INTERNATIONAL JOURNAL OF TECHNICAL AND NATURAL SCIENCES



DOI: https://doi.org/10.69648/NVAX8158

International Journal of Natural and Technical Sciences (IJTNS), 2025; 5(1): 43-75

ijtns.ibupress.com

Online ISSN: 2671-3519

Application: 02.06.2025

Revision: 16.06.2025

Acceptance: 26.06.2025

Publication: 30.06.2025

Saxena, M., & Singh, N. P. (2025). Role of Hybrid Feature Selection Algorithms in Foreseeing Student Performance. International Journal of Natural and Technical Sciences, 5(1), 43-75. https://doi.org/10.69648/NVAX8158

**Mamta Saxena,** School of Engineering and Technology, MVN University Palwal, Haryana, India, https://orcid.org/0009-0007-8476-2836. Email: 17ca9001w@mvn.edu.in

**Netra Pal Singh,** School of Engineering and Technology, MVN University Palwal, Haryana, India, https://orcid.org/0000-0002-3006-9522. Email: deanset@mvn.edu.in

BALKAN JOURNALS

# Role of Hybrid Feature Selection Algorithms in Foreseeing Student Performance

**Mamta Saxena, Netra Pal Singh**

*Abstract*

The interest of analysts in applying analytical algorithms to investigate students' performance data with a view to enhancing their knowledge is growing among data miners. The key factor of this trend is ample multimedia data generated by educational institutions with the usage of technologies, tools of e-learning, and other digital platforms for conducting online courses. Educators could utilize these data to examine and understand students' learning behaviors by using data mining techniques to forecast student achievement, among other things. In data mining models, the difficult task is to select effective strategies that will generate satisfactory forecast accuracy. To achieve such goals, this article uses six hybrid feature selection (FS) algorithms such as hybrid PSO, hybrid GA, filter wrap, wrapper embedded, ensemble based, and filter embedded and five algorithms such as "Random Forest (RF)", "Decision Tree (DT)", "Logistic Regression (LR)", "K-Nearest Neighbor (KNN)", and "Support Vector Machine" (SVM) to forecast the performance of students and to make comparison of performance of five classification processes in the perspective of six FS models. Based on the analysis of the given data set, it is concluded that the performance of classification models improved when 20 important features were part of the data in place of all 30 attributes. Further reduction in the number of attributes did not result in further improvement in the performance of classification models in predicting students' success or failure.

**Keywords: '**Hybrid Feature Selection', 'Random Forest', 'Decision Tree', 'Logistic Regression', 'KNN', 'SVM', 'EDM'

## INTRODUCTION

Every educational institution desires that students who are enrolled with them must achieve academic success to a high standard because it helps in creating brand value for the institutions. Students, on the other hand, value their academic performance since it is central to achieving success, including getting admitted to a prominent university or institution. It will finally fetch a remunerative job with better growth in their career. It will ultimately raise their lifestyle and living. The question is "how does data mining come into the picture?" To mention, researchers from the University of Miami claim that Grade Point Average (GPA) in high school can predict the category of college a student will attend. In addition, GPA in high school can also predict who will complete their degree program and who will not, along with predictions about their future earnings (Marte, 2014). Therefore, it is becoming essential to develop better data mining techniques to estimate the educational accomplishment of the students beforehand, making use of easily accessible data. This new set of improved data mining algorithms will assist analysts in finding masked patterns, classifying the data into predefined categories, and predicting students' academic accomplishment from students' databases at universities.

It is a well-documented fact that for extracting valuable insights from volumes of data that are being captured by all sectors of the economy, data analytical algorithms are used. Some of the key contribution of the researchers of data mining in different sectors are Koh and Tan (2005), Sodhro and Zahid (2021) in healthcare sector, Vazan et al (2017) in manufacturing and engineering, Awoyemi et al (2017) in fraud detection, Lan et al (2018) in bioinformatics, Farzaneh and Fadlalla (2017) in business organizations, Gosh and Chaudhuri (2021) in stock markets, Shah (2022) in remote sensing. In addition, researchers have applied data mining algorithms in many other fields. The educational sector is no exception to these trends. Large volumes of data are collected by the educational institutions as these institutions have invested in the implementation of technologies in the classroom, i.e., usage of smart boards, computers, smartphones, tablets, iPads, learning management systems, virtual reality systems, etc. (Zauzmer, 2020; Winstead, 2025; Jackson, 2013; Velazquez, 2023). These investigations necessitate scraping information from these raw databases and finding patterns or valuable information with educational data mining (EDM) tools.

EDM tools are very useful for making desirable and accurate predictions, whether students will pass or fail, by extracting hidden information from the educational

databases available with educational institutions, rather than waiting for the final results of the examination. Hasan et al (2020) discovered that the success or failure of students could be predicted with 88.3% accuracy with the help of a random forest classification algorithm.

Data pertaining to students is of interest to management of institutions because it can help in predicting performance of students well in advance that can be used in devising strategies for taking corrective measures if needed (Nagahi et al., 2020; Okubo et al.,2017; BYung-Hak, 2018); predicting dropouts (Wang et al., 2017; Whitehill et al., 2017; Xing & Dongping, 2019), identifying undesirable behavior of students (off-task behaviors) (Bohong et al., 2020; Baker et al., 2016; Goldberg et al., 2021, Cetintas et al., 2010, Zaletelj & Kosir, 2017), and tracking psychological state of student(s) in real time with wearable technology and sensors (Antoniou et al., 2020; Ahonen et al., 2018; Magsi et al., 2021). It is a well-known fact that an academic success or failure of a student is influenced by their parents' educational background, personal characteristics, psychological makeup, and other contextual circumstances, such as living environment. EDM helps identify the connection between these factors and the results of students in their examinations. Helal et al. (2018) used academic and demographic data to identify students who are weak in their studies and advised them to put extra effort into their studies so that they can do well in their academic pursuits. Their study is based on four classification algorithms, such as 'Naïve Bayes (NB)', 'Sequential Minimal Optimization (SMO)', 'J48', and 'Repeated Incremental Pruning to Produce Error Reduction (JRIP)'. They developed sub-models for the purpose with better efficiency. Hussain et al. (2018) found that internal evaluation marks are the most significant factors influencing the results of the students. Fernandes et al (2019) used a gradient boosting machine algorithm to predict the success of the students. They concluded that "grades" of previous exams and "absences" in the classroom are key features in forecasting the performance of the students in examinations. They further concluded that the status of the school, age of the students, and neighborhood of home location are prospective features in forecasting the success of the students.

Keeping in view the importance of the hidden knowledge in databases of students, this research article analyzed the data set of students with five machine learning algorithms in combination with six hybrid FS algorithms for selecting the best features to foresee the performance of the students with existing data. The hybridization involves either the FS algorithm or the machine learning algorithm, but in the present research paper, it is in the FS techniques. The research paper is structured

into five sections, starting with an introduction section. It is followed by a review of the literature on classification and hybrid FS algorithms in the context of EDM in Section 2. Six hybrid FS techniques and five classification methods, along with the methodology of the present study, are presented in Section 3. The analysis and results of the analysis, along with a brief interpretation, are embodied in Section 4. Section 5 wraps up the work in the form of a conclusion.

## LITERATURE REVIEW

ML algorithms along with FS algorithms are used in many domains of study for classifying variables of interest (Nagahisarchoghaei et el, 2020; Shah and Du, 2021a; Shah and Du, 2021b; Shah and Du, 2021c). In educational data analysis, ML and Deep Learning (DL) models have been applied in the past to forecast the achievement of students. A review of select articles EDM used for predicting academic achievements of the students is presented in this section. These methods are applied to forecast the academic success of students that can be explained with many factors, i.e., socio-economic characteristics, lifestyle, family structure, peers of students, etc. Additionally, past studies have considered gender of student's friends, the time spent on social media for networking, family characteristics (size and income of the family, parents' education, parent profession), academic examination test score data, lab experiment results, being present in classes, GPA, prior classes grades, and demographic information such as age, gender, race, etc.). In addition, academic achievements before and after enrollment in a given degree program or class are used as a proxy for academic qualities in many studies.

Aulck et al. (2016), Daud et al. (2017), and Garg (2018) used demographic characteristics of the students to predict their academic achievement. To be specific, Aulck et al. (2016) made use of age data, while Kemper et al. (2020) concentrated on the country of origin of students for predicting the success or failure of students. Aluko, et al. (2018) and Luhaybi, et al. (2018) considered the grade point average (GPA) of the previous class for the purpose of predicting performance. GPA is frequently used as a pre-enrolment feature since the academic performance of the students, as reflected in their prior endeavors, always impacts future successes. Post-enrollment features such as marks in tests, midterm examinations, quizzes, and assignments have been used to assess student achievements by Asif et al (2017). Li et al. (2019) reported that participation in extracurricular activities has a big impact on academic achievement, and the success rate is also influenced

by attendance and participation in classroom discussion. Francis and Babu (2019) mentioned that the family characteristics like size and income can have a big impact in predicting performance of the students in many cases.

In a nutshell, EDM is an effective method for forecasting the educational accomplishment of the students. The two machine learning algorithms most frequently utilized in analyzing education data are random forest (RF) and decision tree (DT). Hussain et al (2018) applied J48, "Decision Tree", "JRIP", and "Gradient Boosting" Classifiers. Heuer and Breiter (2018) applied logistic regression, 'Decision Tree', 'Support Vector Machine' (SVM), and random forest (RF) classifiers for classifying the achievement of students as success or failure. To predict dropouts, Haiyang et al (2018) applied a Time Series Forest (TSF) classification model. Rizvi et al (2019) applied a decision tree-based approach for studying the role of demographics in online teaching in order to foresee triumphs of the students. Hlosta et al. (2017) used a framework of Ouroboros and classification models such as 'support vector machine', 'Random Forest', 'Naïve Bayes', 'Tree Boosting XG Boost', and logistic regression algorithms for the identification of students who are not performing well.

