# Optimized classification of student performance outcomes using LEE feature selection in the context of educational data mining

**Kishore Kumar Kamarajugadda[1], Movva Pavani[2], Rani Vanathi Gurusamy[3], Nagarajan Karthikeyan[4], Pavan Kumar Nidumolu[5], Desidi Narsimha Reddy[6], Muniappan Ramaraj[7], Rajasekaran Nithya[8]**

[1]Department of Electronics and Communication Engineering, ICFAI Foundation for Higher Education, Hyderabad, India
[2]Department of Electronics and Communication Engineering, Nalla Malla Reddy Engineering College, Hyderabad, India
[3]Department of Computer Science, Ramakrishna College of Arts and Science for Women, Coimbatore, India
[4]Department of Computer Science and Engineering, Kristu Jayanti College, Bengaluru, India
[5]Department of Computer Science and Engineering, KL University, Andhra Pradesh, India
[6]Data Consultant, Soniks Consulting LLC, Dallas, United States
[7]School of Quantum Science, Computing and AI, Rathinam Global Deemed to be University, Coimbatore, India
[8]School of Computing Science, KPR College of Arts Science and Research, Coimbatore, India

**Article Info**

***Article history:***

Received Oct 23, 2025
Revised Apr 24, 2026
Accepted May 11, 2026

***Keywords:***

Classification methods
Educational data mining
Optimization algorithms
Performance measures
Student dataset

**ABSTRACT**

Student speculative victory is a vital area that needs to be predicted to improve the quality of education and aid the institutional decision making. This research work has to planned to use learning based enhanced evaluation (LEE) feature selection method with real world educational datasets for optimized data mining approach to predict student performance. High dimensionality and irrelevant features are common problems with enhanced models, affecting classification accuracy and efficiency. LEE feature algorithm is used to extract important features, that enhance the performance of the model, reduce the calculation quantity of the model. The methodology consists of pre-processing of the dataset, feature selection using LEE algorithm, and testing four classifiers namely support vector machine (SVM), k-nearest neighbor (KNN), adaptive learning, and naïve Bayes. The incorporation of LEE improves the model's ability by reducing noise and highlighting the influential features. Experimental results show that optimized techniques are better in terms of accuracy and robustness than others. The models are evaluated based on important performance metrics such as accuracy, precision, recall, F1-score, and training time. The enhanced approach will help to add to the literature of the field of educational data mining (EDM), providing a practical and effective way of predicting student performance in real academic settings.



***Corresponding Author:***

Muniappan Ramaraj
School of Quantum Science, Computing and AI, Rathinam Global Deemed to be University
Coimbatore, Tamil Nadu, India
Email: ramaraj.phdcs@gmail.com

## 1. INTRODUCTION

A new emphasis of educational data mining (EDM) has been the ability to envisage academic performance of students. To frustrate this, feature selection techniques play an important role in finding the attributes of utmost importance to affect students' performance [1]. This study proposes an optimized data mining approach, which uses the learning based enhanced evaluation (LEE) feature selection algorithm to enhance the predictive model accuracy and efficiency. The proposed framework, which uses LEE to real world educational data, discards redundant and uninformative features and allows more accurate predictions

[2]. These selected features are then applied in training and testing several classification algorithms such as support vector machine (SVM), k-nearest neighbor (KNN), adaptive learning, and naïve Bayes classification which are well established in classification problems [3]. The traditional models are often found to be sub-optimal in capturing the complex nonlinear relationship in environmental data and thus in predicting the data with an acceptable accuracy [4].

While the potential of machine learning algorithm to predict tunnel ground behavior has been shown in recent studies, there are still several key challenges [5]. There are very few publicly available large datasets of labelled, labeled data which support long-term tunnel performance and only a limited number of empirical models trained on small site-specific data sets which do not have the generality necessary for use across a range of geological and excavation conditions. This variety of student difference in learning styles is an important challenge when designing effective instructional strategies [6]. The visual, auditory, kinesthetic (VAK) model gives a theoretical basis for classification of learners, but conventional assessment methods are based on subjective questionnaires, manual assessment and require much time and are not accurate enough [7]. But the previous study lacks of strong predictive models which can classify learning activities into learning style categories with high accuracy without facing the problem of “feature selection,” “class imbalance,” and “overfitting”. Academic pressure, irregular life and psychological problems are often encountered by university students that have a great impact on their subjective well-being (SWB) [8].

An important challenge in creating accurate predictive models is the fact that student performance data sets are often imbalanced, with a majority of students typically in the “average” or “pass” classes, and a minority of students in classes such as “excellent” or “fail” [9]. This imbalance creates an opportunity for biased model training, causing machine learning algorithms to be more effective at detecting the majority class and leading to inaccurate detection and misclassification of students at risk of academic failure. As a result, the models’ predictive reliability, fairness, and generalizability is considerably hampered [10].

## 2. LITERATURE REVIEW

The recent research area is a growing area of research to gain insights from large-scale educational data, which can be used to improve student outcomes and to guide decision making. Many machine learning and data mining algorithms have been used to predict student performance including SVM, decision trees (DT), KNN, naïve Bayes, and ensemble models. Chen *et al.* [11] has subsidized to the existing body of knowledge by reviewing the application of machine learning techniques in higher education with a focus on predicting students’ performance, learning engagement, and self-efficacy. Bilal *et al.* [12] had illustrated with a methodical preprocessing method to overcome spelling variations, textual noise and lack of grammatical uniformity in Roman-Urdu opinions. Guo *et al.* [13] has indistinguishable momentous of this study that provides large-scale, prospective and multicenter cohort evidence on the health, behavioral and psychosocial development of Chinese adolescents [13]. Siddiqua *et al.* [14] has introduced with new enhances to the artificial intelligence (AI) field of mental health, proposing AIDA—a novel AI-based framework for assessing depression in the context of Bangladeshi students. The proposed methods with associations of socially relevant factors and validated psychological metrics to improve contextual accuracy.

