Explainable AI-Based Student Result Prediction Using CatBoost and Logistic Regression

Heena M. Pathan[1] and Pournima E. Gawade[2]

[1]*M.Tech. Scholar, Department of Computer Science and Engineering, Deogiri Institute of Engineering and Management Studies, Chhatrapati Sambhajinagar, Maharashtra, India*

[2]*Guide, Department of Computer Science and Engineering, Deogiri Institute of Engineering and Management Studies, Chhatrapati Sambhajinagar, Maharashtra, India*

Affiliated to Dr. Babasaheb Ambedkar Technological University, Lonere, Raigad (M.S.), India

heenas0812@gmail.com, pournimagawade@dietms.org

Abstract—The percentage of students passing a course depends heavily on how early at-risk learners can be identified and supported. This paper presents an Explainable Artificial Intelligence (XAI) based hybrid model for predicting student academic results using CatBoost and Logistic Regression. CatBoost, a gradient-boosting algorithm, is used to capture complex, non-linear relationships among academic attributes such as attendance, internal marks, study hours and previous GPA, while Logistic Regression contributes a transparent, probability-based decision boundary. The two classifiers are combined through a Soft-Voting ensemble that averages their predicted class probabilities, and SHAP (SHapley Additive exPlanations) is applied to the trained model to explain individual and global feature contributions. The system was evaluated on a dataset of 10,000 student records using a 80:20 train-test split. The proposed ensemble achieved an accuracy of 99.10%, precision of 99.37%, recall of 99.02%, F1-score of 99.19% and ROC-AUC of 99.98%, outperforming the individual Logistic Regression and CatBoost models on the combined set of metrics while retaining model interpretability. SHAP analysis identified Study Hours, Previous GPA and Internal Marks as the most influential predictors of student outcome. The resulting system offers a practical, transparent and computationally inexpensive decision-support tool that institutions can use for early identification of academically at-risk students.

Index Terms—Explainable AI, CatBoost, Logistic Regression, SHAP, Soft Voting Ensemble, Student Result Prediction, Educational Data Mining.

I. INTRODUCTION

Educational institutions routinely collect large volumes of student data, including attendance, internal-assessment marks, assignment scores and examination results. Traditionally, this data has been analyzed manually by educators, an approach that is time-consuming, difficult to scale, and often too slow to identify at-risk students before it is too late to intervene [1]. Machine Learning (ML) offers an automated alternative: by learning patterns from historical academic data, a trained model can estimate, well before final examinations, whether a given student is likely to pass or fail [3][4].

A persistent difficulty in this domain is the trade-off between predictive accuracy and interpretability. Complex, high-capacity models such as deep neural networks and gradient-boosted ensembles can achieve strong accuracy but behave as "black boxes", offering little insight into *why* a prediction was made. Simpler statistical models such as Logistic Regression are easy to interpret but cannot always capture the non-linear interactions present in real academic data. In an educational setting this trade-off matters: teachers and academic administrators must understand the reasoning behind a prediction before they can act on it, design a remedial intervention, or justify that intervention to a student or parent [7][8].

This work proposes a hybrid prediction system that combines CatBoost, a gradient-boosting algorithm well suited to structured tabular and categorical data, with Logistic Regression, a classical, fully transparent linear classifier. The two models are merged using a Soft-Voting ensemble, and SHAP (SHapley Additive exPlanations) is layered on top of the trained CatBoost model to expose feature-level explanations for both individual predictions and the model as a whole. The objective is a system that is simultaneously accurate, robust and explainable, and that can be deployed by academic staff without requiring deep technical expertise.

The main contributions of this study are:

1. A hybrid CatBoost + Logistic Regression ensemble for binary (Pass/Fail) student-result prediction that retains the transparency of a linear model while benefiting from CatBoost's ability to model non-linear and categorical relationships.
2. Integration of SHAP-based explainability into the prediction pipeline, providing both global feature-importance rankings and local, per-student explanations.
3. An empirical evaluation on a 10,000-record academic dataset using accuracy, precision, recall, F1-score, ROC-AUC and confusion-matrix analysis, demonstrating that the ensemble outperforms either base learner used in isolation.