According to Wasif et al. (2019) student demographic and enrollment characteristics are helpful to ascertain academic success. They concluded that the Bayesian network performs better than the Decision Tree algorithm in predicting the achievement of students. To envisage academic accomplishment of the students, Alberto et al. (2021) employed methods such as multilayer perceptron (MLP), DT, RF, and extreme gradient boosting. The best performance was of MLP, and concluded that MLP accuracy was highest, i.e., 78.2%. Azizah et al. (2018) applied 'Naive Bayes' (NB) and 'Decision Tree' to forecast who will pass and who will fail. He concluded that NB had the best accuracy (63.8%).

Hybrid Models are applied for guessing students' future performance in different combinations. The following Table 1 provides a clear view of classification hybrid models along with references. However, the present study has not used these hybrid models. Hybridization is done for FS algorithms in this study.

**Table 1:**

*List of Hybrid Classification Algorithms*

| Hybrid Model | Research Reference |
|---|---|
| Decision Tree + Logistic Regression | Kumar & Soni (2019). "Predicting Student Performance using Hybrid Classification Techniques." *IJCSIT*. |
| Random Forest + Support Vector Machine | Amado, et al. (2015). "A Hybrid Model for Predicting Student Performance in Educational Data Mining." *JEDM*. |
| Neural Network + Decision Tree | Silva, et al. (2017). "A Hybrid System for Predicting Students' Academic Performance." *IJCSE*. |
| K-Nearest Neighbours + Naive Bayes | Shabir & Arshad (2017). "Hybrid Machine Learning Model for Predicting Student Performance." *ICCCET*. |
| Gradient Boosting + Random Forest | Deshpande, & Agarwal. (2021). "Performance Prediction of Students using a Hybrid Ensemble Learning Model." *IJACSA*. |
| Support Vector Machine + Naive Bayes | Yaseen, Z., et al. (2020). "Predicting Student Performance Using Hybrid SVM and Naive Bayes." *IJCAI*. |
| KNN + Neural Network + Random Forest | Zhang, et al. (2019). "A Hybrid Ensemble Model for Student Performance Prediction." *AIED*. |
| XGBoost + Logistic Regression | Ali, & Qureshi (2020). "Predicting Academic Performance of Students using Hybrid Machine Learning Algorithms." *LJSS*. |
| Bagging + Boosting (Random Forest + AdaBoost) | Ebrahim, et al. (2018). "A Hybrid Ensemble Model for Academic Performance Prediction of Students." *JES*. |

## RESEARCH METHODOLOGY

The process of employing data mining tools to forecast and examine academic records of the students involves a methodical approach of gathering, preprocessing and examining a range of data sources pertaining to the demographics and academic credentials of pupils. Data cleansing, FS, classification model selection, training, and evaluation are important phases. This research article uses hybrid

FS algorithms to identify features with high predictability and classification algorithms to classify the students as good or bad performer. The research objectives, description of data sets, etc., are presented in the subsequent sections.

## OBJECTIVE / QUESTIONS

The main research objective is to examine the impact of hybrid FS algorithms on the effectiveness of classification algorithms applied to the best set of predictors. Two research questions will be addressed by this study:

**RQ1**. Which features of data sets are crucial for predicting students' academic success?
**RQ2.** Which assemblage of hybrid FS and classification models works best together toward forecasting the success or failure of students?

### *Description of the Dataset:*

The dataset comprises 395 students' records with 30 features for each record. This dataset has been used in many studies and is available publicly on many data repositories such as Kaggle and the UCI Data Repository. It was previously used to check the students' academic success and passing rates. There are three categories of attributes in this dataset (i) demographic features (sex, age, address, family size, Parent status, health), (ii) academic background features (type of school, failed attempts, time od studying in a day or week, support extended by school, paid activities, nursery, higher studies, absences from clases) and (iii) social-economic features (Academic qualification of Mother & Father, Job of Mother & Father, family support, reason for not performing, care of guardian, travelling time from home to school, internet used, romantic, family relation, free time, gout for outing, Weekday alcohol consumption, Weekend alcohol consumption). These characteristics are described in Table 2.

**Table 2:**

*Description of Dataset*

| SN | Attribute | Type (Feature) | Measure | Description |
|---|---|---|---|---|
| 1 | School | Categorical | Nominal | 'GP' – "Gabriel Pereira" or 'MS' – "Mousinho da Silveira" |
| 2 | Gender | Categorical | Nominal | Female or Male |
| 3 | Age (Years) | Categorical | Nominal | ≥ 16' OR '< 16' |
| 4 | Address | Categorical | Nominal | 'U' – "Urban" or 'R' – "Rural" |
| 5 | "Famsize" | Categorical | Nominal | ' ≤ 3 or – > 3 |
| 6 | "Pstatus" | Categorical | Nominal | "Living with parents": Yes or No |
| 7 | "M_edu" | Categorical | Nominal | "Education of Mother": '$10^{th}$, $12^{th}$ , 'Graduate' or 'Post Graduate' or 'Research'. |
| 8 | F_edu | Categorical | Nominal | "Education of Father" : $10^{th}$, $12^{th}$, 'Graduate' or 'Post Graduate' or 'Researcher". |
| 9 | "M_job" | Categorical | Nominal | "Job type of Mother" : 'Teaching', "health care professional", "Civil 'servant'" (e.g. "admin or police"), 'Home Maker' or "Other" |
| 10 | F_job | Categorical | Nominal | Father Job: 'teacher', 'health' care related, civil 'services' (e.g. administrative or police), 'at_home' or 'other' |
| 11 | Reason | Categorical | Nominal | Reason to choose this school: Nearness to 'home', 'reputation of the school', 'course' preference or 'other' |
| 12 | Guardian | Categorical | Nominal | "1-mother", "2-father" or "3-Other"" |
| 13 | Travel_time | Categorical | Binary | <= 2 or > 2 Divided in to two categories based on hours |
| 14 | Study_time | Categorical | Binary | <= 2 or > 2 . Divided in to two categories based on hours |
| 15 | Failures | Categorical | Binary | Number of past class failures: <=3 or > 2 |
| 16 | Schoolsup | Categorical | Binary | Additional School Support: '1-yes' or '2-no' |
| 17 | Famsup | Categorical | Binary | Family Educational Support: '1-yes' or '2-no' |
| 18 | Paid | Categorical | Binary | Extra paid classes within the course subject (Math or Portuguese) '1-yes' or '2-no' |

| | | | | |
|---|---|---|---|---|
| 19 | Activities | Categorical | Binary | Extracurricular Accomplishments: '1-yes' or '2-no' |
| 20 | Nursery | Categorical | Binary | Attended Nursery School: 'yes' or 'no' |
| 21 | Higher | Categorical | Binary | Want to take Higher Education: 'yes' or 'no' |
| 22 | Internet | Categorical | Binary | Internet Access at Home: 'yes' or 'no' |
| 23 | Romantic | Categorical | Binary | With A romantic Relationship: 'yes' or 'no' |
| 24 | Famrel | Categorical | Binary | Quality of Family Relationships: (1 ("very bad") – 5 "Excellent") - converted to binary: < = 3 or : > 3 |
| 25 | Freetime | Categorical | Binary | Free Time After School ("1- Very Bad" – "5- Excellent") converted to binary: <= 3 or > 3 |
| 26 | Goout | Categorical | Binary | Going out with friends ("1- very bad" – "-5 Excellent") converted to binary: :<= 3 or : > 3 |
| 27 | Dalc | Categorical | Binary | Work Day Alcohol Consumption (1 very bad – 5 Excellent) but converted to binary: :<= 3 or : > 3 |
| 28 | Walc | Categorical | Binary | Weekend Alcohol Consumption ("1- very bad" – "5- Excellent") converted to binary: <= 3 or : > 3 |
| 29 | Health | Categorical | Binary | Current health Status: ("1- very bad" – "5- Excellent") Changed to binary: :<= 3 or > 3 |
| 30 | Absences | Categorical | Binary | Number of Absences: Altered in to Binary : ≤ 30 or : > 30 |
| 31 | Passed (Target) | Categorical | Binary | "Yes" or "No" |

"Source UCI Data Repository :https://archive.ics.uci.edu/dataset/320/student+performance". Note Zip Codes are not considered.

## Experimental Setup:

### *Environment*

To evaluate the rank/importance of attributes in the given dataset with select FS/ hybrid FS methods and to achieve the research objective, a set of six hybrid FS and five classification algorithms is identified based on a review of the literature.

For the purpose of analysis and implementation of a combination of two sets of algorithms, Python libraries are used as listed in Table 3. The steps involved in the computation are listed below.

**Step 1: Selection of the Dataset**

The dataset, as described in Table 2, is selected for applying hybrid FS algorithms and classifiers. The dataset is identified and downloaded from the existing repositories of datasets available at many sources on the Internet.

**Step 2: Data Preprocessing**

This step involves functions like loading the dataset, handling missing values, encoding categorical variables, and dividing the dataset into a training and a test dataset. The authors have used 80% data for training the models and 20% for testing the goodness of the model. Balancing of the training data set from the perspective of the target variable is also ascertained at this stage before taking the next step.

**Step 3: Selection of Algorithms for FS and Prediction of the Target Variable of the Dataset**

This step consists of the selection of six hybrid FS and five classification algorithms based on a review of the literature in the research domain.

**Step 4: Generate Base Model Prediction Results**

Results are calculated using five classification algorithms by considering all features of the given data set.

**Step 5: Evaluation of the Performance of Hybrid FS and Classification Algorithms Using Quality Metrics/ Parameters**

Computations on the selected dataset are done with five classification algorithms to forecast the performance of students. This is done for a different set of attributes, having their importance derived using hybrid FS algorithms. Metrics are applied and compared to identify which combination of algorithms achieves better values of quality metrics for a given set of attributes.

## *Libraries Used*

This section presents a list of libraries that can be used for the hybrid FS method, classification algorithms, visualization, and evaluation matrices. The details are given in Table 3.

