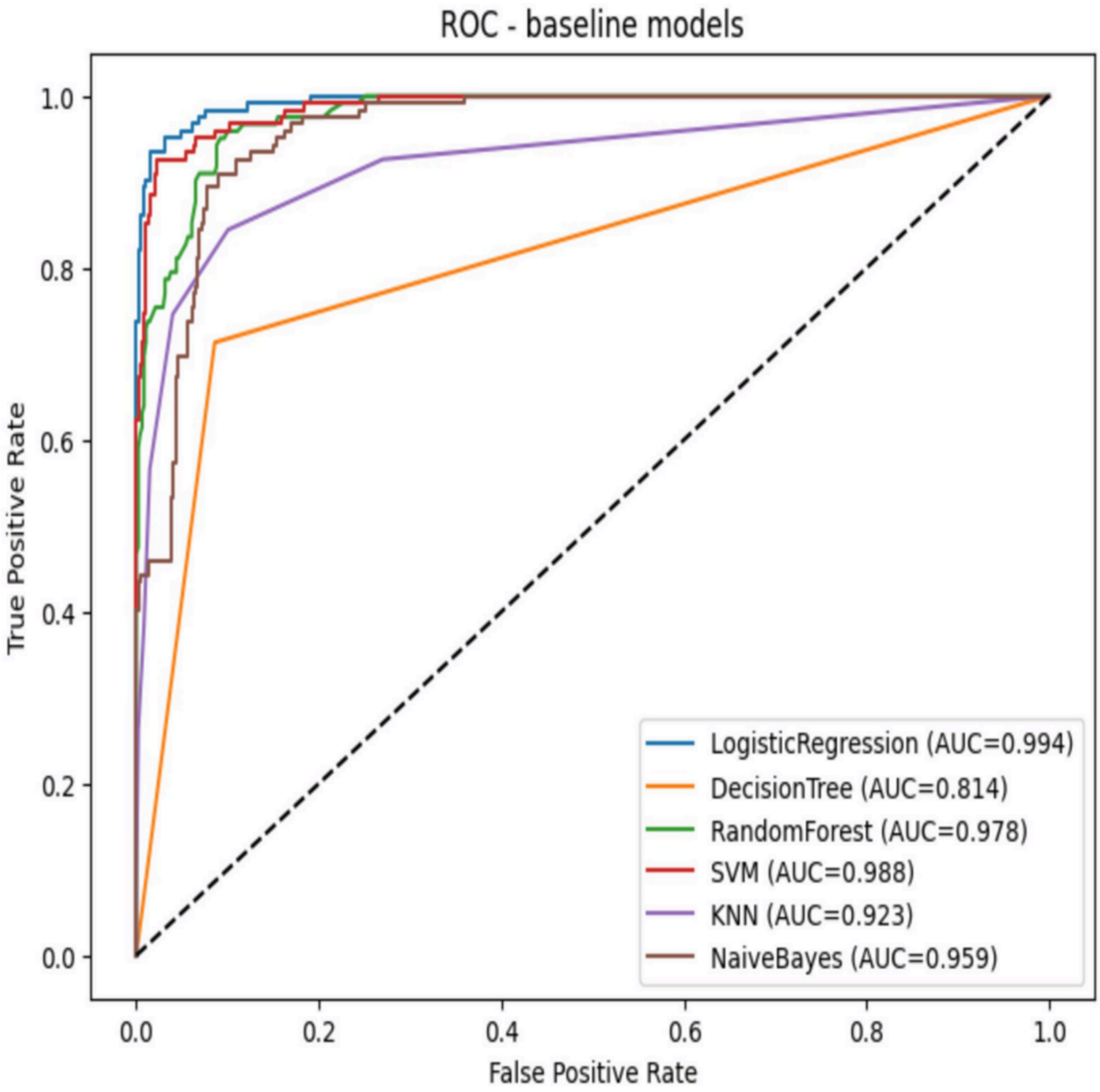

FIGURE 2: ROC baseline models

AUC, Area Under the ROC Curve; KNN, K-Nearest Neighbors; ROC, Receiver Operator Characteristic; SVM, Support Vector Machine

