INDIAN JOURNAL OF SCIENCE AND TECHNOLOGY

RESEARCH ARTICLE

OPEN ACCESS

Received: 03-12-2024

Accepted: 23-12-2024

Published: 15-01-2025

Integrative Ensemble Learning Algorithm for Predicting Students' Performance

A Emima[1]*, D I George Amalarethinam[2]

1 Research Scholar, PG and Research Department of Computer Science, Jamal Mohamed College (Autonomous), Affiliated to Bharathidasan University, Tiruchirappalli, 620020, Tamil Nadu, India

2 Principal and Head, Associate Professor, PG and Research Department of Computer Science, Jamal Mohamed College (Autonomous), Affiliated to Bharathidasan University, Tiruchirappalli, 620020, Tamil Nadu, India

Citation: Emima A, Amalarethinam DIG (2025) Integrative Ensemble Learning Algorithm for Predicting Students' Performance. Indian Journal of Science and Technology 18(1): 72-84. https://doi.org/10.17485/IJST/v18i1.3718

* **Corresponding author**.

emimastaecey@gmail.com

Funding: None

Competing Interests: None

Published By Indian Society for Education and Environment (iSee)

ISSN
Print: 0974-6846
Electronic: 0974-5645

Abstract

Objectives: To create a stable student performance prediction model utilizing ensemble learning methods. **Methods:** The study uses boosting techniques such as CatBoost, Extreme Gradient Boosting (XGBoost), and Light Gradient Boosting Machine (LightGBM) as simple classifiers, which are then combined into a composite classifier to improve predictive accuracy. During the training phase, a 5-level hyperparameter optimization for the basic classifiers is performed using ETLBO Optimization IELA's distinguishing feature is its Stacking ensemble method, which functions as an ensemble technique, combining the expected outcomes of the simple classifiers to build the final prediction model ETLBO algorithm is applied for hyper parameter optimization to discover the best hyperparameters of 28 features from 33 features in the base classifiers to yield the better result in experiment. This approach improves prediction efficiency and accuracy by combining the capabilities of separate models using boosting algorithms and stacking-based ensemble techniques. **Findings:** Student achievement Dataset in secondary education for Mathematics are taken from Public repository is used in this research work. It comprises features such as student grades, demographic information, social aspects, and school-related data. The working of boosting and stacking approaches with the classifier in the dataset consisting of LightGBM and CatGB are performed higher than those of either classifier when used separately without the ensemble technique. Additionally, CatGB operates substantially better when combined with XGBoost. With an accuracy of 86.43% and an F-score of 84.98%, the composite classifier that combines the three simple classifiers achieves the highest gain. **Novelty:** This research is unusual to combine the boosting and stacking ensemble techniques, which improves prediction accuracy and efficiency when compared to previous models for predicting student performance.

Keywords: Educational Data Mining; Ensemble Algorithm; Boosting Algorithm; Machine Learning Algorithm; Base Classifier; Composite Classifier

1 Introduction

Predicting student performance has received a lot of attention in recent years because of its ability to improve educational outcomes and enable early intervention for at-risk kids[1]. However, constructing a strong and accurate prediction model remains difficult due to the complexity of factors influencing student accomplishment, such as individual learning styles, external socioeconomic impacts, and diverse academic backgrounds. Traditional machine learning techniques, while beneficial in some circumstances, frequently exhibit low generalizability and prediction accuracy when applied to other datasets or educational contexts.To overcome these issues, ensemble learning approaches have developed viable tools for integrating the strengths of many models to improve prediction performance. These strategies, including boosting and stacking, can provide a more robust prediction model by utilizing a variety of individual classifiers. This research focuses on the application of advanced ensemble methods to create a more accurate and dependable student performance prediction model. Ensemble algorithms improve prediction accuracy by combining the results of numerous basic classifiers. While this study predominantly uses the UCI Student Performance dataset from secondary school students in Portugal, we accept that our model's generalizability is limited due to the dataset's geographical and domain specificity. As a result, we intend to extend our model in future work to include datasets from various countries, regions, and educational environments. This will help to evaluate the ensemble model's adaptability and usefulness in forecasting student performance across a variety of situations, such as subjects, age groups, and school types. Furthermore, we acknowledge the importance of conducting a full quantitative comparison with modern models, such as neural networks and deep learning approaches, in order to benchmark our ensemble model. Although past research has evaluated ensemble approaches qualitatively, a quantitative examination of accuracy, F-score, and other pertinent metrics will help us grasp our model's relative merit[2]. Mahmoud Ragab et. al[3] used different datasets that show how students engage with the educational system. Numerous classifiers including decision trees, random forests, support vector systems, k-nearest neighbor, naive Bayes trees, logistic regression, and artificial neural networks are used to evaluate the predictive performance. To improve the performance of the classifiers, well-known procedures like Boosting, Random Forest, Bagging, and Voting Algorithms are also incorporated. Mohammad Noor Injadat et.al.[4] Conducted an extensive investigation and examination of two different datasets at two discrete points in the course delivery process (20% and 50% completion). They provided insights on feature selection which helps to choose suitable machine learning algorithms and the parameters. It implemented techniques for methodically selecting an ensemble learner from a set of six possible machine learning algorithms by utilizing the p-value and the Gini index. However, experiments were not conducted in unbalanced dataset.

Mohammad Noor Injadat et.al.[5] used two different undergraduate datasets from different universities in order to predict students' success at two different phases in the course delivery process (20% and 50% completion). Furthermore, they used a systematic multi-split technique, as well as the Gini index and p-value, to optimize a bagging ensemble learner built from various combinations of six alternative basis machine learning algorithms. However, they did not specify any method to optimize the topology of the neural network. Singh et.al.[6] proposed an ensemble algorithm on 1000 instances and 22 attributes that are used to evaluate student performance and used machine learning algorithms namely Decision Tree (DT), Nave Bayesian (NB), K-Nearest Neighbors (KNN), and Extra Tree (ET). A model is built to combine the results of each individual base learner, employing both Bagging and Boosting ensemble approaches. To discover and select the most successful model, the outcomes of bagging and boosting ensemble strategies are compared. However, they did not use any boosting machine learning algorithms to enhance the performance. Elaf et.al. [7] proposed a students' performance prediction model using classifiers namely Artificial Neural Network, Naïve Baysean and Decision Tree. Ensemble techniques such as Bagging, boosting and Random search were used to enhance the effectiveness of the classifiers. However, the model is evaluated only with 25 new students.

