An Explainable AI-Driven Early Warning and Intervention Framework for Risk-Stratified Student Attrition Prediction in Education

Ms. Rupali Ambalal Jadhav

PhD Scholar, Atmiya University, Rajkot Assistant Professor, Thakur Institute of Management Studies, Career Development & Research, Mumbai

Dr. Rupal Parekh

Assistant Professor
Atmiya University
Rajkot

Abstract: Learning analytics always has trouble with student engagement and academic performance due to students dropping out. Predictive algorithms that have been around for a long time can find students who are likely to drop out, but they can't say the reason why the students dropout. Because there isn't enough transparency, teachers and academic decision-makers have a hard time understanding the main risk factors that lead to students dropping out.
This study presents an Explainable Artificial Intelligence (XAI)-based early warning and intervention framework to help with the problems that come with predicting risk-stratified student attrition. Ensemble machine learning models such as Random Forest, Gradient Boosting, and Stacking classifiers can be employed to predict the likelihood of an individual's dropout based on academic, behavioural, and socioeconomic variables. SHapley Additive exPlanations (SHAP) are used to explain each prediction in a clear way at the feature level.
A systematic intervention framework connects each level of risk to certain academic and mentoring programs. Students are put into groups based on how likely they are to take risks: low, medium, or high. frame work can support to link risk level to academic and monitoring approaches. The interventions are planned to allow for changes and risk assessments that can adapt through a system of constant monitoring.

The model does well on Accuracy, F1-score, Recall, and ROC-AUC tests, and it's still easy to understand. This study presented a framework for learning analytics that converts predictive analytics into a proactive and comprehensible decision-support system.

Keywords: Learning Analytics; Student Attrition; Explainable AI; Ensemble Learning; Early Warning System; Risk-Based Intervention Framework.

Introduction

Student dropout is still a big problem for learning analytics and educational systems. It has big effects on both individuals and society as a whole. A lot of the time, students who drop out of studies have fewer social and job perspectives in their lives [1]. People in societies are also less likely to want to work and more likely to need help [2]. We need to deal with dropouts as soon as possible if we want to improve academic and economic outcomes in the long run.

One way to find students who are in trouble and help them right away is to use Early Warning Systems (EWS). Many government programs have created systems that use data to look for signs of disengagement and guess who might leave [3][4]. But many of the predictive models we use now aren't good enough to give risk ratings without also explaining why students leave school. Because of this lack of openness, it's harder to find focused, evidence-based treatments, and stakeholders are less sure of themselves.

Some ensemble learning methods that have helped find complicated, non-linear patterns in multi-dimensional student data in educational data mining are Random Forest, Gradient Boosting, XGBoost, and Stacking classifiers. Not enough research has been done to put explainability, structured risk stratification, and systematic intervention mechanisms all in one framework. These models still use different base learners to make their predictions more accurate. This work tackles the problem by making an Explainable AI-based early warning and intervention system that can tell which students are most likely to drop out based on their risk levels. It does this by using ensemble learning and SHAP-based interpretability to put students into groups of low, medium, and high risk. Then, it connects each level of risk to a group of planned intervention strategies [5]. When you always keep an eye on things, change how they are judged, and give help that is specific to the situation, predictive analytics becomes a proactive and easy-to-understand decision-support system.

II. Literature review

The literature review looked at new developments in AI-powered early warning systems and how to predict how many students will drop out in 2024 and 2025. There are three parts to the study: creating a framework for intervention, adding explainability, and coming up with new methods.

2.1 Early Warning Systems for Learning Analytics

Early Warning Systems (EWS) look at students' academic grades, demographics, and behavioural data to find students who might not finish school. In 2024 and 2025, new techniques like temporal engagement modelling and ensemble learning were used to make predictions more accurate. But most systems don't have clear rules for how to intervene and keep an eye on things; instead, they mostly focus on predicting risks.

2.2 Explainable AI in Education

Because machine learning models are getting more complicated, it is now more important to be open in educational analytics. SHAP is an Explainable AI method that lets people see what factors were important in making predictions by looking at the models. Things have gotten better, but there still isn't enough research that looks at all three of these things at once: explainability, risk assessment, and intervention methods

2.3 Research Gap:

Sr. No	Author(s) & Year	Dataset / Context	Methodology / Model Used	Key Features Considered	Performance Metrics	Major Contribution	Research Gap Identified
1	Psyridou et al. (2024)[6]	Longitudinal Finnish student data (Primary to Secondary)	Random Forest, Gradient Boosting	Cognitive, school-related, social-emotional factors	Grade 6 AUC = 0.61; Grade 9 AUC = 0.65	Early identification of latent dropout predictors; emphasis on emotional depth	Moderate accuracy; limited advanced temporal modeling
2	Elbouknify et al. (2025)[7]	Moroccan secondary education (multi-district dataset)	XGBoost, LightGBM, Random Forest + SHAP	Academic, demographic, institutional variables	XGBoost Accuracy = 94.2%	High predictive accuracy with SHAP-based explainability	Limited behavioral temporal analysis
3	Cheng et al. (2025)[8]	Chinese university LMS behavioral logs	Dual-Modal Sliding Window (DMSW) framework	Temporal LMS engagement patterns	15% improvement in early recall	Early detection (4 weeks prior); behavioral volatility modeling	Institution-specific dataset limits generalization

