Feature selection techniques and classification algorithms for student performance classification: a review

Muhamad Aqif Hadi Alias, Najidah Hambali, Mohd Azri Abdul Aziz, Mohd Nasir Taib, Rozita Jailani
School of Electrical Engineering, College of Engineering, Universiti Teknologi MARA, Shah Alam, Malaysia

Article Info

Article history:

Received Oct 20, 2023
Revised Feb 3, 2024
Accepted Feb 9, 2024

Keywords:

Artificial neural networks
Classification
Decision tree
Feature selection
K-nearest neighbors
Linear models
Student performance

ABSTRACT

The process of categorizing students' performance based on input data, encompassing demographic information and final exam results, is recognized as student performance classification. Educational data mining has gained traction in assessing students' performance. However, this study entails the need to analyze the diverse attributes of students' information within an educational institution by using data mining techniques. This study thoroughly examines both previous and current methodologies presented by researchers, addressing two main aspects: data preprocessing and classification algorithms applied in student performance classification. Data preprocessing specifically delves into the exploration of feature selection techniques, encompassing three types of feature selection and search methods. These techniques aim to identify the most significant features, eliminate unnecessary ones, and reduce data dimensionality. In addition, classification algorithms play a crucial role in categorizing or predicting student performance. Models such as k-nearest neighbors (KNN), decision tree (DT), artificial neural networks (ANN), and linear models (LR) were scrutinized based on their performance in prior research. Ultimately, this study highlights the potential for further exploration of feature selection techniques like information gain, Chi-square, and sequential selection, particularly when applied to new datasets such as students' online learning activities, utilizing a variety of classification algorithms.

Corresponding Author:

Najidah Hambali
School of Electrical Engineering, College of Engineering, Universiti Teknologi MARA
Shah Alam, Selangor, Malaysia
Email: najidah@uitm.edu.my

1. INTRODUCTION

In recent years, an emerging topic that has been concerned by each educational institution is the students' performance. Anticipating students' performance early on proves to be a valuable asset in enhancing their learning experience. Identifying at-risk students in the initial phases of the course allows for ample time to implement interventions and strategies aimed at improving their academic outcomes [1]–[7]. Undeniably, it is considered as a major factor to uplift the quality of the institutions and the students themselves [8]–[11]. In order to better understand and improve the learning process and the surroundings in which it takes place, educational data mining has recently gained relevance and pace where it is crucial in forecasting students' academic success [12]–[17]. The phrase "educational data mining" refers to the use of data mining techniques to improve educational quality, pinpoint students who need to improve and uncover factors influencing student academic achievement [18]. This field of study involves examining various attributes to analyze student information within an educational institution [19], [20]. It is believed that data mining is still relatively new in education even though there has been significant use in commercial sector [21].

The data acquired must be recognized as the factor that most significantly affects the students' performance in order to create the prediction model efficiently. The increasing volume of educational data underscores the imperative to extract valuable insights from patterns in learning behavior [22]. Specifically, educational data mining focuses on developing the algorithms that can uncover the hidden patterns in educational data since the study involves with numerous features of students' information that need to be analyzed [23]–[26]. However, most of the acquired data are comprehensive which also contain the unwanted features whereby without data preprocessing, some misinterpretations might be made by the model which indicate inaccuracy in predicting students' performance [27], [28]. Attributes in the dataset with minimal variance, where the values exhibit negligible differences, are excluded as they contribute insignificantly to the mining process [29]. Several feature selection techniques, namely genetic algorithms (GA), Gain ratio (GR), relief, and information gain (IG) were presented in evaluating the undergraduate students' academic performance to analyze their practicality and performance alongside various classification algorithms [24].

Other than that, there has been a growth in the use of artificial intelligence in education [30]–[32], particularly machine learning, where it is projected to provide effective methods to improve education in general in the near future. Intelligent m-learning systems have lately seen a surge in popularity as a means of providing more effective education and adaptable learning that is suited to each student's learning capacity [33]. The early attempts to enable such systems, for creating tools to help students and learning in a conventional or online context, through the use of machine learning techniques focused on anticipating student achievement in terms of grades attained [34], [35]. Classification stands out as the predominant technique for predicting students' academic performance utilizing some classification algorithms encompass decision tree (DT), k-nearest neighbor (KNN), support vector machine (SVM), naive Bayes (NB), and artificial neural network (ANN) [36]. Using a dataset containing board results and 12 attributes associated with a class comprising 172 students of various genders and statuses, the findings indicated that the ANN outperformed the KNN algorithm, particularly concerning relative squared error and mean absolute error [37].

In drafting this review article, our motivation is to explore the application of various data mining techniques, involving feature selection and machine learning algorithms used in classification. Our research area centers on the investigation of implementing data mining techniques in academic environments, involving the classification of students' performance. Some published papers had covered the topics, employing feature selection methods alongside classification algorithms in predicting students' performance. Contrastly, these studies only focused on a few methods of feature selection categorized as filter based [29], [36] while the study in 2018 only revealed the use of classification algorithms without applying feature selection [37]. We contend that constructing a precise classification model necessitates the implementation of an appropriate preprocessing technique, including feature selection method. The sections of this article are grouped as follows: section 2 presents an overview of previous research in employing diverse methods of feature selection. Section 3 delves into the machine learning algorithms used in classification, followed by the summary's discussion of the previous studies in section 4. Finally, section 5 encapsulates the conclusion drawn from our exploration.