Pourhosseini *et al.* [15] has significant novelty of this work is the novel hybrid support vector machine (NHSVM) framework have been established and optimized for accurate prediction of total dissolved solids (TDS) in water systems. The anticipated hybrid SVM approach combines the strengths of conventional predictive models with optimization techniques and intelligent learning mechanisms, boosting the robustness and predictive accuracy of the models. Dashtgoli *et al.* [16] had illustrate of this research is based on data-driven methods to accurately and efficiently predict using conventional numerical or empirical methods, which demands complex assumptions, high computational cost, and limited adaptability to the heterogeneous soil conditions. Song *et al.* [17] has a context for multidimensional investigation is obtainable, coalescing idiosyncratic reaction and objective environmental parameters, to predict thermal comfort in library spaces where user satisfaction and concentration are highly sensitive to microclimatic conditions. The primary innovation of this work is the design of a modified and explicable assessment and enhancement of methodology for neurosurgical skills based on virtual reality (VR) simulations [18]. Its suggestions were thorough review of the current methodologies in EEG-based stress assessment, with a focus on trends, progress, and the efficacy of different signal processing and machine learning [19].

Mumenin *et al.* [20] had illustrate on the existing method subsidizes to the field of mental health assessment in higher education by developing an effective and data-driven approach for early detection of depression among university students. Shoaib *et al.* [21] has illustrate on the key role is in uniting multi-attribute analysis with the development of advanced predictive modelling to facilitate the correct identification of graduates’ career change, job stability and sectoral mobility. The study presents a systematic combination of several predictive algorithms such as SVM, KNN, naïve Bayes and ensemble models, which accounts for complex patterns in educational data [22]. The algorithm's applicability in real-world academic environments is illustrated by the work of Khan *et al.* [23] which shows how the algorithm can be applied in a variety of academic contexts, making it a powerful and scalable tool for educational institutions to improve student success and retention. Waheed *et al.* [24] had shown how well neural network architectures handle complex, non-linear relationships inside educational data, providing a scalable and flexible approach for a variety of e-learning platforms. The work presented in this contribution highlights the promise of intelligent models that can yield higher quality personalized learning and better enable educational support in self-paced learning environment. Ismail and Yusof [25] had classifies different data preprocessing techniques, feature selection techniques, and model optimization techniques used in science, technology, engineering, and mathematics (STEM) education datasets and provides insightful information about factors that affect the prediction accuracy. One of the main drawbacks identified in previous research is inclusion of irrelevant and redundant features in education datasets, affecting the efficiency of the models and the computational cost.

## 3. METHODS

The anticipated approach syndicates with enhanced data mining framework with feature selection for improving the prediction of student performance. First, the raw teaching data is collected and pre-processed in order to guarantee the quality and consistency of the data. It considers with cut down the dimensionality and to remove irrelevant attributes, LEE feature selection algorithm is used which is used to identify the most important predictors leading to the student's academic outcomes systematically and systematically. Enhanced LEE feature selection algorithm is one of the enhanced versions of the traditional LEE method that considered to recognize the most pertinent structures while plummeting redundancy and computational cost. It involves an iterative process of learning, evaluation, and elimination and is well suited to educational data sets with a number of academic, demographic, and behavioral attributes. It contains socio-economic, gender and demographic factors that aid in discerning trends and disparities at the population level. The basic data in the dataset are usually academic attributes, such as marks in the internal assessments, marks in assignments, attendance percentage, previous semester grades, and examination grades. The target variable is typically a numeric score or a pass/fail mark, or a grade category (excellent, good, average, and poor). Before the model is trained, there are preprocessing steps to be taken for the data set, including filling in missing data, normalizing the data, encoding categorical data and feature selection (such as modified LEE feature selection).

### 3.1. Optimized learning based enhanced evaluation feature selection with SVM classifier

The proposed work syndicates the optimized LEE feature selection with SVM classifier, to achieve better accuracy and efficiency in predicting student performance. The optimized feature space is then fed to an SVM classifier which forms a decision boundary that separates the categories of students' performance with a maximum margin. The process of optimized LEE feature selection with SVM classifier for student performance prediction (SPP) is shown in Figure 1. It starts with the preprocessing procedures, which include the imputation of missing values, normalization, and encoding. The evaluation of each feature is performed by LEE function that combines feature relevance with feature redundancy with weight parameters α and β. The best feature subset is then chosen to maximize the discriminative power. This simplified representation is given to the SVM classifier, and kernel functions are used to calculate the decision boundaries.

#### 3.1.1. Learning based enhanced evaluation feature selection

LEE feature selection technique is used to select important features of student data to retain only the most informative predictors for performance modelling.

$$\text{LEE}(f_j) = \alpha \times Rel(f_j, y) - \beta \times Red\,(f_j,\ F) \quad (1)$$

$$F^* = argmax \sum Lee\,(f_j)\ for\, f_j\ \in X \quad (2)$$

Where $Rel(f_j, y)$ relevance of feature $f_j$ with respect to target $y$ (measured using mutual information or correlation). $Red\,(f_j,\ F)$ redundancy of $f_j$with already selected feature set F. $\alpha, and\ \beta$ weight parameters to balance relevance and redundancy. $F^*$represents the optimized subset of features.

#### 3.1.2. Optimized learning based enhanced evaluation based SVM classifier

After feature selection, the refined dataset is classified using SVM. SVM aims to find the optimal hyperplane H that maximizes the margin between classes.

$$f(x) = sign\ (\sum \alpha i, yi, K\ (xi, x) + b) \quad (3)$$

Here, $\alpha i$ are lagrange multipliers, $yi$ is a class label values, $K\ (xi, x)$ is the kernel function, b is the bias term.