4	Arno et al. (2025)[9]	India's UDISE+ national education database	DropWrap (Neural Network framework)	Institutional-level indicators (infrastructure, teacher-student ratio, dropout history)	Macro-level institutional prediction	Scalable, policy-aligned AI for national governance	Lacks student-level personalization
5	A. Jacquline Dorothy & M. Sangeetha (2025)[10]	Multi-dimensional academic & behavioral dataset	Hybrid Deep Learning Framework (AI + Behavioral Data)	Academic performance, behavioral engagement indicators	Strong performance across Accuracy, F1-score, Recall, ROC-AUC	Integrated AI with behavioral analytics; hybrid deep learning for early dropout prediction	Limited focus on structured risk stratification and intervention mapping

Table I: Comparative Review of Existing Research on Student Dropout Prediction

The studies examined indicate that AI-driven dropout prediction has significantly advanced, particularly in the realms of explainable AI, behavioural analytics, and ensemble learning. When risk categorisation, organised intervention mapping, and ongoing adaptive monitoring fail to integrate, the necessity for a coherent and reliable early warning system becomes evident. This study suggested a risk-based intervention model using XAI to solve these problems.

III. Statement of the Research Problem

A lot of students still drop out of college and university. It was harder for schools to teach and for students to find jobs. A lot of old-fashioned ways to keep an eye on things depend on looking back at how well someone did in school, which makes it harder to act quickly. Even with new predictive analytics algorithms, they still can't find students who are at risk because many of them aren't clear and don't have organised ways to help. We need one framework that makes it easier to find problems early, figure out what the risks are in a way that makes sense, and plan targeted interventions. This will help us do better in school.

Iv. Research Questions, Objectives, and Hypotheses

4.1 Research Questions

To tackle the identified research issue, the study is directed by the subsequent research enquiries:

RQ1: Which factors most precisely forecast student attrition rates and academic challenges in higher education?

RQ2: How well do predictive analytics models find students who are at risk compared to other ways of keeping an eye on them?

RQ3: Is there a strong link between the number of students who leave and the number who come?

RQ4: Are metrics that gauge student engagement effective in forecasting dropout risk?

RQ5: How can predictive insights be used to make targeted interventions that help students who are at risk?

4.2 Goals of the Research

The study's goals are as follows:

O1: To create a model that can predict which students are at risk so that they can be found early.

O2: To look at the most important variables and data points that are most closely linked to student dropouts.

O3: To create a structured intervention framework based on the results of the classification.

O4: To use real institutional data to see how well the intervention framework works.

4.3 Research Hypotheses

To empirically examine the research questions, the following hypotheses are formulated:

H1: Academic performance indicators significantly influence student attrition in higher education.

H2: There is a significant negative correlation between attendance and student attrition.

H3: Student engagement metrics significantly predict dropout risk.

H4: Predictive analytics models demonstrate significantly higher performance in identifying at-risk students compared to traditional monitoring methods.

H5: Risk-based targeted interventions significantly improve student retention outcomes.

V . Research Design Alignment Framework

Research Question (RQ)	Research Objective (O)	Hypothesis (H)	Variables Involved	Evaluation Method
RQ1: Determine which variables best predict academic challenges and student attrition in higher education.	**O1:** Develop a predictive model for early detection **O2:** Analyze key correlated variables.	**H1:** Academic performance significantly predicts attrition. **H2:** Attendance negatively correlates with attrition .**H3:** Engagement metrics significantly influence dropout risk.	Independent: Academic scores, attendance, engagement metrics.Dependent: Attrition risk	Correlation analysis, Feature importance (SHAP), Logistic/Ensemble model evaluation
RQ2: How effective are predictive analytics models compared to traditional methods?	**O1:** Develop predictive model.	**H4:** Ensemble models significantly outperform traditional methods.	Model performance metrics	Accuracy, F1-score, Recall, ROC-AUC comparison

RQ3: Is attendance and student attrition significantly correlated?	**O2:** Analyze key correlated variables.	**H2:** Attendance is negatively correlated with attrition.	Attendance vs Attrition	Pearson/Spearman correlation, Regression coefficients
RQ4: Can engagement metrics predict dropout risk effectively?	**O1, O2:** Model development and variable analysis.	**H3:** Engagement metrics significantly predict dropout risk.	LMS activity, submission frequency, participation	Feature importance analysis, Model coefficients
RQ5: How can predictive insights design targeted interventions?	**O3:** Design intervention framework. **O4:** Evaluate effectiveness of intervention.	**H5:** Risk-based intervention improves retention outcomes.	Risk category vs Retention outcome	Pre–post intervention comparison, Retention rate improvement

Table II: Conceptual Mapping of Research Questions, Objectives, Variables, and Evaluation Methods

VI. Proposed Framework Overview

The first thing that will be done in the planned research is to gather information about academics, behaviour, and socioeconomic status. Cleaning and normalising the data makes sure it is ready for modelling. We use an ensemble model, such as Random Forest, Gradient Boosting, or Stacking, to guess how likely it is that students will drop out. SHAP Explanation gives a clear, feature-level explanation of what each prediction means.
We use scores from predictive modelling to group students into low, medium, and high-risk groups. Use intervention mapping to connect each type of risk to a certain way to help. This study shows how the early warning system combines a constantly changing monitoring loop with proactive, flexible response and risk assessment in real time.