2. FEATURE SELECTION TECHNIQUES

As concerned by some of the researchers, data preprocessing is essential for improving data quality and impacting its reliability for data mining algorithms [38] whereby failing to do so will allow the erroneous conclusions to be made by the prediction model since the raw data contains a lot of unwanted features and noise [27]. Researchers in [25] emphasized that the data mining quality is mainly affected by the acquired data and features. Coherently, data-level solution using oversampling technique and two feature selection methods; wrapper and filter based were used as the benchmark methods in this study to overcome the problem of imbalanced multi-classification dataset [39]. Feature selection (FS) is one of the techniques that can be used in the data preprocessing step where it is used to identify the most important features and remove the unwanted features along with reducing the dimensionality of data. Some researchers had highlighted on implementing feature selection techniques into some of the classification algorithms for improving the prediction model [12], [24], [27], [38], [40], [41]. Three types of feature selection will be discussed in this part which includes filter-based, wrapper-based, and embedded-based.

2.1. Filter-based

Filter-based technique is employed as a preprocessing step based on the results of statistical tests relating to the correlation with the dependent variable. It is used to find irrelevant features and generates a dataset with the best feature columns based on their scores. Since it does not require model training, this approach is deemed faster and has minimal computing complexity. For instance, some researchers had applied several filter-based feature selection methods such as information gain (IG) [25], [27], [40], [41], gain ratio (GR) [24], [27], [38], [42], Pearson correlation [43]–[45], Chi-square [27], [42], and minimum redundancy and maximum relevancy (mRMR) [36], [46]. Below are several methods of filter-based feature selection:

2.2. Information gain/mutual information

A filter-based approach called IG, which employs statistical tests to find the most important characteristics was used by [40], [36], [47], [48]. A feature selection technique called mutual information (MI) was applied in [40], to attain the ideal feature set where it is a filter technique that estimates entropy reduction by comparing the information gain of each independent feature to the information gain of the dependent feature and choosing the feature with the highest information gain. By implementing feature selection techniques, studies in [36], [48] found that IG performed better in signifying the important features in each study.

In addition to its notable performance with certain classifiers, IG emerged as a robust feature selection method in studies such as [36], [48]. Notably, IG demonstrated exceptional efficacy, particularly in conjunction with classifiers like ANN and DT, as evidenced in [36]. Furthermore, in the context of detecting Internet cheaters among students [48], MI showcased its effectiveness by outperforming other methods when coupled with the random forest (RF) classifier. The study selected the top 5 features using MI, revealing their enhanced compatibility and performance synergy with the RF classifier. Additionally, the utilization of MI alongside analysis of variance (ANOVA) in the same study underscored the versatility of these feature selection techniques, each contributing distinctively by selecting 5 features out of a pool of 13.

In the development of multi-class prediction models for students' grade prediction, a main concern from the study was the imbalanced multi-class issue and the overfitting problem [39]. To prevent the issues, an oversampling technique known as synthetic minority oversampling technique (SMOTE) alongside two feature selection methods, namely *WrapperSubsetEval*, *ClassifierSubsetEval*, and IG, were introduced in which will be evaluated with several classification algorithms. The findings of apply in metric proposed method alongside six classification algorithms, including DT (J48), NB, KNN, SVM, logistic regression and RF, had significantly shown that all the FS performed approximately the same across all classifiers, with above 90% of each metrics. Based on the results, it seems that there is no need to apply wrapper methods since it requires much computation complexity and processing capacity. In this case, information gain was quite commendable based on its performance, which is not much different alongside the wrapper methods.

2.3. Gain ratio

In the investigation conducted by researchers [49], it was discerned that GR exhibited superior efficiency, boasting a remarkably low time complexity of just 0.08 milliseconds. This stark contrast was observed when comparing GR to alternative methods, namely Chi-Square and IG. The findings underscored the computational expediency of GR, emphasizing its potential as a time-efficient solution for tasks where rapid processing is paramount. Based on the performance of filter-based approaches, GR and Pearson correlation had the highest rank scores (ranging from 0.2 to 1), indicating that these findings were solely influenced by individual characteristics [50]. In the formulation of various data mining techniques for predicting students' performance [29], the GR feature selection method was integrated and paired with seven classifiers. Among these classifiers, GR exhibited superior performance when coupled with the decision table classifier, achieving a recorded accuracy of 76.57%.

As reported in [24], GR chose 10 features out of 14 and was integrated with multiple classifiers, achieving the highest accuracy when combined with the KNN classifier. In response to the initial research question posed in [51], various feature selection methods, encompassing wrapper, correlation, and GR, were assessed alongside baseline classifiers like NB, J48, and RF. Out of all combinations, GR yielded the highest F1-Score when paired with NB, reaching 80.1% with 10 attributes retained.

2.4. Correlation-based

The correlation-based feature selection (CFS) is a filter-based feature selection technique that is independent of the final classification model. It quantifies the strength of the linear relationship between two variables where it has a numerical value between -1 and 1, where -1 represents a high negative linear correlation, 0 denotes no correlation, and +1 suggests a strong positive correlation. CFS technique was used in some studies in analyzing the correlation between two numerical attributes in order to obtain a minimal set of features [43]–[45] whereby just the top 10 features evaluated by correlation attribute evaluator (CAE) was considered [43] and three learning behaviors were removed out of 28 variables [44] while 2 features from experience application programming interface (xAPI) dataset were removed based on the correlation analysis [45].