Safira Begum et.al. [8] used ensemble classifiers to develop a predictive model for student outcomes. The first step entailed data preprocessing, followed by an analysis of the correlation between entrance attributes to detect potential redundancies caused by too large positive correlations. The refined attributes are trained and tested using classifiers such as Boosting, Bagging, and Random Subspace. A genetic algorithm is used to optimize the three classifiers to improve the predictive model's accuracy. It is intended for problem-solving optimization to identify the best option from a given collection of possibilities to meet the desired goal. However, the accuracy of the result is a little less when compared to the other model. Amit Malik et.al. [9] proposed a machine learning approach for predicting student adaptability in online entrepreneurship education. The Kaggle Educational dataset was used to assess several algorithms such as Random Forest, C5.0, CART, and Artificial Neural Network. These algorithms performed well, to prove the effectiveness of the machine learning techniques in predicting students' receptivity to entrepreneurial online training. However, the predictor variables have been assigned to a limited level with G-Index.

This research paper proposes a unique Integrative Ensemble Learning Algorithm (IELA) which uses Extreme Gradient Boosting (XGBoost), Catboost (CB) and Light Gradient Boosting Machine (LightGBM) as simple classifiers under the Composite Classifier ensemble design. The specific strengths of each algorithm drove the choices of Catboost (CB), XGBoost, and LightGBM as simple learners. CatBoost is used as one of the simple classifiers because of its ability to handle categorical characteristics, large datasets, and assisting in increasing the efficiency of models. LightGBM is able to handle categorical parameters, rapid processing speed, efficient processing of large quantities of data, and high accuracy rates [10]. XGBoost helps by minimizing overfitting and managing missing data properly. Furthermore, all three algorithms, including GB, have a proven track record of providing very accurate findings. ETLBO algorithm is applied for hyperparameter optimization in the basic classifiers to yield the better result in experiment.

2 Methodology

Educational Data Mining (EDM) predicts student performance by using machine learning approaches. Boosting model is a progressive build procedure that focuses on reducing errors from earlier models while increasing the influence of outstanding performance [11]. The proposed research made use of the UCI Machine Learning Repository's Student Performance Dataset. This dataset includes demographics, academic, and behavioral information from Portuguese secondary schools, making it ideal for forecasting student outcomes. The Features included such as study time, education of parents, and socioeconomic status. To prepare the dataset for machine learning techniques, it was preprocessed using data cleaning, normalization, and feature selection. Boosting models including Light Gradient Boosting Machine (LightGBM), CatBoost, Extreme Gradient Boosting (XGBoost), and the Integrative Ensemble Learning Algorithm (IELA) were used. The accuracy and F-score of the model were used to assess predictive reliability and the balance of precision and recall. Although the dataset yielded useful insights, its regional emphasis and static nature limited generalizability. Cross-validation approaches were used to address these concerns and improve model resilience. This study demonstrates the effectiveness of boosting algorithms for properly predicting student achievement in educational data mining.

2.1 XGBoost

eXtreme Gradient Boosting is known as XGBoost. It is an example of a Gradient Boosting Machine (GBM), a method mostly used to create predictive modeling issues for both regression and classification [12]. The majority of GBM models that are now in use have demonstrated superior performance on several machine learning reference data sets, hence it outperforms other machine learning techniques [13]. With the use of fresh models made to adjust the residuals or errors of earlier models, which are then integrated to provide the final forecast, XGBoost is an ensemble method. Comparatively speaking, XGBoost performs better in experiments than a lot of other ensemble classifiers (like AdaBoost) [14]. The XGBoost algorithm has gained widespread recognition for its influence in numerous machine learning and data mining difficulties, where it is becoming more frequently used.

Objective Function=∑i=1nl(y_i,y^i(t))+∑i=1tΩ(f_i)

- The quantity of training instances is n.
- t denotes the quantity of boosting rounds, or iterations.
- y_i is the i-th instance of true label.
- y^i(t)is the predicted label for the i-th instance at iteration t.

- l is the loss function which measures the difference between the predicted label and the true label.
- $\Omega(f_i)$ is the regularization term of the individual trees.

2.2 LightGBM

LightGBM uses boosting techniques and outperforms XGBoost in terms of training time and scalability for larger datasets [15]. Although their accuracies are equivalent, LightGBM is optimized for parallel learning, which provides increased efficiency with large datasets. Notably, it employs a leaf-wise growth method, focusing on leaves with the greatest delta loss, resulting in higher accuracy when compared to level-wise algorithms. However, with smaller datasets, vigilance is required to avoid overfitting, which can be addressed by adjusting the max depth option to regulate the split depth.

Objective Function=∑i=1nloss(yi,y^i)+∑k=1KΩ(fk)

Where n is meant for the number of samples, y^i is the i^{th} sample predicted output, K is denoted as number of leaves and Ω(fk) is represented as regularization term for the leaf.

2.3 CatBoost

The CatBoost classifier is an extremely efficient machine learning method created exclusively to predict categorical features. It has automatic regularization, and can easily manage missing data. It requires little preprocessing of categorical data and produces quick and accurate predictions [16].

Objective Function =∑i=1nloss(yi,y^i)+∑k=1KΩ(fk)

Where, n is the number of samples,

loss is the loss function, that can depend on the particular task

y^iis $i\text{-}^{th}$ sample for the predicted output.

K Denotes the total number of leaves in the tree.

Ω(fk) is the leafs regularization term.