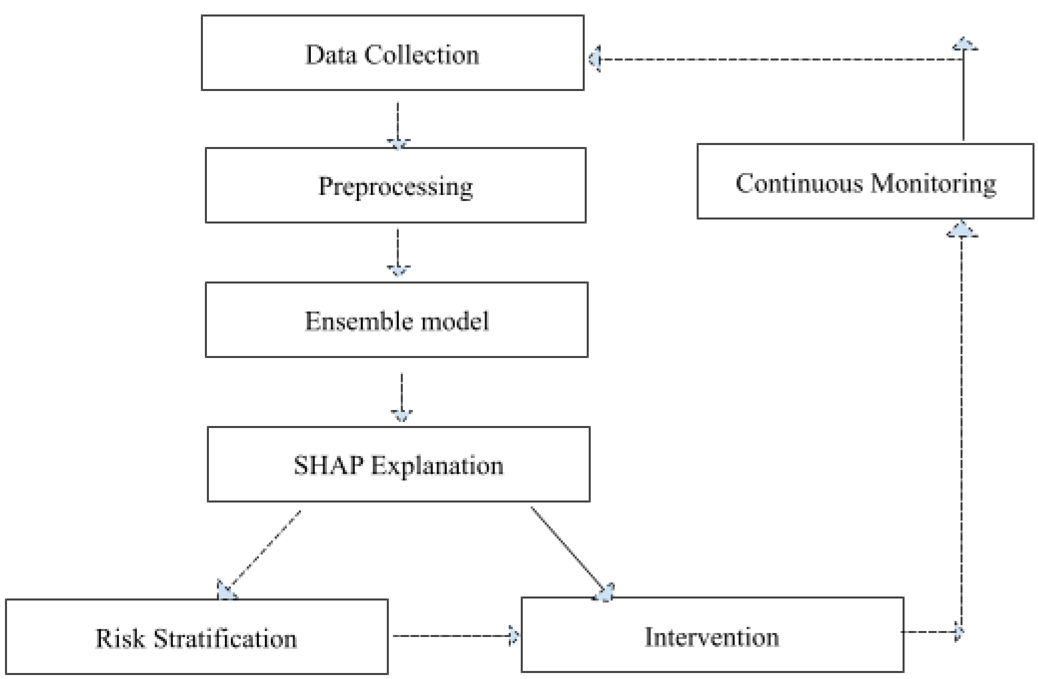


Figure I: Proposed XAI-Based Early Warning Framework for Risk-Stratified Student Attrition Prediction

IV Methodology

Data set description: In order to develop a predictive model for early student attrition identification, the current research used a dataset of 4000 student records that included behavioural, academic, and demographic characteristics.

Demographical Features:

- **Gender:** Gender of the student
- **Location:** Residential area of the student
- **Family Income Level:** Annual income category (Low / Middle / High)
- **Part-Time Work:** Student engaged in part-time employment (Yes/No).
- **Family Responsibilities:** Student has family responsibilities (Yes/No)

Academic Background Features:

- **Type of Admission:** Admission category (CAP / Management / Minority)
- **CET Score:** score in the Entrance examination
- **SSC Percentage:** 10th grade percentage
- **HSC Percentage:** 12th grade percentage
- **Course Enroll:** Enrolled program
- **Department:** Department name like Academic

Behavioural Features:

- **Attendance Percentage:** The number of students who came to class as a percentage of the total number of students
- **Internal Assessment Marks (Out of 20):** Internal marks should be greater than 9 for passing
- **Engagement Hours per Week:** Weekly academic engagement hours.
- **Assignment Submission Rate:** The amount of work that was turned in.

Target Variable:

- **Result:** Academic result (Fail or Pass).
- **Attrition Status:** 0 = Continued, 1 = Dropped.

Models Implemented

This study utilised sophisticated machine learning techniques to forecast the number of students likely to discontinue their education. We used a technique called Random Forest (RF) to make things more stable and better at predicting what would happen next. Bootstrap aggregation helps RF make a lot of decision trees.. Then, it figures out what they all say on average. XGBoost is a type of gradient boosting that uses regularisation to speed up the process, make predictions more accurate, and stop overfitting. It learned how to find patterns that weren't straight.. We then used Random Forest and XGBoost as the base learners to make a Stacking Classifier. Using Logistic Regression as the meta-learner to combine its predictions made the model better overall. You can fully weigh the pros and cons of both bagging and boosting strategies with this multi-model approach by using the meta-level interpretability of logistic regression.

