iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 14 **Issue:** XI **Month of publication:** November 2026

DOI: https://doi.org/10.22214/ijraset.2026.85582

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue XI Nov 2026- Available at www.ijraset.com

Student Performance Prediction Using Machine Learning Techniques for Early Academic Intervention

Dr. Rahul M. Dhokane[1], Pratiksha Kisan Shirsath[2], Shreya Balasaheb Gite[3], Aditi Uttam Ugale[4], Kajal Ravsaheb Avhad[5]
[1,2,3,4,5]*Department of Computer Engineering, Sir Visvesvaraya Institute of Technology, Nashik, Maharashtra, India*

Abstract: ***Identifying weak students only after the final examination leaves little room for corrective action. Machine learning (ML) can analyse attendance, internal marks, assignment records, previous academic results and study behaviour to flag at-risk students while the course is still running. This paper reviews ten recent studies (2025–2026) on student performance, dropout-risk and early-warning prediction, covering classical classifiers, boosting ensembles, hybrid and stacked models, fuzzy decision support, deep learning and graph-based approaches. The studies are compared with respect to datasets, preprocessing, algorithms, evaluation metrics and reported outcomes. Recurring limitations are identified: small single-institution datasets, class imbalance, target leakage, optimistic validation, limited interpretability and a weak link between prediction and actual intervention. Based on these findings, a decision-support framework is proposed in which Logistic Regression, Decision Tree and Random Forest models are trained on academic and behavioural data, students are placed into risk tiers, and teachers receive alerts through a dashboard. The framework is designed to support, not replace, teacher judgement. Directions for future work are also outlined.***

Keywords: ***Student performance prediction, machine learning, educational data mining, early academic intervention, at-risk students, random forest, learning analytics, decision support system.***

I. INTRODUCTION

In most institutions a student's difficulties become visible only when the final examination result is declared. By then the semester is over, and counselling, extra classes or additional study material can no longer change the outcome. Traditional evaluation depends mainly on examination marks, which describe where a student ended up but say little about the trend that led there. When a class is large, teachers also cannot manually track the attendance, marks, assignments and behaviour of every student. Educational data mining (EDM) and learning analytics address this gap by applying data-driven models to academic records and learning-platform logs. A growing body of work now predicts grades, pass/fail status, dropout risk and academic level from demographic, academic, behavioural and engagement data [1]–[10]. Reported approaches range from interpretable decision trees and fuzzy rule bases to stacked ensembles, recurrent networks and hypergraph models. However, the studies use different datasets, targets and metrics, so their headline numbers cannot be compared directly, and many do not show how a prediction turns into a concrete action by a teacher. This paper makes three contributions. First, it reviews ten recent studies and compares their data, methods, evaluation and reported results. Second, it summarises the technical and methodological problems that recur across them. Third, it proposes a practical decision-support framework for early academic intervention that uses Logistic Regression, Decision Tree and Random Forest models together with a teacher-facing dashboard. The remainder of the paper is organised as follows: Section II reviews related work, Section III describes datasets, preprocessing and evaluation practice, Section IV summarises the algorithms, Section V presents the proposed framework, Section VI discusses findings and challenges, and Sections VII and VIII give future scope and conclusions.

II. RELATED WORK

A. Classical and Ensemble Models

Wang [1] combined Decision Tree, Random Forest, SVM and a neural network through soft voting and stacking and evaluated them on the UCI Student Performance, OULAD and NELS:88 datasets. The hybrid model is reported to beat every individual learner on all three datasets, with a peak accuracy of 98.8% on UCI. These figures are self-reported, and several tables in the paper are labelled "adjusted" by the author, so independent replication would be needed before relying on them.