In order to enhance prediction accuracy to an acceptable level, a hybrid or heterogeneous method combining CAE, ensemble learning and seven distinct machine learning algorithms was presented [43]. According to the results, any classification algorithm constructed using heterogeneous ensemble learning and CAE outperformed methods implemented using ensemble learning without CAE. Based on the performance of filter-based methods, GR and Pearson correlation obtained most of the features in high rank scores (between 0.2 and 1) where these findings were merely on the influence of individual features [50]. In the analysis of learning behavior of students' college and its learning effect [44], the researchers used a threshold value based on the dependency value produced by Pearson correlation. In this case, a 0.50 value was considered for analyzing the correlation of learning behavior towards its learning effect. Similarly, in a study focusing on identifying the minimal set of features essential for effective analysis [45], a threshold value of 0.7 was strategically applied. This threshold served as a criterion for assessing the correlation between various features, facilitating the identification and subsequent exclusion of highly correlated features. By setting the threshold at 0.7, the study aimed to streamline the feature set, eliminating redundancy, and enhancing the efficiency of subsequent analyses. The careful consideration of threshold values in both studies underscores the importance of methodological precision in uncovering meaningful insights from complex datasets.

2.5. Chi-Square

The chi-square approach is a prominent feature selection method. It is a statistical test used to assess how much observed values differ substantially from predicted results, and it is used to determine the predictor variable [49]. The researchers in [27] found that Chi-square and IG algorithms outperformed the others, according to the analysis of the Kappa statistic and F-measure. Both [45], [52] had conducted a statistical test on the features, namely Chi-square test to analyze the significance of the features. As a reference, p-value with 0.05 was considered to measure the features' significance where any values that are above it will be discarded.

2.6. Minimum redundancy and maximum relevancy

This method chooses a subset of features that have the highest correlation with the output and the lowest correlation among themselves. It ranks features based on mutual information using the minimal-redundancy-maximal-relevance criterion. Different classification algorithms and feature selections that have been examined reveal that classification using appropriate classifiers for specific category data and proper feature selection enhance the prediction model's accuracy [36]. Alongside IG, mRMR also obtained high accuracy with the use of DT and ANN classifiers based on several feature combinations. In [46], the researchers developed a framework of study that focuses on the accuracy of matching between four feature selection techniques and four classification models for student performance prediction. When pairing with KNN algorithm, mRMR and GR performed about the same where the results of 7 features selected from each method scored about 90% accuracy.

2.7. Wrapper-based

Feature selection procedure for wrapper method is based on a specific machine learning algorithm that will be applied to a certain record. It employs a greedy search strategy, assessing all potential feature combinations depending on the evaluation criterion. The GA was used by [12], [53] and defined by binary representation of individual solutions, simple crossover and mutation operators, and a proportional selection mechanism in order to determine the optimal feature combinations and to minimize the amount of calculation as well as removing the uncorrelated features. The results showed that GA can increase the fitness of gene sequences to some extent whereby the data dimension reduced from 7,070 to 3,579, indicating that 3,491 features were considered uncorrelated [53].

A binary genetic approach (BGA) was utilized as a feature selection algorithm in the study [54], with each solution supplied as a vector of a binary string. Except for the NB technique, the BGA feature selection algorithm improved the models' performance. In [55], a wrapper-based FS technique was used, which known as binary teaching-learning based optimization (BTLBO), that comprises of two primary components which are search algorithm and evaluation classifier. BTLBO exhibited the ability to enhance the overall performance of machine learning algorithms when combined with linear discriminant analysis (LDA) which improved by 3% and 8% for both datasets assessed based on the area under curve (AUC) values.

In [56], a study was introduced to evaluate the efficacy of various feature selection approaches on some classification algorithms using educational datasets. Three methods of wrapper-based feature selection including sequential forward selection (SFS), sequential backward selection (SBS) and differential evolution (DE) were implemented. Based on the values of prediction accuracy mean, these three methods performed slightly better than other filter-based methods used in the study where specifically DE scored the highest mean. In [46], greedy forward selection algorithm had selected the fewest features from 15 features whereas the other three methods which are mRMR, chi-square and IG-ratio selected 9, 10, 10 features respectively. The Greedy forward selection algorithm was found to be performed better with the use of ANN classifier.

In predicting the students' final grades at the early stages of a course, a wrapper feature selection method, namely Boruta algorithm, was used which employs RF algorithm [57]. Through an iterative process, it assesses the significance of the original attributes compared to the shadow counterparts, generated through the shuffling of the original attributes. Attributes with lower importance than their respective shadow counterparts are omitted, whereas those with higher importance are acknowledged as confirmed attributes. As demonstrated in their findings for the Mid-March data subset, the RF-based algorithm exhibited an average accuracy of 78%, whereas it decreased to 72.7% and 74.7% when employing the NB-based and KNN-based algorithms, respectively.

2.8. Embedded-based

With an embedded technique, feature selection is integrated into the classification algorithm in which the classifier modifies its internal parameters and calculates the proper weights/importance for each feature to generate better classification accuracy. One of the methods in selecting features for the dataset was considered in [58] which is basically based on classification, namely Random n-class classifier. It contains the number of redundant features, informative features which were provided as 0 and 1 and the total number of features. These features were created as random linear combination of informative features.

In the realm of supervised learning methods, the study in [59] initiated the logistic regression approach as a feature selection method, marking the inception of their exploration into choosing relevant features and categories. The preliminary findings from this endeavor highlighted the identification of 19 significant features within the dataset, as ascertained by the logistic regression technique. These features were deemed critical for discerning patterns associated with the normal class, shedding light on the method's efficacy in pinpointing key contributors to the classification task at hand.

In [50], two ensemble techniques, namely bagging and boosting were used to be integrated with classification models. In the experiment, only seven classification models were chosen whose performance was improved by employing 10-fold cross-validation. RF-IG and DT-IG were found to perform better when combined with ensemble approaches especially boosting method by achieving the highest scores (0.93, 0.753, 0.833) and (0.91, 0.76, 0.822) respectively.