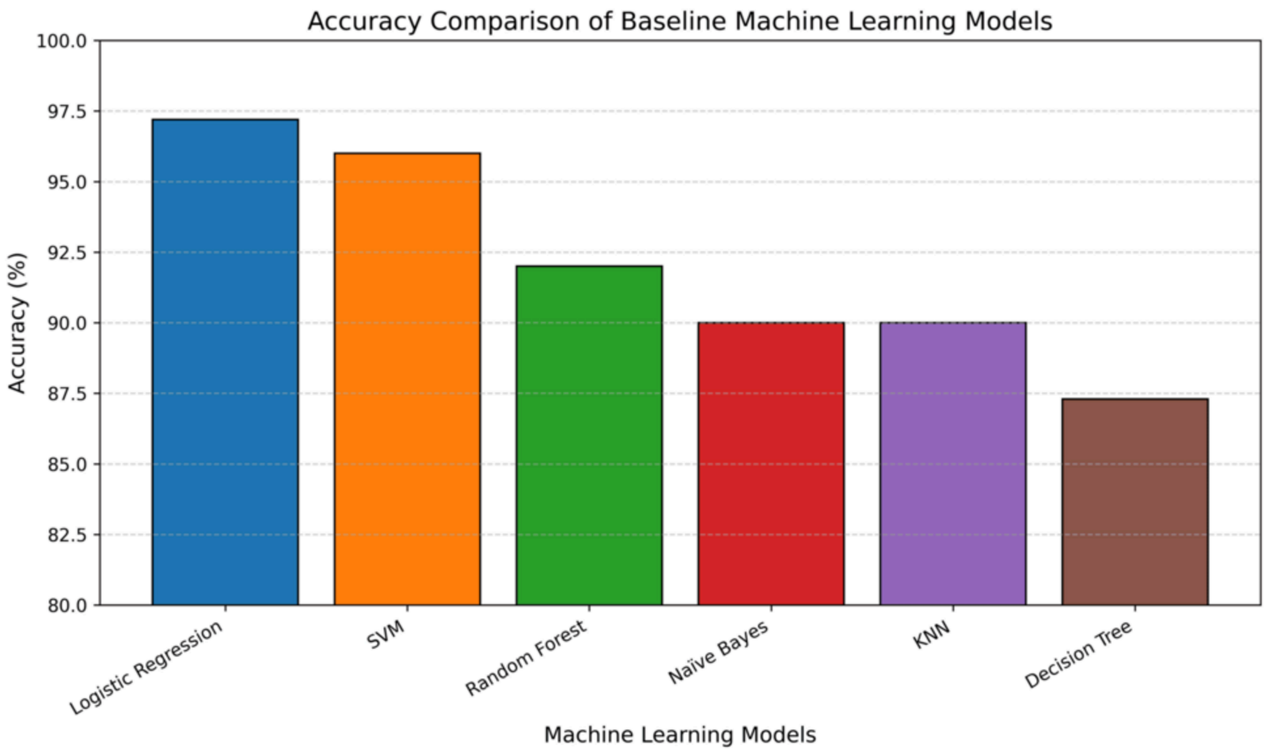

FIGURE 3: Model accuracy comparsion

KNN, K-Nearest Neighbors; SVM, Support Vector Machine

The LightGBM model improves overall prediction quality, achieving high AUC (0.98+) and robust recall for dropout instances. The proposed hybrid stacked ensemble further strengthens performance, achieving the best results with an accuracy of 96.5%, a recall of 87.7% for dropout class, and an AUC of 0.989. This confirms its ability to generalize well and reduce individual model weaknesses. The confusion matrix of the proposed hybrid stacked ensemble model on the test dataset is shown in Figure *4*.

