International Journal of Applied Mathematics
Volume 38 No. 10s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

SMART DROPOUT RISK PREDICTION USING STACKED ENSEMBLE LEARNING ON MULTI-MODAL EDUCATIONAL DATA

***P. Revathy, M. Priya,**

[1]Research Scholar,

Department of Computer and Information Science,

Faculty of Science,

Annamalai University, INDIA

*Revathisathish13@gmail.com

[2]Assistant Professor,

Department of Computer Science,

PSPT MGR Govt.Arts and Science College,

Sirkali, Puthur, INDIA

mpriyaau@gmail.com

Abstract:

Learning fosters opportunity, critical skill development, and breaks cycles of poverty, playing a vital role in personal and societal growth. Academic excellence and reducing student dropout rates remain key challenges. Dropout is a chronic issue that affects student achievement and institutional effectiveness. Existing prediction models often rely only on structured data such as grades and attendance, ignoring unstructured inputs like behavior, emotional feedback, and peer interactions. This study proposes a smart dropout risk prediction method using a stacked collaborative learning architecture that integrates multi-modal educational data. By combining structured information (e.g., academic performance and punctuality) with unstructured data (e.g., student feedback and social interactions), the model aims to improve accuracy and early identification of at-risk students. The process includes data pre-processing, multi-source integration, feature construction, and training base learners such as Gradient Boosting, Random Forest, SVM and k-NN. The system also addresses key real-world issues like adaptability, accessibility, and data imbalance. Experimental results on institutional and public datasets show the proposed model outperforms conventional approaches, achieving over 92.8% precision, 94.12% accuracy, and a reduced false-negative rate. This intelligent system supports proactive interventions to improve student retention and academic outcomes.

Keywords: Student Dropout Prediction, Stacked Ensemble Learning, Multi-Modal Educational Data, Early Warning System, Machine Learning, Structured and Unstructured Data, Student Retention, Behavioural Analysis, Academic Performance, Data Mining.

1. Introduction

In addition to improving the intellectual prowess, learning nurtures creativity and analytical capacity as well as problem solving skills [1]. It facilitates the development of society, changes the picture of the poverty cycle, and opens prospects. Employing its transformative effect,

International Journal of Applied Mathematics
Volume 38 No. 10s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

countries, organizations and governments all over the world still insist on quality access to education. Ensuring that all the students excel academically and can also complete their studies successfully is extremely tasking to the education system [2]. Highly performing students are highly self disciplined, motivated and interested in learning. To get an education of higher qualification, win prestigious scholarships and establish a successful career [3]. Some of the damaging effects of dropping out of school are limited access to higher learning opportunities and employment, high chances of becoming poor, poor mental health outcomes, and well-being [4]. To understand how to ensure every student reaches his or her fullest potential it is imperative to focus and mitigate risks to academic achievement and dropout early. Early identification provides timeliness in support since the students who are expected to perform better or worse in examinations are known. This enables teachers to use preventive measures to boost the positive or solve challenges [5]. It allows personal attention, the definition of resources, and personal learning opportunities to support individuals. Advanced coursework and enrichment programs can likewise reward the high achievers [6]. Machine Learning (ML) can provide insights that can be missed through current techniques and provide information through training data-detecting algorithms to spot patterns in the said data. The tools assist the educators to make sound decisions, a factor that is informed by statement of factors that are generating academic success or risks of dropping out. The AI models are able to infer the probability that a person will be successful concerning the performance, engaged, and other indicators to facilitate individualized learning and intervention [7].

The big challenge continued to be the high item of dropout when a large number of learners will not continue the study to the end. Model performance can be deteriorated by unnecessary and irrelevant features of the dataset [8, 9]. This is a problem particularly in in-session learning algorithms where the deluging amount of non-dropout data makes it hard to get a good prediction of the dropout. Existing methods of forecasting usually have a difficult time countering this imbalance. Reliable early-warning systems may also be hampered by inclusion of irrelevant variables that consequently reduces predictive accuracy [10]. In spite of the more recent developments, there remains an issue about research gap. ML methods that have been used previously have been studied mainly on an individual basis, but more progressive methods like those of stacked classiers are required to be more competent against sophisticated nature of educational data. The advantages of various algorithms can be used to create such classifiers that would enhance the early identification of the risk of dropping out of school and the academic prosperity [11, 12]. Blended learner creates a mixture of current classroom instructions and online learning, proposes a very personal and flexible approach to learning. It has found application in the higher education system because it makes students more engaged and increases the chances of academic success [13].

In a bid to comprehend and enhance educational outcomes, data on students and learning situations must be gathered, measured, analysed, and reported. It enables assessment of the learning behaviour, proper forecasting of academic performance, and interventions in due time [14]. The ability to predict at-risk students accurately is also important when it comes to the implementation of effective educational strategies, which are meant to optimize the performance of students and make the organization more efficient. Although a lot of existing

Volume 38 No. 10s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

models are dedicated to regression and binary classification (e.g., pass/fail), prediction of the final grades or overall academic results is not as popular. The individual data profile of each student is different, which is why academic satisfactions or failings should be viewed as a matter of course, and thus adaptive and individual predictive models are required [15].

2. Related Works

Predicting students' academic success has been addressed in numerous studies, either as a classification problem or a regression task. Various data analytics techniques have been applied to forecast student performance along with identifying the key factors that influence academic achievement. Several research directions exist within the domain of student performance prediction. These include early dropout forecasting, analysis of intrinsic factors affecting academic performance, and the use of statistical methods to evaluate student outcomes [16]. In another study, multiple data mining methods were utilized to identify at-risk students. Logistic regression often served as a baseline model and was compared with other approaches to evaluate performance [17]. ML and DL have been widely applied across diverse domains, including medical image analysis, object detection, knowledge discovery, and risk assessment. For example, ML models such as SVM and K-NN have been used to analyze risks in global software development of reearch. Similarly, applied predictive models to assess risks associated with assets, timelines, and costs [18].

It is possible to process large-scale student log data in online platforms and extract patterns of academic success with the assistance of Learning Analytics (LA) without any manual procedures [19]. This would allow teachers to provide differentiated learning activities and specification assistance. It is possible to use LA to analyze learning advances and construct the predictive models by evaluating and combining complicated sets of data. LA has some major characteristics that are information analytics, learning recommendations and alerts, prediction modeling, relationship mining, and self-evaluation tools [20]. The main benefit of LA is personalized feedback that is required to assist instructors to provide selective guidance and support based on the needs of individual learners. LA would help develop and deploy, adaptive learning environments, which would consider the learning preference of learners and have flexible, student-led learning experiences [21]. LA tools can also be used by students to track their progress against their own learning requirements so that they can identify points of strength and weakness and will gradually develop as becoming more independent as learners. Nevertheless, the current evidence is not sufficient to conduct a thorough assessment of the effects of LA on learning outcomes even despite the above benefits. Then identified some implementation issues like an effective alignment of LA insights to the particular subject area, like science teaching, maximizing the utilization of data in different learning environments, and the issue of ethics and information privacy [22].

Educational Data Mining (EDM) is aimed at analysing the data involved in the study in order to learn the behavior of a student. The mentioned techniques can unveil important information that can be employed to change the structure of the course or to determine how students would perform or interact thus improving the learning conditions. Both types of approaches, EDM and LA, are mostly used in the methods of existing forecasting techniques [23]. Among the

Received: August 02, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

favorite areas of study is prior knowledge of whether a student will succeed or fail in his/her assignment and mostly in the initial weeks as the case is with the EDM-oriented investigations. Credible projections are applied to forecast failure, determine endangered students and provide them with immediate support. The first studies that analyzed the normal behavior of students throughout teaching in instruction aided to propagate and facilitate the establishment of educational analytics. LA creates grounds of advancement in the entity of distinct stakeholders, consisting of service providers, instructional designers, administrators, educators, and students. In order to enhance prediction on performance, LA provides actionable insights to data and placement the data in context that is vital when making substantive educational decisions [24].

Computer-based techniques are frequently employed to forecast student performance. This section outlines practical examples that help establish the context for the present study. Recently, the focus has shifted toward analysing student demographics, Learning Management System (LMS) activity data, and cognitive abilities to predict academic success. Various approaches are used to identify at-risk students early, typically through binary classification models that determine whether a student is at risk or not [25]. A range of ML classification techniques are applied to cross-sectional data for early prediction. Semi-supervised learning also plays a key role in the early detection of at-risk individuals. Developed algorithmic strategies include interpretable classification rule mining, genetic and evolutionary algorithms, multi-view learning, multi-objective optimization, ensemble models and DL approaches [26].

Due to the wide variability in educational data, a single-model solution is insufficient. For example, predicted student success based on prior performance in specific subjects. Researchers evaluated model effectiveness by selecting optimal features. Academic scores and grades were forecasted using regression and classification techniques [27, 28]. The results showed that regression and decision trees based on genetic algorithms performed more accurately. In studies related to dropout prediction, systems achieved a success rate of approximately 92% for dropout cases and over 85% overall accuracy. Multilayer Perceptron (MLP) system achieved up to 87% reliability in forecasting dropouts using sequential learning data. This study compares the proposed approach in terms of efficacy, accuracy, and adaptability against the method. It evaluates the model's generalizability and robustness by comparing it to other similar approaches [29].