Input: Educational dataset $D$, feature set $X$, class labels $Y$.
Output: Predicted student performance.
Initialize
  Handle missing values, normalization, and encoding.
For each feature $f_j$ in $X$ do (Lee feature selection of $F^*$)
  Lee $(f_j) = \alpha * Rel(f_j, y) - \beta * Red\ (f_j,\ F)$
Update the optimal feature selection of Lee function
  $F^* = argmax\ \sum Lee\ (f_j)\ for\ f_j\ \in X$
  Train SVM classifier on reduced feature set $F^*$, K is kernel value
Compute the decision of SVM model
  $f(x) = sign\ (\sum \alpha i, yi, K\ (xi, x) + b)$
  Assign the predicted label $\hat{y}$
  End for Loop
Evaluation of measures like - Accuracy, Precision, Recall, F1-score, AUC
Return predicted label $\hat{y}$ and evaluation measures

Figure 1. Optimized LEE based support vector machine (LSVM) algorithm used to SPP

### 3.2. Enhanced learning based enhanced evaluation feature selection with KNN

The features are selected using the enhanced LEE feature selection method and classified using the KNN classifier with an aim of enhancing classification accuracy by eliminating irrelevant and redundant features. The optimized LEE feature selection and the KNN rule for predicting student performance is illustrated as in Figure 2. Data preprocessing steps start with the treatment of missing data, normalization, and encoding. The LEE function is used for each feature to determine its relevance to class labels and minimize redundancy. The optimal feature subset F* is the one that maximizes the value of LEE. This simplified feature set is then fed into the KNN classifier, and the Euclidean distance is calculated to find the K nearest neighbors. The predicted label is given by the majority voting with neighbor.

#### 3.2.1. Optimized learning based enhanced evaluation based KNN classifier rule

The optimized LEE based k-nearest neighbor (LKNN) classifier rule classifies a new data point according to the maximum class of its nearest neighbors. It is based on the distance between the new instance and all the training instances and choose the k nearest ones and use majority voting (or weighted voting) for deciding which label to assign to the new instance.

$$d(x, x_i) = \sqrt{\sum_{j=1}^{m} \left(x_j - x_{ij}\right)^2} \quad (4)$$

$$\hat{y} = \underset{c \in C}{argmax} \sum_{i \in N_k} 1(y_i = c) \quad (5)$$

Where c is class label values and $y_i$ is denoted by class of neighbor i, then 1 is denoted by indicator function (1 if true, 0 otherwise). The enhanced LEE method systematically selects features that are simultaneously informative, non-redundant, and stable across validation folds.

### 3.3. Optimized learning based enhanced evaluation feature selection with adaptive learning algorithm

The optimized LEE feature selection with adaptive learning algorithm has integrated the advantages of dimensionality reduction and dynamic learning to improve prediction accuracy. However, the LEE method selects most informative and relevant features from the dataset, removing the redundant and noisy features and hence reducing the computational complexity. The optimized LEE feature selection integrated with adaptive learning algorithm for SPP is shown in Figure 3. The first step in the procedure is preprocessing of the educational data, such as dealing with missing values, normalizing the data, and encoding the data. These features are scored with the LEE function which is a balance measure of relevance to the target (student performance) and redundancy with other features. This adaptive mechanism will stabilize training, improve convergence and avoid overfitting.

Input: Educational dataset $D$, feature set $F$, class labels $\lambda$.
Output: Predicted student performance.
Initialize
Handle missing values, normalization, and encoding.
For each feature $f_j$ in $X$ do (Lee feature selection of $F^*$)
Lee $(f_j) = \alpha * Rel(f_j, y) - \beta * Red\ (f_j,\ F)$
Update the optimal feature selection of Lee function

$$F^* = argmax \sum_{f_j \in x} LEE(f_j)$$

Train KNN classifier on reduced feature set $F^*$, K is kernel value
Applied on the KNN Rule

$$d(x, x_i) = \sqrt{\sum_{j=1}^{m} (x_j - x_{ij})^2}$$

Assign the predicted label $\hat{y}$, $\hat{y} = {}^{argmax}_{c \in C} \sum_{i \in N_k} 1(y_i = c)$
End for Loop
Evaluation of measures like - Accuracy, Precision, Recall, F1-score, AUC
Return predicted label $\hat{y}$ and evaluation measures

Figure 2. Enhanced LEE feature selection with KNN applied on SPP

Input: Educational dataset $D$, Learning rate $\eta$, Decay factor $\alpha$, No of neighbor K.
Output: Predicted student performance.
Initialize
Handle missing values, normalization, and encoding.
For each feature $f_j$ in $X$ do (Lee feature selection of $F^*$)

$$\text{Lee}\ (f_j) = \frac{I(f_j;Y)}{1+R(f_j,F\backslash f_j)}$$

Update the optimal feature selection of Lee function

$$F^* = \{f_j |\ lee(f_j) \geq \theta\}$$

Applied on the Adaptive Learning Classifier

$$\eta_{t+1} = \eta_t \cdot \frac{1}{1+\alpha.t}$$

Predict labels for test set using trained model.
End for loop
Evaluation of measures like - Accuracy, Precision, Recall, F1-score, AUC

Figure 3. Optimized LEE feature selection with adaptive learning algorithm using SPP

#### 3.3.1. Optimized adaptive learning classifier

The optimized adaptive learning classifier (AdaBoost classifier algorithms are applied into student dataset to predict the student academic performance then integrating with feature selection techniques that incorporates with LEE techniques to reduce the noise and improve the efficiency) is a machine learning approach where the model dynamically adjusts its learning process based on the data, environment, or training progress. Adaptive learning classifiers change their learning rate, weights, and/or decision boundaries during learning to maximize performance and stabilize learning, whereas static classifiers do not adjust their parameters during learning and use a fixed learning rate.

$$L(\theta) = \frac{1}{n}\sum_{i=1}^{n} \iota\,(f(x_i;\theta).\, y_i) \quad (6)$$

$$\eta_{t+1} = \eta_t \cdot \frac{1}{1+\alpha.t} \quad (7)$$

Here, $\alpha$ is a decay factor and $t$ is the iteration. This prevents overshooting in early stages and allows fine-tuning later.