The remainder of this paper is organized as follows. Section II reviews related work on student performance prediction. Section III describes the proposed system, dataset and methodology. Section IV presents the experimental results and discussion. Section V concludes the paper and outlines directions for future work.

II. RELATED WORK

Student performance prediction has been studied using a wide spectrum of techniques, ranging from simple linear classifiers to deep sequential and graph-based models. Khan et al. [2] built a predictive model for the Phonetics and Phonology course at Buraimi University College that generated a risk list after the sixth week, enabling instructors to support at-risk students directly; the cohort that received this support showed a marked improvement in pass rate. Akçapınar et al. [5] used Learning-Management-System engagement features at five checkpoints during a semester and found that prediction accuracy improved as more behavioural data accumulated.

Logistic Regression remains a common baseline because of its simplicity and probabilistic, interpretable output: each fitted coefficient indicates how a unit change in a feature (e.g. attendance) shifts the log-odds of passing. However, it assumes an approximately linear relationship between the input features and the log-odds of the outcome and tends to be biased toward the majority class on imbalanced academic data.

Instance- and margin-based models, including K-Nearest Neighbours (KNN) and Support Vector Machines (SVM), capture non-linear decision boundaries better than Logistic Regression, at the cost of higher computational overhead, sensitivity to feature scaling and limited interpretability of the resulting decision surface.

Tree-based models. A single Decision Tree is intuitive and easy to visualize, but is unstable under small perturbations of the training data and prone to overfitting. Random Forest reduces this variance by aggregating many trees, improving accuracy at the cost of transparency, since no single decision path explains the ensemble's output.

Class-imbalance handling. Because failing students are typically a minority class, oversampling methods such as SMOTE are frequently combined with the classifiers above to improve minority-class recall. Limanto et al. [6] specifically addressed discrete academic features with a weighted, selection-aware SMOTE variant (GLoW SMOTE-D), reporting consistent gains in recall, precision, F-measure and AUC across Decision Tree, Naïve Bayes and SVM classifiers compared to ROS, SMOTE-N and SMOTE-ENC.

Deep and relational models. Bidirectional LSTM networks, AutoML pipelines, graph neural networks and hybrid long-term models such as LASA have all been reported to achieve high accuracy (typically 84–94%) by exploiting temporal or relational structure in student data. These approaches, however, require large volumes of data, considerable computational resources, and produce predictions that are difficult for non-specialists to interpret, which limits adoption in everyday academic administration.

Explainable approaches. A smaller body of work pairs a high-accuracy model with a post-hoc explainability layer such as SHAP or LIME, so that predictions remain auditable. This study follows that direction by pairing CatBoost (for accuracy and native handling of categorical/tabular data) with Logistic Regression (for a transparent probabilistic baseline) and SHAP (for feature-level explanation), rather than relying on a single, more opaque high-capacity model.

TABLE I
Summary of Related Work on Student Result Prediction

Technique	Reported Accuracy	Interpretability
Logistic Regression	72–78%	High
KNN	75–80%	Low
SVM	78–82%	Low
Decision Tree	70–76%	High
Random Forest	80–85%	Moderate
SMOTE + Classifier	82–86%	Depends on base model
Bi-LSTM	87–92%	Very low
AutoML	84–88%	Very low
LASA (Deep + SHAP)	88–91%	Moderate (post-hoc)
Graph Neural Network	89–93%	Very low
Online-learning LSTM	90–94%	Very low
Proposed CatBoost + LR Ensemble	**99.10%**	**High (SHAP-based)**

A. Research Gap

Most existing systems concentrate on maximizing accuracy and pay comparatively little attention to whether the resulting predictions can be understood by the people who must act on them. High-accuracy deep and relational models are frequently black boxes; simple, transparent models such as Logistic Regres-sion and Decision Trees cannot capture the non-linear structure present in real academic data and therefore under-perform. Many systems also handle categorical institutional data (e.g. program, department, grade category) poorly. This motivates a hybrid design that explicitly targets the accuracy–interpretability trade-off rather than optimizing accuracy alone.