©IJRASET: All Rights are Reserved

Alkan et al. [4] surveyed 613 secondary-school students with an 84-item questionnaire, selected 30 variables with the Boruta algorithm, handled class imbalance with SMOTE and compared C5.0, CART, SVM and Random Forest. Random Forest was the most stable under 10-fold cross-validation (74.3% accuracy), whereas C5.0 scored best on the held-out test split (75.4%) and CART fitted the training data best but generalised poorly. The authors converted the Random Forest coefficients into a simple additive score that flags students below a threshold. A validation on only 30 students is acknowledged as a limitation.

Gul et al. [5] compared eight regression-oriented models, including Decision Tree, Random Forest, Lasso, KNN, XGBoost, CatBoost, AdaBoost and Gradient Boosting, using MAE, MSE and RMSE. CatBoost gave the best accuracy (87.46%), followed by Gradient Boosting (87.28%) and Decision Tree (82.42%). Agal [6] addressed data scarcity by releasing a synthetic dataset of more than 100,000 student records with 28 attributes, and benchmarked eleven algorithms on two tasks. LightGBM reached a macro F1-score of 0.842 for student-level prediction, and an ordinal neural network reached 0.846 with a quadratic weighted kappa of 0.892, showing the benefit of modelling ordered classes explicitly. Prior academic achievement dominated feature importance, and SHAP was used for local explanations.

B. Engagement, Class Imbalance and Dropout Risk

Althaqafi et al. [8] used the xAPI-Edu-Data learning-management-system dataset and showed that oversampling with SMOTE, grid-search tuning and a stacking ensemble with Random Forest as the meta-learner raised weighted precision from 0.60 to 0.83. Individual models also improved after enrichment (SVM 0.81, Logistic Regression 0.80, Decision Tree 0.75). Their results indicate that engagement indicators such as raised hands, resource visits and discussion participation are useful predictors.

Hegde et al. [3] built a dropout-risk model from 253 undergraduate survey responses covering CGPA, attendance, stress, financial difficulty, health, course interest, thoughts of discontinuation and family problems. Random Forest achieved accuracy 0.86, precision 0.96, recall 0.88, F1-score 0.92 and AUC 0.911, although Logistic Regression had a higher accuracy (0.92). Predicted probabilities were converted into High, Medium and Low risk tiers and shown on a Flask-based dashboard. The authors openly note that the dropout label was engineered from the same survey answers, and that the "discontinue" feature is part of the label definition, which can make the results optimistic.

C. Interpretable Decision Support

Liu and Wen [2] proposed a Mamdani fuzzy decision support system for the xAPI-Edu-Data dataset (480 students). Attendance, engagement and assignment completion are converted into linguistic terms, a 27-rule base infers a risk score, and the score is mapped to monitoring, targeted support or intensive intervention. On a 30% test split the system reached accuracy 0.86, macro F1 0.84, AUC 0.91 and High-Risk recall 0.89. The authors state that Random Forest scored slightly higher and that the main advantage of their system is transparency, not superior accuracy. Their "early warning" experiment simulates 25%, 50% and 75% data availability on a static dataset, so it should not be read as true weekly monitoring.

D. Deep Learning and Graph-Based Models

Qi [7] proposed AMP-Net, a deep network that fuses historic grades, classroom engagement, socio-economic class and sentiment from digital footprints into one vector. Chen et al. [9] tuned LSTM hyper-parameters with a chaotic niche alpha-evolution optimiser (CNAE-LSTM) and report gains over LSTM, SVM, CNN and Transformer baselines on a secondary-education dataset of 480 instances. Yang et al. [10] modelled higher-order similarity between students from campus behaviour data of 14,565 students with a hypergraph and classified performance with TabNet, reporting an F1-score of up to 93.10% for the best configuration. These models can capture non-linear and relational patterns, but they need more data, more tuning and offer less transparency to teachers.