2.9. Search techniques

In [50], filter-based incorporates some search techniques, namely 'ranker' and 'greedy stepwise' for 'attribute evaluator'. A study regarding predicting the intention of using social media in online blended learning presented by [52], where data was obtained with 61 attributes, which were then minimized to 24 and 5 attributes, following the use of greedy technique and the wrapper method respectively. Two feature selection methods in the study [60] which are Information gain and wrapper method were implemented, where BestFirst was the search method used by wrapper method while Ranker Search method was applied for information gain, to rank the attributes based on its gain value. Similarly, the Ranker Search method was utilized in [29], [61] along with several feature selection techniques CAE, IG, and GR. It is used to determine the best attribute from the student's performance dataset where only the top 10 features were chosen to determine the accuracy of the classification methods [29] while in developing a model to predict students' final grades in an introductory programming course early in the semester, the Ranker search method was included for two feature selection methods which are correlation-based and information-gain, in which a significance cutoff of 0.20 was used and any features below this mark will be disregarded [61].

3. CLASSIFICATION ALGORITHMS

Machine learning is critical in educational data mining, providing the specific purpose of predicting students' performance in order to improve the overall quality of learning. There are four types of machine learning algorithms which are supervised, semi-supervised, unsupervised and reinforcement machine learning where in this part, we will discuss more on supervised machine learning such as KNN, DT, ANN, and linear models. Researchers had introduced several studies regarding the evaluation of students' performance in the learning process by using supervised machine learning. For example, in developing an early prediction of students at risk of failing a face-to-face course in power electronic systems, the scrutinized classifiers have demonstrated notable effectiveness in the identification of students at risk of course failure. Indeed, significant accuracy and sensitivity values ranging from 70% to 81% were observed, even when exclusively considering attributes from the students' background [62]. Thus, in this section, we will review some classification algorithms in displaying their application in classification tasks:

3.1. K-nearest neighbor

The KNN algorithm is a supervised machine learning method that estimates the chance that a data point will belong to one of two groups depending on the feature similarity. Several classification algorithms were employed by [24], [41], which will be evaluated based on their performance of efficiently predicting student academic performance. Among all comparative findings, the GA feature selection approach using KNN had the highest accuracy of 91.37% [24] and by evaluating sets of feature selection methods and classification algorithms, it significantly demonstrated that mRMR feature selection approach with 10 selected features produced 91.12% accuracy with the KNN classifier [41].

Working on the development of an early warning system, involving various socio-cultural, structural, and educational factors that directly influence a student's choice to discontinue their education [63], several classification algorithms, namely SVM, RF, stochastic gradient descent (SGD) and KNN, were employed as predictive models for the dataset. According to their findings, the KNN algorithm demonstrated superior performance by achieving the lowest losses mean absolute error (MAE) and root mean square error (RMSE) and consequently securing the highest accuracy score (R2). Specifically, it surpassed 99.5% accuracy for the training set and exceeded 99.3% for the test set.

In a study, [12] implemented the modified K-nearest neighbor (M-KNN) approach to categorize students' performance and compared its results with the conventional KNN method. The accuracy score provided by the classification techniques, KNN and M-KNN, was employed as the assessment criterion in this study. M-KNN accuracy increased by using GA whereby it recorded 82.6% whereas KNN accuracy was only 73.6%. KNN is one of the classification algorithms included in [56], [64], which was used to be assessed its performance alongside other classifiers such as SVM, NB, DT, and discriminant analysis (DISC), with the use of feature selection methods and KNN was found to have a significant impact in both studies. The goodness of subsets was measured with varying cardinalities in terms of prediction accuracy and the number of selected features for 11 wrapper-based feature selection algorithms using the KNN and SVM as baseline classifiers [64]. In terms of exploration and exploitation abilities (fitness), the sunflower optimization (SFO) algorithm with KNN and SVM performed better since it only determined four features out of 20 whereas KNN classifier outperformed other classifiers on the student data based on the findings obtained [56].

3.2. Decision tree

A decision tree is a straightforward structure in which each non-terminal node reflects a test or decision on the data item under consideration. Some researchers had included the use of decision tree algorithm by proposing it in predicting students' academic performance [18], [23], [29], [61], [65] in which it showcased notable performance compared to other classifiers. DT and RF are two of the classification methods that were compared in the study [23]. Their findings demonstrated that Decision Tree outperformed Random Forest in terms of classification performance with 66.85% accuracy. In introducing a study of predicting academic performance of student using classification techniques, some classifiers such as NB, decision tree (J48), and multilayer perceptron (MLP) were employed [18]. It revealed that J48 had the highest accuracy at 73.92%. By utilizing four supervised educational data mining approaches, namely NB, MLP, J48, and RF, a dataset was analyzed by [65]. Results depicted that decision tree J48 outperformed other educational data mining algorithms on all subsets of the dataset, excluding the 2-level classification subsets for student social activities.

Based on several combinations of classifiers and feature selection methods, J48 produced the second highest accuracy of up to 75.34% when compared to other combinations including NB, RF, J48, MLP, decision table, JRip, and logistic regression classifiers [29]. In the process of formulating a prognosticative model designed to apprise students of their anticipated academic outcomes in the early stages of the semester, 13 machine learning algorithms from 5 categories were tested and applied [61]. It can be seen that J48 had reached an accuracy of 88%, followed by NB with 84% and decision Table with 83% accuracy. In comparison to other types of algorithms, the decision tree family of algorithms had generally attained better accuracy.