**Table 3:**

*Libraries used for Computation and Visualization*

| **Libraries of FS Methods** | | |
|---|---|---|
| **Hybrid FS Method** | **Libraries** | **Comments** |
| Hybrid PSO | Pyswarm | It combines with filter and wrapper methods |
| Filter-Embdded | Scikitlearn, MLxtend, TPOT, Borutpy | It combines filter and embedded methods |
| Hybrid GA | Pygad, inspired, Deap | GA searched the best subset of features. The hybrid part combines it with filter, wrapper, and embedded methods. |
| Wrapper-Embedded | Scikitlearn, MLxtend, TPOT, Borutapy | It combines the wrapper method with embedded methods. |
| Filter-Wrapper | Scikitlearn, MLxtend, TPOT, Borutpy | It combine filter with wrapper method. |
| Ensemble Based Mathods | Scikitlearn, mlxtend,Borutapy | It calculates importance features using multiple FS methods and ensemble best on voting |
| **Libraries of Classification Algorithms** | | |
| Algorithm | Libraries | Comments |
| Random Forest | Sklearn.Emsemble | Library imports Random Forest Classifier |
| SVM | Sklearn.svm | Library imports SVM classifier |
| Decision Tree | Sklearn.decision-treeclassifier | Library imports decision tree classifier |
| Logistic Regression | Sklearn.linear_model | Library imports logistic regression classifier |
| KNN | Sklearn.neighbors | Library imports KNN Classifier |
| **Libraries for Evalution Metrices and Visualization** | | |
| Evaluation Metrices | Scikit_learn.metrices | Calculates “accuracy (ACC)”, “precision (PRE)”, “recall (REC)”, “F1 Score (F1_S), ROC Curve (ROCC), and AUC curve (AUCC) |
| Visualization | Matplotlib, seaborn | For plotting graphs |

### *Evaluation Metrics*

Five key measures are chosen to evaluate the prediction ability of classification models. These measures are ACC, PRE, F1_S, PRE, REC, AUCC, and AUCC. All measures inferred together offer a more complete view of the model's performance. To find out how successfully the models applied in the research article detect real positives and negatives in more detail, with these measures.

## Interpretation of Results

The present section presents ranks (a measure of the importance of the attributes) using six hybrid FS algorithms. In the next stage of computation, five classification algorithms are used to forecast the results of the students. This is achieved in four parts, i.e., considering all attributes, followed by considering the 20 most important attributes, the 15 most important attributes, and the 10 most important attributes. This is being done to conclude a tradeoff between loss of information (resulting in the selection of fewer attributes) and its influence on the outcome of classifiers in predicting the results of the students in the examination. Generally, it is concluded that by removing highly non-significant predictors, the performance of the classifier improves and reduces the requirement of computing power. But by eliminating more variables, it negatively impacts the performance of the classifier. The results are presented in the same sequence in subsequent sub-sections.

### Ranks of Features as Calculated Using Six Hybrid FS Algorithms

This section presents ranks of attributes with six hybrid FS algorithms that are showcased in Table 4, along with a reference.

**Table 4:**

*Ranks of all attributes with Hybrid Methods*

| Feature | Hybrid PSO Method (Xue, et al, 2016) | Hybrid GA Method (Xue, et al, 2016) | Filter Wrapper Method (Dash & Liu, 1997) | Wrapper-Embedded Method (Saeys et al, 2007) | Ensemble-Based Method (Saeys, et al, 2008) | Filter-Embedded Method (Guyon & Elisseeff, 2003) |
|---|---|---|---|---|---|---|
| | Rank | Rank | Rank | Rank | Rank | Rank |
| school | 6 | 22 | 6 | 27 | 25 | 30 |
| Sex | 20 | 19 | 26 | 12 | 20 | 25 |
| Age | 24 | 5 | 3 | 13 | 14 | 4 |
| Address | 10 | 4 | 5 | 15 | 10 | 22 |
| Famsize | 16 | 26 | 27 | 20 | 22 | 20 |
| Pstatus | 17 | 29 | 23 | 28 | 28 | 29 |
| Medu | 9 | 25 | 9 | 5 | 6 | 2 |
| Fedu | 3 | 13 | 17 | 4 | 2 | 7 |
| Mjob | 8 | 21 | 30 | 2 | 3 | 5 |
| Fjob | 13 | 30 | 12 | 1 | 5 | 11 |
| Reason | 30 | 18 | 29 | 3 | 7 | 10 |
| Guardian | 27 | 27 | 21 | 6 | 9 | 15 |
| Traveltime | 12 | 17 | 28 | 29 | 29 | 28 |
| Studytime | 2 | 16 | 4 | 17 | 11 | 3 |
| Failures | 14 | 1 | 1 | 7 | 1 | 1 |
| Schoolsup | 28 | 24 | 25 | 21 | 8 | 8 |
| Famsup | 5 | 11 | 24 | 22 | 26 | 21 |
| Paid | 22 | 9 | 14 | 16 | 15 | 23 |
| Activities | 11 | 8 | 22 | 11 | 17 | 9 |
| Nursery | 15 | 6 | 8 | 23 | 23 | 24 |
| Higher | 26 | 2 | 10 | 26 | 16 | 17 |
| Internet | 19 | 10 | 20 | 25 | 24 | 26 |
| Romantic | 1 | 28 | 13 | 19 | 21 | 12 |
| Famrel | 7 | 3 | 19 | 14 | 12 | 14 |
| Freetime | 23 | 20 | 18 | 18 | 18 | 19 |
| Goout | 21 | 7 | 7 | 10 | 4 | 13 |
| Dalc | 18 | 12 | 16 | 30 | 30 | 27 |
| Walc | 29 | 14 | 15 | 24 | 27 | 6 |
| Health | 25 | 15 | 2 | 8 | 13 | 16 |
| Absences | 4 | 23 | 11 | 9 | 19 | 18 |

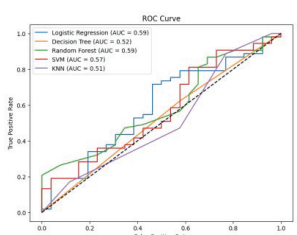
From Table 4, it can be seen that each model generates different ranks of the attributes. The reasons could be the assumption of each FS algorithm in the context of the nature of the features/ attributes of the data. Secondly, the subroutines used in Python do not consider pre-processing of data for the assumption (Sikri et al., 2023). This results in a wrong ranking in many cases

## Results of Five Classification Algorithms & Quality Indexes

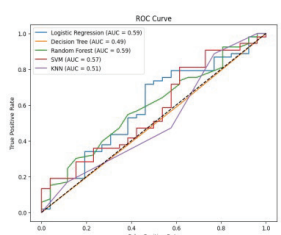
This section presents the performance of five classification algorithms in predicting the success or failure of the students. These algorithms are logistic Regression (LR), Decision Tree (DT), Random Forests (RF), Support Vector Machines (SVM), and K-nearest neighbor (KNN). Six quality parameters are used for making a comparison of the performance of five classification algorithms in conjunction with six hybrid FS methods. These quality parameters are (i) ACC, (ii) PRE, (iii) REC, (iv) F1_S, (v) AUCC, and (vi) ROCC. The values of these parameters for five classification algorithms without any FS are presented in Table 5. The results presented in Table 3 are treated as base model results and will be used for making comparisons when non-significant attributes are eliminated. The ROCCe of five classification models for six FS algorithms are shown in Fig. 1(a) to Fig. 1(f). It is worth mentioning that there is no role for the FS algorithm since all features are used for the purpose of classifying students into two categories, i.e., pass or fail. Evaluation matrices of Table 5 are considered only to bring symmetry in the presentation in relation to other results, wherein fewer numbers of features are selected to run classification algorithms, i.e., 20,15, and 10 attributes.

**Fig 1:**

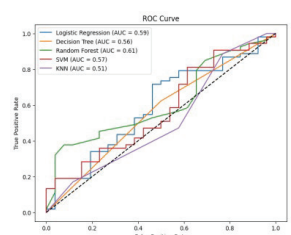
*ROC Curves of classification models for all Features*



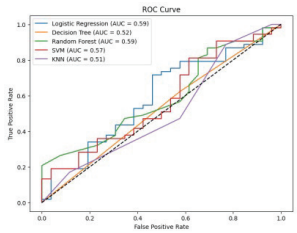

ROC Curve of Hybrid PSO Model
(c)




ROC Curve of Hybrid GA Model
(b)

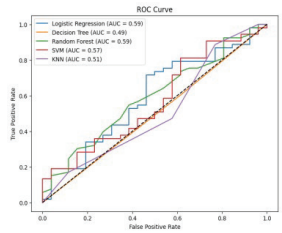



ROC Curve of Filter-Wrap Model
(a)



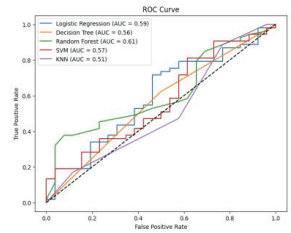

ROC Curve of Wrapper-Embedded Model

(d)




ROC Curve of Ensemble Model

(e)




ROC Curve of Filter-Embedded Model

(f)