### 3.4. Optimized learning based enhanced evaluation feature selection with naïve Bayes algorithm

Optimized LEE feature selection with naïve Bayes algorithm improves the SPP by selecting most relevant features and eliminating noise and redundancy in the data. The LEE approach determines the subset of attributes that is most useful in maximizing classification accuracy and naïve Bayes is used on this streamlined feature space because of its simplicity and efficiency. The optimized LEE feature selection with naïve Bayes algorithm for SPP is shown in Figure 4. It starts with preprocessing, missing value treatment, normalization and encoding to make sure the data is consistent. Each feature is then examined by the LEE function, and the significance is calculated based on the correlation between the feature and the target class and on the redundancy of the features. This lowered dimensionality is then passed on to the naïve Bayes classifier, where the posterior probabilities of each class is computed based on Bayes' theorem.

#### 3.4.1. Naïve Bayes classifier algorithm

Naïve Bayes classifier is a simple, yet powerful probabilistic machine learning algorithm based on Bayes' theorem assuming conditional independence of features. It approximates the probability of each class, given a set of features and assigns the most likely class to the set of input features.

$$P(C_k| X) = \frac{P(C_k)\prod_{i=1}^{n} P(x_i|C_k)}{P(X)} \quad (8)$$

$$Y = {}^{argmax}_{C_k} P(C_k) \prod_{i=1}^{n} P(x_i|C_k) \quad (9)$$

Here, $C_k$ is denoted with class label values and predict the student information, x is an iteration vector of the given data, $P(C_k)$ is denoted with the prior probability of class of $C_k$, then likelihood function is $x_i$ and give class of $C_k$.

Input: Educational dataset $D$, threshold value $\theta$
Output: Predicted student performance.
Initialize
Handle missing values, normalization, and encoding.
For each feature $f_j$ in $X$ do (Lee feature selection of $F^*$)

$$\text{Lee}\ (f_j) = \frac{I(f_j;Y)}{1+R(f_j,F\backslash f_j)}$$

Update the optimal feature selection of Lee function

$$F^* = \{f_j \mid lee(f_j) \geq \theta\}$$

Applied on the Naïve Bayes Classifier algorithm

$$P(C_k| X) = \frac{P(C_k)\prod_{i=1}^{n} P(x_i|C_k)}{P(X)}$$

Predict labels for test set using trained model

$$Y = {}^{argmax}_{C_k} P(C_k) \prod_{i=1}^{n} P(x_i|C_k)$$

Predict student performance on the test dataset.
End for loop
Evaluate results using Accuracy, Precision, Recall, F1-score, and AUC.

Figure 4. Optimized LEE feature selection with naïve Bayes algorithm using SPP

## 4. RESULTS AND DISCUSSION

The data set contains more than 2,765 students' information and is used to assess the impact of LEE-based feature selection on academic performance prediction. The use of LEE greatly improved the accuracy of the models by choosing the most pertinent attributes. LEE-based SVM and adaptive learning models achieved high accuracy, precision and F1-score among the classifiers used: SVM, KNN, adaptive learning, and naïve Bayes. The results show that it is possible to improve the prediction accuracy and decrease the training time if the irrelevant and noisy features are removed.

### 4.1. Accuracy

In optimized LEE feature selection framework, accuracy is calculated post feature selection (most suitable features) and after classification (KNN, SVM, naïve Bayes, and adaptive learning). The LEE function eliminates redundant and irrelevant features, which can help to reduce noise and improve model efficiency. Hence, accuracy is used as a baseline criterion to assess the performance of each classifier in predicting student performance.

$$\overline{AC_c} = \frac{1}{k}\sum_{k=1}^{k} ACC_c^{(k)} \tag{10}$$

$$\sigma_c = \sqrt{\frac{1}{k-1}\sum_{k=1}^{k}\left(ACC_c^{(k)} - \overline{Acc_c}\right)^2} \tag{11}$$

For comparison across algorithms, present a table of $\overline{AC_c}$, $\sigma_c$. Optionally perform a paired statistical test on fold-wise accuracies to check whether the observed $\overline{Ac_c}$ is statistically significant. This procedure ensures a fair, reproducible comparison of how optimized LEE affects each classifier's predictive performance on the student-performance dataset.

In the Table 1 shows that the overall comparison of the SPP, the prediction with LKNN has the highest precision (0.9356) and recall (0.925), which makes it have the best F measure (0.9236) and area under the curve or AUC (0.952). The overall balance of LEE based naïve Bayes (LNB) classifier is better with AUC =0.965 and it is precision and recall were consistently high with 0.9244 and 0.905 respectively. In the meantime, LSVM achieved reliable results with a precision of 0.9057 and an AUC of 0.956; likewise, the LEE based adaptive learning (LAL) model still achieved competitive results with a precision of 0.9081 and an AUC of 0.960. The results suggest that all algorithms are successful, with the LNB and LKNN algorithms providing the highest level of predictive student performance accuracy.

The comparative performance of the different algorithms (LSVM, LKNN, LNB, and LAL) using the precision, recall and F-measure for predicting student performance is shown in Figure 5. The LKNN model yielded the best results with a precision of nearly 0.94, a recall rate exceeding 0.92, and an F-measure of 0.923, demonstrating balanced and accurate classification. The LNB classifier also showed good performance with a precision of 0.923, the recall of 0.905 and a competitive F-measure of 0.908. However, LSVM and LAL achieved slightly lower results with a precision of about 0.90 and a recall of less than 0.88 and F-measure below 0.88, which suggests that they lack the sensitivity of identifying real student outcomes.