TABLE I
SUMMARY OF REVIEWED STUDIES

Ref.	Data	Method	Reported Result
[1]	UCI, OULAD, NELS:88	DT, RF, SVM, ANN with soft voting and stacking	Accuracy 98.8% (UCI), 96.9% (OULAD), 97.5% (NELS:88)
[2]	xAPI-Edu-Data, 480 students	Mamdani fuzzy inference	Accuracy 0.86, macro F1 0.84, AUC 0.91, High-Risk recall 0.89

Ref.	Data	Method	Reported Result
[3]	Survey, 253 undergraduates	LR, DT, RF	RF: accuracy 0.86, F1 0.92, AUC 0.911; LR accuracy 0.92
[4]	Survey, 613 secondary students	C5.0, CART, SVM, RF with Boruta and SMOTE	10-fold CV accuracy: RF 74.3%, C5.0 73%
[5]	Public student records	Eight models incl. CatBoost, GB, RF, DT	CatBoost 87.46%, GB 87.28%, DT 82.42%
[6]	Synthetic, over 100,000 records	Eleven algorithms, ordinal NN	Ordinal NN macro F1 0.846, QWK 0.892; LightGBM 0.842
[7]	Grades, engagement, socio-economic, sentiment	AMP-Net (deep learning)	Accuracy 0.85 and AUC-ROC 0.87 reported for different input groups
[8]	xAPI-Edu-Data	SVM, LR, DT, stacking (RF) with SMOTE	Weighted precision 0.60 to 0.83 with stacking
[9]	Secondary education, 480 instances	CNAE-LSTM	Reported to outperform LSTM, SVM, CNN, Transformer
[10]	Campus behaviour, 14,565 students	Hypergraph with TabNet	Best F1-score 93.10%

III. DATASETS, PREPROCESSING AND EVALUATION

A. *Commonly Used Datasets*

The UCI Student Performance dataset of Cortez and Silva [11] contains 649 records and 33 attributes from two Portuguese secondary schools, and is described in [1] together with OULAD (32,593 students on an online university platform) and NELS:88 (about 24,599 eighth-grade students followed longitudinally). The xAPI-Edu-Data set [12] holds 480 student records with raised-hands, resource-visit, announcement-view and discussion counts plus a Low, Middle or High performance class, and is used in [2] and [8]. Survey data collected by the authors themselves are used in [3] and [4], while [6] offers a synthetic alternative that avoids privacy restrictions.

B. *Preprocessing and Feature Selection*

The reviewed studies share a common pipeline: mean or mode imputation of missing values, one-hot or ordinal encoding of categorical answers, and standardisation or min–max scaling for scale-sensitive models such as SVM and neural networks [1], [2], [3]. Redundant features are removed using correlation thresholds ($|r| > 0.85$) [1] or the variance inflation factor (VIF > 5) [4], and relevant features are retained using Random Forest importance [1], the Boruta wrapper [4] or expert review [2]. Random Forest [13] is widely used both as a classifier and as a feature-ranking tool.

C. *Handling Class Imbalance*

At-risk students are usually a minority, so a model that predicts “pass” for everyone can still show high accuracy. Studies counter this with SMOTE [14] oversampling [4], [8], with class-weighted learning [3], and by reporting per-class recall [2]. Oversampling must be applied only to the training split, otherwise synthetic samples leak into the test data.

D. *Evaluation Metrics*

Classification studies report accuracy, precision, recall, F1-score ($F1 = 2PR/(P + R)$) and AUC-ROC, while regression-style studies use MAE, MSE and RMSE [5]. Ordinal targets are better judged with quadratic weighted kappa [6]. For early intervention, recall of the at-risk class is the most important figure, because a missed student receives no help, whereas a false alarm costs only a short conversation. Stratified k-fold cross-validation [1], [4] gives more reliable estimates than a single split, particularly for datasets of a few hundred records.

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue XI Nov 2026- Available at www.ijraset.com

IV. MACHINE LEARNING ALGORITHMS

Table II summarises the algorithms that appear most often in the reviewed work and are relevant to the proposed system. Logistic Regression gives probabilities and readable coefficients and is a strong baseline; in [3] it even achieved the highest accuracy. Decision Trees express decisions as simple conditions on attendance and marks, which teachers understand easily, but a single deep tree overfits, as CART did in [4]. Random Forest averages many trees, reduces overfitting and reports feature importance, and it was the most consistent model in [3] and [4]. SVM suits high-dimensional data but is sensitive to kernel and cost settings, and boosting methods such as CatBoost and LightGBM often lead on larger tabular data [5], [6] at the cost of tuning effort and transparency.