2.4 Stacking Ensemble

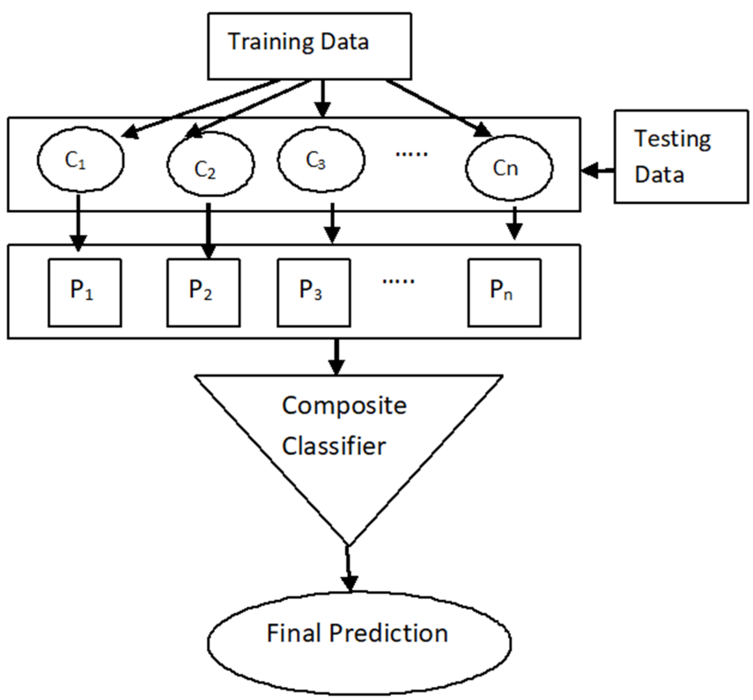

Fig 1. Stacking ensemble workflow of training data with multiple base classifiers and a composite classifier for final prediction

Stacking an ensemble algorithm involves training simple classifiers. By employing cross-validation on the results of the basis classifiers, the composite-learner will be trained. The Stacking ensemble approach employs a two-step learning procedure [17]. In Figure 1 the first phase, a collection of simple-level classifiers C1, C2,..., Cn is formed from an independent and evenly distributed collection of samples, Si = (X_i, Y_i), where X_i is the features and Y_i is class values. The first level data is used to generate a meta-dataset with P_i predicted value of the n basic classifiers. In second phase, the composite dataset is constructed, and a Composite classifier is getting trained with its data and then utilized for prediction.

2.5 Integrative Ensemble Learning Algorithm (IELA)

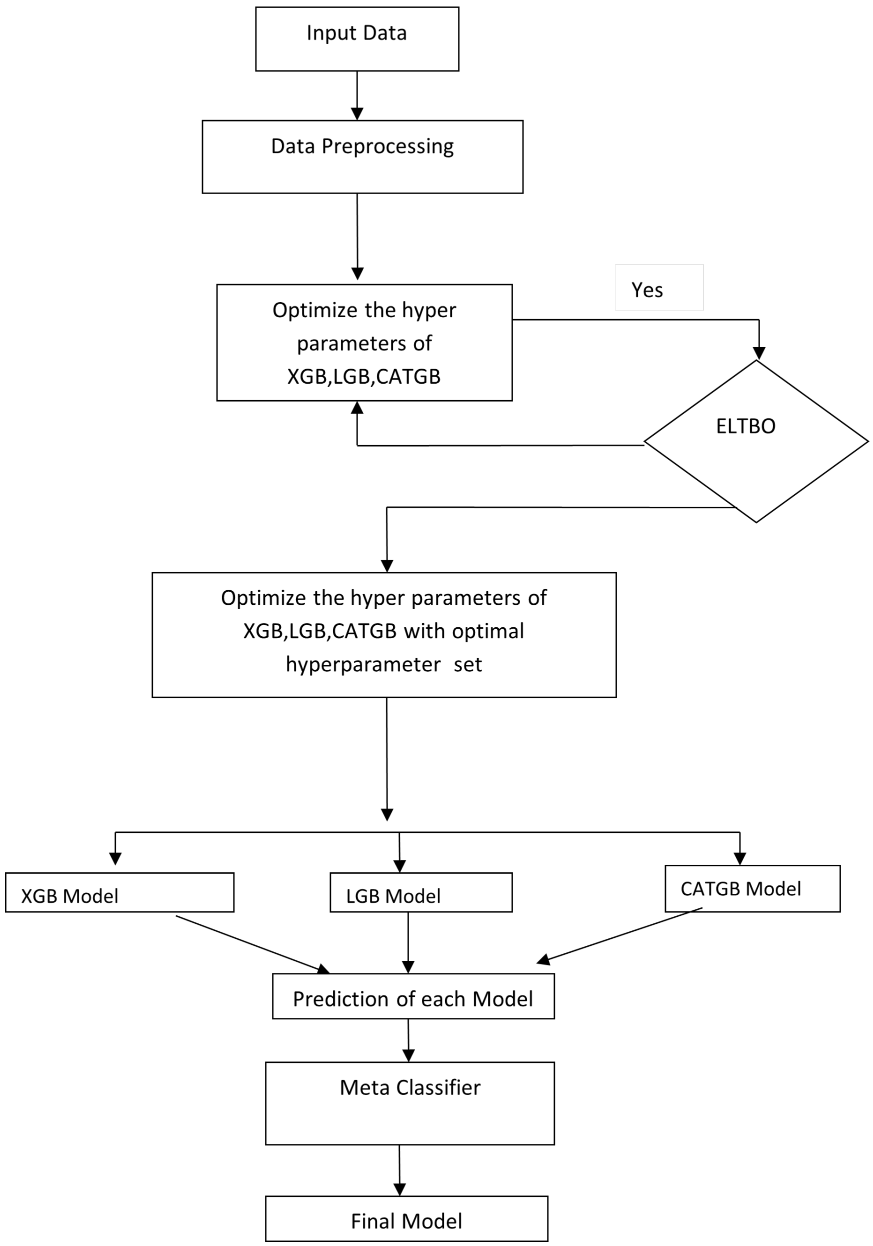

Fig 2. Integrative Ensemble Learning Algorithm (IELA)-Workflow of hyperparameter optimization and ensemble modeling with XGB, LGB, and CATGB, finalized by a meta-classifier