Table 1. Overall comparison of LEE performance measures

| Algorithms | Precision | Recall | F-measures | ROC/AUC |
|---|---|---|---|---|
| LSVM | 0.905684 | 0.875 | 0.858762 | 0.956 |
| LKNN | 0.935686 | 0.925 | 0.923626 | 0.952 |
| LNB | 0.924397 | 0.905 | 0.907986 | 0.965 |
| LAL | 0.908082 | 0.875 | 0.85956 | 0.960 |

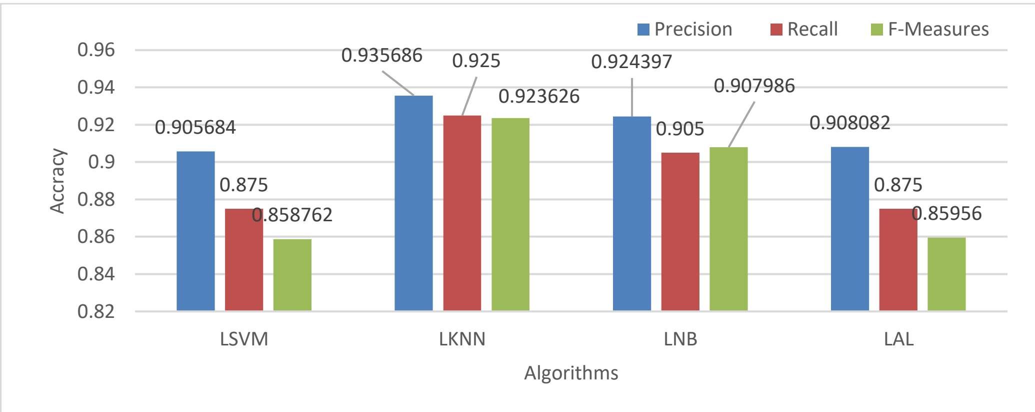


Figure 5. Performance comparison of optimization approach

## 4.2. Precision, recall, and F-measure

The optimized LEE feature selection is done on the student performance dataset to select the most discriminative and least redundant attributes, and then fed into various classifiers including KNN, SVM, naïve Bayes, and adaptive learning. Each algorithm is tested with precision, recall, and F-measure, providing more insight into the quality of classification than accuracy. For each algorithm, after the LEE selection, compute these metrics on the LEE optimized feature set (or on the LEE optimized algorithm-specific feature set if you optimize per algorithm). Typically, better improvements from LEE will result in an increase in true positive (TP) and decrease in false positive (FP) and/or false negative (FN), leading to an increase in precision and/or recall and a larger F-measure. The receiver operating characteristic (ROC) curves of selected features using LEE for various classifiers (KNN, SVM, naïve Bayes, and adaptive learning) are shown in Figure 6. The models have good discriminative power, with an AUC above 0.95 in all model. The highest AUC (0.965) is for the naïve Bayes classier which has the highest classification power for distinguishing successful and unsuccessful students. Adaptive learning is next with an AUC of 0.960, indicating good adaptability and prediction. SVM records AUC of 0.956, which indicates good decision boundaries while KNN records AUC of 0.952, which is slightly less but still useful for pattern recognition applications.

$$Precision = \frac{\Sigma_i TP_i}{\Sigma_i (TP_i + FP_i)} \quad (12)$$

$$Recall = \frac{\Sigma_i TP_i}{\Sigma_i (TP_i + FN_i)} \quad (13)$$

$$F - measures = \frac{1}{c}\sum_{i=1}^{c} F, i. \quad (14)$$

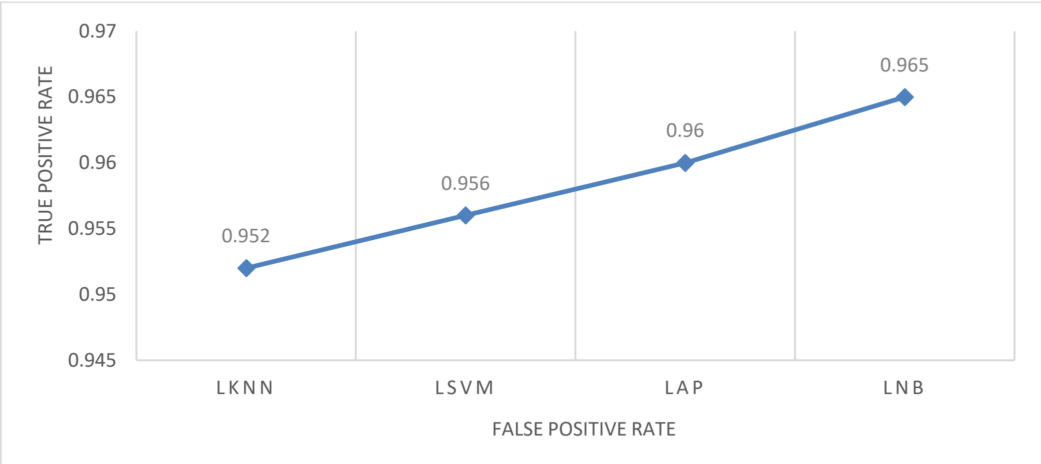


Figure 6. ROC curves for optimize function applied on SPP

### 4.3. Time and ROC with AUC

Typically, two measures are reported for optimized LEE: i) time (wall clock or CPU time) that reflects the cost of feature selection and model training/inference and ii) the discrimination performance in terms of the ROC curve and its summary measure AUC. In practice, you report the total running time, which includes the time spent in feature selection and the time spent computing the classifier output (it is continuous scores, which are probabilities or margins, on the test set), and you report the ROC/AUC calculated on the probabilities/margins output by the classifier on the test set.

$$T_{LEE} = O(n.d.c_{LEE}) \quad (15)$$

$$ROC = \sum_{k=1}^{m-1} \frac{TPR_{k+1} + TPR_k}{2} .(FPR_{k+1} - FPR_k) \quad (16)$$

A larger AUC means better classification ability to distinguish students' success vs failure. The optimized LEE-based KNN, SVM, naïve Bayes, and adaptive learning models are therefore compared and contrasted based on both execution time and ROC/AUC in order to make a balance between cost and accuracy of SPP.

The performance of various algorithms with LEE feature selection approach is shown in Table 2, in which LKNN algorithm gives highest accuracy (0.925) with the lowest computation time (0.001354s) followed by LNB with accuracy (0.905) and computation time (0.001535s). The moderate accuracy of 0.875

was achieved by LSVM, whereas the comparatively higher execution times were 0.009047s and 0.007889s for LSVM and LAL, respectively. The results underline the efficiency of LKNN and LNB to achieve correct prediction of students' performance while training time is kept to a minimum. As shown in the Figure 7, LKNN algorithm has the highest accuracy (0.925) and the lowest computation time (0.001354s), followed by LNB with the accuracy of 0.905 and time of 0.001535s. However, LSVM and LAL gave lesser accuracy, i.e., 0.875, and higher execution time was 0.009047s and 0.007889s, respectively. The results show that LKNN and LNB can achieve better prediction accuracy while maintaining the demand for computational resources, and achieve better results than LSVM and LAL.