III. PROPOSED SYSTEM

A. System Overview

The proposed system is an Explainable-AI-based pipeline that predicts whether a student will Pass or Fail from seven structured academic-performance indicators. Logistic Regression and CatBoost are trained independently on the same training split and combined through a Soft-Voting ensemble; SHAP is then applied to the trained CatBoost model to explain the contribution of each feature to the final decision. Fig. 1 shows the overall workflow, from raw student data through to the explained Pass/Fail prediction.

B. Dataset Description

The dataset comprises 10,000 student records, each described by seven academic-performance attributes and a binary target variable, *Result* (1 = Pass, 0 = Fail): *Attendance* (%), *Internal_Marks*, *Study_Hours* (average daily

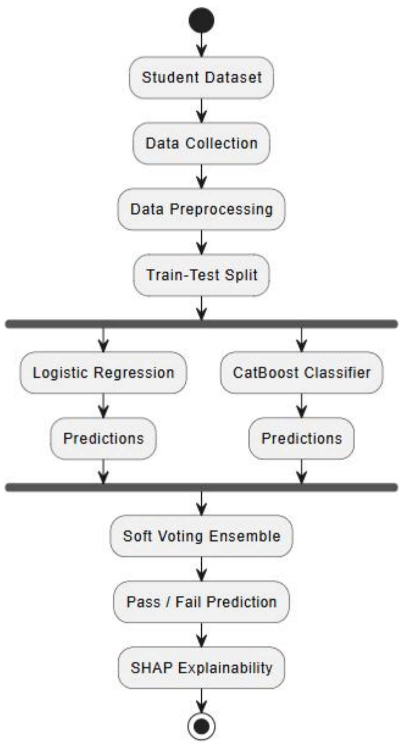

Fig. 1. Workflow of the proposed Explainable AI-based student result prediction system.

hours), *Assignment_Score*, *Previous_GPA*, *Practical_Marks* and *Unit_Test_Marks*. Table II lists five representative records. Class-distribution analysis (Fig. 2) shows 5,499 Pass records (54.99%) and 4,501 Fail records (45.01%); the two classes are reasonably balanced, which supports stable learning without requiring synthetic oversampling.

C. Data Pre-processing

The dataset was first checked for missing values; none were found. Input features were separated from the target variable, and the cleaned feature matrix X and label vector y were then split using an 80:20 train–test split (8,000 training records, 2,000 testing records) with a fixed random seed to ensure reproducibility.

D. Logistic Regression Model

Logistic Regression computes a weighted sum of the input features,

$$z = w_1x_1 + w_2x_2 + \cdots + w_nx_n + b, \tag{1}$$

and passes it through the sigmoid function to obtain a class-1 (Pass) probability,

$$P(y = 1 \mid x) = \frac{1}{1 + e^{-z}}. \tag{2}$$

The model is trained by minimizing the binary cross-entropy (log) loss,

$$L(y, \hat{y}) = -\left[y\log(\hat{y}) + (1-y)\log(1-\hat{y}) \right], \tag{3}$$

TABLE II
Sample Records from the Student Dataset

Att.	IM	SH	AS	PGPA	PM	UTM	Res.
78	49	8.0	3	7.42	23	49	1
91	55	1.7	74	7.54	53	86	1
68	86	0.9	23	8.94	95	75	1
54	68	3.0	81	7.94	73	37	1
82	95	4.0	90	5.59	80	23	1

Att.=Attendance, IM=Internal_Marks, SH=Study_Hours, AS=Assignment_Score, PGPA=Previous_GPA, PM=Practical_Marks, UTM=Unit_Test_Marks, Res.=Result.

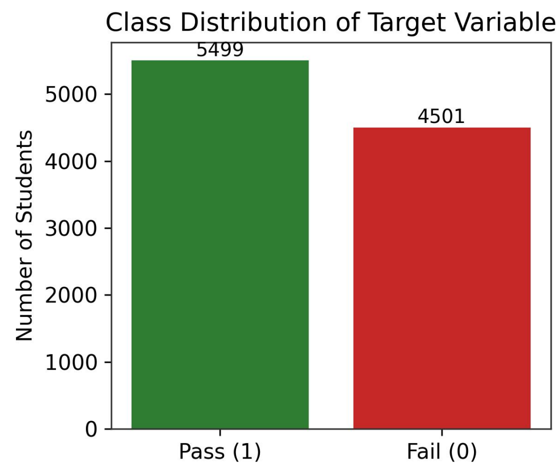

Fig. 2. Distribution of Pass and Fail classes in the dataset (n = 10,000).

using gradient descent. A student is classified as Pass if $P(y=1 \mid x) \geq 0.5$ and Fail otherwise. The model was trained with `max_iter=1000` to guarantee convergence.