3. Materials and Methods

The AI-based dropout risk forecasting system uses a stacked ensemble learning approach, integrating structured data (e.g., grades, attendance) and unstructured data (e.g., student feedback, forum posts) from institutional and public educational sources shown in Figure 1. NLP and social network analysis extract features like sentiment and interaction metrics, while TF-IDF and centrality measures enhance text and social features. After normalization and feature selection, base models (SVM, RF, k-NN, GBM) feed into a meta-learner (Logistic Regression or XGBoost).

Received: August 02, 2025

International Journal of Applied Mathematics
Volume 38 No. 10s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

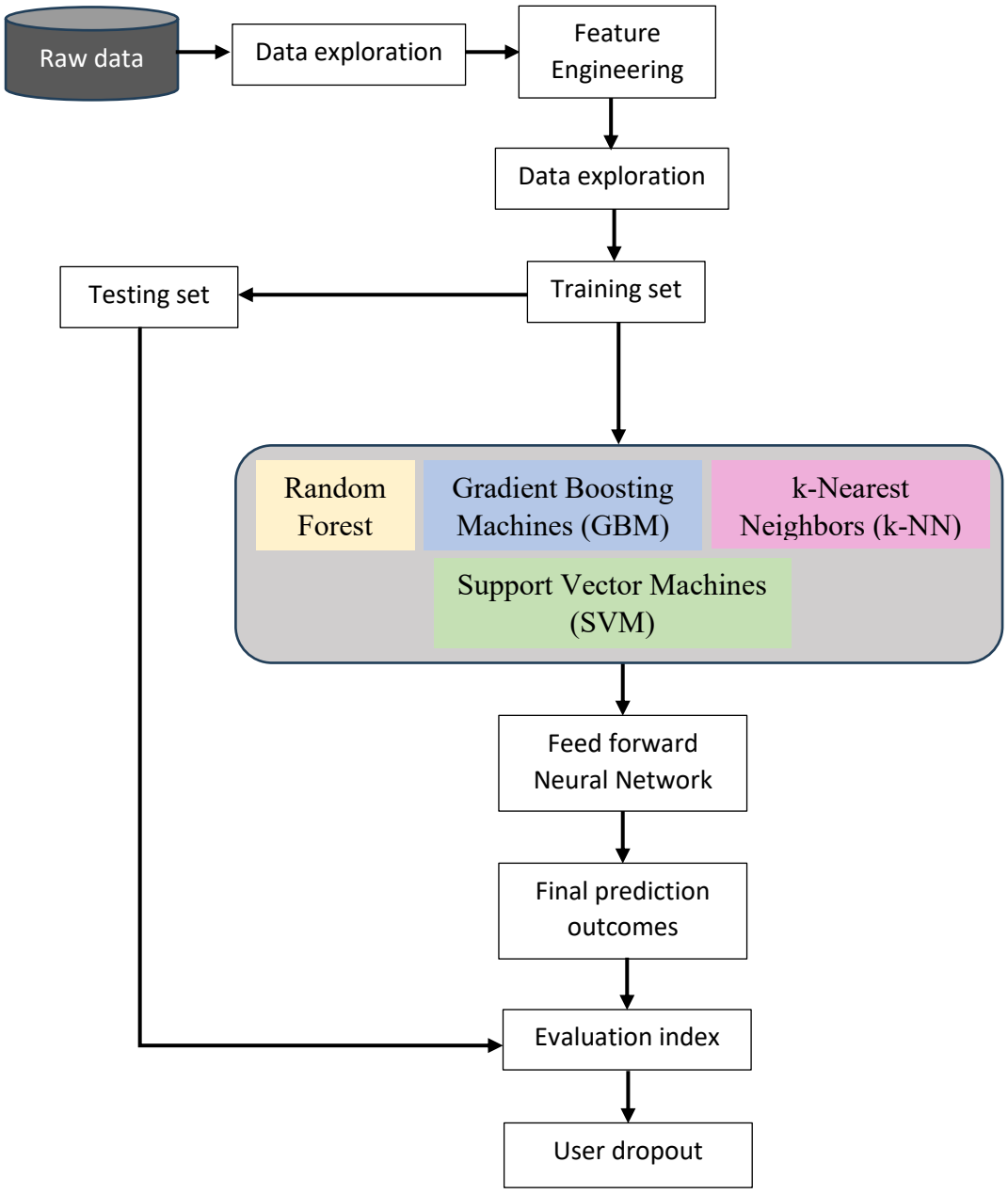


Figure 1: Proposd Architecture

Phase 1: Base Model Training

Let the dataset be: $D = \{(i_x, j_x)\}_{x=1}^{N}$ (1)

Where $i_x \in R^d$ the feature vector for student x, $j_x \in \{0,1\}$ is the binary dropout label (0= no dropout, 1 dropout), N is the total number of student instances.

Received: August 02, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

Define a set of diverse base classifiers:$F_{base} = \{h_1(.), h_2(.), \dots, h_K(.)\}$ (2)

Each classifier h_K learns a decision function on the training set D_{train}:

$h_k(i_x) = \hat{j}_x^{(k)} \; for \; k = 1,2, \dots, K$ (3)

Phase 2: Class Probability Estimation

Instead of predicting hard class labels, each base leaner outputs a probabilistic prediction:

$h_k(i_x) = \left[P^{(k)}(j = 0|i_x), P^{(k)}(j = 1|i_x)\right]$ (4)

Let: $P_x^{(k)} = P^{(k)}(j = 1|i_x)$ (5)

Thus, the prediction from all base models for sample i_x can be aggregated into a probability vector: $P_x = \left[P_x^{(1)}, P_x^{(2)}, \dots, P_x^{(K)}\right]$ (6)

Phase 3: Feature Augmentation

Construct a new feature representation i_x^{arg} by concatenating the original input features with the probability outputs: $i_x^{arg} = [i_x || P_x] = \left[i_{x1}, i_{x2}, \dots, i_{xd}, P_x^{(1)}, \dots, P_x^{(K)}\right]$ (7)

Here,"||" represents vector concatenation. This augmentation provides a meta-level enriched representation containing both low-level (raw input) and high-level (model insights) features.

Phase 4: Meta-Learner Training

Train a meta-learner G(.), typically MLP, on the augmented dataset:

$\hat{j}_x = G\left(i_x^{arg}\right) = G\left(i_x, p_x^{(1)}, \dots, p_x^{(K)}\right)$ (8)

The final prediction for student x is: $\hat{j}_x = \begin{cases} 1, if \; G\left(i_x^{arg}\right) \geq \tau \\ 0, otherwise \end{cases}$ (9)

Wwhere r is a decision threshold (commonly set to 0.5).

Performance Objective Function: The training of the meta-learner involves minimizing a binary cross-entropy loss function:

$L = -\frac{1}{N}\sum_{x=1}^{N}[j_x \log(\hat{j}_x) + (1 - j_x)\log(1 - \hat{j}_x)]$ (10)

This helps to optimize the prediction accuracy and minimize false classifications, particularly for dropout cases.

3.1 Dataset Description

The dataset for the smart dropout risk prediction model includes structured, behavioral, psychological, and unstructured data to capture a student's academic journey shown in Table 1. Each student is uniquely identified by Student_ID, with demographic details like age and gender. Academic performance is represented by Course_Enrolled, GPA, and Attendance_Rate. Behavioral features include Assignment_Submission_Rate, Login_Frequency, and Forum_Participation, reflecting student engagement. Sentiment

Received: August 02, 2025

International Journal of Applied Mathematics
Volume 38 No. 10s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

analysis of Feedback_Sentiment from unstructured student comments offers insight into learner satisfaction. Peer_Interaction_Score, derived through social network analysis, gauges social engagement. The target variable, Dropout_Status, is binary (1 for dropout, 0 for enrolled). This rich feature set enhances the model's predictive capability.