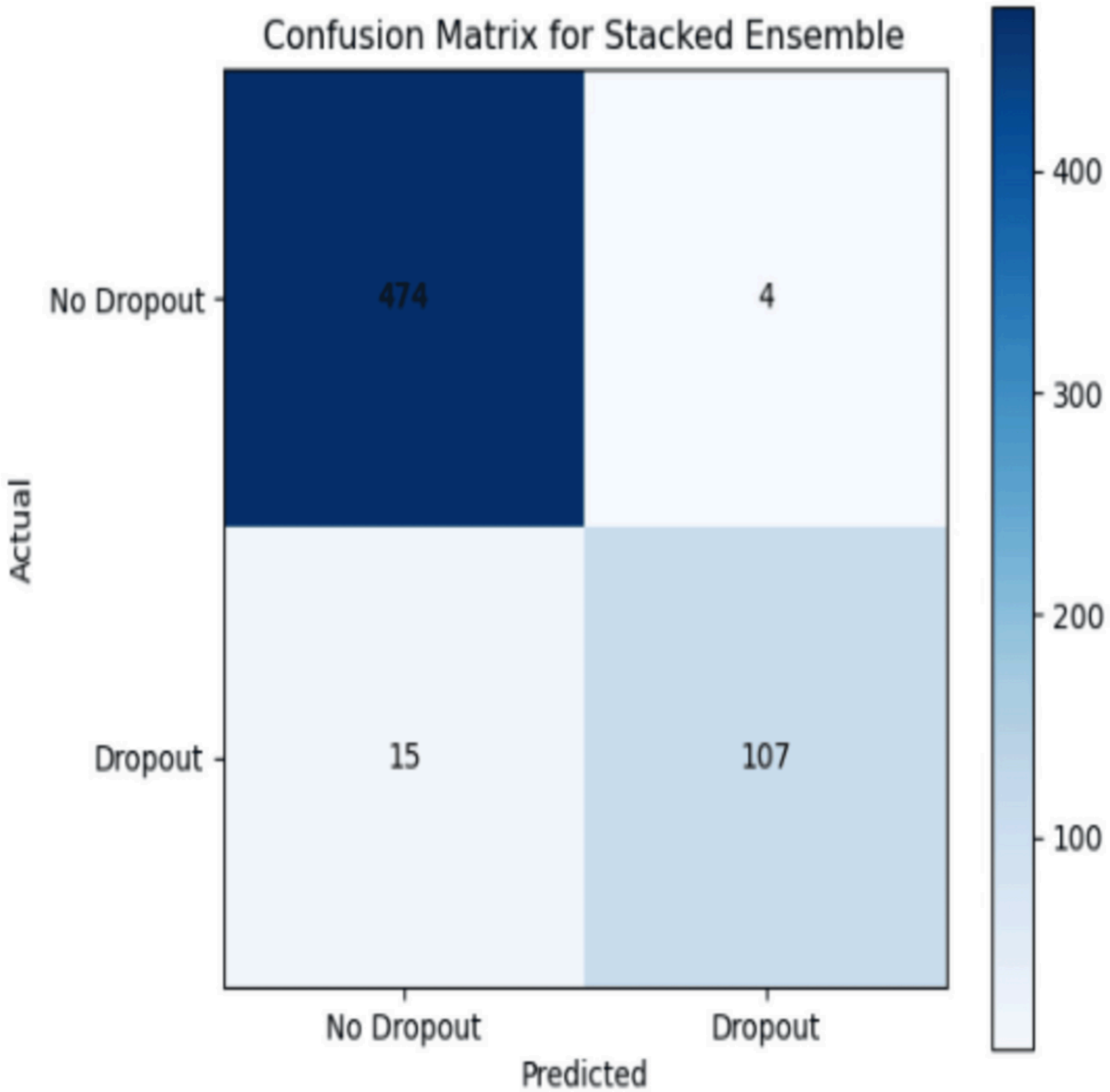

FIGURE 4: Confusion matrix of the proposed hybrid stacked ensemble on the test dataset

Explainability results show that factors such as days since enrollment, family income, assignment submission rate, prior GPA, engagement score, and attendance rate significantly influence dropout risk. SHAP and LIME visualizations clearly highlight personalized risk drivers for individual students, enabling educators to understand specific vulnerabilities and plan targeted support. The Streamlit-based interactive student dropout prediction dashboard is shown in Figure *5*.