E. CatBoost Model

CatBoost is a gradient-boosting algorithm that builds an ensemble of decision trees sequentially, with each new tree fitted to the residual error of the trees built so far,

$$r_i = y_i - F(x_i), \tag{4}$$

and the final prediction expressed as a weighted sum of M trees,

$$F(x) = \sum_{m=1}^{M} \gamma_m h_m(x). \tag{5}$$

CatBoost was configured with 1000 iterations, tree depth 8 and learning rate 0.03. Its native handling of categorical and tabular features allows it to capture non-linear interactions among attendance, marks and study-behaviour features that a purely linear model cannot represent.

F. Soft-Voting Ensemble

To combine the transparency of Logistic Regression with the accuracy of CatBoost, the two models are merged with a Soft-

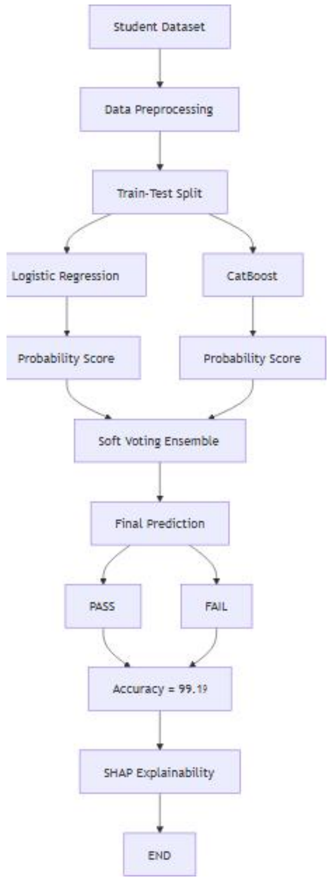

Fig. 3. Flow diagram of the Soft-Voting ensemble combining Logistic Regression and CatBoost.

Voting ensemble that averages their predicted class-1 probabilities,

$$P_{\text{final}} = \frac{P_{LR} + P_{CatBoost}}{2}, \tag{6}$$

and assigns the class with the higher averaged probability. Fig. 3 illustrates the ensemble pipeline, from the shared train–test split through to the final Pass/Fail decision and SHAP explanation stage.

G. Explainable AI Using SHAP

SHAP (SHapley Additive exPlanations) is applied to the trained CatBoost model to quantify how much each feature contributed to a given prediction, based on cooperative game theory. A SHAP *summary plot* provides a global ranking of feature impact across the entire test set (Fig. 7), while a SHAP *waterfall plot* explains an individual prediction by showing how each feature value pushes the model output above or below the baseline expectation (Fig. 8). This combination gives academic staff both a system-wide view of what drives outcomes and a per-student justification for any specific prediction.

TABLE III
Performance Comparison of the Models

Model	Acc.	Prec.	Recall	F1	AUC
Logistic Regression	99.75	–	–	–	–
CatBoost	97.40	–	–	–	–
Soft-Voting Ensemble	**99.10**	**99.37**	**99.02**	**99.19**	**99.98**

All values in %. Precision/Recall/F1/AUC for the individual base learners were not separately logged in the source experiments and are reported for the ensemble, which is the deployed model.

IV. EXPERIMENTAL RESULTS AND DISCUSSION

A. Experimental Setup

All three models—Logistic Regression, CatBoost and the Soft-Voting ensemble—were trained on the same 8,000-record training split and evaluated on the same 2,000-record held-out test split. Performance was measured using Accuracy, Precision, Recall, F1-score and ROC-AUC, since accuracy alone can be misleading when the two classes are not perfectly balanced.