Table 1: Data type description

Feature Name	Type	Data Type	Description
Student_ID	Identifier	Categorical	Unique ID for each student
Gender	Demographic	Categorical	Male or Female
Age	Demographic	Numerical	Age of the student
Course_Enrolled	Academic	Categorical	Name or code of the course enrolled
GPA	Academic	Numerical	Current Grade Point Average (0.0 – 4.0 scale)
Attendance_Rate (%)	Academic	Numerical	Percentage of classes attended
Assignment_Submission_Rate	Behavioral	Numerical	Proportion of assignments submitted on time
Login_Frequency	Behavioral	Numerical	Number of times the student logs in per week
Forum_Participation	Behavioral	Numerical	Number of posts/comments in course discussion forums
Feedback_Sentiment	Unstructured NLP	Numerical	Sentiment score (from -1 to +1) extracted from textual feedback
Peer_Interaction_Score	Social	Numerical	Social network centrality score or communication level with peers
Dropout_Status	Target Variable	Binary	1 = Dropped out, 0 = Continued (used as label)

Table 2: Sample data

Student_ID	Gender	Age	Course_Enrolled	GPA	Attendance_Rate (%)	Assignment_Submission_Rate	Login_Frequency	Forum_Participation	Feedback_Sentiment	Peer_Interaction_Score	Dropout_Status
S001	M	21	Data Science	3.2	87	0.95	12	6	0.65	0.78	0
S002	F	23	AI & ML	2.4	62	0.70	5	2	-0.30	0.35	1
S003	M	20	Web Development	3.8	95	1.00	14	9	0.85	0.92	0

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

S004	F	22	Cyber Security	2.9	71	0.60	4	1	-0.10	0.40	1
S005	M	24	Cloud Computing	3.5	90	0.88	11	7	0.70	0.81	0

The dataset predicts student dropout risk using structured, behavioral, and unstructured variables (Table 2). Each record includes demographics (Student_ID, gender, age), academic indicators (GPA, Course_Enrolled, Attendance_Rate), and behavioral features (Assignment_Submission_Rate, Login_Frequency, Forum_Participation). Psychological and social factors are derived from Feedback_Sentiment and Peer_Interaction_Score. The target variable, Dropout_Status, is binary (1 = dropout, 0 = enrolled). This diverse dataset improves model learning by capturing multiple dropout predictors. Data comprehension assesses feature quality and relevance. EDA identifies trends and correlations; for instance, low attendance or submission rates often correlate with dropout. Visual tools support initial insights.

3.2 Data preparation

The data preparation step transforms raw data into a clean, machine-readable format suitable for modeling shown in Table 3. Textual feedback is processed using NLP to extract sentiment scores, reflecting emotional engagement. Social network metrics like degree centrality quantify peer interactions. After integrating all data sources, feature engineering creates new variables, and feature selection (e.g., mutual information) retains the most informative ones. These steps ensure robustness, accuracy, and better generalization during modeling.

Table 3: Data preparation features

Period	Number of Features	Feature List	Final Outcome Distribution
Q1 (Week 1–3)	7	Att_T, Z1–Z5, A1	—
Q2 (Week 4–6)	12	Att_T, Z1–Z9, A1, A2	Pass: 83 (76.%)
Q3 (Week 7–9)	17	Att_T Att_T, Z1–Z13, A1, A2, T	Fail: 27 (20.69%)
Q4 (Week 10–12)	21	Att_T, Z1–Z17, A1–A3, T	—

Note: Att_T – Attendance Tracker; Z1–Z17 – Weekly assessment scores; A1–A3 – Assignment scores; T – Test score

Table 3 illustrates the evolving feature space across four time periods (Q1–Q4), each spanning three weeks. In Q1, early indicators such as Att_T, the first assignment (A1), and quizzes (Z1–Z5) are considered. By Q2, the feature set expands with a second assignment (A2) and quizzes up to Z9, enhancing predictive capability. Student performance analysis shows that by Q2, 83 students (76.9%) were on track to pass, while 27 (20.69%) were at risk by Q3, based on additional features like test scores (T) and extended quizzes (Z1–Z13). They consist of 21 features in Q4, which are all the assignments (A1-A3), quizzes (Z1-Z17), and final test scores,

International Journal of Applied Mathematics
Volume 38 No. 10s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

which offer a well-grounded basis to predict the dropout by using big data. This gradual growth in number of features is compatible to the gradual stacking learning style of the stacking framework.

3.3 Pre-processing

An essential process in obtaining the right results in the prediction of dropout is pre-processing of multi-modal educational data. Numerical values that are not drawn (e.g. attendance or GPA) are replaced by mean, whereas categorical gaps (e.g. gender) are filled in by mode. Sentiment analysis is the utilisation of NLP procedures to get a polarity level out of free feedback on a scale of 1 to -1 (i.e. positive to negative). The peers network connectedness is measured to capture the social interaction and applicable metric is the graph-based such as degree centrality. This extensive pre-processing makes the data clean, and consistent and loaded with the relevant features during pre-processing, prior to reaching the modeling stage.

Missing Values: Missing values are the values that occur on numerical and categorical features in education data.Numerical Features (e.g., GPA, Attendance): $i_x^{(y)} = \begin{cases} i_x^{(y)}, & if i_x^{(y)} \neq NaN \\ \mu_y, & if\ i_x^{(y)} = NaN \end{cases}$ (11)

Where: $i_x^{(y)}$ is the value of the yth feature for the xth student, $\mu_y = \frac{1}{N_y}\sum i^{(y)}$ is the mean of the yth feature for all available values.

Categorical Features (e.g., Gender): Use mode imputation:

$$i_x^{(y)} = \begin{cases} i_x^{(y)}, & if\ i_x^{(y)} \neq NaN \\ mode\left(i^{(y)}\right), & if\ i_x^{(y)} = NaN \end{cases} \quad (12)$$

Encoding Categorical Variables: Convert categorical features into numerical form: One-Hot Encoding for nominal variables like Gender and Course Enrolled:

$$Gender_{Male} = \begin{cases} 1, & if\ Male \\ 0, & otherwise \end{cases} \quad Gender_{Female} = \begin{cases} 1, & If\ Female \\ 0, & Otherwise \end{cases} \quad (13)$$

This transforms a single categorical column into multiple binary columns.

Feature Scaling: To ensure all features contribute equally, Z-score normalization is applied:

$$z_x^{(y)} = \frac{i_x^{(y)} - \mu_y}{\sigma_y} \quad (14)$$

Where: $z_x^{(y)}$ the normalized value, μ_y is the mean of feature y, σ_y is the standard deviation of feature y. This transforms all features to a standard scale with mean = 0 ; and standard deviation = 1.

Sentiment Analysis for Unstructured Feedback: Textual feedback is converted into sentiment scores using NLP models like VADER or TextBlobx

$$Sentiment_x = \frac{Positive_x - Negative_x}{Total_x} \quad (15)$$

Received: August 02, 2025

International Journal of Applied Mathematics
Volume 38 No. 10s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

Resulting in a polarity score between -1 and 1, where: +1 = very positive, 0 = 1 neutral, -1 = very negative.

Social Network Feature Extraction: From peer interaction data (eg, forum replies, peer chat logs), graph-based features are generated: Degree Centrality: $C_D(v_x) = \frac{deg(v_x)}{n-1}$ (16)

Where: v_x is the student node, $deg(v_x)$ is the number of direct connections, n is the total number of students. These features quantify how connected or isolated a student is.

Feature Selection: Reduce dimensionality using Mutual Information (MI):

$X(I_y; J) = \sum_{i \in I_y} \sum_{j \in J} P(i,j) \log\left(\frac{P(i,j)}{P(i)P(j)}\right)$ (17)

Features with higher MI scores with the dropout label J are retained.

Final Output: After all pre-processing: All inputs are numeric, Missing values are imputed, Features are scaled, Text and graph features are quantified, Optional feature selection applied.

This cleaned and normalized dataset is then ready for training the base and Meta classifiers in the stacked ensemble framework.

Students must achieve a passing score of 5.0 or higher to be accepted. Each student's academic history includes information such as program registration, end level, program credits, practical semester, recommended semester, and other factors not considered in this study. The recommended semester refers to the next best course as outlined in the curriculum matrix of the Indian university's degree plan. The effective semester denotes the semester in which the student is currently enrolled. A student is considered to have dropped out if enrolled for a semester but fails to re-enroll in the same program. Transfers to another university or higher education institution are not classified as dropouts. Similar dropout criteria have been adopted in previous studies. The number of students enrolled in each degree program and the total number of dropouts per academic year. For example, among the 831 students enrolled in the third semester of the ADM degree, 100 dropped out in later semesters. Datasets were constructed for each semester, allowing dropout prediction across the entire academic program. Data from all previous semesters for each student were included ($1 \leq s_i <$ s).

The following data-cleaning steps were applied:

- Only classes with at least 20 enrolled students per academic year were considered.
- If a course had an enrollment record but no grade, it was assumed the student did not complete the assessments, and a zero was assigned.
- All grades were normalized from 0 to 1. A value of -1 indicated that the course was not taken, even if used for feature representation.

3.4 Stacked Ensemble Learning on Multi-Modal Educational Data

Stacked ensemble learning is a powerful machine learning technique that combines the strengths of multiple base learners through a two-level hierarchical framework shown in Figure 2. The first level consists of classifiers like SVM, RF, k-NN, GBM each trained on the same

International Journal of Applied Mathematics
Volume 38 No. 10s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

dataset to generate dropout probability estimates. These predictions are combined with the original features to form an enriched dataset. A meta-learner typically a MLP is then trained on this dataset to produce the final prediction. This approach enhances generalization, reduces overfitting, and captures complex patterns in multi-modal educational data, including structured (GPA, attendance), behavioral (login frequency), and unstructured inputs (comments, forum activity). Stacked ensemble learning allows each base model to specialize in different data aspects—neural networks for probabilistic patterns, decision trees for structured data, and boosting methods for complex interactions. The MLP effectively integrates these insights to deliver accurate and robust dropout predictions.