TABLE II
COMPARISON OF COMMON ALGORITHMS

Algorithm	Strengths	Limitations
Logistic Regression	Simple, fast, probabilistic output, interpretable coefficients	Assumes linear decision boundary
Decision Tree	Rule-like and easy to explain to teachers	Unstable; overfits when deep
Random Forest	Stable, reduces overfitting, gives feature importance	Less transparent than a single tree
SVM	Effective in high dimensions	Sensitive to kernel and cost; hard to interpret
Boosting (CatBoost, LightGBM)	Often highest accuracy on tabular data	More tuning; weaker interpretability
Neural / deep models	Capture non-linear and sequential patterns	Need more data; black-box behaviour

V. PROPOSED FRAMEWORK

A. Problem and Objectives

The proposed system predicts a student's academic performance early in the course so that teachers can act before the final examination. Its objectives are to analyse academic and behavioural data, build ML models for performance prediction, identify at-risk students at an early stage, give teachers data-driven insight, and thereby improve intervention strategies and the overall success rate.

B. Workflow

The workflow has seven steps, shown in Fig. 1. Attendance, internal marks, assignment performance, previous academic performance and study behaviour are collected first. The data are then cleaned, missing values are handled, categorical fields are encoded and numerical values are normalised where required. Relevant features are selected, and Logistic Regression, Decision Tree and Random Forest models are trained and compared using accuracy, precision, recall and F1-score. No model is declared the best from accuracy alone. The selected model predicts Pass/Fail status, grade level or a risk category, and the results are shown on a dashboard that also sends alerts about weak or at-risk students.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

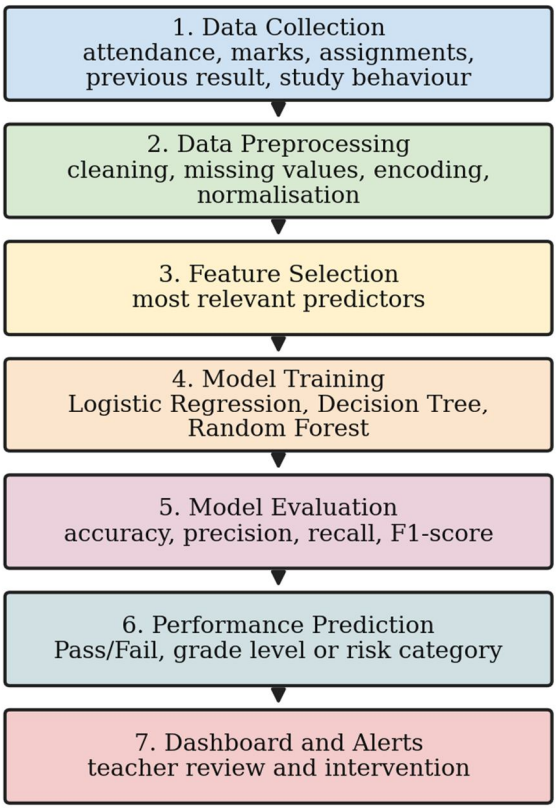


Fig. 1 Workflow of the proposed student performance prediction system

C. Technology Stack and Architecture

The prediction component is implemented in Python with Pandas and NumPy for data handling and Scikit-learn for model training and evaluation, using Google Colab for experiments. The dashboard is built with React.js and connects through a FastAPI backend to a PostgreSQL or MongoDB database, and code is managed on GitHub. This layered arrangement (Fig. 2) keeps the data, model, service and presentation concerns separate so that each can be improved independently.