Table 1. Dataset Features

S.no	Feature	Description
1	school	Student's school (Binary: 'GP' - Gabriel Pereira or 'MS' - Mousinho da Silveira)
2	sex	Student's sex (Binary: 'F' - Female or 'M' - Male)
3	age	Student's age (numeric: from 15 to 22)
4	address	Student's home address type (Binary: 'U' - urban or 'R' - rural)
5	famsize	Family size (Binary: 'LE3' - less or equal to 3 or 'GT3' - greater than 3)
6	Pstatus	Parent's cohabitation status (Binary: 'T' - living together or 'A' - apart)
7	Medu	Mother's education (numeric: 0 - none, 1 - primary education (4th grade), 2 - 5th to 9th grade, 3 - secondary education or 4 - higher education)
8	Fedu	Father's education (numeric: 0 - none, 1 - primary education (4th grade), 2 - 5th to 9th grade, 3 - secondary education or 4 - higher education)
9	Mjob	Mother's job (nominal: 'teacher', 'health care related', 'civil services' (e.g. administrative or police), 'at_home' or 'other')
10	Fjob	Father's job (nominal: 'teacher', 'health care related', 'civil services' (e.g. administrative or police), 'at_home' or 'other')
11	reason	Reason to choose this school (nominal: 'close to home', 'school reputation', 'course preference' or 'other')
12	guardian	Student's guardian (nominal: 'mother', 'father' or 'other')
13	traveltime	Home to school travel time (numeric: 1 - <15 min., 2 - 15 to 30 min., 3 - 30 min. to 1 hour, 4 - >1 hour)
14	studytime	Weekly study time (numeric: 1 - <2 hours, 2 - 2 to 5 hours, 3 - 5 to 10 hours, 4 - >10 hours)
15	failures	Number of past class failures (numeric: if 1 <= n <= 3, else 4)
16	schoolsup	Extra educational support (binary: yes or no)
17	famsup	Family educational support (binary: yes or no)
18	paid	Extra paid classes within the course subject (Math or Portuguese) (binary: yes or no)
19	activities	Extra-curricular activities (binary: yes or no)
20	nursery	Attended nursery school (binary: yes or no)
21	higher	Wants to take higher education (binary: yes or no)
22	internet	Internet access at home (binary: yes or no)
23	romantic	With a romantic relationship (binary: yes or no)
24	famrel	Quality of family relationships (numeric: from 1 - very bad to 5 - excellent)
25	freetime	Free time after school (numeric: from 1 - very low to 5 - very high)
26	goout	Going out with friends (numeric: from 1 - very low to 5 - very high)
27	Dalc	Workday alcohol consumption (numeric: from 1 - very low to 5 - very high)
28	Walc	Weekend alcohol consumption (numeric: from 1 - very low to 5 - very high)
29	health	Current health status (numeric: from 1 - very bad to 5 - very good)
30	absences	Number of school absences (numeric: from 0 to 93)
31	G1	First period grade (numeric: from 0 to 20)
32	G2	Second period grade (numeric: from 0 to 20)
33	G3	Final grade (numeric: from 0 to 20, output target)

A newly developed Integrative Ensemble Learning Algorithm (IELA) is proposed in this research to predict academic achievement among students. Dataset of Secondary school students' performance are taken to predict students' performance using 5-level classification techniques. Figure 2, The proposed method is made up of two types of classifiers: Simple classifiers and Composite classifiers. XGBoost, LightGBM and CatBoost are used as Simple classifier. In the Composite classifier, the base algorithm that performs best with default hyper parameters are used in it. Then, the Stacking ensemble learning algorithm is used to increase the performance of the simple classifiers. This study makes use of a students performance dataset acquired during the 2005-2006 academic year, involving secondary school students' academic progress in math classes from the UCI repository. The dataset contains 37 questions in total. It includes demographic, social/emotional, and school-related variables specified in Table 1. The dataset was chosen for its applicability to student accomplishment analysis. The constraints include possible errors in self-reported data and the dataset's specialized setting, which may impact generalizability.

Data pre-processing is an important step in preparing a dataset for classification methods, as it significantly impacts the final results. In the initial phase, pre-processing is performed on the student dataset. During this stage, nominal values of the dataset parameters are converted into numerical values. The Label Encoder class from the Preprocessing sub-library of the Scikit-learn library is used. The Label Encoder assigns a numerical value to each categorical and nominal data entry. TLBO is a population-based strategy that uses a population of solutions to arrive at a global solution. The population is called a group or class of learners. The TLBO method is separated into two phases: the 'Teacher Phase' and the 'Learner Phase'. The 'Teacher Phase' refers to learning by the teacher, whereas the 'Learner Phase' refers to learning through interaction among learners. The enhanced ETLBO seeks to determine the most suitable feature subset by utilizing the Euclidean distance formula in the evaluation of the fitness value and common control parameters. Notably, no missing values are identified in the dataset. For 5-level classification, all attributes in the dataset are utilized. Following pre-processing, the dataset is partitioned into training and test sets, with 70% of the data allocated for training and 30% for testing. The training data is employed for training each of the base classifiers, while the test data is used to assess the model's performance.

Optimizing simple classifier hyper-parameters: Selecting the hyper-parameters is necessary prior to training a model. The proper values of the hyper-parameters need to be adjusted in order to produce an effective model. The 5-level classification tasks are carried out using the XGBoost, LightGBM and CatBoostmodels after the best hyperparameters for the simple classifiers have been determined. Each basic classifier is trained using the training data, and predictions are created using these models. The composite-classifier is created by identifying the best-performing method on the datasets from XGBoost, LightGBM and CatBoost. The composite-classifier is optimized using k-fold cross-validation on predictions derived from these simple classifiers, with the dataset partitioned into 10 subsets. Iteratively optimizing hyperparameters and evaluating model performance. The Stacking ensemble learning algorithm's composite-classifier incorporates predictions from XGBoost, LightGBM and CatBoost. resulting in an effective hybrid methodology. After completing the specified approaches, the effectiveness of IELA is evaluated using 30% of the test data in 5-level classification tasks. Recognizing the constraints of accuracy in data sets with imbalances, the evaluation includes the f-score as a measure alongside accuracy. IELA's performance is evaluated using accuracy and f-score.