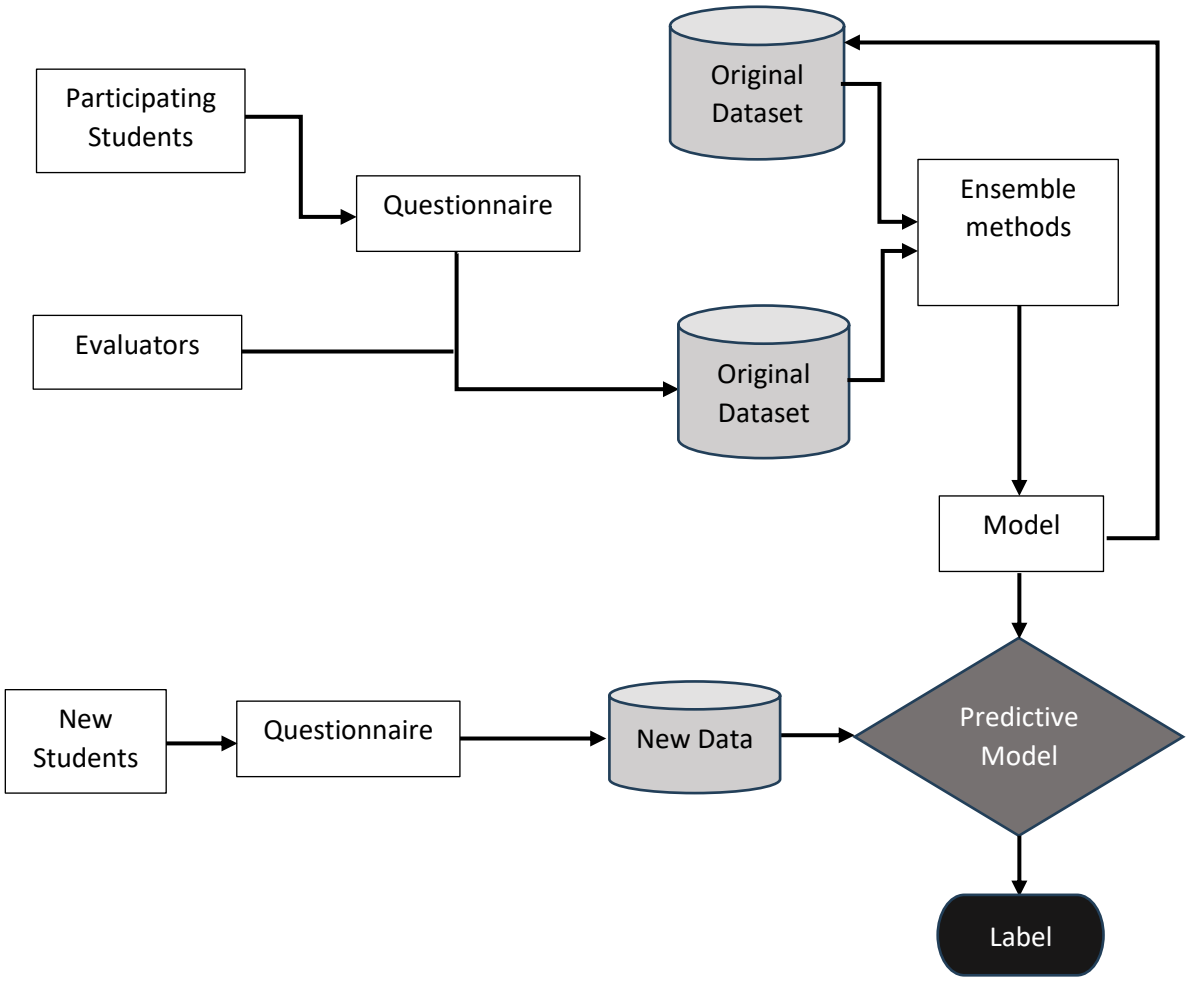
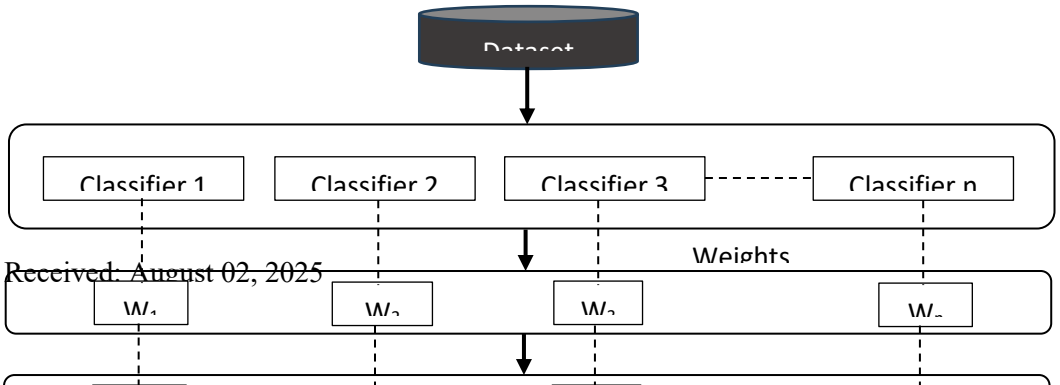


Figure 2: Stacked Ensemble Learning method



Received: August 02, 2025

International Journal of Applied Mathematics
Volume 38 No. 10s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

Figure 3: Flowchart of Weighed Classifier Ensemble Method

To further improve performance, this study employs a weighted ensemble approach, where the contribution of each base model is proportionally adjusted based on its predictive strength. This method allows the ensemble to leverage the strengths of individual classifiers, resulting in a more accurate and robust final prediction, as illustrated in Figure 3.

Algorithm: Intelligent Dropout Risk Prediction using Stacked Ensemble Learning

Input: Multi-modal educational dataset $D = \{(i_x, j_x)\}_{x=1}^{N}$ (18)

Where $i_x \in R^d$ are feature vectors and $j_x \in \{0,1\}$ are dropout labels.

Output: Predicted dropout label $\hat{j}_x \in \{0,1\}$ for each student.

Step 1: Data Pre-processing

Handle missing values: $i_x^{(y)} = \begin{cases} i_x^{(y)} \; if \; available \\ \mu_y \; or \; mode \;\; otherwise \end{cases}$ (19)

Encode categorical variables (e.g., one-hot encoding).

Normalize numerical features using Z-score normalization: $z_x^{(y)} = \frac{i_x^{(y)} - \mu_y}{\sigma_y}$ (20)

Extract: Sentiment from feedback: $s_x = Sentiment\ Analysis\ (f_x) \in [-1,1]$ (21)

Social score via degree centrality: $C_D(v_x) = \frac{deg(v_x)}{n-1}$ (22)

Step 2: Base Model Training

Train K base learners $h_k(.)$ on training set D_{train}: $F_{base} = \{h_1, h_2, \dots, h_K\}$ (23)

Each base learner produces class probability outputs: $P_x^{(k)} = h_k(i_x) = P^{(k)}(j = 1|i_x)$ (24)

Stack these into a probability vector: $P_x = \left[P_x^{(1)}, P_x^{(2)}, \dots, P_x^{(K)}\right]$ (25)

Step 3: Feature Augmentation

Received: August 02, 2025

Volume 38 No. 10s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

Augment original features i_x with base predictions: $i_x^{arg} = [i_x||P_x]$ (26)

This forms the input for the meta-leamer.

Step 4: Meta-Learner Training

Train meta-learner G (.) e.g., a Multi-Layer Perceptron (MLP): $\hat{j}_x = G\left(i_x^{arg}\right)$ (27)

Prediction thresholding: $\hat{j}_x = \begin{cases} 1 \ if \ G\left(i_x^{arg}\right) \geq \tau \\ 0 \ otherwise \end{cases}$ (28)

Step 5: Model Optimization: Optimize meta-leaner using binary cross-entropy loss:

$L = -\frac{1}{N}\sum_{x=1}^{N}[j_x \log(\hat{j}_x) + (1 - j_x)\log(1 - \hat{j}_x)]$ (29)

Apply early stopping or cross-validation to prevent overfitting.