Explainability in Attrition Prediction Using SHAP:

Because SHAP is added to the explainability layer of the early warning system that was suggested, the predictions of the ensemble-based attrition model are easy to understand.. SHAP finds feature-level contribution scores that show how academic, behavioural, and socioeconomic traits affect the predicted risk after the model looks at how likely each learner is to drop out. These reasons help us figure out why the most students dropped out, both for each student and for the whole dataset. SHAP connects prediction outputs with insights that are easy to understand, which makes it easier to sort risks into clear groups. It also encourages personalised, data-driven planning for interventions within the learning analytics framework.

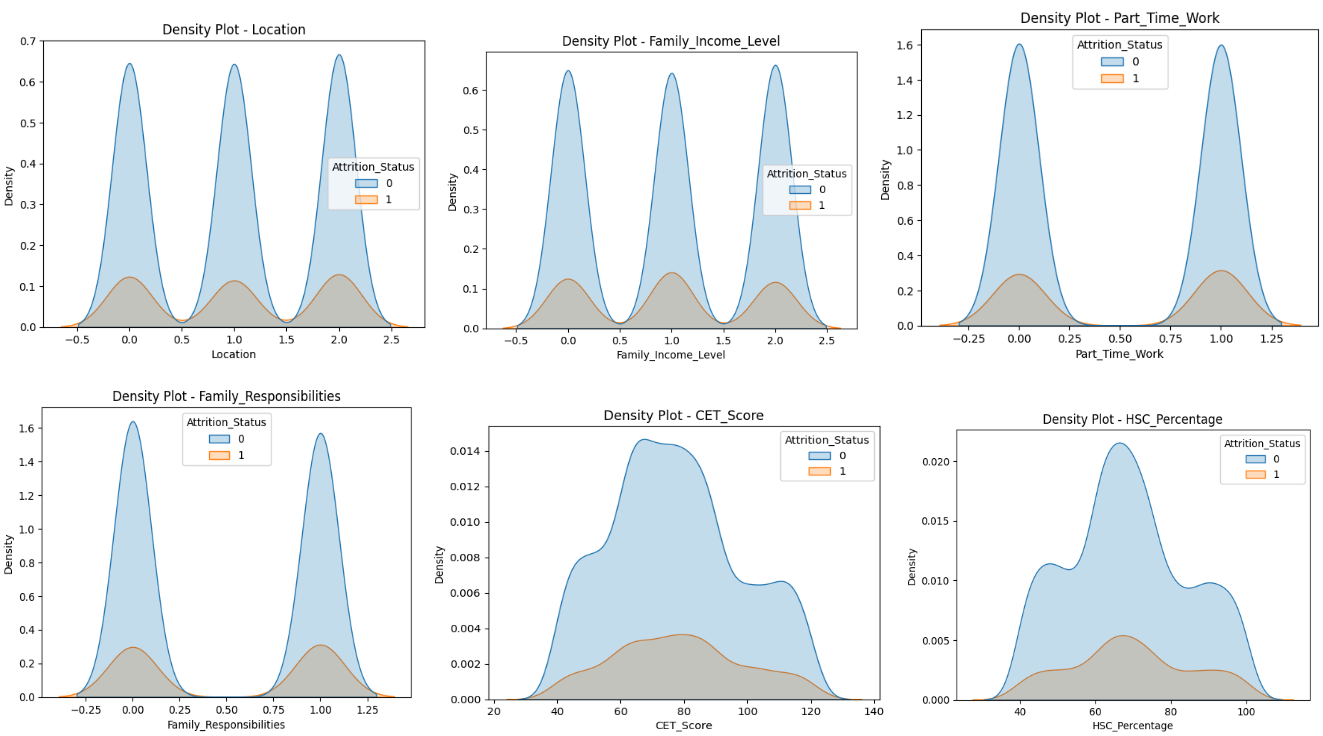
Mechanism for Risk Stratification

The risk stratification mechanism divides students into groups based on how likely the ensemble model thinks they are to drop out of school. The suggested method doesn't just tell students what their risk is. It doesn't do that. But it doesn't put them all together. It divides them into three groups: Low Risk, Medium Risk, and High Risk. This makes it easier to plan interventions in a way that is organised and sets priorities.

We use some rules to figure out how likely it is that something will hurt you. People who were low-risk learners and less likely to succeed had to be watched all the time. Students with medium risk scores should get one-on-one academic help that helps them get involved in school activities. Students who are at high risk are more likely to drop out of school, so they need quick, personalised help like academic support, counselling, and regular check-ins.

This model changes data that isn't useful into useful information. This makes it easier to use resources and decide what to do. It makes sure that decisions made in the learning analytics environment are based on data by combining explainability and risk classification.