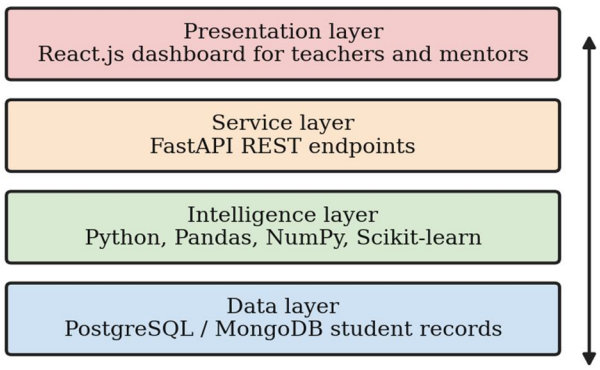


Fig. 2 Layered architecture of the proposed system

D. Risk Tiers and Intervention

Model output is mapped to three risk tiers, in the same spirit as the tiered schemes in [2] and [3]. Table III lists the indicative signals and the teacher action for each tier. The probability cut-offs between tiers are institution-specific and should be calibrated on local data and agreed with teachers, because heuristic thresholds without calibration are only a guide [3].

TABLE III
RISK TIERS AND TEACHER ACTIONS

Risk Tier	Indicative Signals	Teacher Action
Low	Good attendance, satisfactory internal marks, assignments submitted	Routine monitoring and feedback
Medium	Falling attendance or borderline internal marks	Mentoring, extra study material, follow-up
High	Low attendance, low internal marks, incomplete assignments, weak previous performance	Counselling, extra classes, personal meeting with student

E. Role of the Teacher

The system is a decision-support tool. A prediction is an indication of risk and not a final judgement, and no ML model is 100% accurate. Teachers must verify each alert and consider the student's actual circumstances before acting, a principle also stressed in [1] and [3].

VI. DISCUSSION AND CHALLENGES

1) *Results are not directly comparable:* The reviewed studies differ in dataset, target variable (pass/fail, grade, level, dropout), split strategy and metric, so a figure such as 98.8% [1] cannot be ranked against 0.86 [2] or 74.3% [4]. A fair comparison needs a common dataset and protocol, which is one motivation for the synthetic benchmark in [6].
2) *Label construction and leakage:* When the target is derived from the same inputs, as acknowledged in [3], or when oversampling touches the test data, reported performance is inflated. Targets should be defined independently, for example from later observed outcomes.
3) *Small, single-institution data:* Most studies use a few hundred to a few thousand records from one institution [2]–[4], [9], which limits generalisation. External validation, such as the 30-student check in [4], is still very small.
4) *Interpretability versus accuracy:* Boosted and deep models can reach higher scores [5], [10] but are harder for teachers to trust. Fuzzy rules [2], decision trees and feature-importance or SHAP explanations [6] help to explain why a student was flagged.
5) *From prediction to action:* Several systems stop at a risk score. Linking each tier to a defined action, as in [2] and Table III, and testing whether the action improves outcomes is still largely unexplored.
6) *Data quality, privacy and fairness:* Student data are sensitive, records may be incomplete, and predictions may disadvantage groups if socio-economic variables dominate. Institutions need consent, access control, bias audits and human oversight of every automated alert [1], [6].

VII. FUTURE SCOPE

The system can be extended in several directions: real-time student monitoring with streaming data; integration with learning management systems such as Moodle and Google Classroom; a mobile application; an AI chatbot for student support; personalised learning recommendations; automatic study-plan generation; student progress tracking; a parent notification system; and explainable AI techniques such as SHAP for per-student explanations [6]. Libraries such as PyTorch, Hugging Face Transformers and LangChain may be used for the chatbot and natural-language feedback features, but they are future-scope technologies and are not part of the current implementation. Future studies should also validate the framework on larger multi-institution data and measure whether alerts actually improve student results.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