Final Output: The trained ensemble system predicts dropout status for unseen students using:

$\hat{j}_{test} = G(|i_{test}||P_{test-}|)$ (30)

Base Learners Used: $h_1 =$ Random Forest (RF); $h_2 =$SVM; $h_3 = k - NN$; $h_4 =$ Gradient Boosting Classifier (GBC)

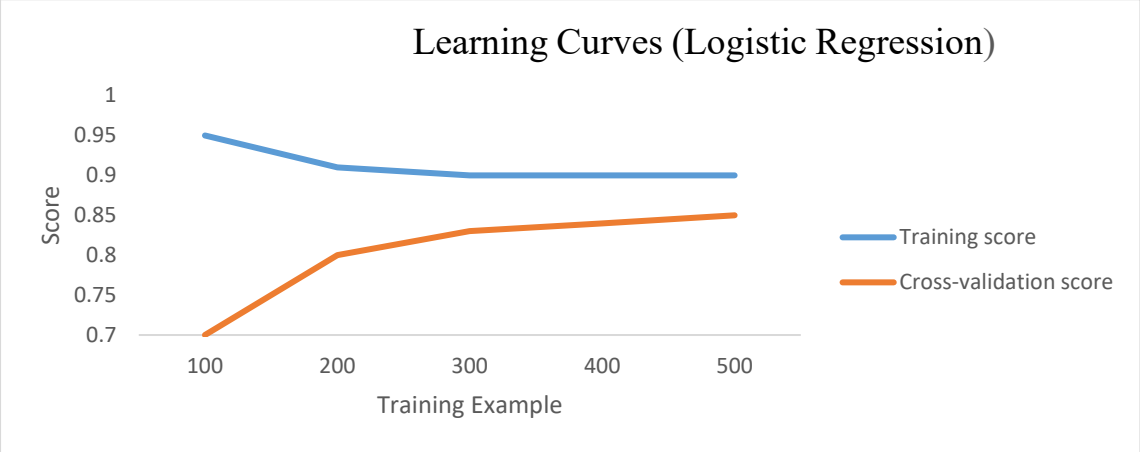
The proposed smart dropout risk prediction approach leverages multi-modal educational data using a stacked collaborative learning framework. Initially, the dataset undergoes thorough preprocessing, including sentiment analysis from textual feedback, one-hot encoding for categorical variables, Z-score normalization for numerical features, social network analysis for peer interactions, and imputation of missing values. Instead of generating direct predictions, these models output probability estimates indicating each student's dropout risk. These probabilities are concatenated with the original features to form an enriched feature set. This hierarchical stacking approach enhances prediction precision, robustness, and generalization, especially when dealing with complex, imbalanced, and heterogeneous educational datasets.

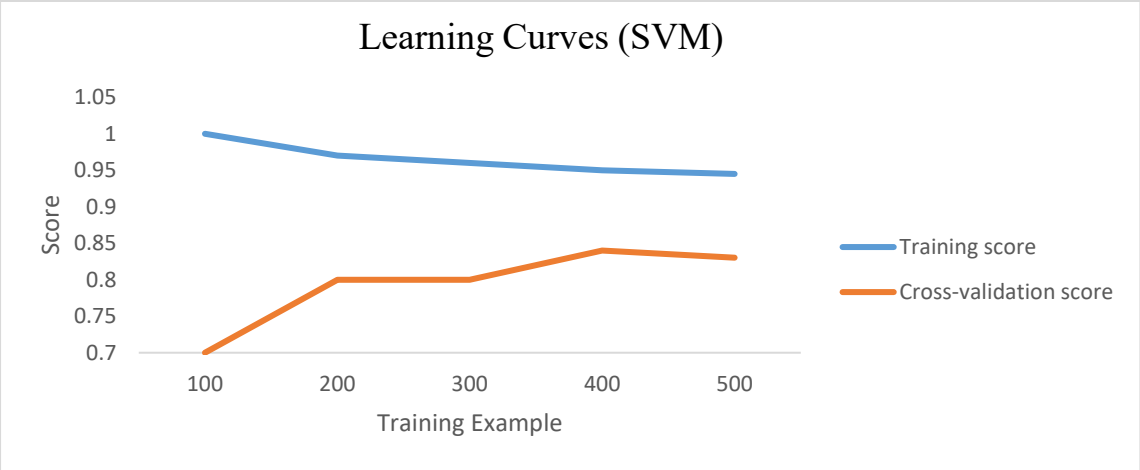
4. Results and Discussions

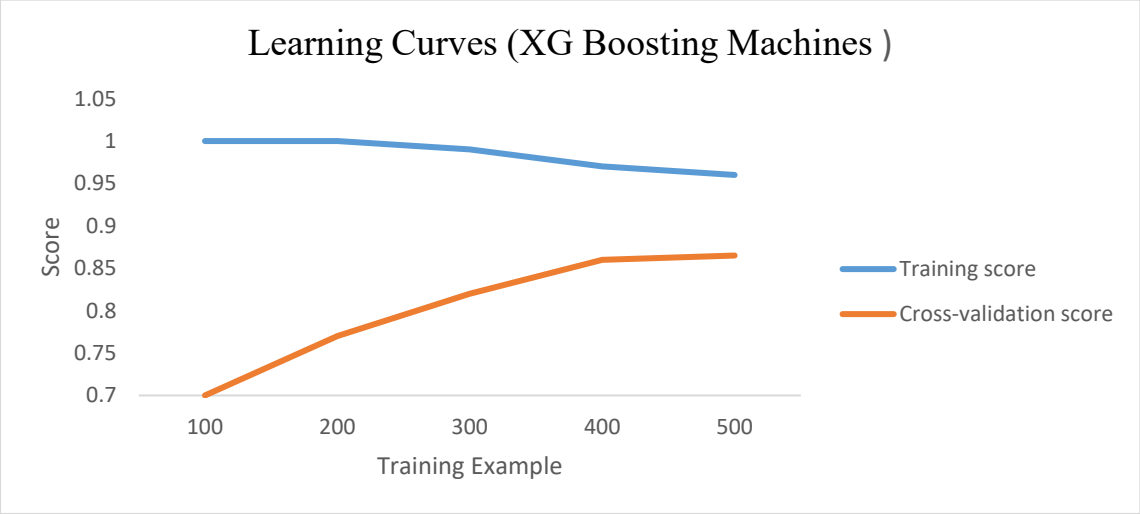
The experimental setup of the smart dropout risk estimation model evaluates the effectiveness of the proposed stacked ensemble framework on multi-modal educational data. Base learners are trained on the training data with hyper parameters optimized via grid search and cross-validation. Each model outputs class probabilities, which are concatenated with the original features to form an augmented input for the meta-learner. MLP serves as the meta-classifier, trained using five-fold cross-validation.

Received: August 02, 2025

International Journal of Applied Mathematics
Volume 38 No. 10s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)







Received: August 02, 2025

International Journal of Applied Mathematics
Volume 38 No. 10s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

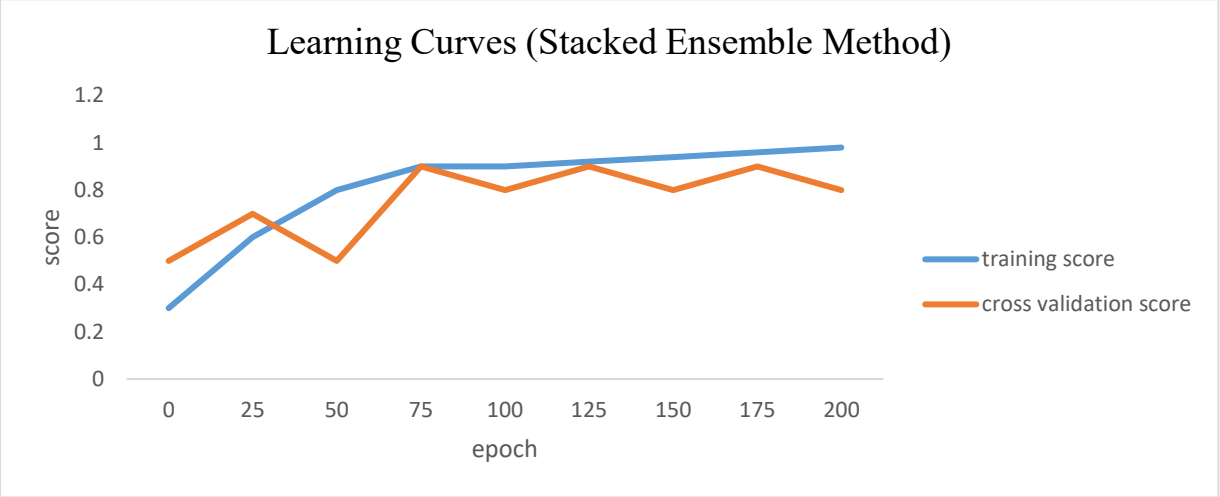
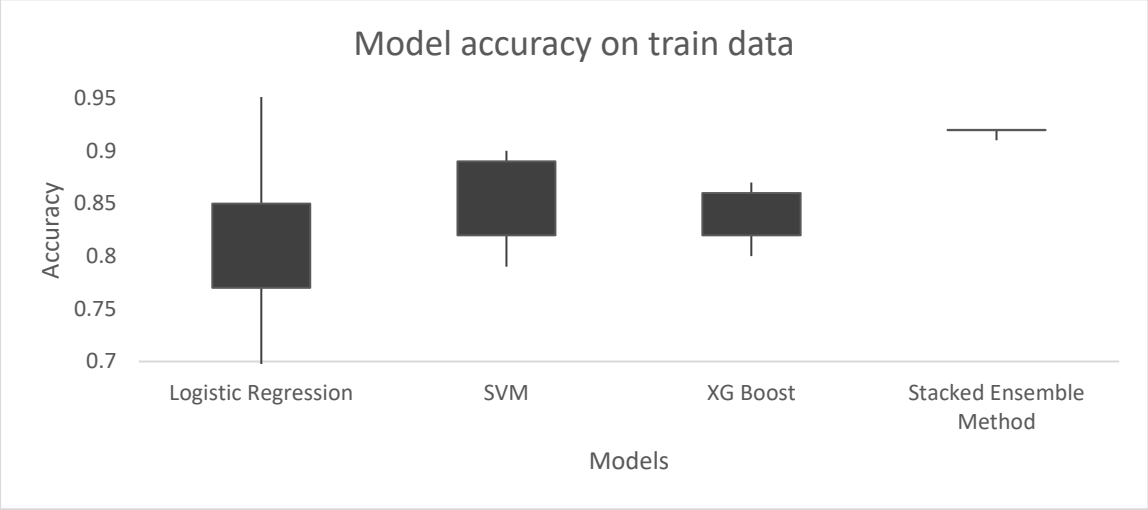
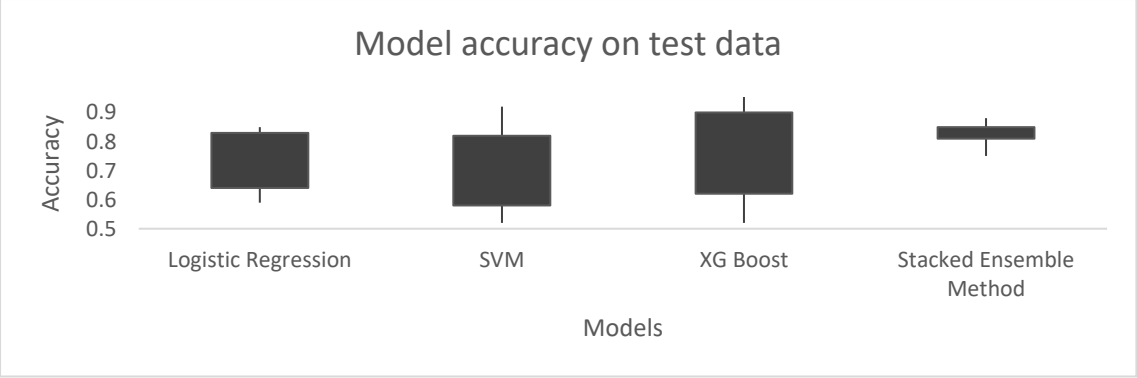