V Experimental Results



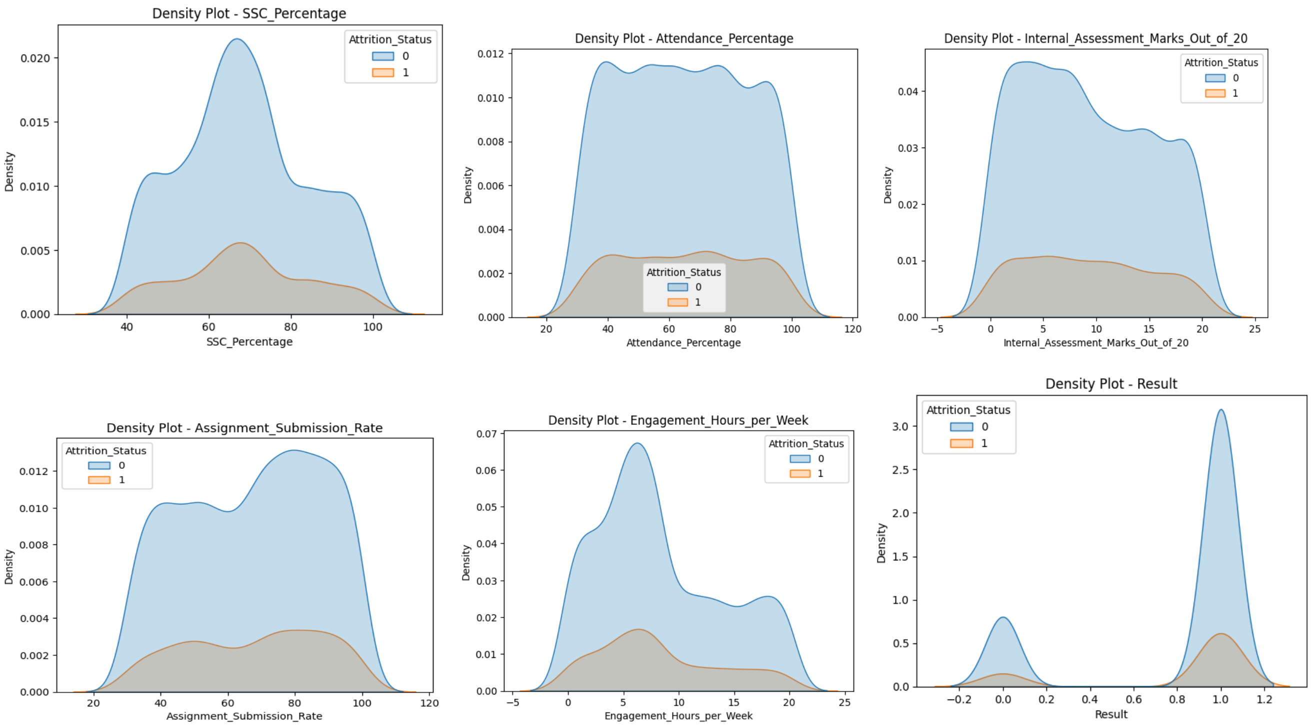


Figure II: Distribution of Key Features' Density in Relation to Student Attrition

The kernel density plots in the picture show how demographic, academic, and engagement-related factors are spread out in relation to the goal variable Attrition Status (0 = Retained, 1 = Dropout). The picture shows clearly what the difference is between students who stayed and those who left based on their grades. You should think about things like how often you go to class, how well you do on tests, and how well you do your homework. Some of the socioeconomic factors that clearly affect the likelihood of attrition are family income, part-time work, and family responsibilities. Family income, part-time work, and family responsibilities are some of the socioeconomic factors that clearly affect the likelihood of attrition. These findings demonstrate that the Explainable AI-based framework can identify students at high risk and implement early intervention strategies to reduce dropout rates.

5-Fold Cross-Validation Performance

Model	**Cross-Validation Accuracy**
Random Forest	81.00%
Gradient Boosting	74.56%
Stacking	81.89%

Table III: Model Performance Based on 5-Fold Cross Validation

We tested the framework with 5-fold cross-validation to make sure it was strong enough to be used in other situations. The Stacking ensemble model was clearly better than all of the individual base models because it had an accuracy of 81.89%.The Random Forest model was 81.00% accurate, and the Gradient Boosting model was 74.56% accurate. This means that it was very good at making predictions.These results show that stacking is a good way to improve a model by using the strengths of several base classifiers. This makes the system for predicting student dropouts more accurate