VIII. CONCLUSION

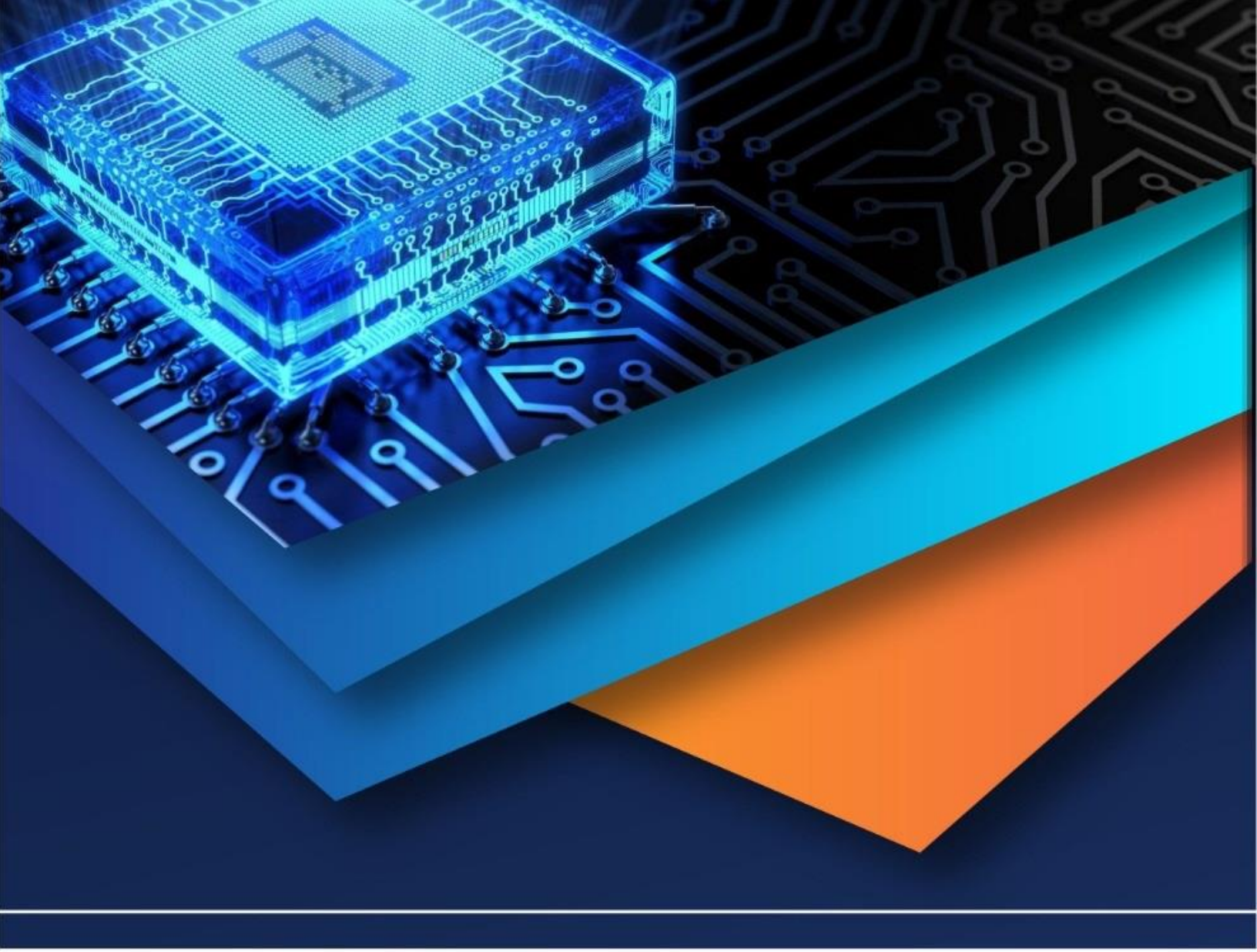
This paper reviewed ten recent studies on ML-based student performance prediction and early intervention and proposed a decision-support framework for teachers. The review shows that ensemble methods, especially Random Forest and boosting, perform consistently well, that handling class imbalance and avoiding label leakage matter as much as the choice of algorithm, and that interpretability and a clear link to intervention are essential for real adoption. The proposed framework combines Logistic Regression, Decision Tree and Random Forest models with risk tiers and a dashboard so that weak students can be supported before the final examination, while the teacher keeps the final decision.