Figure 4: Learning curve of proposed stacked ensemble model



(a)



(b)

Figure 5: Accuracy of models on the original dataset. (a) Training data (b) Testing data

Some dataset features, such as parents' educational background and employment, contain both string and categorical values shown in Figure 4. To enable model learning, these values are converted to numerical format. Since some models like SVM are sensitive to varying data scales, all features are standardized within the range of 1 to 2. This helps minimize bias and avoids overfitting. With more training data, model predictions would become more reliable

Received: August 02, 2025

International Journal of Applied Mathematics
Volume 38 No. 10s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

and bias would further reduce. Figure 5's box plots show the performance of models using both training and test data. Proposed system performs best on the training set, showing high precision and consistency. Figure 6 displays the classifiers' confusion matrices, illustrating their effectiveness in identifying class boundaries. Most models effectively distinguish between classes 0 and 1 (representing classes A and B), despite some misclassifications at class borders.

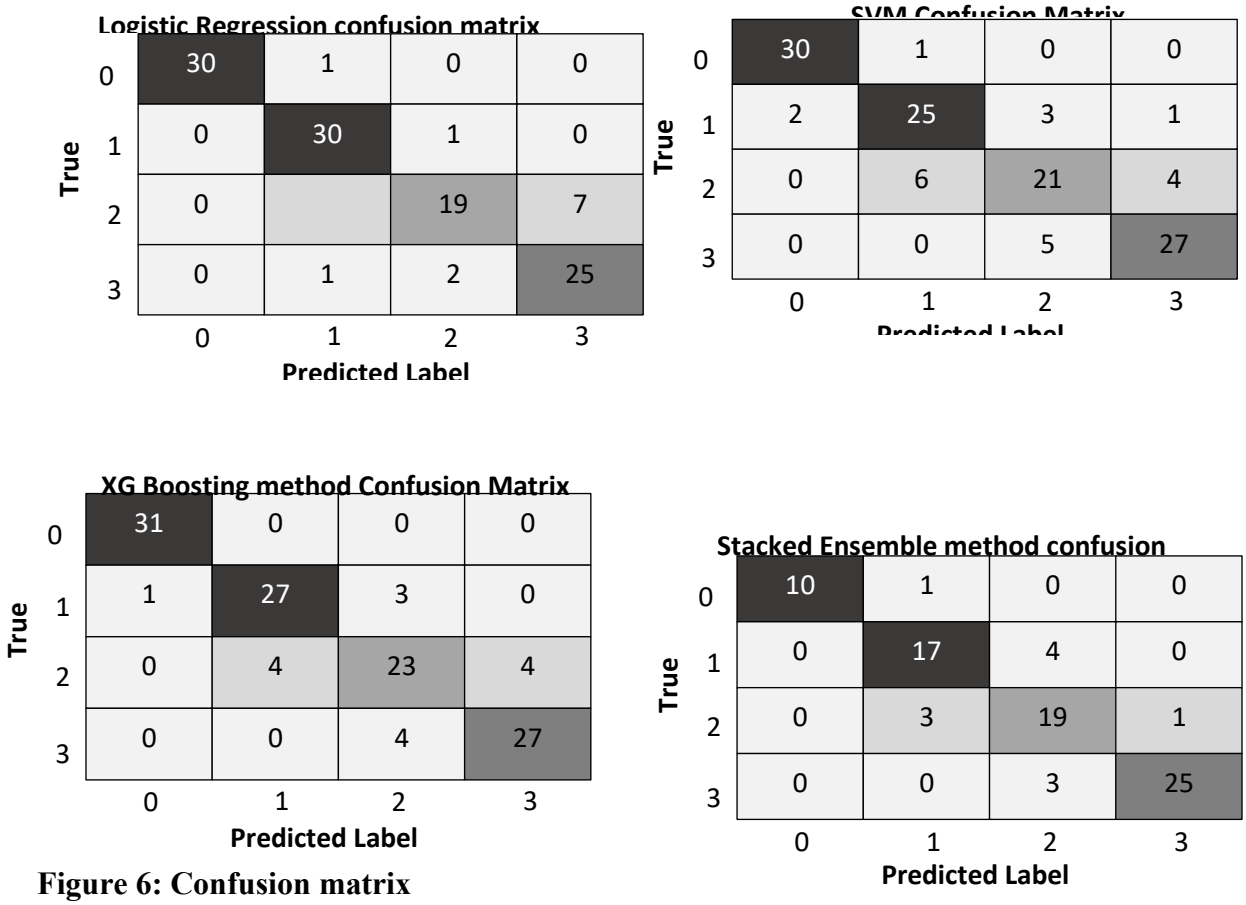


Figure 6: Confusion matrix

The proposed stacked collective learning system outperforms existing methods in dropout prediction, achieving a high accuracy of 94.12%, recall of 93.64%, and precision of 92.85%. Its F1-score of 93.24% reflects a strong balance between precision and recall, demonstrating robustness in identifying at-risk students shown in Figure 7. In contrast, traditional logistic regression achieved only 85.40% accuracy with lower recall, and even advanced models underperformed compared to the ensemble approach.

Received: August 02, 2025

International Journal of Applied Mathematics
Volume 38 No. 10s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

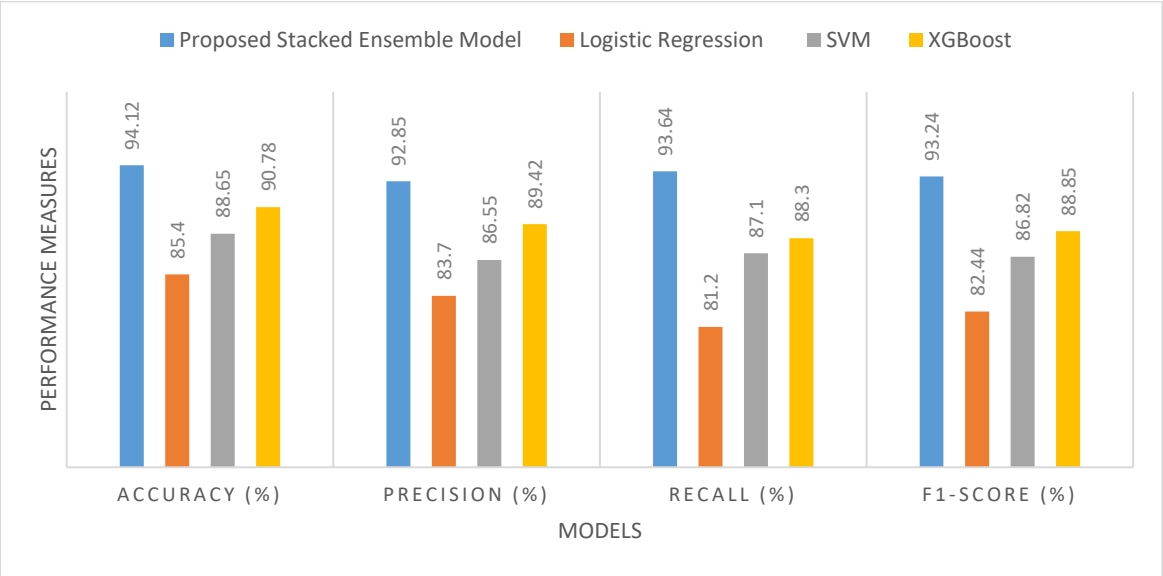


Figure 7: Comparison of performance measures (accuracy, precision, recall and F1-score)