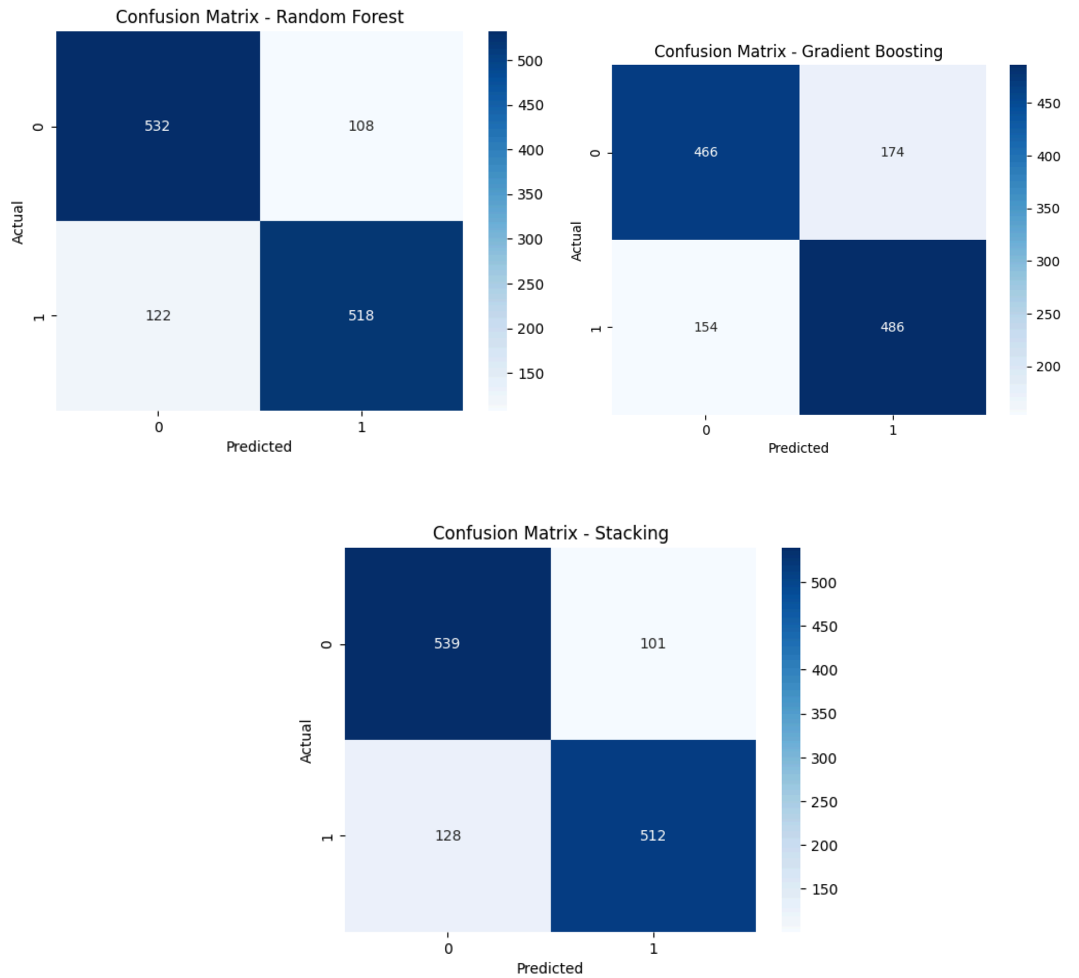


Figure III: Confusion Matrix Results

Model	True Negative (TN)	False Positive (FP)	False Negative (FN)	True Positive (TP)
Random Forest	532	108	122	518
Gradient Boosting	466	174	154	486
Stacking	539	101	128	512

Table V: Comparison of Model Performance for Predicting Student Attrition

The results show that the Stacking ensemble model sorts students better because it keeps the same number of students who are at risk of dropping out and those who are not. This makes it easier for schools to find both types of students and gives them better ways to warn them early.

Model	Accuracy	Precision	Recall	F1 Score
Random Forest	0.8203	0.8275	0.8094	0.8183
Gradient Boosting	0.7438	0.7364	0.7594	0.7477
Stacking	0.8211	0.8352	0.8000	0.8172

Table V: Model Performance Comparison for Student Attrition Prediction

The Stacking ensemble model was the best at finding students who were leaving school. It was the most accurate (82.11%) and the most precise (83.52%).The Random Forest model also did very well, with an accuracy of 82.03% and the highest F1 score (0.8183). This means that it could find the right answers and keep them in mind. The Gradient Boosting model, on the other hand, did a lot worse on all of the tests. The results indicate that ensemble-based techniques, particularly stacking, enhance the accuracy of predictions regarding students' likelihood of failure.