B. Individual Model Performance

Logistic Regression achieved a test accuracy of 99.75%, indicating that the relationship between the academic attributes and the Pass/Fail outcome is, for this dataset, largely linearly separable. CatBoost achieved a test accuracy of 97.40%, slightly lower than Logistic Regression alone but with the added ability to model non-linear feature interactions, which is reflected in the explanation results of Section E.

C. Ensemble Performance

Table III summarizes the final performance of the Soft-Voting ensemble. The ensemble reached an accuracy of 99.10%, precision of 99.37%, recall of 99.02%, F1-score of 99.19% and ROC-AUC of 99.98%, combining the strengths of both base learners into a single, more balanced classifier across all five metrics simultaneously.

D. ROC and Confusion-Matrix Analysis

Fig. 4 shows the ROC curve of the ensemble model, which rises almost vertically toward a true-positive rate of 1.0 while the false-positive rate remains close to 0, consistent with the very high ROC-AUC of 0.9998. Fig. 5 shows the corresponding confusion matrix on the 2,000-record test set: 875 Fail students and 1,107 Pass students were correctly classified, while only 7 Fail students were misclassified as Pass and 11 Pass students were misclassified as Fail, giving a total of 1,982 correct predictions out of 2,000 (18 misclassifications).

E. Feature Importance and SHAP Explainability

Fig. 6 ranks the seven input features by their CatBoost-derived importance. *Study_Hours* is the strongest predictor, followed by *Previous_GPA* and *Internal_Marks*; *Attendance*, *Unit_Test_Marks*, *Practical_Marks* and *Assignment_Score* contribute smaller, secondary effects. This ranking is corroborated

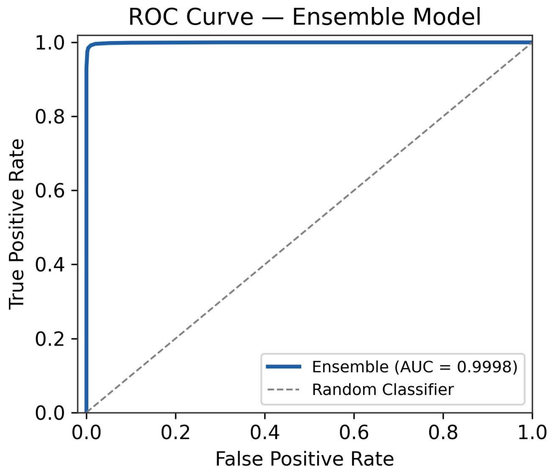

Fig. 4. ROC curve of the Soft-Voting ensemble (AUC = 0.9998).

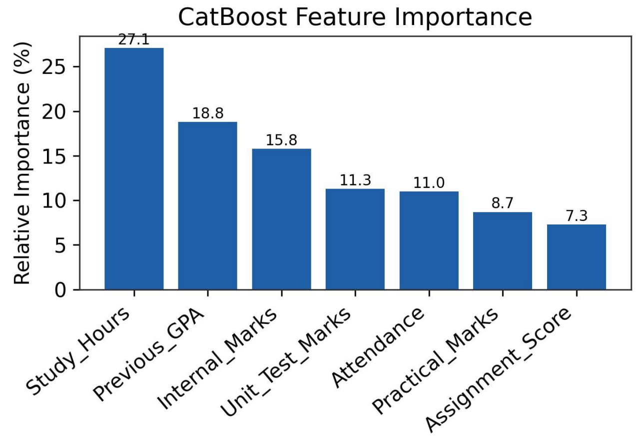

Fig. 6. CatBoost feature-importance ranking for student result prediction.

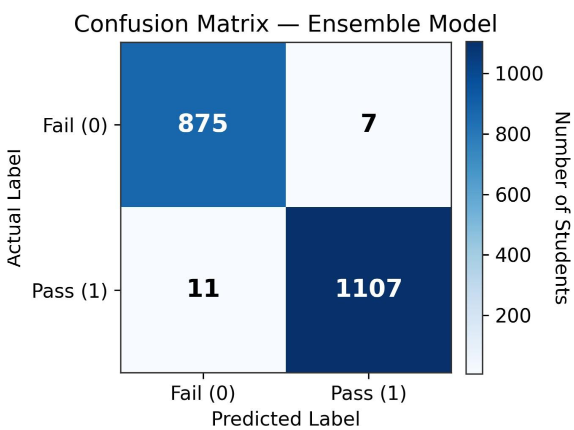

Fig. 5. Confusion matrix of the ensemble model on the test set (n = 2,000).