Several regression models, including linear regression, DT, NB, sequential minimum optimization (SMO), ANN, KNN, REPTree, and partial decision trees (PART), and RF, have been devised to forecast students' academic performance [38], [66]. Notably, RF emerged as the most effective model for predicting students' performance, demonstrating superior performance due to its composition of multiple decision trees [66], where the study in [38] also observed a substantial enhancement in the accuracy of predicting students' academic performance by employing the RF model, achieving precision, recall, and F-measure rates of 94.70%, respectively. Similarly, by utilizing classification techniques in constructing a drop out classification model supplemented by RF algorithm and imbalance dataset methodology, named SMOTE, the RF+SMOTE method demonstrated better performance when k=2 (referring to the number of folds utilized), with the highest accuracy, recall, and f-measure of 93.43%, 92.27%, and 92.99%, respectively [67].

Multiple feature selection approaches were employed to analyze an educational dataset from a national test in order to identify the significant feature subsets [27]. Based on the use of 3 feature selection methods, the Classification and Regression Trees (CART) classifier obtained the highest average F-measure which is 0.835 followed closely by MLP with 0.829. Machine learning techniques using DT, namely C4.5, Iterative Dichotomiser 3 (ID3), and improved ID3, were implemented by Patil *et al.* [68] on the training database in stage three. A comparison of DT generating methods C4.5, ID3, and improved ID3 was performed where the improved ID3 algorithm outperformed the conventional ID3 and C4.5 algorithms.

3.3. Artificial neural network

Artificial neural networks (ANN) are gaining popularity in a variety of fields, including education. The ANN structure is made up of a series of linked artificial neurons, each with its own weight. It is composed of three layers of organized nodes: input, hidden, and output. An ANN, in general, is an adaptive system that alters its structure based on the internal and external information involved in the process during the learning process. The most typical feed-forward neural network, known as MLP, transmits data from the input to the neurons in the output layer. Both studies [21], [37] incorporated MLP usage in their research, representing the ANN model where a significant positive effect on accuracy was observed, achieving a value of 93% [21].

Several studies had consistently highlighted ANN as a particularly prominent and effective classifier, showcasing better performance when compared to alternative approaches [25], [36], [41], [46], [60], [69], [70]. A comprehensive investigation into the classification stage was conducted, employing five distinct classifiers: DT, KNN, ANN, NB, and SVM [36]. The objective was to ascertain the most proficient classifier that will be well-combined of each in conjunction with various feature selection techniques. The results of this analysis unequivocally underscored ANN and DT as the foremost classifiers, demonstrating not only higher accuracy but also notable excellence in precision, recall, and F1-Score, surpassing their counterparts [36], [69], while ANN model outperformed other data mining approaches, NB and DT, recorded 73.8% accuracy with the behavioral features used and 55.6% for non-behavioral features [25].

On the other hand, some studies also showcased another view of the ANN's performance by pairing up with several feature selection methods [41], [46], [60]. In a multi-class classification of students' performance, four feature selection methods, including GA, mRMR, IG, and SVM, were applied to the dataset to remove un-relevant features [41]. The GA feature selection method, with 10 selected features, demonstrated the highest accuracy of 90.6% with the use of ANN classifier. In contrast, the rest of feature selection methods, namely mRMR, IG, and SVM, had a commendable accuracy with the application of KNN. ANN recorded the highest correctly classified instances for about 78.3% before applying feature selections and 79.375% after implementing IG method, which indicate 376 instances and 381 instances were correctly classified, respectively [60]. Meanwhile, an initial 6,882 records with 15 attributes including admittance student data and grade from engineering core course subject [46], this study proposed 4 feature selection methods, consisting of greedy algorithm, GR, chi-square, and mRMR, into a multi-class classification of students' performance. The findings discovered that the greedy forward selection approach had better accuracy of 91.16% with ANN classifier.

3.4. ANN training

Backpropagation algorithm [71], [72] and cross-validation [21], [73] were significantly used in developing ANN model in some studies where backpropagation algorithm is used to make the connections between neurons sufficient, by changing the weights of these connections in order to build a proper neural network. According to the results, utilizing MLP has provided more accurate values than DT, with accuracy percentages ranging from 42% to 97% [71] while by implementing both backpropagation algorithm with cross-validation, Tomasevic *et al.* [72] obtained the overall highest precision with ANN by feeding the student engagement data and past performance data and also tested for different number of hidden layers. By using cross-validation in MLP, the model could properly predict the dataset where 223 students out of 524 and 83 out of 178 in percentage split were predicted [21] and also useful during the process of fine-tuning [73] where 5-fold cross-validation was used on the training set to find the optimal values for each model.

3.5. ANN hyperparameters

In a study [59], a 5-layer neural network with three hidden layers was implemented. The neural network simulation results for three distinct examples, involving 2, 3, and 4 hidden layers, demonstrated that the most favorable outcomes were achieved with three hidden layers, avoiding over/under-fitting issues. In [74], through a meticulous tuning process, the highest accuracy in the ANN model was achieved by configuring it with 200 neurons, utilizing the logistic function as the activation function. The L-BFGS-B solver optimized the model for convergence, while regularization with an alpha value of 7.10^{-4} prevented overfitting. This parameter ensemble led to an MLP with an impressive R2 value of 0.938, reflecting a robust alignment between the model's predictions and observed academic performance. In the context of classification using ANN, Imdad *et al.* [37] identified the optimal configuration with two hidden layers, a momentum value of 0.2, and a learning rate of 0.3. At this configuration, their data achieved 100% accuracy with fewer errors per epoch, along with reduced time and errors. In another instance [73], grid search and randomized search were employed to determine the optimal hyperparameter values for classifiers like ANN, SVM, and RF. After fine-tuning, the accuracy of the ANN model improved from 90.94% to 92.00%, precision from 88.29% to 89.07%, F1-Score from 91.25% to 92.29%, and recall from 94.41% to 95.76%. Researchers in [75] utilized Bayes's theorem and ANN to create models predicting students' chances of graduating from a tertiary institution. The study revealed that ANN outperformed Bayes's theorem in terms of performance accuracy. Significantly, the accuracy of the ANN improved as the number of hidden layers increased. The best result was found when four hidden layers were used, with an accuracy of 99.97% on the training dataset.