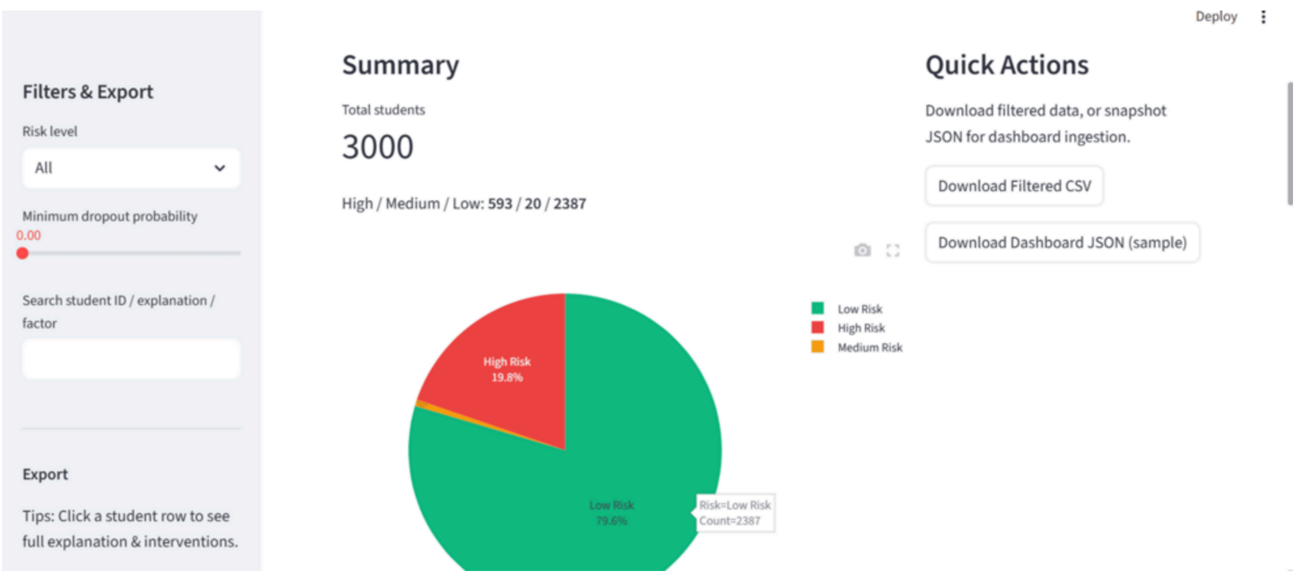

FIGURE 5: Streamlit dashboard - student dropout prediction

External validity and generalizability

The current study utilizes a synthetic/multidimensional educational dataset to evaluate the proposed framework. While this allows controlled experimentation and robust model comparison, real-world institutional datasets may exhibit distributional shifts, missing values, and demographic variability. Therefore, external validation across multi-institutional datasets is recommended to assess cross-domain generalizability.

Conclusions

This study presents an integrated machine learning and explainable AI framework for early student dropout prediction. The combination of systematic preprocessing, baseline evaluation, a high-performance LightGBM model, and a hybrid stacked ensemble demonstrates improved predictive performance and model robustness. The incorporation of SHAP and LIME enhances transparency, enabling educators to interpret both global feature importance and instance-level explanations. Moreover, the three-level risk categorization framework provides structured support for differentiated intervention strategies, enhancing the practical relevance of the proposed approach. The accompanying interactive dashboard further illustrates the potential applicability of the framework within institutional decision-support systems.

Future research will focus on validating the proposed framework using real-world, multi-institutional academic datasets to strengthen external validity and assess cross-domain generalizability. Cross-institution evaluation will enable systematic assessment of robustness across diverse demographic, socioeconomic, and curricular contexts. From a methodological perspective, future extensions may incorporate temporal deep learning architectures, such as Long Short-Term Memory networks, to capture longitudinal academic and behavioral trends. Additionally, integrating richer psychological and socio-emotional indicators may further improve early detection performance. To enhance deployment readiness, domain adaptation and transfer learning strategies will be explored to facilitate adaptation across institutions with varying data distributions. Furthermore, development of an automated intervention recommendation module and potential deployment as a scalable cloud-based decision-support system may support large-scale educational environments seeking proactive, data-driven dropout prevention strategies.

Additional Information

Author Contributions

All authors have reviewed the final version to be published and agreed to be accountable for all aspects of the work.

Concept and design: Subhiksha S R

Acquisition, analysis, or interpretation of data: Subhiksha S R

Drafting of the manuscript: Subhiksha S R

Disclosures

Human subjects: All authors have confirmed that this study did not involve human participants or tissue. **Animal subjects:** All authors have confirmed that this study did not involve animal subjects or tissue. **Conflicts of interest:** In compliance with the ICMJE uniform disclosure form, all authors declare the following: **Payment/services info:** All authors have declared that no financial support was received from any organization for the submitted work. **Financial relationships:** All authors have declared that they have no financial relationships at present or within the previous three years with any organizations that might have an interest in the submitted work. **Other relationships:** All authors have declared that there are no other relationships or activities that could appear to have influenced the submitted work.