Figure 3 depicts Spearman's Rank Correlation Heatmap which is used to visualize the relationship between various numerical variables. Here is a breakdown of the main elements in heatmap: Color coding: The heatmap uses colors to indicate the strength and direction of the correlation. Dark red denotes a strong positive correlation, dark blue indicates a strong negative correlation, and white indicates no correlation. Each cell in the heatmap represents a correlation coefficient between two variables. For example, there is a strong positive correlation (red square) between G2 and G1, indicating that students' grades in these two periods are very similar. The diagonal from the top left to bottom right shows a perfect positive correlation of 1.00 between each variable and itself. This heatmap is an effective tool for determining which factors are more strongly linked to student grades, and it can be especially useful for educators and researchers seeking to improve academic outcomes. Table 2 and Table 3 indicate the features that have a positive effect on the final grade (G3). The higher the correlation coefficient, the more significant the relationship with the final grade. And features have a negative impact on the overall grade (G3). The more negative the correlation coefficient, the more pronounced the inverse relationship with the final grade.

Table 2. Features with Negative Correlation

Index	Feature	Correlation_with_G3
0	school	0.064571
2	age	0.173438
5	Pstatus	0.044262
11	guardian	0.075010
12	traveltime	0.120530
14	failures	0.361224
15	schoolsup	0.150598
16	famsup	0.050482
22	romantic	0.091460
24	freetime	0.004994
25	goout	0.166119
26	Dalc	0.120944
27	Walc	0.104459
28	health	0.047790

Table 3. Features with Positive Correlation

Index	Feature	Correlation_with_G3
1	sex	0.103151
3	address	0.119457
4	famsize	0.066279
6	Medu	0.225036
7	Fedu	0.170049
8	Mjob	0.090138
9	Fjob	0.047784
10	reason	0.108885
13	studytime	0.105170
17	paid	0.060924
18	activities	0.026090
19	nursery	0.031752
20	higher	0.172418
21	internet	0.107816
23	famrel	0.054977
29	absences	0.017731
30	G1	0.878001
31	G2	0.957125

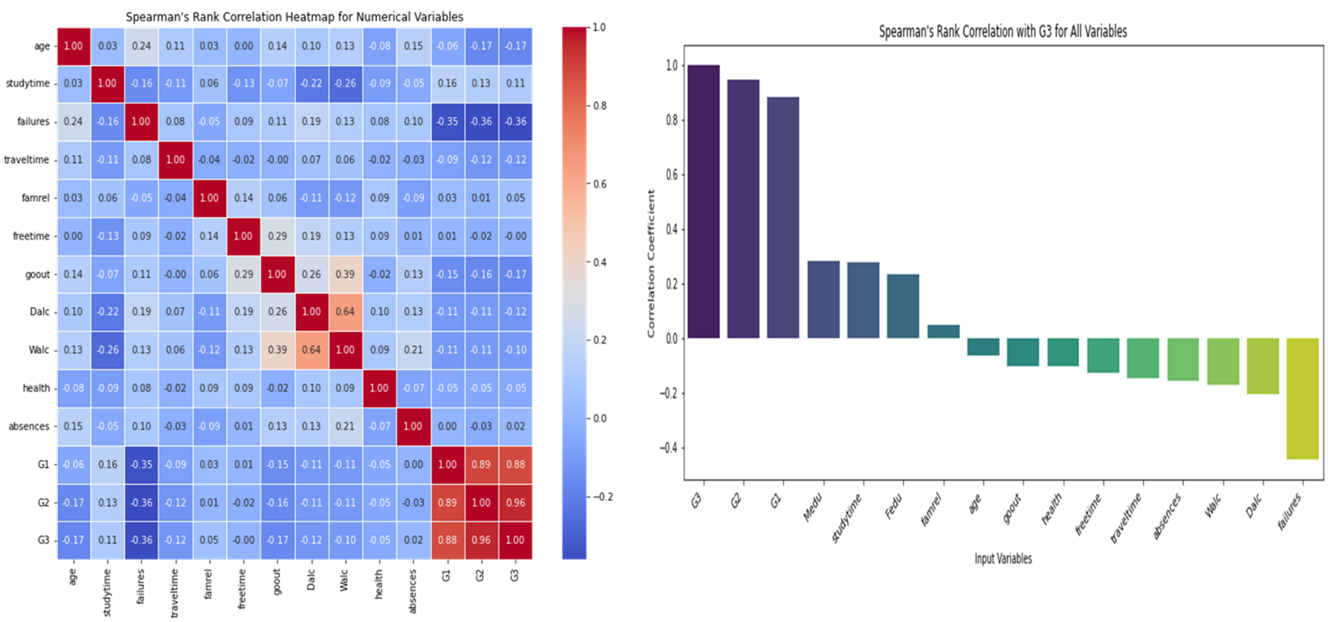

Fig 3. Spearman's rank correlation heatmap for numerical variables and bar chart of correlation coefficients with the target variable (G3)

3 Results and Discussions

This study uses classification techniques, namely XGBoost, LightGBM and CatBoost algorithms, to predict secondary school students' academic achievement. Using LabelEncoder, categorical variables are first converted to numerical values as part of the preprocessing step. Wrapper based Feature selection using Enhanced Teaching Learning Optimization algorithm (WFS-ELTBO) WFS-ETLBO is used to select a feature subset of 28 from 33 attributes to build a better classification model. Subsequently, 70% of the dataset is used for training the model, while 30% is used for testing and assessment. simple classifiers (XGBoost, LightGBM and CatBoost) are built using default parameters during training.