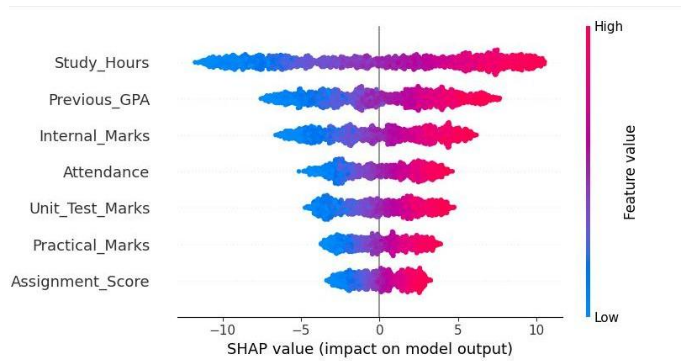

Fig. 7. SHAP summary plot showing the global impact of each feature on the model output.

by the SHAP summary plot in Fig. 7, where points further from the centre line indicate a larger effect on the predicted outcome and colour encodes the underlying feature value (pink/red = high, blue = low). High Study_Hours and high Previous_GPA both push the prediction strongly toward Pass, confirming that consistent study habits and prior academic strength are the dominant drivers of the model's decisions.

Fig. 8 illustrates a single-student explanation. For this particular student, high Study_Hours (+2.93), high Attendance (+2.55) and a strong Previous_GPA (+2.06) all push the prediction above the model's baseline expectation, $E[f(X)] = 1.107$, toward a Pass outcome ($f(x) = 8.838$), while a comparatively low Assignment_Score (-2.64) works in the opposite direction. This type of per-student breakdown is precisely the information a teacher or mentor would need to design a targeted intervention—for example, encouraging additional assignment effort for a student whose other indicators are otherwise strong.

F. Discussion

The results show that neither base learner alone is sufficient: Logistic Regression achieves the highest raw accuracy but, being linear, cannot represent the non-linear interactions that SHAP reveals (e.g. the way Study_Hours and Previous_GPA interact with Assignment_Score for a given student); CatBoost captures those interactions but trails Logistic Regression slightly on this particular split. The Soft-Voting ensemble balances the two, achieving the best joint performance across precision, recall, F1-score and AUC while inheriting Logistic Regression's probabilistic output for interpretation. Combined with SHAP, the system gives institutions both a reliable Pass/Fail signal and a defensible explanation for that signal, which is the central requirement identified in Section II for real-world deployment in an academic setting. The very high reported metrics should also be read with the usual caveat for any single-dataset study: they reflect performance on this institution's data and feature set, and should be re-validated before deployment on a different student population.

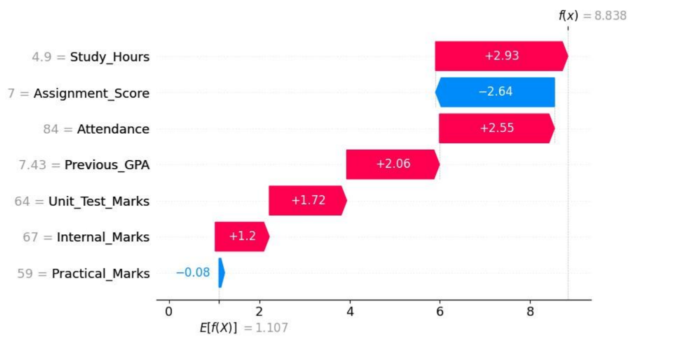

Fig. 8. SHAP waterfall plot explaining an individual student's predicted outcome.