**Table 5:**

*Evaluation Metrics for all features using different Classification Algorithms*

| **Hybrid PSO Model** | | | | | |
|---|---|---|---|---|---|
| **Algorithm Name** | **ACC** | **PRE** | **REC** | **F1_S** | **AUCC** |
| LR | 0.64557 | 0.686567 | 0.867925 | 0.766667 | 0.592888 |
| DT | 0.607595 | 0.729167 | 0.660377 | 0.693069 | 0.580189 |
| RFt | 0.632911 | 0.671429 | 0.886792 | 0.764228 | 0.59688 |
| SVM | 0.670886 | 0.675325 | 0.981132 | 0.800000 | 0.569666 |
| KNN | 0.670886 | 0.701493 | 0.886792 | 0.783333 | 0.508345 |
| **Hybrid GA Model** | | | | | |
| **Algorithm Name** | **ACC** | **PRE** | **REC** | **F1_S** | **AUCC** |
| LR | 0.64557 | 0.686567 | 0.867925 | 0.766667 | 0.592888 |
| DT | 0.582278 | 0.700000 | 0.660377 | 0.679612 | 0.541727 |
| RF | 0.683544 | 0.712121 | 0.886792 | 0.789916 | 0.595428 |
| SVM | 0.670886 | 0.675325 | 0.901132 | 0.800000 | 0.570029 |
| KNN | 0.670886 | 0.701493 | 0.886792 | 0.783333 | 0.508345 |
| **Filter Wrap Method** | | | | | |
| **Algorithm Name** | **ACC** | **PRE** | **REC** | **F1_S** | **AUCC** |
| LR | 0.64557 | 0.686567 | 0.867925 | 0.766666748 | 0.592888 |
| DT | 0.594937 | 0.714286 | 0.660377 | 0.686274448 | 0.560958 |
| RF | 0.658228 | 0.691176 | 0.886792 | 0.776859033 | 0.66582 |
| SVM | 0.670886 | 0.675325 | 0.911132 | 0.775703619 | 0.569666 |
| KNN | 0.670886 | 0.701493 | 0.876792 | 0.779407332 | 0.508345 |

| **Wrapper-Embedded Method** | | | | | |
|---|---|---|---|---|---|
| **Algorithm Name** | **ACC** | **PRE** | **REC** | **F1_S** | **AUCC** |
| LR | 0.64557 | 0.686567 | 0.867925 | 0.766666748 | 0.592888 |
| DT | 0.582278 | 0.708333 | 0.641509 | 0.673266937 | 0.551524 |
| RF | 0.683544 | 0.705882 | 0.90566 | 0.793388062 | 0.582003 |
| SVM | 0.670886 | 0.675325 | 0.901132 | 0.772056539 | 0.569666 |
| KNN | 0.670886 | 0.701493 | 0.826792 | 0.759006076 | 0.508345 |
| **Ensemble-Based Method** | | | | | |
| **Algorithm Name** | **ACC** | **PRE** | **REC** | **F1_S** | **AUCC** |
| LR | 0.64557 | 0.686567 | 0.867925 | 0.766667 | 0.592888 |
| DT | 0.531646 | 0.666667 | 0.603774 | 0.633664 | 0.494194 |
| RF | 0.658228 | 0.691176 | 0.886792 | 0.776859 | 0.643324 |
| SVM | 0.670886 | 0.675325 | 0.981132 | 0.800000 | 0.569666 |
| KNN | 0.670886 | 0.701493 | 0.886792 | 0.783333 | 0.508345 |
| **Filter-Embedded Method** | | | | | |
| **Algorithm Name** | **ACC** | **PRE** | **REC** | **F1_S** | **AUCC** |
| LR | 0.64557 | 0.677419 | 0.792453 | 0.730435 | 0.592888 |
| DT | 0.556962 | 0.686567 | 0.867925 | 0.766667 | 0.522859 |
| RF | 0.620253 | 0.687500 | 0.622642 | 0.653466 | 0.604862 |
| SVM | 0.670886 | 0.682540 | 0.811321 | 0.741380 | 0.569666 |
| KNN | 0.670886 | 0.675325 | 0.931132 | 0.782862 | 0.508345 |

## Ranks of Top 20 attributes as identified by six hybrid FS algorithms

This section presents a list of the 20 most important attributes as identified by six hybrid FS algorithms. It can be seen that different features are identified by six different hybrid FS algorithms as given in Table 6. Justification for the variation of ranks is given in section 4.2.

**Table 6:**

*Ranks of Top 20 attributes with Hybrid Methods*

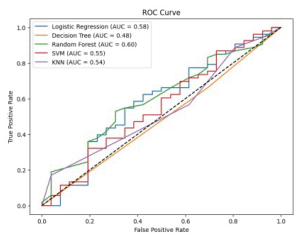
| **By Using Hybrid PSO Method** | | **Hybrid GA Method** | | **Filter-Wrap Method** | | **Wrapper-Embedded Method** | | **Ensemble Method** | | **Filter-Embedded Method** | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Feature | Rank | Feature | Rank | Feature | Rank | Feature | Rank | Feature | Rank | Feature | Rank |
| Age | 15 | School | 16 | School | 19 | Sex | 16 | Age | 16 | School | 18 |
| Address | 19 | Sex | 17 | Age | 2 | Age | 11 | Address | 20 | Sex | 16 |
| Famsize | 4 | Age | 1 | Address | 14 | Address | 18 | Famsize | 18 | Age | 12 |
| Pstatus | 12 | Address | 2 | Pstatus | 12 | Medu | 4 | Medu | 3 | Famsize | 13 |
| Medu | 8 | Famsize | 18 | Medu | 16 | Fedu | 2 | Fedu | 4 | Pstatus | 19 |
| Mjob | 6 | Pstatus | 3 | Fedu | 9 | Mjob | 1 | Mjob | 2 | Medu | 2 |
| Fjob | 9 | Medu | 19 | Mjob | 20 | Fjob | 5 | Fjob | 7 | Mjob | 1 |
| Reason | 7 | Fedu | 20 | Fjob | 18 | Reason | 3 | Reason | 5 | Guardian | 3 |
| Guardian | 10 | Mjob | 4 | Reason | 17 | Guardian | 7 | Guardian | 6 | Studytime | 15 |
| Traveltime | 20 | Traveltime | 5 | Traveltime | 8 | Studytime | 9 | Studytime | 10 | Failures | 4 |
| Failures | 1 | Studytime | 6 | Studytime | 4 | Failures | 6 | Failures | 1 | Schoolsup | 14 |
| Paid | 3 | Failures | 7 | Failures | 3 | Famsup | 13 | Famsup | 11 | Famsup | 10 |
| Activities | 13 | Schoolsup | 8 | Schoolsup | 5 | Paid | 8 | Paid | 9 | Paid | 9 |
| Nursery | 18 | Famsup | 9 | Activities | 13 | Activities | 14 | Activities | 17 | Activities | 6 |
| Internet | 16 | Activities | 10 | Nursery | 10 | Romantic | 15 | Romantic | 13 | Higher | 17 |
| Romantic | 2 | Internet | 11 | Higher | 15 | Famrel | 19 | Famrel | 19 | Romantic | 5 |
| Freetime | 14 | Freetime | 12 | Famrel | 11 | Freetime | 20 | Freetime | 12 | Famrel | 8 |
| Goout | 5 | Goout | 13 | Dalc | 6 | Goout | 12 | Goout | 15 | Freetime | 11 |
| Health | 17 | Dalc | 14 | Walc | 7 | Health | 17 | Health | 14 | Goout | 7 |
| Absences | 11 | Health | 15 | Health | 1 | Absences | 10 | Absences | 8 | Dalc | 20 |

## ROC Curves and Evaluation Matrix of Classification Models in Conjunction With Hybrid FS Models For Top 20 Features/ Attributes Of Data.

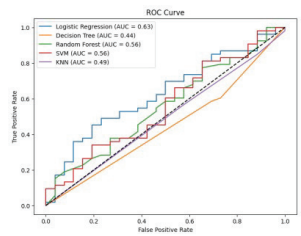
The results of five classifiers for the top 20 attributes as identified by different hybrid FS models are given in Table 7, and a graphical presentation of the ROC curve is shown in Figure 2 (a-f). Based on the values of accuracy, precision, recall, F1 Score, and AUC, and the ROC curve, the following can be inferred.

**Fig 2:**

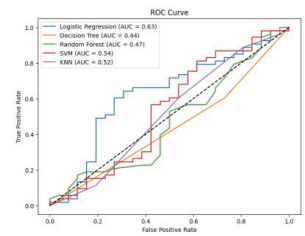
*ROC Curves of five classification models and six hybrid FS with top 20 Features*



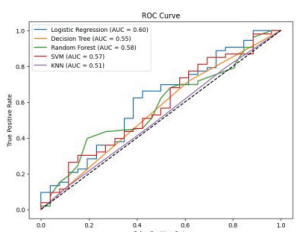

ROCC of Hybrid PSO Model
(a)



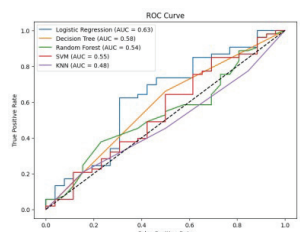

ROCC of Hybrid GA Model
(b)



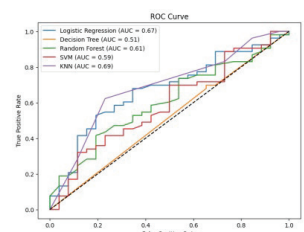

ROCC of Filter-Wrap Model
(c)




ROCC of Wrapper-Embedded Model
(d)




ROCC of Ensemble Model
(e)




ROCC of Filter-Embedded Model
(f)

It can be seen from the quality parameters given in Tables 5 and 7 that the values of precision, recall, and F1-score for the 20 top attributes are slightly higher in comparison to when all attributes are considered by the classification models. Therefore, it can be concluded that removing non-significant attributes not only improves the performance of classification models but also reduces the need for computing power. In addition, it reduces the need for capturing more data in the future (Dahiya et al., 2017; Singh & Singh, 2019). Secondly, the ROC curve had

improved for two FS methods, i.e., wrapper-embedded model and filter-embedded algorithms. It is evident from making a comparison between fig 1(d) and 2(d), and fig1(f) and 2(f).