Table 4 compares three machine learning models—XGBoost, LightGBM and catboost —on key parameters such as accuracy and F-Score. XGBoost has the accuracy and F-Score (74.58% and 76.53%, respectively). LightGBM has a little lower accuracy of 71.88% and an F-Score of 72.53%. Cat Boosting has the highest accuracy of 79.00% and an F-Score of 78.01%. These findings provide useful information for picking the best model based on a balance of prediction performance and computational efficiency.

Table 4. Level classification with default hyper-parameter

	Accuracy	F-Score
XGBoost	74.58	76.53
LightGBM	71.88	72.25
CatBoost	79.00	78.01%

Table 5 represents the hyperparameter search areas for three distinct gradient boosting algorithms: XGBoost (xgb), LightGBM (Lightgb), and CatBoost. Each algorithm has unique hyperparameters that must be tweaked for peak performance. These hyperparameters include the learning rate, maximum depth, minimum child weight, number of estimators, tree-specific colsamples, number of leaves, subsamples, and leaf regularization. The ranges supplied for each hyperparameter reflect the values that will be tested during the hyperparameter tuning procedure. For example, learning rates can range from 0.01 to 0.5, maximum depths from 3 to 11, and other parameters can vary. This systematic research seeks to determine the best combination of hyperparameters for each algorithm, hence improving their predictive capabilities for the given datasets.

Table 5. Default hyper-parameter for simple classifier algorithms

Xgb	Lightgb	catboost
Max Depth: [3, 5, 7, 9, 11]	**Max Depth:** [3, 5, 7, 9, 11]	**Max Depth:** [3, 5, 7, 9, 11]
min_child_weight [1, 5, 10, 20, 50]	**min_child_samples:** [5, 10, 20, 50, 100]	**12 leaf reg:** [50, 100, 150, 200]
colsample_bytree [0.5, 0.6, 0.7, 0.8, 0.9, 1.0]	**num_leaves** [20, 30, 50, 100, 200]	**Subsamples (n_estimators):** [0.5, 0.6, 0.7, 0.8, 0.9, 1.0]

The design and behavior of the XGBoost, LightGBM and CatBoost models are determined by the hyperparameters given in Table 6. XGBoost is configured with a max_depth of 10, 500 boosting steps (n_estimators), and a gamma value of 0.5 in the given hyperparameter setups. This determines the minimal loss reduction for node splitting. LightGBM uses a high gamma of 150, which indicates a strict necessity for loss reduction in node partitioning, yet it still has the maximum depth of 10 and 10 boosting stages. Furthermore, LightGBM determines the step size during training by using a learning rate of 0.1. In contrast, CatBoost uses a moderate learning rate of 0.1 and sets a max_depth of 10 and 100 boosting stages (n_estimators), similar to XGBoost. Interestingly, CatBoost adds unique features not found in any of the other boosting algorithms: a l2_leaf_reg of 1.0, indicating regularization of L2 on leaf weights. The above configurations demonstrate the decisions made about algorithm specific hyperparameter tuning to maximize model performance for LightGBM, CatBoost, and XGBoost.

Table 6. Optimal hyperparameters for 5-level classification

	Hyper-parameter	Data Set
XGBoost	max_depth	10
	n_estimators	500
	gamma	0.5
LightGBM	max_depth	10
	n_estimators	10
	gamma	150
	Learning rate	0.1
CatBoost	max_depth	10
	n_estimators	100
	learning_rate	0.1
	l2_leaf_reg	1.0

After tuning the hyperparameters of the XGBoost, LightGBM and CatBoost algorithms, the generated models are used in the Super Learning (SL) method of the proposed Integrated Ensemble Learning Approach (IELA). The meta-classifier is the model that performs best with the default hyperparameters on both datasets. Specifically, XGBoost and LightGBM are used as meta-classifiers in the math dataset's 2- and 5-level classifications. In the Portuguese dataset, LightGBM and XGBoost are chosen as meta-classifiers in the 5-level classification processes. During the Stacking ensemble process, several combinations of GB, XGBoost, and LightGBM are tested to find the best ensemble for improved predictive performance.

Table 7 compares the performance metrics of multiple ensemble models made up of distinct base learners and an integrated meta-classifier (XGBoost). Thesimple classifiers are Gradient Boosting (GB), XGBoost, and LightGBM. The accuracy, F-score of each ensemble configuration are presented. It shows the performance of both individual classifiers and composite classifiers, which are created by merging different simple classifiers. Across all combinations, composite classifiers appear to have higher accuracy and F-scores than their individual simple classifiers. Notably, combining multiple base classifiers produces the best results, with XGBoost, LightGBM, and CatGB achieving the highest accuracy of 79.93% and F-score of 79.24%. This implies that combining the characteristics of numerous classifiers using ensemble methods can significantly improve predicted performance in classification problems.

Table 7. The performance of IELA for 5-level classification withLightGBM Composite Classifier

Simple Classifier	Composite Classifier	Accuracy	F-Score
XGBoost	LightGBM	77.38 %	78.32%
LightGBM	LightGBM	78.82%	77.02%
CatGB	LightGBM	78.64%	78.92%
LightGBM+ CatGB	LightGBM	78.39%	77.68%
XGBoost + LightGBM	LightGBM	79.82%	78.87%
XGBoost + CatGB	LightGBM	79.46%	78.78%
XGBoost+LightGBM+CatGB	LightGBM	79.93%	79.24%

Improved accuracy and F-score are achieved by combining various simple classifiers into each composite classifier shown in Table 8. XGBoost and LightGBM, when applied independently as simple classifiers, show competitive accuracy and F-scores, with LightGBM somewhat beating XGBoost on these measures. However, when compared to the other two simple classifiers, CatGB gives somewhat lower accuracy and F-score. Performance is noticeably improved, though, when CatGB is combined with other classifiers, especially in the composite classifiers. The combined classifier made up of LightGBM and CatGB performs better than either of them used alone in terms of accuracy and F-score. Additionally, CatGB performs significantly better when paired with XGBoost which shows that multiple simple classifier ensemble approaches, especially CatGB in conjunction with XGBoost and LightGBM, can greatly improve classification task predicting performance.