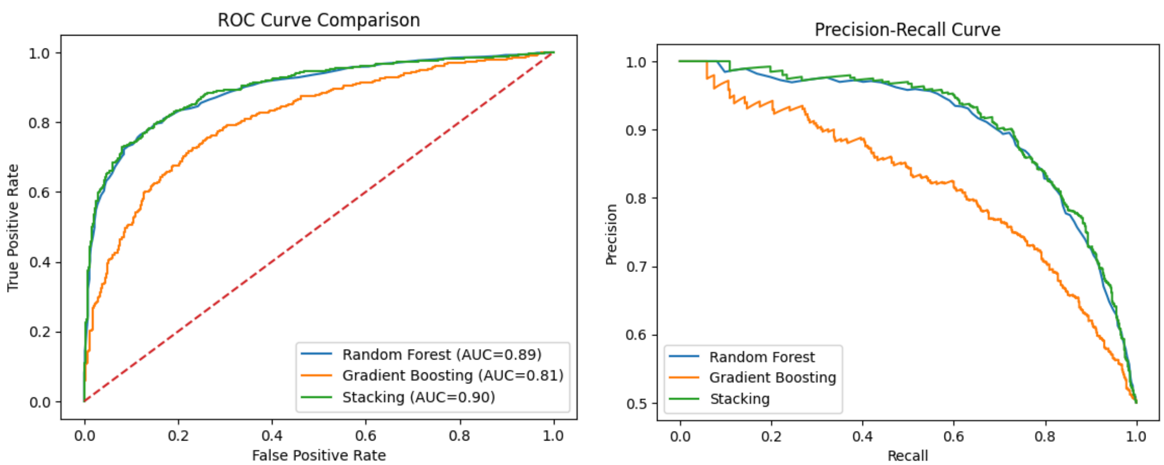


Figure IV: A Comparison of Machine Learning Models' ROC and Precision–Recall Curves

The picture shows how well the Random Forest, Gradient Boosting, and Stacking models did when they were compared using ROC and Precision Recall curves. The Stacking model had the best AUC (Area Under the Curve) = 0.94 in the ROC analysis. This means it was better at figuring out who stayed and who left. The Random Forest model worked very well (AUC = 0.89), but the Gradient Boosting model didn't work as well (AUC = 0.81).

The accuracy-Recall curves also show that the Random Forest and Stacking models stay more accurate even when the recall level goes up or down. This means they are better at finding students who are having problems and making fewer mistakes. The results show that the stacking model and other ensemble-based methods work better for predicting when students will drop out when used in the early warning framework that was suggested.

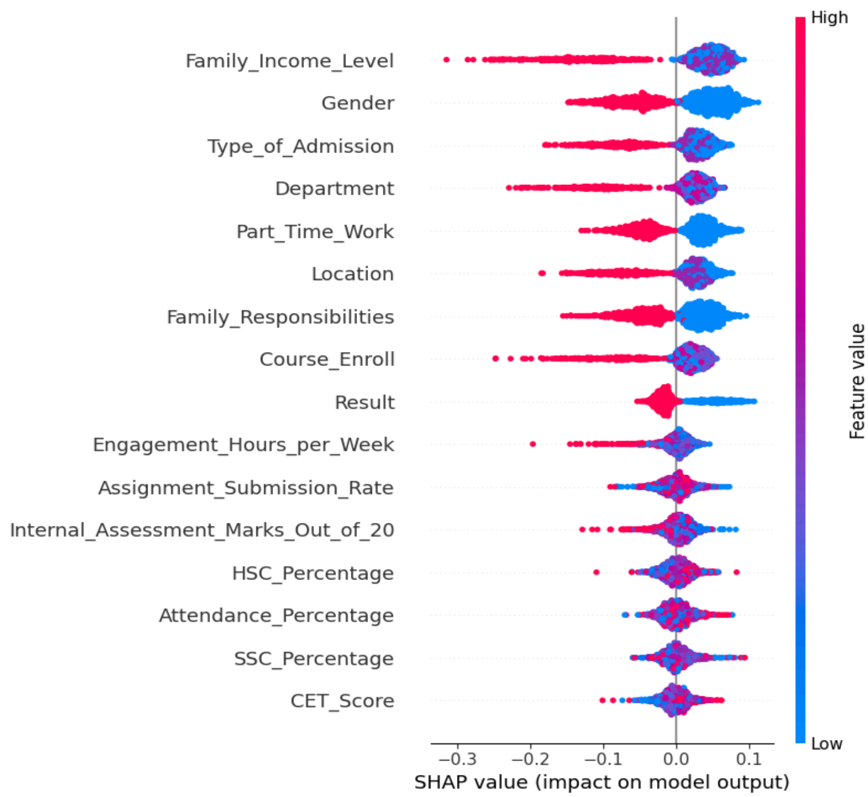


Figure V: SHAP Summary Plot Showing How Important Each Feature Is for Predicting Student Dropout

The SHAP summary picture shows how many things make it hard for the model to guess how many students will leave school. We put the features in order of how much they help us guess. Family income level, gender, type of admission, department, and part-time job are all important sociodemographic factors that affect the risk of dropping out, according to the results.

Grades, hours spent in class each week, assignment submission rates, internal evaluation scores, and attendance percentages all have an effect on the model's predictions. The graph shows that the red dots have more important features than the blue dots. The horizontal positions of the points show how much they change the prediction and how they do it.

The SHAP analysis helps schools figure out what makes students more likely to leave school. This makes the suggested AI framework easier to understand and gives schools the tools they need to start using some early intervention strategies.