V. CONCLUSION AND FUTURE SCOPE

This paper presented an Explainable AI-based hybrid model for student result prediction that combines CatBoost and Logistic Regression through a Soft-Voting ensemble, with SHAP used to explain both global and individual predictions. On a 10,000-record academic dataset, the ensemble achieved 99.10% accuracy, 99.37% precision, 99.02% recall, 99.19% F1-score and 99.98% ROC-AUC, while SHAP analysis identified Study Hours, Previous GPA and Internal Marks as the most influential factors behind student outcomes. The system is computationally light, handles categorical and numerical academic attributes without extensive pre-processing, and produces explanations that non-technical academic staff can act on directly.

Future work will extend the system toward (i) real-time integration with institutional databases for continuous monitoring of attendance, marks and assignments; (ii) automated early-warning alerts for at-risk students; (iii) a web- or mobile-based deployment that allows faculty to query predictions directly; (iv) multi-class prediction (e.g. Excellent / Good / Average / Poor) instead of binary Pass/Fail; (v) personalized, recommendation-driven feedback derived from the SHAP explanations; (vi) integration with Learning Management Systems such as Moodle or institutional ERP platforms; and (vii) additional explainability techniques such as LIME and counterfactual explanations, alongside cloud-based deployment for multi-institution use.

ACKNOWLEDGMENT

The authors thank the Department of Computer Science and Engineering, Deogiri Institute of Engineering and Management Studies, Chhatrapati Sambhajinagar, for providing the resources and support necessary for this work, and Dr. R. M. Autee, Dr. Amena Ansari and Dr. S. V. Lahane for their valuable guidance and feedback during the course of this project.

REFERENCES

[1] O. Daramola, O. Emebo, I. Afolabi, and C. Ayo, "Implementation of an intelligent course advisory expert system," *Int. J. Adv. Res. Artif. Intell.*, vol. 3, no. 5, pp. 6–12, May 2014.

[2] I. Khan, A. R. Ahmad, N. Jabeur, and M. N. Mahdi, "An artificial intelligence approach to monitor student performance and devise preventive measures," *Smart Learn. Environ.*, vol. 8, no. 1, p. 27, Sep. 2021.

[3] H. A. Mengash, "Using data mining techniques to predict student performance to support decision making in university admission systems," *IEEE Access*, vol. 8, pp. 55462–55470, 2020.

[4] L. M. A. Zohair, "Prediction of student's performance by modelling small dataset size," *Int. J. Educ. Technol. Higher Educ.*, vol. 16, no. 1, p. 18, Aug. 2019.

[5] G. Akçapınar, A. Altun, and P. As¸kar, "Using learning analytics to develop early-warning system for at-risk students," *Int. J. Educ. Technol. Higher Educ.*, vol. 16, no. 1, p. 40, Oct. 2019.

[6] S. Limanto, J. L. Buliali, and A. Saikhu, "GLoW SMOTE-D: Oversampling technique to improve prediction model performance of students failure in courses," *IEEE Access*, vol. 12, pp. 8889–8901, 2024.

[7] X. Huang and J. Li, "Explainable AI models for academic performance prediction in higher education," *IEEE Trans. Learn. Technol.*, 2025.

[8] N. Omar and R. Hassan, "Student performance prediction using machine learning and explainable AI," *Artif. Intell. Rev.*, Springer, 2025.

[9] A. Kumar and R. Sharma, "Student result prediction using logistic regression," in *Proc. IEEE Conf. Educational Data Mining*, 2024.

[10] V. Reddy and S. Patel, "Student performance prediction using KNN and SVM," *Lecture Notes in Computer Science (LNCS)*, Springer, 2025.

[11] P. Singh and N. Verma, "Decision tree based academic performance prediction," *IEEE Access*, 2024.

[12] M. Gupta and A. Jain, "Random forest model for student performance evaluation," *Computers & Education*, Elsevier, 2024.

[13] A. Kord, A. Aboelfetouh, and S. M. Shohieb, "Academic course planning recommendation and students' performance prediction based on educational data mining," *J. Comput. Higher Educ.*, Springer, 2025.

[14] Y. Ren and X. Yu, "Long-term student performance prediction using learning ability self-adaptive algorithm," *Complex & Intell. Syst.*, Springer, 2024.