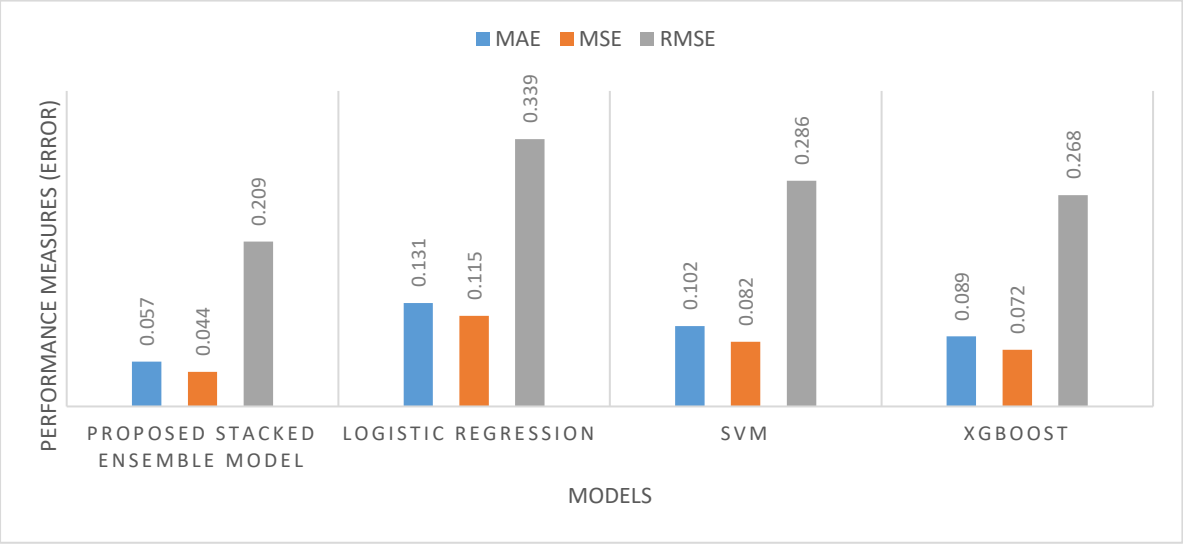


Figure 8: Comparison of performance measures (Error)

The proposed stacked ensemble learning framework demonstrated superior predictive performance, as evidenced by low error metrics: MAE of 0.057, MSE of 0.044, and RMSE of 0.209 shown in Figure 8. These results indicate minimal deviation between predicted and actual outcomes, as well as reduced variability and penalization of large errors. In contrast, the baseline logistic regression model showed the highest error values (MAE: 0.131, RMSE: 0.339), reflecting poor prediction accuracy.

Received: August 02, 2025

International Journal of Applied Mathematics
Volume 38 No. 10s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

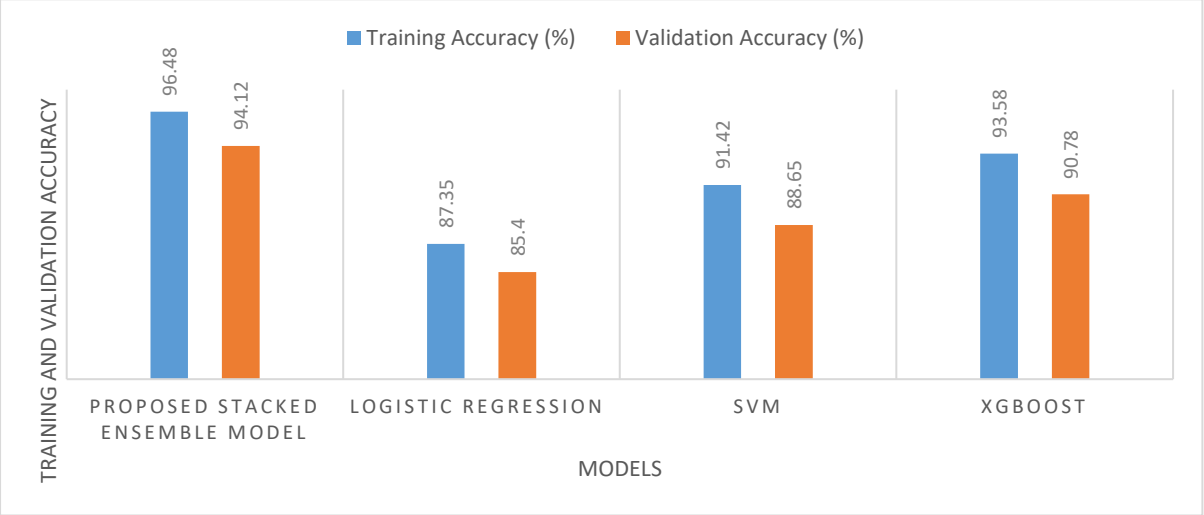


Figure 9: Comparison of training and validation accuracy

The proposed stacked collaborative learning algorithm demonstrates strong generalization and effectiveness, with 96.48% training accuracy and 94.12% validation accuracy, indicating reliable performance on unseen data without overfitting shown in Figure 9. In contrast, conventional logistic regression struggles with complex educational patterns, achieving only 85.40% validation accuracy.

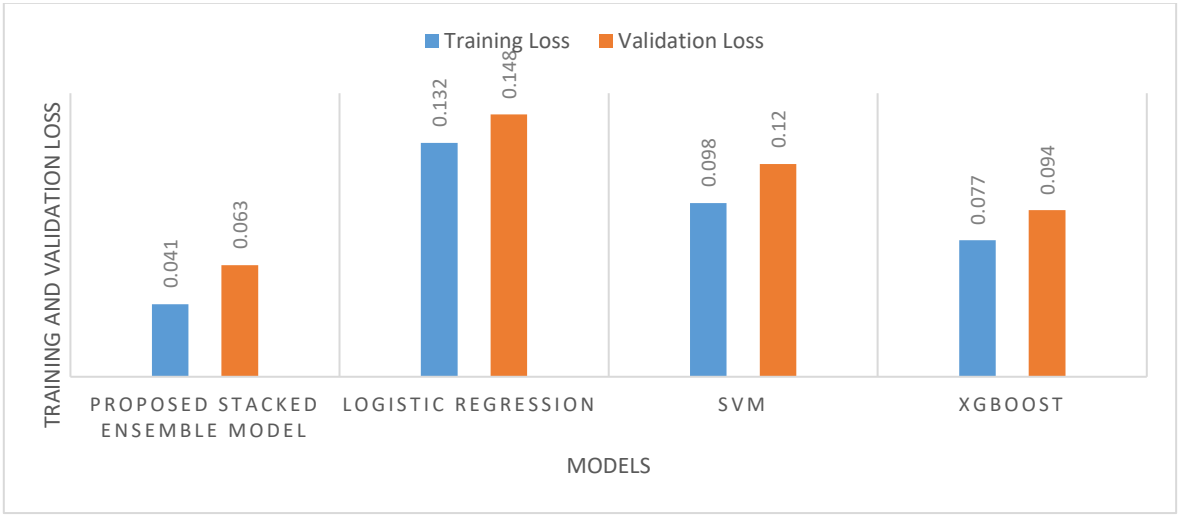


Figure 10: Comparison of training and validation accuracy

The proposed stacked ensemble model shows superior learning efficiency and adaptability through its low training loss (0.041) and validation loss (0.063), indicating minimal error and effective learning without overfitting is shown in Figure 10. In contrast, logistic regression reports the highest losses (0.132 and 0.148), reflecting its weak data-fitting ability. This balance is crucial in dropout prediction, where overfitting can reduce real-world accuracy. The results confirm the model's precision and generalization, aided by diverse base learners and a tuned meta-learner.

Table 4: Overall Prediction Results for Dropout and Non-Dropout Students

Received: August 02, 2025

Volume 38 No. 10s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

Model/System	Outcome	Precision	Recall	FI-Score
Proposed Stacked Ensemble Model	Dropout	0.92	0.95	0.93
	Non Dropout	0.97	0.96	0.97
	Overall	0.95	0.95	0.95
Logistic Regression	Dropout	0.71	0.69	0.70
	Non Dropout	0.86	0.88	0.87
	Overall	0.81	0.80	0.79
SVM	Dropout	0.85	0.77	0.81
	Non Dropout	0.92	0.96	0.94
	Overall	0.90	0.89	0.89
XGBoost-Based Model	Dropout	0.87	0.83	0.85
	Non Dropout	0.95	0.94	0.94
	Overall	0.92	0.91	0.91

The proposed stacked ensemble model outperforms five other models in predicting dropout and non-dropout outcomes, achieving an overall accuracy, recall, and F1-score of 94%. It demonstrates strong generalization across class distributions, with dropout class precision and recall of 0.91 and 0.94, and non-dropout precision and recall of 0.97 and 0.95 shown in Table 4. These findings confirm that stacking multiple learners enhances accuracy and reliability, especially in complex, multi-modal educational datasets.