**Table 7:**

*Evaluation Metrics for Top 20 features using different Classification Algorithms*

| **Hybrid PSO Model** | | | | | |
|---|---|---|---|---|---|
| **Algorithm** | **ACC** | **PRE** | **REC** | **F1_S** | **AUCC** |
| LR | 0.658228 | 0.703125 | 0.879057 | 0.781310 | 0.582729 |
| DT | 0.544304 | 0.670377 | 0.660665 | 0.665486 | 0.484035 |
| RF | 0.658228 | 0.703125 | 0.889057 | 0.785235 | 0.600508 |
| SVM | 0.680126 | 0.680000 | 0.972264 | 0.800283 | 0.552975 |
| KNN | 0.698228 | 0.721176 | 0.896762 | 0.799441 | 0.535559 |
| **Hybrid GA Model** | | | | | |
| **Algorithm** | **ACC** | **PRE** | **REC** | **F1_S** | **AUCC** |
| LR | 0.632911 | 0.691429 | 0.886792 | 0.777019 | 0.628447 |
| DT | 0.493671 | 0.692653 | 0.684906 | 0.688758 | 0.441219 |
| RF | 0.632911 | 0.728752 | 0.890189 | 0.801422 | 0.556967 |
| SVM | 0.698228 | 0.680556 | 0.924528 | 0.784001 | 0.558055 |
| KNN | 0.687595 | 0.711765 | 0.899057 | 0.794523 | 0.486575 |
| **Filter Wrap Method** | | | | | |
| **Algorithm** | **ACC** | **PRE** | **REC** | **F1_S** | **AUCC** |
| LR | 0.645577 | 0.696056 | 0.90566 | 0.787144 | 0.626996 |
| DT | 0.493671 | 0.717451 | 0.673774 | 0.694928 | 0.436502 |
| RF | 0.607595 | 0.737419 | 0.832453 | 0.782060 | 0.469158 |
| SVM | 0.670886 | 0.675325 | 0.961132 | 0.793271 | 0.535559 |
| KNN | 0.658228 | 0.711176 | 0.886792 | 0.789334 | 0.516328 |
| **Wrapper-Embedded Method** | | | | | |
| **Algorithm** | **ACC** | **PRE** | **REC** | **F1_S** | **AUCC** |
| LR | 0.670886 | 0.686567 | 0.867925 | 0.766666748 | 0.603774 |
| DT | 0.607595 | 0.708333 | 0.641509 | 0.673266937 | 0.550798 |
| RF | 0.645570 | 0.705882 | 0.905660 | 0.793388062 | 0.575472 |
| SVM | 0.670886 | 0.675325 | 0.901132 | 0.772056539 | 0.574746 |
| KNN | 0.607595 | 0.701493 | 0.826792 | 0.759006076 | 0.513425 |

| **Ensemble-Based Method** | | | | | |
|---|---|---|---|---|---|
| **Algorithm** | **ACC** | **PRE** | **REC** | **F1_S** | **AUCC** |
| LR | 0.683544 | 0.705882 | 0.905660 | 0.793388 | 0.62627 |
| DT | 0.607595 | 0.729167 | 0.660377 | 0.693069 | 0.580189 |
| RF | 0.645570 | 0.686567 | 0.867925 | 0.766667 | 0.538099 |
| SVM | 0.683544 | 0.684211 | 0.981132 | 0.806202 | 0.545718 |
| KNN | 0.569620 | 0.690794 | 0.873585 | 0.771510 | 0.478955 |
| **Filter-Embedded Method** | | | | | |
| **Algorithm** | **ACC** | **PRE** | **REC** | **F1_S** | **AUCC** |
| LR | 0.658228 | 0.685714 | 0.905660 | 0.780487 | 0.671263 |
| DT | 0.569620 | 0.679245 | 0.679245 | 0.679245 | 0.510160 |
| RF | 0.632911 | 0.714286 | 0.754717 | 0.733945 | 0.607402 |
| SVM | 0.645570 | 0.676056 | 0.905660 | 0.774193 | 0.590711 |
| KNN | 0.645570 | 0.698413 | 0.830189 | 0.758621 | 0.686865 |

## ROCC and Evaluation Metrics of Classification Models in Conjunction With Top 15 Features Identified by Hybrid FS Models

This section presents the top 15 attributes of the data set which are a high association with the target variable. The attributes are shown in Table 8. It can be seen that there are differences in the ranks of the attributes due to inherent weaknesses of the black box approach of computing using the built-in Python subroutine. Secondly, all methods cannot be applied to all kinds of datasets, which is an inherent weakness of machine learning algorithms.

**Table 8:**

*Name and Ranks of Top 15 attributes as identified with six Hybrid Methods*

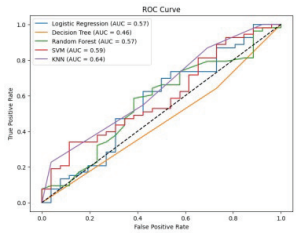
| **By Using Hybrid PSO Method** | | **By Using Hybrid GA Method** | | **By Using Filter-Wrap Method** | | **By Using Wrapper-Embedded Method** | | **By Using Ensemble Method** | | **By Using Filter-Embedded Method** | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Feature | Rank | Feature | Rank | Feature | Rank | Feature | Rank | Feature | Rank | Feature | Rank |
| Sex | 12 | Address | 1 | Age | 2 | Age | 11 | Medu | 3 | School | 13 |
| Famsize | 3 | Famsize | 2 | Address | 14 | Medu | 4 | Fedu | 4 | Sex | 9 |
| Pstatus | 11 | Pstatus | 3 | Pstatus | 12 | Fedu | 2 | Mjob | 2 | Age | 11 |
| Medu | 8 | Medu | 4 | Fedu | 9 | Mjob | 1 | Fjob | 7 | Famsize | 6 |
| Mjob | 14 | Fjob | 5 | Travel-time | 8 | Fjob | 5 | Reason | 5 | Guard-ian | 1 |
| Fjob | 9 | Guard-ian | 6 | Study-time | 4 | Reason | 3 | Guard-ian | 6 | Study-time | 7 |
| Reason | 6 | Study-time | 7 | Failures | 3 | Guard-ian | 7 | Study-time | 10 | Failures | 8 |
| Guard-ian | 7 | Failures | 8 | School-sup | 5 | Study-time | 9 | Failures | 1 | School-sup | 12 |
| Study-time | 15 | School-sup | 9 | Activi-ties | 13 | Failures | 6 | Famsup | 11 | Famsup | 3 |
| Failures | 1 | Paid | 10 | Nursery | 10 | Famsup | 13 | Paid | 9 | Paid | 4 |
| Paid | 2 | Activi-ties | 11 | Higher | 15 | Paid | 8 | Roman-tic | 13 | Higher | 14 |
| Activi-ties | 10 | Higher | 12 | Famrel | 11 | Activi-ties | 14 | Freetime | 12 | Roman-tic | 5 |
| Roman-tic | 4 | Roman-tic | 13 | Dalc | 6 | Roman-tic | 15 | Goout | 15 | Famrel | 10 |
| Goout | 5 | Famrel | 14 | Walc | 7 | Goout | 12 | Health | 14 | Goout | 2 |
| Absenc-es | 13 | Free-time | 15 | Health | 1 | Absenc-es | 10 | Absences | 8 | Dalc | 15 |

The ROCC for five classification algorithms, i.e., LR, RF, DT, SVM, and KNN, along with different Hybrid FS algorithms wherein the top 15 features of the dataset are selected for analysis, are shown in Fig. 3(a-f).

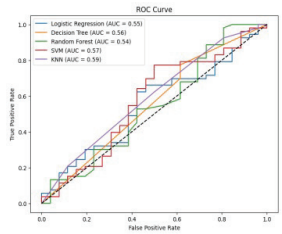
wing can be inferred.

**Fig 3:**

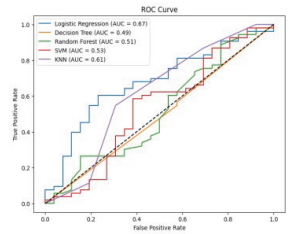
*ROCC of five classification models and six hybrid FS with the top 15 Features*



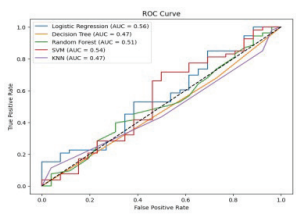

ROCC of Hybrid PSO Model
(a)



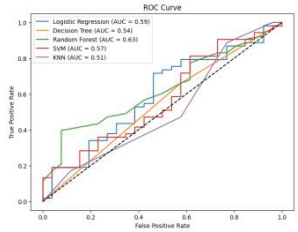

ROCC of Hybrid GA Model
(b)



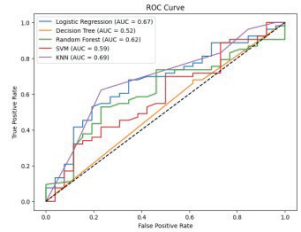

ROCC of Filter-Wrap Model
(c)




ROCC of Wrapper-Embedded Model
(d)




ROCC of Ensemble Model
(e)




ROCC of Filter-Embedded Model
(f)

The evaluation matrices with the top 15 ranked attributes are given in Table 9. It can be seen from the values of quality parameters that in the case of the hybrid PSO model of the FS algorithm, there is a slight improvement in the performance of the SVM classification algorithm. For the Hybrid GA model of FS, the performance of the decision tree has improved, but AUCC did not indicate better performance. Similar changes are seen for other FS models. ROCC had improved only for the filter-ensemble algorithm, as evident from Fig. 1(f), Fig. 2 (f), and Fig. 3(f).