Data Availability Statements

The datasets (and/or code) supporting this study are available from the corresponding author upon reasonable request. The datasets and program code supporting the findings of this study are available from the corresponding author upon reasonable request.

Acknowledgements

The author declares that this research was conducted independently. The dataset used in this study is a synthetic dataset generated exclusively for research and experimental purposes. It does not contain real student information and was designed to simulate educational attributes relevant to dropout prediction. While synthetic data enable controlled experimentation and methodological evaluation, future validation using real-world institutional datasets is necessary to assess generalizability. The datasets and study materials used in this study are available from the author upon reasonable request.

References

1. Tinto V: Dropout from higher education: A theoretical synthesis of recent research. Review of Educational Research. 1975, 45:89-125.
2. Cabrera AF, Nora A, Castañeda MB: College persistence: Structural equations modeling test of an integrated model of student retention. The Journal of Higher Education. 1993, 64:123-139. 10.1080/00221546.1993.11778419
3. Kovačić ZJ: Early prediction of student success: Mining student enrollment data. Proceedings of Informing Science & IT Education Conference (InSITE) 2010. 2010, 647-665.
4. Márquez-Vera C, Cano A, Romero C, Ventura S: Predicting student failure at school using genetic programming and different data mining approaches with high dimensional and imbalanced data. Applied Intelligence. 2013, 38:315-330. 10.1007/s10489-012-0374-8
5. Lundberg SM, Lee SI: A unified approach to interpreting model predictions. 31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA. 2017, 1-10.
6. Ribeiro MT, Singh S, Guestrin C: "Why should I trust you?": Explaining the predictions of any classifier. KDD '16: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2016, 1135-1144. 10.1145/2939672.2939778
7. Bunkar K, Singh UK, Pandya B, Bunkar R: Data mining: Prediction for performance improvement of graduate students using classification. 2012 Ninth International Conference on Wireless and Optical Communications Networks (WOCN), Indore, India.

2012, 1-5. 10.1109/WOCN.2012.6335530

8. Bhat M, Long D: Designing interactive explainable AI tools for algorithmic literacy and transparency. DIS '24: Proceedings of the 2024 ACM Designing Interactive Systems Conference. 2024, 939-957. 10.1145/3643834.3660722
9. Conati C, Porayska-Pomsta K, Mavrikis M: AI in education needs interpretable machine learning: Lessons from open learner modelling. arXiv. 2018, 10.48550/arXiv.1807.00154
10. Shahiri AM, Husain W, Abdul Rashid N: A review on predicting student's performance using data mining techniques. Procedia Computer Science. 2015, 72:414-422. 10.1016/j.procs.2015.12.157
11. Márquez-Vera C, Cano A, Romero C, Mohammad Noaman AY, Fardoun HM, Ventura S: Early dropout prediction using data mining: a case study with high school students. Expert Systems. 2016, 33:107-124. 10.1111/exsy.12135
12. Lee S, Chung JY: The machine learning-based dropout early warning system for improving the performance of dropout prediction. Applied Sciences. 2019, 9:3093. 10.3390/app9153093
13. Krüger JGC, de Souza Britto Jr A, Barddal JP: An explainable machine learning approach for student dropout prediction. Expert Systems with Applications. 2023, 233:120933. 10.1016/j.eswa.2023.120933
14. Kim S, Choi E, Jun Y-K, Lee S: Student dropout prediction for university with high precision and recall. Applied Sciences. 2023, 13:6275. 10.3390/app13106275
15. Kim S, Choi E, Jun Y-K, Lee S: All-year dropout prediction modeling and analysis for university students. Applied Sciences. 2023, 13:1143. 10.3390/app13021143
16. Gunning D, Stefik M, Choi J, Miller T, Stumpf S, Yang G-Z: XAI—Explainable artificial intelligence. Science Robotics. 2019, 4:eaay7120. 10.1126/scirobotics.aay7120
17. Enrichr. https://maayanlab.cloud/Enrichr/.

How to cite this article:
S R S (February 27, 2026) A Machine Learning and Explainable Artificial Intelligence Approach to Student Dropout Prediction Using Multidimensional Educational Data. Cureus J Comput Sci 3 : es44389-026-00036-8. DOI https://doi.org/10.7759/s44389-026-00036-8