Table 2. SPP based on LEE selection approach

| Algorithms | Time | Accuracy |
|---|---|---|
| LSVM | 0.009047 | 0.875 |
| LKNN | 0.001354 | 0.925 |
| LNB | 0.001535 | 0.905 |
| LAL | 0.007889 | 0.875 |

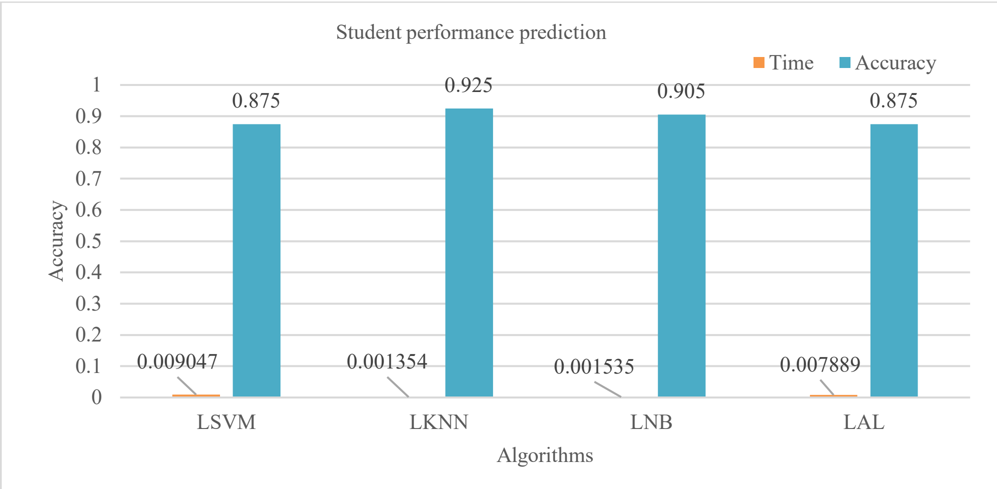


Figure 7. Comparison of accuracy and time based on LEE method

## 5. CONCLUSION

The research paper will be providing an improved data mining model to predict student performance using learning-based enhanced evaluation feature selection method and traditional learning classifiers. This article showed that pruning irrelevant and redundant features not only increased the classification rates but also reduced the computational complexity and was an effective and feasible method in an educational environment. Evaluation results thus confirmed that the use of feature optimization is of very critical importance in improving the predictive performance of the classifiers used in LEE based on SVM, KNN, adaptive learning and naïve Bayes. The experimental has acme the need to assimilate intellectual feature selection and predictive models for better decision-making in education. Optimized techniques, however, is robust but there is room for refinement. Experimental results show that the accuracy of the framework is 92.5% and execution time is 0.001354 seconds, and the lowest accuracy is 87.5% with execution time 0.009047 seconds respectively. In future studies, ensemble learning techniques, deep learning models or hybrid optimization methods could be incorporated to better capture more complex student behavior patterns. Moreover, having the data collected from a variety of academic institutions and with the integration of real-time learning analytics, generalizability and adaptability would be enhanced further. This would have even greater implications for the use of the approach and would give teachers even better tools for boosting student success.

## ACKNOWLEDGMENTS

The author gratefully acknowledges the support to receive from the Department of Biotechnology (DBT), Government of India under the DBT Star College Scheme (HRD-11011/10/2024-HRD-DBT) for providing the necessary infrastructure and resources to carry out this research work in the laboratory of Department of Computer Science at Rathinam College of Arts and Science, Coimbatore.

## AUTHOR CONTRIBUTION STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

| Name of Author | C | M | So | Va | Fo | I | R | D | O | E | Vi | Su | P | Fu |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Kishore Kumar Kamarajugadda | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | | ✓ | ✓ | ✓ |
| Movva Pavani | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | | ✓ | ✓ | ✓ |
| Rani Vanathi Gurusamy | ✓ | | ✓ | ✓ | ✓ | | ✓ | | | ✓ | ✓ | | ✓ | ✓ |
| Nagarajan Karthikeyan | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | | ✓ | ✓ | ✓ | | | ✓ | ✓ |
| Pavan Kumar Nidumolu | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | | ✓ | ✓ | ✓ | | | ✓ | ✓ |
| Desidi Narsimha Reddy | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | | ✓ | ✓ | ✓ | | | ✓ | ✓ |
| Muniappan Ramaraj | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | | ✓ | ✓ | ✓ |
| Rajasekaran Nithya | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | | ✓ | ✓ | ✓ | | | ✓ | ✓ |

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

## CONFLICT OF INTEREST STATEMENT

The authors declare that they have no conflicts of interest related to this work.

## DATA AVAILABILITY

Data availability is not applicable to this paper as no new data were created or analyzed in this study.

## REFERENCES


[1] J. C. M. Bustillo, "Optimization-based techniques prediction model in determining employee turnover," *Procedia Computer Science*, vol. 252, pp. 440–449, 2025, doi: 10.1016/j.procs.2025.01.003.

[2] J. Pan *et al.*, "Prediction of plaque progression using different machine learning models of pericoronary adipose tissue radiomics based on coronary computed tomography angiography," *European Journal of Radiology Open*, vol. 14, Jun. 2025, doi: 10.1016/j.ejro.2025.100638.

[3] A. Kanwar, R. M. Singh, K. Lata, A. Dhiman, and P. Singh, "An effective methodology for identifying adverse drug reactions using firefly algorithm," *Procedia Computer Science*, vol. 258, pp. 4060–4069, 2025, doi: 10.1016/j.procs.2025.04.657.

[4] B. Alnasyan, M. Basheri, and M. Alassafi, "The power of deep learning techniques for predicting student performance in virtual learning environments: a systematic literature review," *Computers and Education: Artificial Intelligence*, vol. 6, Jun. 2024, doi: 10.1016/j.caeai.2024.100231.