**Table 9:**

*Evaluation Metrics for Top 15 features of five Classification and six hybrid FS Algorithms*

| **Hybrid PSO Model** | | | | | |
|---|---|---|---|---|---|
| **Algorithm Name** | **ACC** | **PRE** | **REC** | **F1_S** | **AUCC** |
| LR | 0.658228 | 0.696970 | 0.867925 | 0.773109 | 0.570392 |
| DT | 0.518987 | 0.641509 | 0.641509 | 0.641509 | 0.455372 |
| RF | 0.607595 | 0.677419 | 0.792453 | 0.730435 | 0.568215 |
| SVM | 0.683444 | 0.689189 | 0.962264 | 0.803155 | 0.592888 |
| KNN | 0.683544 | 0.718755 | 0.867925 | 0.786325 | 0.636067 |
| **Hybrid GA Model** | | | | | |
| **Algorithm** | **ACC** | **PRE** | **REC** | **F1_S** | **AUCC** |
| LR | 0.64557 | 0.676056 | 0.90566 | 0.774194 | 0.549347 |
| DT | 0.594937 | 0.698113 | 0.698113 | 0.698113 | 0.555878 |
| RF | 0.670886 | 0.707692 | 0.867925 | 0.779661 | 0.541364 |
| SVM | 0.658228 | 0.680556 | 0.924528 | 0.784444 | 0.573295 |
| KNN | 0.683544 | 0.700000 | 0.924528 | 0.796748 | 0.594702 |
| **Filter Wrap Method** | | | | | |
| **Algorithm** | **ACC** | **PRE** | **REC** | **F1_S** | **AUCC** |
| LR | 0.658228 | 0.680556 | 0.924528 | 0.784444 | 0.674891 |
| DT | 0.518987 | 0.666667 | 0.566038 | 0.612245 | 0.489115 |
| RF | 0.632911 | 0.687555 | 0.830189 | 0.752137 | 0.513062 |
| SVM | 0.658228 | 0.680556 | 0.924528 | 0.784444 | 0.533382 |
| KNN | 0.683544 | 0.718755 | 0.867925 | 0.786325 | 0.61357 |
| **Wrapper-Embedded Method** | | | | | |
| **Algorithm** | **ACC** | **PRE** | **REC** | **F1_S** | **AUCC** |
| LR | 0.64557 | 0.692308 | 0.849057 | 0.762712 | 0.558055 |
| DT | 0.544304 | 0.654545 | 0.679245 | 0.666667 | 0.474238 |
| RF | 0.620253 | 0.676923 | 0.830189 | 0.745763 | 0.508345 |
| SVM | 0.683544 | 0.684211 | 0.981132 | 0.806202 | 0.544993 |
| KNN | 0.594937 | 0.652174 | 0.849057 | 0.737705 | 0.473149 |

| **Ensemble-Based Method** | | | | | |
|---|---|---|---|---|---|
| **Algorithm** | **ACC** | **PRE** | **REC** | **F1_S** | **AUCC** |
| LR | 0.64557 | 0.686567 | 0.867925 | 0.766667 | 0.592888 |
| DT | 0.582278 | 0.723431 | 0.660377 | 0.690467 | 0.541727 |
| RF | 0.64557 | 0.686567 | 0.867925 | 0.766667 | 0.62881 |
| SVM | 0.670886 | 0.675325 | 0.981132 | 0.800000 | 0.570029 |
| KNN | 0.670886 | 0.701493 | 0.886792 | 0.783333 | 0.508345 |
| **Filter-Embedded Method** | | | | | |
| **Algorithm** | **ACC** | **PRE** | **REC** | **F1_S** | **AUCC** |
| LR | 0.658228 | 0.685714 | 0.935660 | 0.791422 | 0.671263 |
| DT | 0.569621 | 0.686275 | 0.660377 | 0.673077 | 0.519956 |
| RF | 0.607595 | 0.689655 | 0.754717 | 0.720721 | 0.617562 |
| SVM | 0.645571 | 0.676056 | 0.905667 | 0.774196 | 0.590711 |
| KNN | 0.645572 | 0.698413 | 0.830189 | 0.758621 | 0.686865 |

## ROC Curves and Evaluation Metrics of Classification Models in Conjunction With Top 10 Features Identified by Hybrid FS Models.

The ranks of the top 10 attributes with hybrid six FS algorithms are given in Table 10. It was expected that the top 10 out of 31 attributes would have more similarities in the ranks, but in this case, the different attributes are also identified as top 10 attributes by these six algorithms. The ROC curve for five classification algorithms along with six hybrid FS methods is given in Fig. 4 (a-f).

**Table 8:**

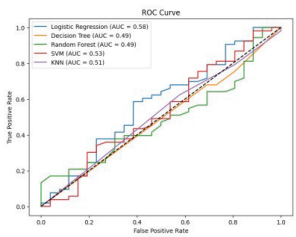
*Showing Ranks of Top 10 attributes with Hybrid Methods*

| **Hybrid PSO Method** | | **Hybrid GA Method** | | **Filter-Wrap Method** | | **Wrapper-Embedded Method** | | **Ensemble Method** | | **Filter-Embedded Method** | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Feature | Rank | Feature | Rank | Feature | Rank | Feature | Rank | Feature | Rank | Feature | Rank |
| Sex | 6 | Sex | 1 | Famsize | 3 | Medu | 4 | Medu | 3 | Sex | 6 |
| Famsize | 3 | Address | 2 | Fedu | 2 | Fedu | 2 | Fedu | 4 | Address | 7 |
| Mjob | 7 | Pstatus | 3 | Fjob | 10 | Mjob | 1 | Mjob | 2 | Famsize | 5 |
| Reason | 10 | Fedu | 4 | Reason | 7 | Fjob | 5 | Fjob | 7 | Mjob | 1 |
| Failures | 1 | Mjob | 5 | Travel-time | 8 | Reason | 3 | Reason | 5 | Failures | 2 |

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Paid | 2 | Fjob | 6 | Study-time | 4 | Guard-ian | 7 | Guard-ian | 6 | Activi-ties | 3 |
| Activi-ties | 8 | Reason | 7 | Failures | 5 | Study-time | 9 | Study-time | 10 | Famrel | 4 |
| Roman-tic | 4 | Guard-ian | 8 | Famsup | 9 | Failures | 6 | Failures | 1 | Free-time | 8 |
| Goout | 5 | Study-time | 9 | Paid | 6 | Paid | 8 | Paid | 9 | Goout | 9 |
| Health | 9 | Failures | 10 | Roman-tic | 1 | Absenc-es | 10 | Absences | 8 | Dalc | 10 |

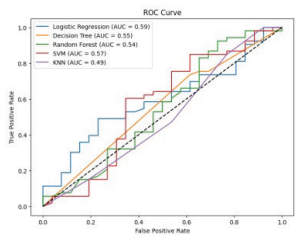
**Fig 4:**

*ROCC of five classification models and six hybrid FS with top 10 Features*



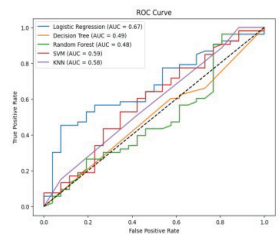

ROCC of Hybrid PSO Model

(a)



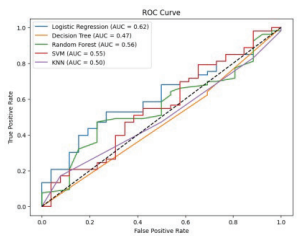

ROCC of Hybrid GA Model

(b)



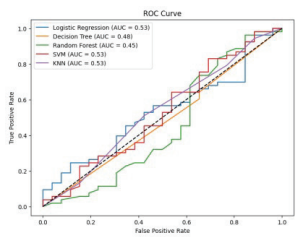

ROCC of Filter-Wrap Model

(c)



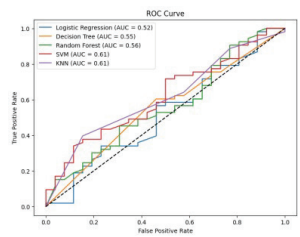

ROCC of Wrapper-Embedded Model

(d)




ROCC of Ensemble Model

(e)




ROCC of Filter-Embedded Model

(f)

It is evident from Figure 4(a-f) that the area under the ROCC curve has reduced when 15 or 20 attributes are considered for classification purposes. In addition, in many cases, the classification algorithm curves are below the straight line. This indicates that the performance of classification algorithms had not improved with FS for a small set of 10 features in place of 15, 20, or all features. Similar trends are exhibited by other parameters calculated and given in Table 11 for making a comparison in the performance of the five classification models.

**Table 11:**

*Evaluation Metrics for Top 10 features using five Classification Algorithms and six hybrid FS methods*

| **Hybrid PSO Model** | | | | | |
|---|---|---|---|---|---|
| **Algorithm** | **ACC** | **PRE** | **REC** | **F1_S** | **AUCC** |
| LR | 0.670886 | 0.701493 | 0.886792 | 0.783333 | 0.575472 |
| DT | 0.544304 | 0.654545 | 0.679245 | 0.666667 | 0.485123 |
| RF | 0.556962 | 0.645161 | 0.754717 | 0.695652 | 0.489115 |
| SVM | 0.670886 | 0.684932 | 0.943396 | 0.793651 | 0.527576 |
| KNN | 0.594937 | 0.666667 | 0.792453 | 0.724138 | 0.512337 |
| **Hybrid GA Model** | | | | | |
| **Algorithm** | **ACC** | **PRE** | **REC** | **F1_S** | **AUCC** |
| LR | 0.645578 | 0.701493 | 0.886792 | 0.783333 | 0.589623 |
| DT | 0.620253 | 0.654545 | 0.679245 | 0.666666 | 0.554064 |
| RF | 0.620253 | 0.645161 | 0.754717 | 0.695652 | 0.542091 |
| SVM | 0.658228 | 0.684932 | 0.943396 | 0.793651 | 0.568578 |
| KNN | 0.648576 | 0.666667 | 0.792453 | 0.724138 | 0.493832 |
| **Filter Wrap Method** | | | | | |
| **Algorithm Name** | **ACC** | **PRE** | **REC** | **F1_S** | **AUCC** |
| LR | 0.658228 | 0.680556 | 0.924528 | 0.784001 | 0.674528 |
| DT | 0.544304 | 0.673469 | 0.622642 | 0.647059 | 0.49492 |
| RF | 0.582278 | 0.666667 | 0.754717 | 0.707965 | 0.484761 |
| SVM | 0.658228 | 0.680556 | 0.894528 | 0.773008 | 0.585269 |
| KNN | 0.658228 | 0.691176 | 0.886792 | 0.776859 | 0.57656 |