3.6. Linear models

Linear regression is a supervised machine learning model that determines the best fit linear line between the independent and dependent variables, or the linear connection between the dependent and independent variables. For the purpose of predicting student academic performance in a course, Uskov *et al.* [76] had included some machine learning algorithms to be analyzed which are linear regression, logistic regression, KNN, NB, ANN regression and classification, DT, RF and SVM. With just a 3.7% average difference between projected and real student total final scores, the linear regression algorithm displayed better accuracy.

The likelihood of the category dependent variable can be predicted using the binary classification procedure known as logistic regression. The logistic function transforms a linear combination of independent variables into a probability score ranging from 0 to 1, which is used to categorize the dependent variable into one of two potential outcomes. Based on two studies by [55], [77], logistic regression was found to be significant by comparing to several machine learning algorithms applied in predicting student performance such as, RF, NB, logistic regression, KNN, SVM, DT, and LDA. To categorize students as 'high risk' or 'low risk', Ramaswami *et al.* [77] figured out that logistic regression model had the highest F1-Score compared to other classifiers while logistic regression and LDA performed better than other classifiers based on AUC value [55].

In a study of preserving the integrity of students in online assessments [48], some machine learning algorithms, namely RF, logistic regression, SVM, KNN and NB were implemented along with two feature engineering methods namely MI and ANOVA. Based on the results of classifiers' performance from the top five selected features by MI, logistic regression was second best performing classifier with an approximate accuracy of 82% and an F-Score of 72%. However, LR recorded the lowest accuracy and F-Score when using the selected features by ANOVA, with 78.33% and 64.57% respectively.

4. DISCUSSION

In this section, a summary of the previous works will be discussed to obtain the significant knowledge gaps which can be further study in future. This summary table is supposed to unveil any gaps that may seem significant to further study. Each feature selection method section was reviewed and about 2 papers from each of it were taken to be organized in the table. Table 1 presents the techniques used, their performance, dataset and portrays any limitations or advantages from each source. By having this table, we can reveal the knowledge gaps of these studies complying with our focus of this review paper.

Firstly, some studies prominently found that IG had significant performance in selecting features out of the dataset [27], [36], [48], [50]. It can be seen that IG had better performance when coupled with some classifiers like DT [27], [36], [50], ANN [36], and RF [48], [50]. Another one method that seems to be performing well is Chi-square, as in two studies found that it somehow contributed to the predictive model performance [27], [45]. In analyzing the xAPI dataset, Chi-square had secured about 5 features out of 16 in which those features then were considered as the significant features (SF) but all models deemed to have a drop in all evaluation metrics by using the SF [45]. However, Chi-square and IG found that both methods selected the same number of features, which is 6 out of 20 [27]. In this case, 85.9% of F-measure was recorded for both using DT (C4.5).

In other context, Pearson correlation seems to be quite functional in analyzing the features' correlation as found in [45] where several features were analyzed and then discarded for being redundant and highly correlated with other features. As for using the public dataset known as Student-mat, Pearson correlation had a significant impact towards analyzing the features which in later stage, the classification models like RF, ANN, and SVM obtained a commendable accuracy of above 80%. The same situation was seen in [44], leveraging data from the literacy learning behavior questionnaire and the performance records of an information literacy course for 320 junior students, our analysis led to the exclusion of three learning behavioral features with correlations below 0.500, along with the demographic attribute ‘Gender.’ Ultimately, 25 features were retained out of the initial 29. The rationale for omitting these three learning behaviors is rooted in their comparatively lower integration with college students’ study routines, daily life, and the prevailing learning environment when contrasted with other attributes of learning behavior. Although this method seems quite performing, there is a gap between these studies that we can unveil as they only utilized the single method of feature selection and a threshold value of 0.5 was used [44]. In this case, we could consider using a diverse of feature selection methods and another threshold value instead of 0.5.