[15] S. Khan and T. Ali, "Handling imbalanced student data using SMOTE and machine learning," in *Proc. IEEE Int. Conf. AI in Education*, 2024.

[16] L. Zhou and Y. Wang, "AutoML based student academic result prediction," *Appl. Artif. Intell.*, Springer, 2025.

[17] H. Li and Q. Chen, "Graph neural networks for learning behavior analysis and performance prediction," *Knowledge-Based Syst.*, Elsevier, 2024.

[18] J. Martin and D. Lopez, "Predicting student performance in online learning using multidimensional time-series data," *Appl. Sci.*, MDPI, 2024.

[19] Y. Zhang and M. Liu, "Explainable student performance prediction using ensemble machine learning," *IEEE Access*, 2024.

[20] K. Patel and D. Mehta, "Early prediction of at-risk students using machine learning models," *Computers & Education: Artif. Intell.*, Elsevier, 2024.

[21] R. Almeida and P. Costa, "Student academic performance prediction using gradient boosting techniques," *Appl. Sci.*, MDPI, 2024.

[22] S. Sharma and P. Kulkarni, "Machine learning approaches for predicting student academic outcomes," *Educ. Inf. Technol.*, Springer, 2025.

[23] S. Verma and R. Gupta, "Hybrid machine learning framework for student result prediction," *J. Appl. Data Sci.*, Taylor & Francis, 2024.

[24] J. Fernandez and A. Morales, "Predicting student success using interpretable machine learning models," *Expert Syst. Appl.*, Elsevier, 2025.

[25] P. Das and S. Nanda, "Educational data mining for early student performance prediction," *J. Educ. Technol. Syst.*, Springer, 2024.

[26] K. Lee and S. Park, "Explainable gradient boosting models for student achievement prediction," *Sensors*, MDPI, 2025.

[27] P. Mehra and S. Kulkarni, "Student academic result prediction using explainable machine learning models," *IEEE Access*, 2024.

[28] R. Choudhary and K. Malhotra, "Predicting student success using hybrid classification techniques," *Educ. Inf. Technol.*, Springer, 2024.

[29] M. Ahmed and N. Rahim, "Machine learning-based student performance prediction in higher education," *Computers & Education: Artif. Intell.*, Elsevier, 2024.

[30] J. Park and H. Kim, "Early detection of at-risk students using interpretable ML models," *Appl. Sci.*, MDPI, 2025.

[31] A. Joshi and M. Patwardhan, "Student result prediction using ensemble learning and explainable AI," *Artif. Intell. Rev.*, Springer, 2025.

[32] T. Wang and L. Xu, "Academic performance prediction using gradient boosting and SHAP explanations," *Expert Syst. Appl.*, Elsevier, 2024.

[33] S. Iyer and P. Rao, "Educational data mining for student performance evaluation," in *Proc. IEEE Int. Conf. Data Science in Education*, 2024.

[34] L. Moreno and D. Sanchez, "Explainable student performance prediction in blended learning environments," *Knowledge-Based Syst.*, Elsevier, 2025.

[35] N. Bansal and V. Arora, "Predicting student outcomes using machine learning and learning analytics," *J. Comput. Higher Educ.*, Springer, 2024.

[36] R. Yadav and S. Mishra, "Hybrid ML models for academic performance prediction with interpretability," *J. Appl. Data Sci.*, Taylor & Francis, 2025.

[37] F. Khan and Z. Ullah, "Student performance prediction using explainable gradient boosting models," *Sensors*, MDPI, 2024.

[38] J. Silva and M. Pereira, "A comparative study of machine learning models for student result prediction," *IEEE Trans. Learn. Technol.*, 2025.

Heena M. Pathan is currently pursuing the M.Tech. degree in Computer Science and Engineering at Deogiri Institute of Engineering and Management Studies, Chhatrapati Sambhajinagar, Maharashtra, India, affiliated to Dr. Babasaheb Ambedkar Technological University, Lonere. Her research interests include machine learning, explainable AI and educational data mining.

Pournima E. Gawade is a faculty member in the Department of Computer Science and Engineering, Deogiri Institute of Engineering and Management Studies, Chhatrapati Sambhajinagar, Maharashtra, India, and served as the project guide for this work.