Risk Level	Number of Students
High	516
Medium	441
Low	323

Table VI: Distribution of Student Attrition Risk Levels

The risk stratification put students into three groups: High, Medium, and Low, based on how likely it was that they would be in danger. This is based on how likely it is that they will go. 516 students are in the group that is most likely to get hurt, according to the results. This means that a lot of them need help with their homework right away. There are also 441 students at medium risk, which means they need help and someone to watch over them. The other 323 students are in the low-risk group, which means they are doing very well in school. This risk-based categorisation works with the proposed early warning system by letting schools use different intervention measures for students who are at different levels of risk of dropping out

VI Conclusion

This paper examined an Explainable AI-driven early warning and intervention framework designed to forecast the likelihood of college student attrition based on their risk profiles. The trial results show that schools can use basic techniques and predictive modelling to find students who are likely to drop out and help them quickly and in a way that works for them.

We discovered significant factors that may assist in predicting student dropout in relation to RQ1. SHAP analysis revealed that sociodemographic factors such as family income, admission type, department, part-time employment, and familial obligations, along with academic performance metrics including attendance, internal grades, assignments, engagement hours, and final outcomes, significantly influence dropout predictions. As shown by the importance of academic engagement and performance metrics as key indicators, student behaviour is very important for keeping students in school.

The Stacking ensemble strategy had the highest AUC (0.94) and accuracy (82.11%) for RQ2. It found students who were having problems. The Random Forest models show that this group method works better than regular monitoring and can quickly and consistently help students who are in danger.

Our data for RQ3 indicates a significant inverse correlation between the percentage of students who attend and the percentage of students who withdraw. The density distribution and SHAP analysis support the idea that being involved in school activities is the main reason why students go to school regularly. The findings of these studies indicate that students with high rates of absenteeism are significantly more prone to discontinuing their education.

Some good signs that a student is likely to drop out for RQ4 are how many hours they spend on schoolwork each week, how quickly they turn in their work, and their internal evaluation score. You should pay attention to academic signs that are based on how well you behave and how much you participate. This is because students who aren't as involved are more likely to leave school.

The suggested framework sorts students into three groups based on how risky they are: high, medium, and low. This is the answer to the fifth question. There are 516 students in the group with the highest risk, 441 in the group with the medium risk, and 323 in the group with the lowest risk. Schools can use this information about where the risk is to make changes that are specific to that risk. Programs that are supposed to help can keep an eye on students who are at a medium risk. Students who are in danger can get help and advice about their schoolwork right away.

The results also back up the study's goals and ideas. Feature analysis uncovers significant factors contributing to student attrition (O2), while predictive models identify students at risk of early departure (O1).The risk-level categorisation method is also the basis for an intervention framework (O3) that helps people use data to make choices that will keep students in school (O4).

The results show that schools can use Explainable AI-powered predictive analytics frameworks to find students who are having trouble and help them. This means that fewer students drop out of school and more students do well.

References

[1] R.W. Rumberger, "High School Dropouts: A Review of Issues and Evidence", Review of Educational Research, Vol. 57, pp. 101-124, 1987.
[2] J.S. Catterall, "On the Social Costs of Dropping Out of School", The High School Journal, Vol. 71, pp. 19-30, 1987.
[3] R. Balfanz and D.J. Maciver, "Preventing Student Disengagement and Keeping Students on the Graduation Path in Urban Middle-Grades Schools: Early Identification and Effective Interventions", Educational Psychologist, Vol. 42, pp. 223-235, 2007.
[4] M. Dynarski and P. Gleason, "How Can We Help? What We Have Learned from Recent Federal Dropout Prevention Evaluations", Journal of Education for Students Placed at Risk, Vol. 7, pp. 43-69, 2002.
[5] A. Jacquline Dorothy and M. Sangeetha, "Predicting Academic Dropouts Using AI and Behavioral Data: A Hybrid Deep Learning Framework for Student Retention", *International Journal of Environmental Sciences*, June 2025, DOI: 10.64252/6a5y9t48.
[6] Psyridou, M., Halme, J., & Nieminen, J. (2024). Early signs of dropout risk in Finnish primary students: An ML-basedlongitudinal analysis. Scandinavian Journal of Educational Research, 68(1), 123–145,https://doi.org/10.1080/00313831.2024.20148217.
[7] Elbouknify, I., Mahdi, A., & Hassan, F. (2025). SHAP-enhanced ensemble learning for dropout detection in Moroccanschools. International Journal of Educational Technology in Higher Education, 22(1), 44–61,https://doi.org/10.1186/s41239-025-00312-8 (Placeholder)
[8] Cheng, L., Zhang, X., & Yuan, H. (2025). A dual-modal behavioral detection model for early dropout prediction in digitallearning. Computers & Education, 196, 104678, https://doi.org/10.1016/j.compedu.2025.104678
[9] Arno, S., Kumar, R., & Bhattacharya, M. (2025). DropWrap: Scalable neural networks for district-level dropout prediction inIndia. Journal of Educational AI, 12(2), 112–129, https://doi.org/10.1016/j.jeai.2025.02.005 (Placeholder)
[10] Jacquline Dorothy, A., & Sangeetha, M. (2025). Predicting academic dropouts using AI and behavioral data: A hybrid deep learning framework for student retention. *International Journal of Environmental Sciences*. https://doi.org/10.64252/6a5y9t48