Table 1. Summary of feature selection performance

Source [Ref.]	Technique	Performance	Dataset	Remark/Limitations
[36]	IG, mRMR	Above 90% accuracy with DT and ANN	Kaggle repository datasets: Students’ academic performance, containing 480 record and 16 attributes	Too many data category combinations, binary and multi-class grading
[48]	MI, ANOVA	85% with RF	iQuiz integrated with Moodle learning management system (LMS)	Question type and difficulty were prominently chosen
[29]	GR	76.57% with DT	Students’ grades, demographic, and school-related information	No discussion on the selected features and how the dataset was applied
[24]	GR, GA	90.26% with KNN, 91.37% with KNN	800 student records with 14 attributes including identification, attendance, assignments, class tests, lab tests, spot tests, skills, central viva, extracurricular activities, quiz tests, project/presentation, backlog, final semester results, and final CGPA.	Multi-class grading, 10 features selected out of 14 but no discussion on which features were relevant
[50]	IG	93% accuracy with RF and DT	Collected from the online educational system consists of 11,814 students, six categories of features: Demographics (de), Personal (pe), Academic (ac), Psychometric (ps), Family attributes (fa), and Learning Logs (ll).	No discussion on which features were selected and used
[44]	Pearson Correlation	92.50 % accuracy with RF, 91.67% with ANN	Collected from the information literacy learning behavior questionnaire data and information literacy course performance data of 320 junior students	three learned behavioral features with correlations below 0.500 were removed as 25 features were retained, only single feature selection was used
[45]	Pearson, Chi-square	stu-mat: 81% with ANN, 84% with RF, 82% with SVM stu-por: 85% with RF	three public datasets, student-por, student-math, and xAPI	xAPI: 7 features selected out of 16 student-mat: 14 features selected out of 32 student-por: 15 features selected out of 32, limited to only two feature selection methods
[27]	Chi-square and IG	85.9% F-measure for both using C4.5	dataset includes enrolment information of students and examination result, containing 7,723 of permissible volunteers with 20 features	Both methods selected 6 features out of 20, Limited to only using data of enrolment information and examination results
[46]	Greedy forward, mRMR, GR	Above 90% with KNN, ANN	6,882 records with 15 attributes including admittance student data and grade	Greedy forward: 7 features selected, mRMR: 9 features selected, GR: 10 features selected; No significant discussion on the dataset category used
[12]	GA	82.6% with Modified-KNN	Student’s Academic performance dataset obtained from Kaggle, with 16 attributes and 480 instances	Multi-class classification, No significant discussion on the selected features, only used a single type of classifier and feature selection
[56]	SBS, DE	Accuracy: 84.72% with DT, 85.21% with KNN	Dataset obtained from a learning management system called Kalboard 360, containing 500 students with 16 features	No discussion on which features were selected and no details on the features
[58]	Random-n classifier, ANOVA		Data was taken from Kaggle of Students’ performance data, containing 1000 samples and 8 attributes	Those selected features were not described and discussed, Limited feature selection methods, the dataset encompassed students’ demographic and academic scores

Lastly, we will choose another one type of feature selection to be discussed, which is known as wrapper based. Both [46], [56] revealed that their tested feature selection methods had achieved better performance than other methods tested. Sequential feature selection has two methods known as sequential forward and backward selection. As found in [46], greedy forward or also known as SFS, had a commendable performance which selected 7 features of 15 with an accuracy of above 90% when trained with KNN and ANN. Contrastly, SBS and DE had the highest accuracy recorded, above 82%, when pairing up with DT, DISC, and KNN [56]. However, the study did not have a detailed discussion on the selected features and what features they were implying to.

To summarize this section, some studies have shown that there are some research gaps that we can acknowledge encompassing the implementation of diverse feature selection methods alongside various classification algorithms. In section 3, an exploration of each classification algorithm was portrayed which can be seen their feasibility in predicting students' performance. Several classification algorithms had been unveiled their performance in this section whereby the implementation of feature selection alongside the predictive models had better results in revealing the pattern of factors that might contribute to students' performance. Throughout this section, we can see that many of the studies included only the prevalent dataset encompassing the demographic of students, family's background, and examination scores, and some did not provide a detailed discussion on their selected features implying to which category of features. We believe that a dataset as students' learning behaviors can provide a better understanding of their efforts as done in [44], [78], [79].

5. CONCLUSION

In this paper, we conducted a comprehensive examination of various feature selection methods and classification algorithms. Our objective was to enhance our understanding of how these techniques can be effectively applied to classify students' performance. Among the numerous data mining techniques employed in classification tasks, we found that feature selection plays a pivotal role. It assists in identifying the most significant features while reducing computational complexity, thereby streamlining the process. Additionally, our findings indicated that the choice of feature selection approach significantly impacts the prediction of student success. Notably, the outcomes of these approaches may vary when applied to different types of data, despite the multitude of studies conducted by various researchers. Machine learning algorithms have gained widespread use across diverse fields, particularly in classification tasks. Despite the significant findings reported in numerous studies, there remains ample opportunity for further investigation involving various data types and data preprocessing techniques. The selection of appropriate algorithms often hinges on factors such as data structure, training duration, and feature count. This study underscores its continued relevance, especially when considering the implementation of new datasets, such as online learning activities of students, in conjunction with diverse sets of algorithms. As discussed in the prior section, most of the studies used public datasets and focused on such demographic data, test scores and family's background. Thus, online learning activities can be used in future work, providing actual students' efforts in assessing their own academic performance.

REFERENCES

[1] M. Riestra-González, M. del P. Paule-Ruíz, and F. Ortin, "Massive LMS log data analysis for the early prediction of course-agnostic student performance," *Computers and Education*, vol. 163, Apr. 2021, doi: 10.1016/j.compedu.2020.104108.

[2] V. Christou *et al.*, "Performance and early drop prediction for higher education students using machine learning," *Expert Systems with Applications*, vol. 225, 2023, doi: 10.1016/j.eswa.2023.120079.

[3] M. Nachouki, E. A. Mohamed, R. Mehdi, and M. Abou Naaj, "Student course grade prediction using the random forest algorithm: analysis of predictors' importance," *Trends in Neuroscience and Education*, vol. 33, 2023, doi: 10.1016/j.tine.2023.100214.

[4] L. Vives *et al.*, "Prediction of students' academic performance in the programming fundamentals course using long short-term memory neural networks," *IEEE Access*, vol. 4, pp. 1–17, 2024, doi: 10.1109/ACCESS.2024.3350169.

[5] W. Qiu, A. W. H. Khong, S. Supraja, and W. Tang, "A dual-mode grade prediction architecture for identifying at-risk student," *IEEE Transactions on Learning Technologies*, vol. 17, pp. 803–814, 2023, doi: 10.1109/TLT.2023.3333029.

[6] M. Adnan *et al.*, "Predicting at-risk students at different percentages of course length for early intervention using machine learning models," *IEEE Access*, vol. 9, pp. 7519–7539, 2021, doi: 10.1109/ACCESS.2021.3049446.