| **Wrapper-Embedded Method** | | | | | |
|---|---|---|---|---|---|
| **Algorithm Name** | **ACC** | **PRE** | **REC** | **F1_S** | **AUCC** |
| LR | 0.64557 | 0.676056 | 0.885661 | 0.766792 | 0.61611 |
| DT | 0.518987 | 0.647059 | 0.622642 | 0.634616 | 0.46807 |
| RF | 0.544304 | 0.644068 | 0.716981 | 0.678571 | 0.556967 |
| SVM | 0.683544 | 0.649189 | 0.912264 | 0.758565 | 0.552975 |
| KNN | 0.594937 | 0.656716 | 0.830189 | 0.733333 | 0.497097 |
| **Ensemble-Based Method** | | | | | |
| **Algorithm Name** | **ACC** | **PRE** | **REC** | **F1_S** | **AUCC** |
| LR | 0.658228 | 0.660556 | 0.904528 | 0.763526 | 0.530842 |
| DT | 0.518987 | 0.653061 | 0.603774 | 0.627451 | 0.481495 |
| RF | 0.607595 | 0.696429 | 0.735849 | 0.715597 | 0.448839 |
| SVM | 0.683544 | 0.689189 | 0.912264 | 0.785191 | 0.525036 |
| KNN | 0.607595 | 0.677419 | 0.792453 | 0.730435 | 0.534107 |
| **Filter-Embedded Method** | | | | | |
| **Algorithm Name** | **ACC** | **PRE** | **REC** | **F1_S** | **AUCC** |
| LR | 0.658228 | 0.620556 | 0.874528 | 0.725971 | 0.520682 |
| DT | 0.556962 | 0.687555 | 0.622642 | 0.65349 | 0.553701 |
| RF | 0.645571 | 0.698413 | 0.810189 | 0.75016 | 0.559507 |
| SVM | 0.658228 | 0.610556 | 0.874528 | 0.719082 | 0.614296 |
| KNN | 0.670886 | 0.701493 | 0.846792 | 0.767325 | 0.612482 |

It can be seen from the results presented in the earlier sections that the performance of five classification algorithms in conjunction with six hybrid FS algorithms has improved when the top 20 attributes are used. This is based on the values of quality parameters, i.e., precision, recall, and F1-score. Similar trends were seen for many combinations of classification and FS algorithms when the top 15 attributes are used in computation. Further, a reduction in the number of features did not result in an improvement in the goodness of fit of the classification algorithms. The results of ROCC did not indicate improvement in forecasting success of classifiers with a reduction in the number of features, except for when 20 features are selected for the purpose of classification. Accuracy is not always a good measure of the goodness of fit of models; therefore, it is not emphasized in making conclusions. However, there is a need to rerun the algorithm by changing the process of discretization.

## Conclusion

It is always desirable to use FS algorithms before implementing machine learning/supervised classifiers, which enhances the values of goodness-of-fit parameters of models. However, it is right to pre-process the features of data as per the assumption/requirement of a given algorithm. Sikri et al. (2023) had shown a significant difference for the chi-square FS method in the rank of features when computation is done with prior right pre-processing of data with respect to the chi-square method of FS. It is, therefore, necessary to optimize even a simple operation, such as discretization or any other assumptions of the algorithm used for either FS or classification. A discrete predictor with four categories, when used with two categories (if needed as per the assumption of the algorithm), makes not only a difference in the ranking of the attributes but also impacts the performance parameters of classification models. Similarly, an artificial neural network as a predictor used with weights and biases between [-1 to +1] may give different results when the limit of inputs is changed to [0 to +1] or in any other range.

The summary of the best combination of the hybrid FS method and supervised classifiers is presented in Table 12. The results are the best of the ROC curve. It can, therefore, be concluded that Logistic Regression and the Filter wrapper method of FS are the best.

**Table 12:**

*Combination of Feature Selection model and classifier with the Highest ROC*

| No. of Predictors/ Features | Classifiers | Hybrid FS Algorithm | Value of ROC Curve | Comments |
|---|---|---|---|---|
| 30 (All) | Random Forest | Ensemble | 0.64 | Many Cases curve is below straight line |
| 20 | Logistic | Filter-Embedded | 0.67 | -do- |
| 15 | Logistic | Filter-Wrapper | 0.67 | -do- |
| 10 | Logistic | Filter Wrapper | 0.67 | -do- |

## REFERENCES

Ahonen, L., Cowley, B. U., Hellas, A., & Puolamaki, K. (2018). Biosignals reflect pair-dynamics in collaborative work: EDA and ECG study of pair-programming in a classroom environment. *Scientific Reports*, *8*, 3138. https://doi.org/10.1038/s41598-018-21248-1

Alberto, R., Alfonso, G. B., Guillermo, H., Javier, P., & Pablo, C. (2021). Artificial neural network analysis of the academic performance of students in virtual learning environments. *Neurocomputing*, *423*, 713–720. https://doi.org/10.1016/j.neucom.2020.08.105.

Alisha Sikri, N. P. Singh, & Surjeet Dalal (2023)**.** Chi-square method of feature selection: Impact of pre-processing of data**,** International Journal of Intelligent Systems and Applications in Engineering**,** 11(3s): 241-248.

Aluko, R. O., Daniel, E. I., Oshodi, O. S., Aigbavboa, C. O., & Abisuga, A. O. (2018). Towards reliable prediction of academic performance of architecture students using data mining techniques. *Journal of Engineering Design and Technology*, *16*(3), 385–397. https://doi.org/10.1108/JEDT-12-2017-0117.

Antoniou, P. E., Arfaras, G., Pandria, N., Athanasiou, A., Ntakakis, G., Babatsikos, E., & Bamidis, P. (2020). Biosensor real-time affective analytics in virtual and mixed reality medical education serious games: Cohort study. *JMIR Serious Games*, *8*(1), e17823. https://doi.org/10.2196/17823.

Asif, R., Merceron, A., Ali, S. A., & Haider, N. (2017). Analyzing undergraduate students' performance using educational data mining. *Computers & Education*, *113*, 177–194. https://doi.org/10.1016/j.compedu.2017.05.005.

Aulck, L., Velagapudi, N., Blumenstock, J., & West, J. (2016). Predicting student dropout in higher education. In *Proceedings of the ICML Workshop on #Data4Good: Machine Learning in Social Good Applications*, New York, NY, USA, 24 June 2016. https://arxiv.org/pdf/1606.06364.pdf

Awoyemi, J. O., Adetunmbi, A. O., & Oluwadare, S. A. (2017). Credit card fraud detection using machine learning techniques: A comparative analysis. In *Proceedings of the International Conference on Computing Networking and Informatics (ICCNI)*, Lagos, Nigeria, 29–31 October 2017.

Azizah, E. N., Pujianto, U., & Nugraha, E. (2018). Comparative performance between C4.5 and Naive Bayes classifiers in predicting student academic performance in a virtual learning environment. In *Proceedings of the 4th International Conference on Education and Technology (ICET)*, Malang, Indonesia, 26–28 October 2018 (pp. 18–22).

Baker, N. K., D'Mello, S., Ocumpaugh, J., Baker, R., & Shute, V. (2016). Using video to automatically detect learner affect in computer-enabled classrooms. *ACM Transactions on Interactive Intelligent Systems*, *6*(2), 1–26. https://doi.org/10.1145/2908190

Bohong, Y.; Zeping, Y.; Hong, L.; Yaqian, Z.; Jinkai, X. In-classroom learning analytics based on student behavior, topic and teaching characteristic mining. Pattern Recognit. Lett. 2020, 129, 224–231.

Byung-Hak, K., Ethan, V., & Ganapathi, V. (2018). GritNeikikik8t: Student performance prediction with deep learning. *arXiv*. https://arxiv.org/abs/1804.07405.

Cetintas, S., Si, L., Xin, P., & Hord, C. (2010). Automatic detection of off-task behaviors in intelligent tutoring systems with machine learning techniques. *IEEE Transactions on Learning Technologies*, *3*(3), 228–236. https://doi.org/10.1109/TLT.2010.29.

Chadrashekar, G., & Sahni, F. (2014). A survey on feature selection methods. Computers and Electrical Engineering, 40(1), 16-28. DOI: 10.1016/j.compeleng.2013.11.04.

Dahiya, Shashi; S. S. Handa, and N P Singh (2017). A feature selection enabled hybrid bagging algorithm for credit risk evaluation, Expert Systems, 34 (6): 1-11

Dash,M. & Liu, H. (1997). Feature selection for classification, Intelligent Data Analysis, 1(1), 131-156. DOI: 10.1016/s1088-467X(97)00008-5.

Daud, A., Aljohani, N. R., Abbasi, R. A., Lytras, M. D., Abbas, F., & Alowibdi, J. S. (2017). Predicting student performance using advanced learning analytics. In *Proceedings of the 26th International Conference on World Wide Web Companion*, Perth, Australia, 3–7 April 2017. https://doi.org/10.1145/3038912.3052588.

Farzaneh, A., & Fadlalla, A. (2017). Data mining applications in accounting: A review of the literature and organizing framework. *International Journal of Accounting Information Systems*, *24*, 32–58. https://doi.org/10.1016/j.accinf.2017.02.002

Fernandes, E. Holanda, M., Victorino, M., Borges, V., Carvalho, R., Van Erven, G. (2019). Educational data mining: Predictive analysis of academic performance of public school students in the capital of Brazil, Journal of Business Research, 94, 335-343.

Francis, B. K., & Babu, S. S. (2019). Predicting academic performance of students using a hybrid data mining approach. *Journal of Medical Systems*, *43*(6), 162. https://doi.org/10.1007/s10916-019-1350-7.

Garg, R. (2018). Predicting student performance of different regions of Punjab using classification techniques. *International Journal of Advanced Research in Computer Science*, *9*(5), 236–241. https://doi.org/10.26483/ijarcs.v9i5.6031.

Ghosh, I., & Chaudhuri, D.T. (2021). FEB-stacking and FEB-DNN models for stock trend prediction: A performance analysis for pre and post COVID-19 periods. *Decision Making: Applications in Management and Engineering*, *4*, 51–84. https://doi.org/10.5267/j.dmae.2021.2.004

Goldberg, P., Sümer, O., & Stürmer, K. (2021). Attentive or not? Toward a machine learning approach to assessing students' visible engagement in classroom instruction. *Educational Psychology*, *33*(1), 27–49. https://doi.org/10.1080/01443410.2021.1897027

Guyon, I., & Elisseeff, A. (2003). An introduction to variable and feature selection, Journal of Machine Learning Research, 3.1157-1182.