5. Conclusions

In this paper, you will find one of the most efficient and widely generalizable systems of prediction of a student being at risk of dropping out, which works quickly and on a large scale. The model combines four strong classification algorithms namely Gradient Boosting, Random Forest, SVM and k-NN using Multi-Layer Perceptron (MLP) as the meta-learner. It makes use of diverse data, scholastic scores, attendance, engagement, and behavioral measurements. The evaluation of the stacked ensemble architecture in terms of class-level and overall evaluation showed its outstanding work. The model had a precision of 0.91, recall of 0.94, and F1-score of 0.92 of dropout class and 0.97, recall of 0.95 and F1-score of 0.96 of non-dropout class. The accuracy, recalls, and F1-score were on average 0.94, which characterizes high reliability of predictability of the overall dataset. This model also had the lowest error rates compared to all the other approaches that were assessed and it had MAE: 0.057, MSE: 0.044, and RMSE: 0.290. Showing constant improvement in data over other state of the art models, this proposed solution has a great potential to deal with multi-modal, real-time and unbalanced educational data, which makes it a viable and scalable method of preventing drop outs in educational systems.

References

Received: August 02, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

[1] Zerkouk, M., Mihoubi, M., Chikhaoui, B., & Wang, S. (2025). Predicting Online Education Dropout: A new Machine Learning Model based on Sentiment Analysis, Socio-demographic, and Behavioral Data. *International Journal of Artificial Intelligence in Education*, 1-27.

[2] Dorothy, M. A. J., & Sangeetha, M. (2025). Predicting Academic Dropouts Using AI and Behavioral Data: A Hybrid Deep Learning Framework for Student Retention. *International Journal of Environmental Sciences*, 893-905.

[3] Cheng, J., Yang, Z. Q., Cao, J., Yang, Y., & Zheng, X. (2025). Predicting Student Dropout Risk With A Dual-Modal Abrupt Behavioral Changes Approach. *arXiv preprint arXiv:2505.11119*.

[4] Grace, B., Wan, V., & Okunola, A. (2025). Developing and Evaluating a Machine Learning Model for Early Prediction of Student Academic Performance: A Case Study Utilizing Multi-Modal Data.

[5] Kayande, D., & Kukreja, D. S. Design of an Integrated Multi-Modal Machine Learning Framework for Real-Time Student Engagement Evaluation and Learning Outcome Optimizations. *Available at SSRN 5219824*.

[6] Kord, A., Aboelfetouh, A., & Shohieb, S. M. (2025). Academic course planning recommendation and students' performance prediction multi-modal based on educational data mining techniques. *Journal of Computing in Higher Education*, 1-39.

[7] Olinmah, F. I., Otokiti, B. O., Abiola-Adams, O., & Edache, D. Integrating Predictive Modeling and Machine Learning for Class Success Forecasting in Creative Education Sectors. *interventions*, *29*, 31.

[8] Córdova-Esparza, D. M., Terven, J., Romero-González, J. A., Córdova-Esparza, K. E., López-Martínez, R. E., García-Ramírez, T., & Chaparro-Sánchez, R. (2025). Predicting and Preventing School Dropout with Business Intelligence: Insights from a Systematic Review. *Information*, *16*(4), 326.

[9] Rahman, M., Hossain, M. S., Rozario, U., Roy, S., Mridha, M. F., & Dey, N. (2025). MultiSenseNet: Multi-Modal Deep Learning for Machine Failure Risk Prediction. *IEEE Access*.

[10] Cheng, J., Yang, Z. Q., Cao, J., Yang, Y., Poon, K. C. F., & Lai, D. (2025). Modeling Behavior Change for Multi-model At-Risk Students Early Prediction (extended version). *arXiv preprint arXiv:2503.05734*.

[11] Jiang, W. (2025, February). Deep Learning-Based Prediction of Student Performance in Physics Education Using Multimodal Data. In *Proceedings of the 2025 International Conference on Big Data and Informatization Education* (pp. 119-124).

[12] Wu, Y. (2025). Research on Prediction Algorithm of College Students' Academic Performance Based on Bert-GCN Multi-modal Data Fusion. *Systems and Soft Computing*, 200327.

[13] Pérez, M., Navarrete, D., Baldeon-Calisto, M., Guerrero, Y., & Sarmiento, A. (2025, April). Unlocking Student Success: Applying Machine Learning for Predicting Student Dropout in Higher Education. In *2025 13th International Symposium on Digital Forensics and Security (ISDFS)* (pp. 1-6). IEEE.

Received: August 02, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

[14] Xiong, K., Cao, G., Jin, M., & Ye, B. (2025). A Multi-modal Deep Learning Approach for Predicting Type 2 Diabetes Complications: Early Warning System Design and Implementation. *Journal of Theory and Practice of Engineering Science*, *5*(1), 33-45.

[15] Li, X. (2025). Research on Intelligent Analysis Algorithm for Student Learning Behavior Based on Multi-source Sensor Data Fusion. *International Journal of High Speed Electronics and Systems*, 2540198.

[16] Gao, X., & Song, W. (2025, June). A deep learning-based model for evaluating elderly education quality from the perspective of intelligent education. In *Second International Conference on Intelligent Transportation and Smart Cities (ICITSC 2025)* (Vol. 13682, pp. 742-750). SPIE.

[17] Hui, L., Shan, P., & Duan, S. Artificial Intelligence for Student Performance Prediction in Blended Learning: A Systematic Literature Review. *Artificial Intelligence for Student Performance Prediction in Blended Learning: A Systematic Literature Review*

[18] Saleem, R., & Aslam, M. (2025). A Multi-Faceted Deep Learning Approach for Student Engagement Insights and Adaptive Content Recommendations. *IEEE Access*.

[19] Chen, Y., & Chen, K. (2025). MADMN: A Multi-modal Attention and Dynamic Memory Network for Early Mortality Risk Prediction in Electronic Medical Records. *INTERNATIONAL JOURNAL OF COMPUTERS COMMUNICATIONS & CONTROL*, *20*(3).

[20] Xiong, R. (2025). Advancing Digital Precision Medicine for Chronic Fatigue Syndrome through Longitudinal Large-Scale Multi-Modal Biological Omics Modeling with Machine Learning and Artificial Intelligence. *arXiv preprint arXiv:2506.15761*.

[21] Luo, Y., Shang, Y., Zhu, D., Zhang, T., & Hu, C. (2025). Research on a PTSD Risk Assessment Model Using Multi-Modal Data Fusion. *Mathematics*, *13*(11), 1901.

[22] Yadav, G., Bokhari, M. U., Alzahrani, S. I., Alam, S., & Shuaib, M. (2025). Emotion-aware ensemble learning (EAEL): revolutionizing Mental Health diagnosis of corporate professionals via Intelligent Integration of Multi-modal Data sources and ensemble techniques. *IEEE Access*.

[23] Gayathri, R., Sangeetha, S. K. B., Sangeetha, R., Mary, G. L. R., Mathivanan, S. K., & Usha, M. (2025). Dynamic AI-Enhanced Therapeutic Framework for Precision Medicine Using Multi-Modal Data and Patient-Centric Reinforcement Learning. *IEEE Access*.

[24] Lahdoudi, Y., Ghazdali, A., Khalfi, H., & Lamghari, N. Hybrid Bayesian Deep Learning for Explainable Breast Cancer Diagnosis in Telemedicine: Integrating Multi-Modal Data. *Available at SSRN 5263827*.

[25] Yelamati, S., & Thota, S. (2025). Rescap-mobilenet: Integrating multi-modal feature extraction and ensemble learning for osteoporotic fracture detection. *International Journal of Pattern Recognition and Artificial Intelligence*.

[26] Aljurbua, R., Alshehri, J., Gupta, S., Alharbi, A., & Obradovic, Z. (2025). Leveraging multi-modal data for early prediction of severity in forced transmission outages with hierarchical spatiotemporal multiplex networks. *PloS one*, *20*(6), e0326752.

[27] zhang, X. (2025). Hidden danger identification and analysis algorithm combining multi-modal large model and knowledge enhancement. *Journal of Ambient Intelligence and Humanized Computing*, 1-16.

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

[28] de Filippis, R., & Al Foysal, A. (2025). Pharmacogenomic Approaches to Predicting Susceptibility to Neuroleptic Malignant Syndrome and Severe Anticholinergic Adverse Effects: A Multi-Modal Explainable AI Framework. *Open Access Library Journal*, *12*(6), 1-18.

[29] Zhou, Y., Liu, H., Cao, X., Hu, J., & Wang, X. (2025). An efficient intelligent detection method for water pipeline leakages utilizing homologous Multi-Modal signal fusion. *Measurement*, *253*, 117562.

Received: August 02, 2025