Table 8. The performance of IELA for 5-level classification with CatGB Composite Classifier

Simple Classifier	Composite Classifier	Accuracy	F-Score
XGBoost	CatGB	79.95	77.77
LightGBM	CatGB	80.46	78.21
CatGB	CatGB	78.54	76.65
LightGBM+ CatGB	CatGB	81.24	80.45
XGBoost + LightGBM	CatGB	83.52	82.51
XGBoost + CatGB	CatGB	84.38	82.43
XGBoost+LightGBM+CatGB	CatGB	86.43	84.98

The comparison of HELA [11] with IELA reveals considerable improvements in performance and approach, as evidenced by relevant literature. In Table 9, HELA yields 78.46% accuracy and 77.58% F-score utilizing XGBoost and LightGBM, which is consistent with research suggesting simpler ensemble approaches for multi-level classification. However, IELA beats HELA with 86.43% accuracy and 84.98% F-score, as evidenced by study highlighting CatGB's usefulness in complicated datasets only in Mathematics set of students. IELA outperforms others in terms of adaptability and scalability when dealing with hierarchical categorization issues. These findings verify IELA's advanced composite strategy, showing it as a strong framework that addresses HELA's weaknesses and is consistent with data that support varied ensemble methods.

Table 9. Comparison of HELA and IELA

Aspect	HELA [17]	IELA- Proposed Work
Maximum Accuracy	78.46% (XGBoost + LightGBM)	86.43% (XGBoost + LightGBM + CatGB)
Maximum F-Score	77.58%	84.98%
Best Combination	XGBoost + LightGBM	XGBoost + LightGBM + CatGB
Performance Gain	Moderate	Significant

The modified confusion matrix Figure 4 depicts the distribution of cases across five classes (I through V) and compares observed and expected class labels. Each matrix cell reflects a proportion of instances for a certain combination of observed and expected categories. Additional metrics, including precision, accuracy, and recall, provide a complete picture of the model's performance in each class. The matrix provides a quick overview of the model's prediction accuracy and capacity to accurately identify instances under various class labels. The research indicates various performance characteristics for each class of the model. While Class I shows a high precision rate of 83.75%, it also has a low recall rate, suggesting that while most predictions are true, some are missed. Class II, on the other hand, shows balanced recall and precision (83.75% and 83.97%, respectively), indicating dependable performance. Class III exhibits low precision but strong recall, indicating a propensity for false positives. Precision and recall show comparable balanced patterns in Classes IV and V. This detailed analysis offers important information about how well the model performs in distinct classes, which is crucial for a number of applications. Finding the ideal ratio between recall and precision is still crucial to a successful model's deployment.

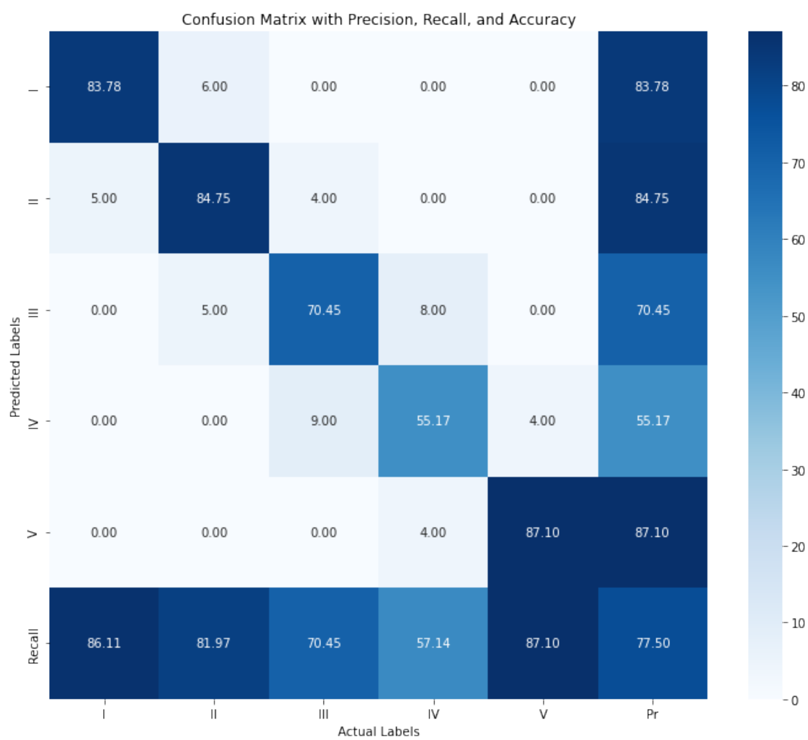

Fig 4. Confusion matrix with precision, recall, and accuracy for multi-class classification model predictions

4 Conclusion

Educational Data Mining (EDM) is emerging as an important tool for promoting education in society. This study presents the Integrative Ensemble Learning Algorithm (IELA), a unique framework for forecasting secondary school students' academic achievement in math classes. The methodology starts with preprocessing the educational dataset, and then uses advanced data mining techniques including Gradient Boosting (GB), XGBoost, and LightGBM as base learners. While this study predominantly uses the UCI Student Performance dataset from secondary school students in Portugal, we accept that our model's generalizability is limited due to the dataset's geographical and domain specificity. As a result, we intend to extend our model in future work to include datasets from various countries, regions, and educational environments. This will help to evaluate the ensemble model's adaptability and usefulness in forecasting student performance across a variety of situations, such as subjects, age groups, and school types. Furthermore, we acknowledge the importance of conducting a full quantitative comparison with modern models, such as neural networks and deep learning approaches, in order to benchmark our ensemble model. The optimal classifier parameters are determined using a 5-level hyperparameter optimization utilizing Enhanced Teaching-Learning-Based Optimization (ETLBO). These improved base learners are sent into a Composite-Level classifier, which employs the Stacking ensemble approach to aggregate predictions and construct the final model. IELA's novelty is its Stacking ensemble technique, which improves predicted accuracy and processing economy by using only the most appropriate base classifiers. Notably, CatBoost paired with XGBoost obtained 86.43% accuracy and an F-score of 84.98%, demonstrating higher performance. IELA is unique in that it combines advanced ensemble approaches with optimization algorithms, resulting in a scalable and efficient prediction system. Future plans include broadening the technique to include temporal datasets for trend analysis and utilizing interpretability tools such as SHapley and LIME to enhance transparency and practical insights. To maximize impact, future research should remove data biases, scale the model for varied educational systems, and create user-friendly tools to empower educators. By pushing the limits of predictive analytics, IELA provides a disruptive approach to improving educational results and making decisions in academic institutions.