Haiyang, L., Wang, Z., Benachour, P., & Tubman, P. (2018). A time series classification method for behaviour-based dropout prediction. In *Proceedings of the IEEE 18th International Conference on Advanced Learning Technologies (ICALT)*, Mumbai, India, 9–13 July 2018. https://doi.org/10.1109/ICALT.2018.00046.

Hasan, R., Palaniappan, S., Mahmood, S., Abbas, A., Sarker, K. U., & Sattar, M. U. (2020). Predicting student performance in higher educational institutions using video learning analytics and data mining techniques. *Applied Sciences*, *10*(11), 3894. https://doi.org/10.3390/app10113894

Helal, S., Jiuyong, L., Lin, L., Esmaeil, E., Shane, D., Duncan, M., & Qi, L. (2018). Predicting academic performance by considering student heterogeneity. *Knowledge-Based Systems*, *161*, 134–146. https://doi.org/10.1016/j.knosys.2018.07.028.

Heuer, H., & Breiter, A. (2018). Student success prediction and the trade-off between big data and data minimization. In *DeLFI 2018—Die 16. E-Learning Fachtagung Informatik* (pp. 219–230). Gesellschaft für Informatik, Bonn.

Hlosta, M., Zdrahal, Z., & Zendulka, J. (2017). Ouroboros: Early identification of at-risk students without models based on legacy data. In *Proceedings of the Seventh International Learning Analytics & Knowledge Conference* (pp. 6–15), New York, NY, USA, 13–17 March 2017. Association for Computing Machinery. https://doi.org/10.1145/3027385.3027408.

Hussain, M., Zhu, W., Zhang, W., & Abidi, S. M. R. (2018). Student engagement predictions in an e-learning system and their impact on student course assessment scores. *Computational Intelligence and Neuroscience*, *2018*, 6347186. https://doi.org/10.1155/2018/6347186.

Hussain, S., Dahan, N. A., Ba-Alwib, F., & Najoua, R. (2018). Educational data mining and analysis of students' academic performance using WEKA. *Indonesian Journal of Electrical Engineering and Computer Science*, *9*(2), 447–459. https://doi.org/10.11591/ijeecs.v9.i2.447-459.

Jackson, L. (2013). Get the 411: Laptops and tablets in the classroom, January 04. Education World. Retrieved on February 02, 2025 from https://www.educationworld.com/a_tech/tech/tech194.shtml.

Kemper, L., Vorhoff, G., & Wigger, B. U. (2020). Predicting student dropout: A machine learning approach. *European Journal of Higher Education*, *10*(1), 28–47. https://doi.org/10.1080/21568235.2019.1613427.

Koh, H., & Tan, G. (2005). Data mining applications in healthcare. Journal of Healthcare Information Management, 19(2), 64–72. https://pubmed.ncbi.nlm.nih.gov/

Lan, K., Wang, D., & Fong, S. (2018). A survey of data mining and deep learning in bioinformatics. *Journal of Medical Systems*, *42*(8), 139. https://doi.org/10.1007/s10916-018-0987-4

Li, F., Zhang, Y., Chen, M., & Gao, K. (2019). Which factors have the greatest impact on student's performance. *Journal of Physics: Conference Series*, *1288*, 012077. https://doi.org/10.1088/1742-6596/1288/1/012077.

Luhaybi, M. A., Tucker, A., & Yousefi, L. (2018). The prediction of student failure using classification methods: A case study. *Computer Science and Information Technology*, *2018*, 79–90. https://doi.org/10.5121/csit.2018.80906.

Magsi, H., Sodhro, A. H., Al-Rakhami, M. S., Zahid, N., Pirbhulal, S., & Wang, L. (2021). A novel adaptive battery-aware algorithm for data transmission in IoT-based healthcare applications. *Electronics*, *10*(3), 367. https://doi.org/10.3390/electronics10030367

Marte, J. (2014). Here's How Much Your High School Grades Predict Your Future Salary, May 20. Retrieved on January 21, 2025 from https://www.washingtonpost.com/ news/wonk/wp/2014/05/20/heres-how-much-your-high-school-grades-predict-how-much-you-make-today/.

Nagahi, M., Jaradat, R., Nagahisarchoghaei, M., Ghanbari, G., Poudyal, S., & Goerger, S. (2020). Effect of individual differences in predicting engineering students' performance: A case of education for sustainable development. In *Proceedings of the International Conference on Decision Aid Sciences and Applications (DASA)*, Online, 8–9 November 2020.

Nagahisarchoghaei, M., Dodd, J., Nagahi, M., Ghanbari, G., & Poudyal, S. (2020). Analysis of a warranty-based quality management system in the construction industry. In *Proceedings of the International Conference on Data Analytics for Business and Industry: Way Towards a Sustainable Economy (ICDABI)*, Online, 26–27 October 2020.

Okubo, F., Yamashita, T., Shimada, A., & Ogata, H. (2017). A neural network approach for students' performance prediction. In *Proceedings of the Seventh International Learning Analytics & Knowledge Conference (LAK '17)*, Vancouver, BC, Canada, 13–17 March 2017 (pp. 274–283). Association for Computing Machinery. https://doi.org/10.1145/3027385.3027401

Rizvi, S., Rienties, B., & Khoja, S. A. (2019). The role of demographics in online learning: A decision tree-based approach. *Computers & Education*, *137*, 32–47. https://doi.org/10.1016/j.compedu.2019.04.010.

Saeys, Y., Inza, I., & Larranaga, P. (2007). A review of feature selection techniques in bioinformatics. Bioinformatics, 23(9), 2507-2517. DOI: 10.1093/bioinformatics/btm344.

Saeys, Y., Abeel, T., & Vande Peer, Y. (2008). Robust feature selection using ensemble feature selection techniques, Machine Learning and Knowledge Discovery in Databases (ECML.PKDD.2008), Lecture Notes in Computer Science, 5212, 313-325, Springer. DOI:10.1007/978-3-590-87481-2_21.

Shah, C., & Du, Q. (2021a). Spatial-aware collaboration-competition preserving graph embedding for hyperspectral image classification. *IEEE Geoscience and Remote Sensing Letters*, *19*(1), 1–5. https://doi.org/10.1109/LGRS.2021.3081070.

Shah, C., & Du, Q. (2021b). Modified structure-aware collaborative representation for hyperspectral image classification. In *Proceedings of the 2021 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, Brussels, Belgium, 11–16 July 2021. https://doi.org/10.1109/IGARSS47720.2021.9554932.

Shah, C., & Du, Q. (2021c). Collaborative and low-rank graph for discriminant analysis of hyperspectral imagery. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *14*, 5248–5259. https://doi.org/10.1109/JSTARS.2021.3082013.

Shah, C., Du, Q., and Xu, Y. (2022). Enhanced TabNet: Attentive interpretable tabular learning for hyperspectral image classification. *Remote Sensing*, *14*(3), 716. https://doi.org/10.3390/rs14030716

Singh, N.P. & Singh, D. (2019). Impact of feature selection methods on the performance of Credit Risk Classification Algorithm. Proceeding of IEEE 13th International Conference Application of Information and Communication Technologies 23-25, October 2019 | Baku, Azerbaijan, pp 101-106.

Sodhro, A. H., & Zahid, N. (2021). AI-enabled framework for fog computing driven e-healthcare applications. *Sensors*, *21*(21), 8039. https://doi.org/10.3390/s21248039

Vazan, P., Janikova, D., Tanuska, P., Kebisek, M., & Cervenanska, Z. (2017). Using data mining methods for manufacturing process control. *IFAC-PapersOnLine*, *50*(1), 6178–6183. https://doi.org/10.1016/j.ifacol.2017.08.111

Velazquez, R. (2023). Virtual reality in education: Benefits, uses, and examples, March 22. Retrieved on April 06, 2025 from https://soeonline.american.edu/blog/benefits-of-virtual-reality-in-education

Wang, W., Yu, H., & Miao, C. (2017). Deep model for dropout prediction in MOOCs. In *Proceedings of the 2nd International Conference on Crowd Science and Engineering (ICCSE'17)*, Beijing, China, 6–9 July 2017.

Wasif, M., Waheed, H., Aljohani, N. R., & Hassan, S. U. (2019). Understanding student learning behavior and predicting their performance. In *Cognitive Computing in Technology-Enhanced Learning* (pp. 1–28). IGI Global. https://doi.org/10.4018/978-1-5225-7191-7.ch001.

Whitehill, J., Mohan, K., Seaton, D., Rosen, Y., & Tingley, D. (2017). Delving deeper into MOOC student dropout prediction. *arXiv*. https://arxiv.org/abs/1702.06404.

Winstead, S. (2025). Using tablets in School: How to implement 1:1 technology using tablets in the classroom, April 06. My eLearning World. Retrieved on April 25, 2025, from https://myelearningworld.com/10-benefits-of-tablets-in-the-classroom/.

Xing, W., & Dongping, D. (2019). Dropout prediction in MOOCs: Using deep learning for personalized intervention. *Journal of Educational Computing Research*, *57*(3), 547–570. https://doi.org/10.1177/0735633119831374

Xue, B. M., Browne, W.N., & Yao, X. (2016). A survey on evolutionary computation approaches to feature selection; IEEE Transactions on Evolutionary Computation, 20 (4), 606-626.

Zaletelj, J., & Košir, A. (2017). Predicting students' attention in the classroom from Kinect facial and body features. *Journal of Image and Video Processing*, *2017*, 80. https://doi.org/10.1186/s13640-017-0252-9.

Zauzmer, J. (2020). More students are learning on laptops and tablets in class. Some parents want to hit the off switch. *The Washington Post*. https://www.washingtonpost.com/local/education/more-students-are-learning-on-laptops-and-tablets-in-class-some-parents-want-to-hit-the-off-switch/2020/02/01/d53134d0-db1e-11e9-a688-303693fb4b0b_story.html