[5] J. Zhou, Y. Zhang, Y. Qiu, K. Peng, and M. Khandelwal, "Enhancing tunnel safety with machine learning models for ground behavior prediction," *Tunnelling and Underground Space Technology*, vol. 165, Nov. 2025, doi: 10.1016/j.tust.2025.106888.

[6] R. Boehme, S. Coors, P. Oster, M. M. -Kiefer, and S. Hilbert, "Machine learning for spelling acquisition: how accurate is the prediction of specific spelling errors in german primary school students?," *Computers and Education: Artificial Intelligence*, vol. 6, Jun. 2024, doi: 10.1016/j.caeai.2024.100233.

[7] D. L. Shivaprasad, D. S. Guru, and R. Kavitha, "A novel approach towards key-point based real-time children emotion prediction," *Procedia Computer Science*, vol. 258, pp. 588–598, 2025, doi: 10.1016/j.procs.2025.04.293.

[8] A. R. Sayed, M. H. Khafagy, M. Ali, and M. H. Mohamed, "Exploring the vak model to predict student learning styles based on learning activity," *Intelligent Systems with Applications*, vol. 25, Mar. 2025, doi: 10.1016/j.iswa.2025.200483.

[9] A. C. Kılıç, A. Karakuş, and E. Alptekin, "Prediction of university students' subjective well-being with sleep and physical activity data using classification algorithms," *Procedia Computer Science*, vol. 207, pp. 2648–2657, 2022, doi: 10.1016/j.procs.2022.09.323.

[10] O. M. El-Deeb, W. Elbadawy, and D. S. Elzanfaly, "The effect of imbalanced classes on students' academic performance prediction," *International Journal of e-Collaboration*, vol. 18, no. 1, pp. 1–17, Jul. 2022, doi: 10.4018/IJeC.304373.

[11] J. Chen, X. Zhou, J. Yao, and S.-K. Tang, "Application of machine learning in higher education to predict students' performance, learning engagement and self-efficacy: a systematic literature review," *Asian Education and Development Studies*, vol. 14, no. 2, pp. 205–240, Mar. 2025, doi: 10.1108/AEDS-08-2024-0166.

[12] M. Bilal, H. Israr, M. Shahid, and A. Khan, "Sentiment classification of roman-urdu opinions using naïve bayesian, decision tree and knn classification techniques," *Journal of King Saud University - Computer and Information Sciences*, vol. 28, no. 3, pp. 330–344, Jul. 2016, doi: 10.1016/j.jksuci.2015.11.003.

[13] X. Guo *et al.*, "Longitudinal machine learning prediction of non-suicidal self-injury among chinese adolescents: a prospective multicenter cohort study," *Journal of Affective Disorders*, vol. 392, Jan. 2026, doi: 10.1016/j.jad.2025.120110.

[14] R. Siddiqua, N. Islam, J. F. Bolaka, R. Khan, and S. Momen, "AIDA: artificial intelligence based depression assessment applied to bangladeshi students," *Array*, vol. 18, Jul. 2023, doi: 10.1016/j.array.2023.100291.
[15] F. A. Pourhosseini, K. Ebrahimi, and M. H. Omid, "Prediction of total dissolved solids, based on optimization of new hybrid svm models," *Engineering Applications of Artificial Intelligence*, vol. 126, Nov. 2023, doi: 10.1016/j.engappai.2023.106780.
[16] D. S. Dashtgoli, M. H. Dehnad, S. A. Mobinipour, and M. Giustiniani, "Performance comparison of machine learning algorithms for maximum displacement prediction in soldier pile wall excavation," *Underground Space*, vol. 16, pp. 301–313, Jun. 2024, doi: 10.1016/j.undsp.2023.09.013.
[17] G. Song, Z. Ai, G. Zhang, Y. Peng, W. Wang, and Y. Yan, "Using machine learning algorithms to multidimensional analysis of subjective thermal comfort in a library," *Building and Environment*, vol. 212, 2022, doi: 10.1016/j.buildenv.2022.108790.
[18] F. Li, Z. Qin, K. Qian, S. Liang, C. Li, and Y. Tai, "Personalized assessment and training of neurosurgical skills in virtual reality: an interpretable machine learning approach," *Virtual Reality and Intelligent Hardware*, vol. 6, no. 1, pp. 17–29, Feb. 2024, doi: 10.1016/j.vrih.2023.08.001.
[19] S. Acharya, A. Khosravi, D. Creighton, R. Alizadehsani, and U. R. Acharya, "Neurostressology: a systematic review of eeg-based automated mental stress perspectives," *Information Fusion*, vol. 124, Dec. 2025, doi: 10.1016/j.inffus.2025.103368.
[20] N. Mumenin *et al.*, "Screening depression among university students utilizing GHQ-12 and machine learning," *Heliyon*, vol. 10, no. 17, Sep. 2024, doi: 10.1016/j.heliyon.2024.e37182.
[21] M. Shoaib, M. U. Afridi, and G. Husnain, "Forecasting algorithm for the vocational mobility of graduates with multiple attributes," *Expert Systems with Applications*, vol. 297, Feb. 2026, doi: 10.1016/j.eswa.2025.129463.
[22] M. Bhushan, U. Verma, C. Garg, and A. Negi, "Machine learning-based academic result prediction system," *International Journal of Software Innovation*, vol. 12, no. 1, pp. 1–14, Dec. 2023, doi: 10.4018/IJSI.334715.
[23] A. Khan, S. K. Ghosh, D. Ghosh, and S. Chattopadhyay, "Random wheel: an algorithm for early classification of student performance with confidence," *Engineering Applications of Artificial Intelligence*, vol. 102, 2021, doi: 10.1016/j.engappai.2021.104270.
[24] H. Waheed, S.-U. Hassan, R. Nawaz, N. R. Aljohani, G. Chen, and D. Gasevic, "Early prediction of learners at risk in self-paced education: a neural network approach," *Expert Systems with Applications*, vol. 213, 2023, doi: 10.1016/j.eswa.2022.118868.
[25] N. Ismail and U. K. Yusof, "A systematic literature review: recent techniques of predicting stem stream students," *Computers and Education: Artificial Intelligence*, vol. 5, 2023, doi: 10.1016/j.caeai.2023.100141.