Contributory statement

Conceptualization of study problem, Algorithm & Data Analysis: A.Emima.

Research Guidance: Dr. D. I. George Amalarethinam.

Acknowledgements

Thankful for the Co-Research Scholars, Assistant professors of Computer Science, Computer Applications of Jamal Mohamed College and Holy Cross College.

References

1) Hananto AR, Rahayu SA, Hariguna T. An ensemble and filtering-based system for predicting educational data mining. *Journal of Applied Data Sciences*. 2021;2(4):157–173. Available from: https://bright-journal.org/Journal/index.php/JADS/article/view/44.
2) Bentaleb A, Abouchabaka J. Ensemble Learning For Mining Educational Data. *Journal of Theoretical and Applied Information Technology*. 2022;100(9):2715–2722. Available from: https://www.jatit.org/volumes/Vol100No9/3Vol100No9.pdf.
3) Ragab M, Aal AMKA, Jifri AO, Omran NF. Enhancement of predicting students performance model using ensemble approaches and educational data mining techniques. *Wireless Communications and Mobile Computing*. 2021;2021:1–9. Available from: https://www.researchgate.net/publication/356859369_Enhancement_of_Predicting_Students_Performance_Model_Using_Ensemble_Approaches_and_Educational_Data_Mining_Techniques.
4) Injadat M, Moubayed A, Nassif AB, Shami A. Systematic ensemble model selection approach for educational data mining. *Knowledge-Based Systems*. 2020;200. Available from: https://doi.org/10.1016/j.knosys.2020.105992.
5) Injadat M, Moubayed A, Nassif AB, Shami A. Multi-split optimized bagging ensemble model selection for multi-class educational data mining. *Applied Intelligence*. 2020;50:4506–4528. Available from: https://link.springer.com/article/10.1007/s10489-020-01776-3.
6) Singh R, Pal S. Machine learning algorithms and ensemble technique to improve prediction of students performance. *International Journal of Advanced Trends in Computer Science and Engineering*. 2020;9(3):1–7. Available from: https://www.warse.org/IJATCSE/static/pdf/file/ijatcse221932020.pdf.
7) Ajibade SSM, upay JD, Ngo-Hoang DL, Oyebode OJ, Sasan JM, Gido NG, et al. Utilization of ensemble techniques for prediction of the academic performance of students. *Journal of Optoelectronics Laser*. 2022;41(6):48–54. Available from: https://www.researchgate.net/publication/361101272_Utilization_of_Ensemble_Techniques_for_Prediction_of_the_Academic_Performance_of_Students.
8) Priyambada SA, Usagawa T, R ME. Two-layer ensemble prediction of students' performance using learning behavior and domain knowledge. *Computers and Education: Artificial Intelligence*. 2023;5:1–12. Available from: https://doi.org/10.1016/j.caeai.2023.100149.
9) Malik A, Onyema EM, Dalal S, Lilhore UK, Anand D, Sharma A, et al. Forecasting students' adaptability in online entrepreneurship education using modified ensemble machine learning model. *Array*. 2023;19. Available from: https://doi.org/10.1016/j.array.2023.100303.
10) Begum S, Padmannavar SS. Genetically optimized ensemble classifiers for multiclass student performance prediction. *International Journal of Intelligent Engineering Systems*. 2022;15:316–328. Available from: https://oaji.net/articles/2022/3603-1645864316.pdf.
11) Keser SB, Aghalarova S. HELA: A novel hybrid ensemble learning algorithm for predicting academic performance of students . *Education and Information Technologies*. 2022;27:4521–4552. Available from: https://link.springer.com/article/10.1007/s10639-021-10780-0.
12) Asselman A, Khaldi M, Aammou S. Enhancing the prediction of student performance based on the machine learning XGBoost algorithm. *Interactive Learning Environments*. 2023;31(6):3360–3379. Available from: https://doi.org/10.1080/10494820.2021.1928235.
13) Zhou T, Jiao H. Exploration of the stacking ensemble machine learning algorithm for cheating detection in large-scale assessment. *Educational and Psychological Measurement*. 2023;83(4):831–854. Available from: https://pubmed.ncbi.nlm.nih.gov/37398846/.
14) Wandera H, Marivate V, Sengeh MD. Predicting school performance using a combination of traditional and non-traditional education data from South Africa. 2019;p. 1–6. Available from: https://www.k4all.org/wp-content/uploads/2019/07/Predicting-school-performance-using-a-combination-of-traditional-and-non-traditional-education-data-from-South-Africa.pdf.

15) He Y, Zhu J, Fu W. A credible predictive model for employment of college graduates based on LightGBM. *EAI Endorsed Transactions on Scalable Information Systems*. 2022;9:1–9. Available from: https://publications.eai.eu/index.php/sis/article/view/348/262.
16) Ibrahim AA, Ridwan RL, Muhammed MM, Abdulaziz RO, Saheed GA. Comparison of the CatBoost classifier with other machine learning methods. *International Journal of Advanced Computer Science and Applications*. 2020;11(11):1–11. Available from: https://thesai.org/Downloads/Volume11No11/Paper_90-Comparison_of_the_CatBoost_Classifier.pdf.
17) Rimal Y, Sharma N. Ensemble Machine Learning Model Improves Prediction Accuracy for Academic Performance: A Comparative Study of Default ML VS Boosting Algorithm. *ICIMMI '23: Proceedings of the 5th International Conference on Information Management & Machine Intelligence*. 2024;p. 1–7. Available from: https://dl.acm.org/doi/10.1145/3647444.3652473.