IX. ACKNOWLEDGMENT

The authors thank their project guide, Dr. Rahul M. Dhokane, and the Department of Computer Engineering, Sir Visvesvaraya Institute of Technology, Nashik, for their guidance and support.

REFERENCES

[1] Y. Wang, "Artificial intelligence in student management systems to enhance academic performance monitoring and intervention," *Sci. Rep.*, vol. 15, Art. no. 35122, 2025, doi: 10.1038/s41598-025-19159-4.
[2] X. Liu and K. Wen, "A fuzzy decision support system for monitoring student academic performance and supporting early intervention," *Discover Artif. Intell.*, vol. 6, Art. no. 1458, 2026, doi: 10.1007/s44163-026-02498-0.
[3] V. Hegde, A. Palaniswamy, D. Bhar, M. Revanna, N. Prakash, A. Prabhu, and P. M. Shanthappa, "An explainable machine learning framework for early student dropout risk prediction and stratification," *Discover Artif. Intell.*, vol. 6, Art. no. 1224, 2026, doi: 10.1007/s44163-026-02288-8.
[4] B. B. Alkan, S. Kuzucuk, N. Alkan, and A. Sinan, "Using machine learning to predict student outcomes for early intervention and formative assessment," *Sci. Rep.*, vol. 15, Art. no. 39797, 2025, doi: 10.1038/s41598-025-23409-w.
[5] M. N. Gul, W. Abbasi, M. Z. Babar, A. Aljohani, and M. Arif, "Data driven decisions in education using a comprehensive machine learning framework for student performance prediction," *Discover Comput.*, vol. 28, Art. no. 153, 2025, doi: 10.1007/s10791-025-09585-3.
[6] S. Agal, "A machine learning framework with a public synthetic data set for early student profiling and outcome prediction," *Sci. Rep.*, vol. 16, Art. no. 25730, 2026, doi: 10.1038/s41598-026-55514-9.
[7] Y. Qi, "A multi-dimensional prediction system for students' academic performance driven by deep learning," *Discover Artif. Intell.*, vol. 6, Art. no. 69, 2026, doi: 10.1007/s44163-025-00744-5.
[8] T. Althaqafi, F. Saleem, and A. Al-Malaise Al-Ghamdi, "Enhancing student performance prediction: the role of class imbalance handling in machine learning models," *Discover Comput.*, vol. 28, Art. no. 79, 2025, doi: 10.1007/s10791-025-09576-4.
[9] H. Chen, Y. Zhou, and Q. Cao, "A hybrid learning framework integrating chaotic Niche alpha evolution for student academic performance prediction," *Sci. Rep.*, vol. 16, Art. no. 5302, 2026, doi: 10.1038/s41598-026-36263-1.
[10] Y. Yang, S. Zhao, S. An, X. Li, and Y. Zhang, "Student academic performance prediction via hypergraph and TabNet," *J. Big Data*, vol. 12, Art. no. 119, 2025, doi: 10.1186/s40537-025-01170-1.
[11] P. Cortez and A. M. G. Silva, "Using data mining to predict secondary school student performance," in *Proc. 5th Annu. Future Bus. Technol. Conf.*, Porto, Portugal, 2008, pp. 5–12.
[12] E. A. Amrieh, T. Hamtini, and I. Aljarah, "Students' academic performance dataset (xAPI-Edu-Data)," Kaggle, 2016. [Online]. Available: https://www.kaggle.com/datasets/aljarah/xAPI-Edu-Data
[13] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.
[14] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: synthetic minority over-sampling technique," *J. Artif. Intell. Res.*, vol. 16, pp. 321–357, 2002.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)