## BIOGRAPHIES OF AUTHORS


**Kishore Kumar Kamarajugadda** received his B.Tech. degree in Electronics and Communication Engineering from Acharya Nagarjuna University, Guntur in 2000, M.Tech. in Digital Systems and Computer Electronics from JNTU College of Engineering, Anantapur, in 2004 and Ph.D. from GITAM, Visakhapatnam, India in 2020. He holds an MBA from ICFAI Business School, Hyderabad. He is an IEEE senior member and fellow, Institute of Engineers (IE). He has been associated with ICFAI Group in various capacities as dean, associate dean, Chairperson-Academic Instructions, head of the Department of Electronics and Communication Engineering. He has about 23+ years of teaching, industry and research experience together. He has over 50+ publications in SCI, SCIE, Scopus listed journals, book chapters, and has presented papers at renowned conferences. He has completed two funding projects and has five patents. He is a reviewer of several reputed international journals and conference proceedings. His areas of interest are internet of things, image processing, machine and deep learning, internet of things, wireless sensor networks, artificial intelligence, and VLSI. He can be contacted at email: drkishorekumarkamarajugadda@gmail.com.



**Movva Pavani** received her B.Tech. degree in Electronics and Communication Engineering from JNTU Hyderabad in 2002, M.E. in Systems and Signal Processing, College of Engineering, Osmania University, Hyderabad, July 2009 and Ph.D. from GITAM, Visakhapatnam, India in 2021. She is an IEEE senior member, life member of the Institute of Engineers (IE) and the Institution of Electronics and Telecommunication Engineers (IETE). Presently working as a professor in the Department of Electronics and Communication Engineering, Nalla Malla Reddy Engineering College, Divya Nagar, Hyderabad. She has about 23+ years of teaching, industry and research experience together. She has over 40 publications in SCI, SCIE, Scopus listed journals, and has presented papers at renowned conferences with 330+ citations. Her areas of interest are internet of things, image processing, machine and deep learning, wireless sensor networks, wireless ad hoc networks, and artificial intelligence. She can be contacted at email: drmovvapavani@gmail.com.



**Rani Vanathi Gurusamy** is currently working as professor and head at Sri Ramakrishna College of Arts and Science for Women. She was associated with the same college for last 24 years. She received her Ph.D. in computer science from Bharathiar University during 2013. Her area of research includes data science, mobile Ad Hoc networks, cloud computing, network security, and online education. She has published more than 30+ papers in national and international journals which includes Scopus, Thomson Reuters, UGC., IEEE, and also presented in various national and international conferences. Her latest book publication is on the topic multimedia systems. She received 6,00,000/- funding assistance from ICSSR IMPRESS during the year 2019. She received "Research award" from Bharathiar University during the National Science Day 2021. She is the current vice-president of SRCW alumni association and the research coordinator of the SRCW. She can be contacted at email: ranics@srcw.ac.in.

**Nagarajan Karthikeyan** working as an assistant professor in Department of Computer Science, Kristu Jayanti College Bengaluru and he completed his doctorate in Bharathiyar University Coimbatore in the area of prediction and pattern mining. He has published more research article in the reputed various national and international journals and also filed the patents in the same field. He can be contacted at email: karthikrohan39@gmail.com.

**Pavan Kumar Nidumolu** is a distinguished academician serving as a professor in the Department of Computer Science and Engineering at KL University, Vaddeswaram, Andhra Pradesh, India. He is actively involved in teaching, research, and academic administration, contributing significantly to the growth and development of engineering education. He is committed to research and innovation, with contributions in reputed journals and conferences. His academic interests span areas such as artificial intelligence, data science, machine learning, and advanced computing technologies. He is actively participating in faculty development programs, workshops, and conferences, reflecting his dedication to continuous learning and academic excellence. He can be contacted at email: nvspavankumar@gmail.com.

**Desidi Narsimha Reddy** is an accomplished professional with an impressive educational background and extensive experience in the field. He holds a postgraduate degree in machine learning and AI from Purdue University, complemented by an MBA in finance and information systems from MG University. Additionally, he has completed a Program on Business Analytics: From Data to Insights from Wharton Management School and is a certified Project Management Professional (PMP) from the PMI Institute. With close to two decades of professional experience, he has carved a niche in Business Intelligence. His proficiency encompasses various domains, including financial reporting applications, data management, master data management, data governance, data science, and artificial intelligence and machine learning. Throughout his career, he has contributed significantly to the field, reflected in the publication of papers in several esteemed journals. His dedication to continuous learning and his diverse skill set makes him a valuable asset in the dynamic landscape of data management and analytics. He is also a member of IEEE. He can be contacted at email: dn.narsimha@gmail.com.

**Muniappan Ramaraj** is working as an assistant professor in the Department of Computer Science at Rathinam Global Deemed to be University, Coimbatore. He holds a Ph.D., degree in Computer Science at Bharathiar University in the year of 2020 with specialization in data mining with image process and also fuzzy logic in the image analysis. His research areas are data mining, image processing, fuzzy logic, pattern recognition, and deep learning concept. He has published more research article in the reputed various national and international journals and also filed the patents in the same field. He has a reviewer of many international journals including with IEEE, ASTESJ, and JERS. He can be contacted at email: ramaraj.phdcs@gmail.com.

**Rajasekaran Nithya** working as an assistant professor in the KPR College of Arts Science and Research, Coimbatore for the last 5 years. She has about 10 years of teaching experience. She got approved as doctorate in computer Science in the year of 2021. She published about 15 Scopus and Web of Science indexed articles in a well reputed journal. She published national level and international level books that are available in Google Kindle, Flipkart and Amazon. She published patterns. Her area of expertise is machine learning, artificial intelligence, software engineering, data mining, and mobile network. She has been serving as a journal reviewer for more than 10 Elsevier journals. She can be contacted at email: nithya5790@gmail.com.