[7] R. Z. Pek, S. T. Ozyer, T. Elhage, T. Ozyer, and R. Alhajj, "The role of machine learning in identifying students at-risk and minimizing failure," *IEEE Access*, vol. 11, pp. 1224–1243, 2023, doi: 10.1109/ACCESS.2022.3232984.

[8] P. Dabhade, R. Agarwal, K. P. Alameen, A. T. Fathima, R. Sridharan, and G. Gopakumar, "Educational data mining for predicting students' academic performance using machine learning algorithms," *Materials Today: Proceedings*, vol. 47, pp. 5260–5267, 2021, doi: 10.1016/j.matpr.2021.05.646.

[9] H. Zeineddine, U. Braendle, and A. Farah, "Enhancing prediction of student success: automated machine learning approach," *Computers and Electrical Engineering*, vol. 89, Jan. 2021, doi: 10.1016/j.compeleceng.2020.106903.

[10] X. Tao *et al.*, "Data analytics on online student engagement data for academic performance modeling," *IEEE Access*, vol. 10, pp. 103176–103186, 2022, doi: 10.1109/ACCESS.2022.3208953.

[11] S. D. Abdul Bujang *et al.*, "Imbalanced classification methods for student grade prediction: a systematic literature review," *IEEE*

Access, vol. 11, pp. 1970–1989, 2023, doi: 10.1109/ACCESS.2022.3225404.
[12] M. Wafi, U. Faruq, and A. A. Supianto, "Automatic feature selection for modified k-nearest neighbor to predict student's academic performance," *Proceedings of 2019 4th International Conference on Sustainable Information Engineering and Technology, SIET 2019*, pp. 44–48, Sep. 2019, doi: 10.1109/SIET48054.2019.8986074.
[13] V. Mhetre and M. Nagar, "Classification based data mining algorithms to predict slow, average and fast learners in educational system using WEKA," in *Proceedings of the International Conference on Computing Methodologies and Communication*, 2017, pp. 475–479, doi: 10.1109/ICCMC.2017.8282735.
[14] E. Alhazmi and A. Sheneamer, "Early predicting of students performance in higher education," *IEEE Access*, vol. 11, pp. 27579–27589, 2023, doi: 10.1109/ACCESS.2023.3250702.
[15] 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, Jun. 2021, doi: 10.1016/j.engappai.2021.104270.
[16] A. I. Adekitan and O. Salau, "The impact of engineering students' performance in the first three years on their graduation result using educational data mining," *Heliyon*, vol. 5, no. 2, 2019, doi: 10.1016/j.heliyon.2019.e01250.
[17] D. Sun *et al.*, "A university student performance prediction model and experiment based on multi-feature fusion and attention mechanism," *IEEE Access*, vol. 11, pp. 112307–112319, 2023, doi: 10.1109/ACCESS.2023.3323365.
[18] S. Roy and A. Garg, "Predicting academic performance of student using classification techniques," in *2017 4th IEEE Uttar Pradesh Section International Conference on Electrical, Computer and Electronics (UPCON)*, Oct. 2017, pp. 568–572, doi: 10.1109/UPCON.2017.8251112.
[19] G. Feng, M. Fan, and Y. Chen, "Analysis and prediction of students' academic performance based on educational data mining," *IEEE Access*, vol. 10, pp. 19558–19571, 2022, doi: 10.1109/ACCESS.2022.3151652.
[20] G. Feng, M. Fan, and C. Ao, "Exploration and visualization of learning behavior patterns from the perspective of educational process mining," *IEEE Access*, vol. 10, pp. 65271–65283, 2022, doi: 10.1109/ACCESS.2022.3184111.
[21] M. Sivasakthi, "Classification and prediction based data mining algorithms to predict students' introductory programming performance," in *Proceedings of the International Conference on Inventive Computing and Informatics*, 2018, pp. 346–350, doi: 10.1109/ICICI.2017.8365371.
[22] G. Feng and M. Fan, "Research on learning behavior patterns from the perspective of educational data mining: Evaluation, prediction and visualization," *Expert Systems with Applications*, vol. 237, 2024, doi: 10.1016/j.eswa.2023.121555.
[23] F. J. Kaunang and R. Rotikan, "Students' academic performance prediction using data mining," in *2018 Third International Conference on Informatics and Computing (ICIC)*, Oct. 2018, pp. 1–5, doi: 10.1109/IAC.2018.8780547.
[24] M. R. Ahmed, S. T. I. Tahid, N. A. Mitu, P. Kundu, and S. Yeasmin, "A comprehensive analysis on undergraduate student academic performance using feature selection techniques on classification algorithms," in *2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*, Jul. 2020, pp. 1–6, doi: 10.1109/ICCCNT49239.2020.9225341.
[25] E. A. Amrieh, T. Hamtini, and I. Aljarah, "Preprocessing and analyzing educational data set using X-API for improving student's performance," in *2015 IEEE Jordan Conference on Applied Electrical Engineering and Computing Technologies (AEECT)*, Nov. 2015, pp. 1–5, doi: 10.1109/AEECT.2015.7360581.
[26] R. Nand, A. Chand, and E. Reddy, "Data mining students' performance in a higher learning environment," in *3rd Novel Intelligent and Leading Emerging Sciences Conference, Proceedings*, 2021, pp. 241–245, doi: 10.1109/NILES53778.2021.9600504.
[27] H. Z. Hashemi, P. Parvasideh, Z. H. Larijani, and F. Moradi, "Analyze students performance of a national exam using feature selection methods," in *2018 8th International Conference on Computer and Knowledge Engineering, ICCKE 2018*, Dec. 2018, pp. 7–11, doi: 10.1109/ICCKE.2018.8566671.
[28] X. Wang, Y. Zhao, C. Li, and P. Ren, "ProbSAP: a comprehensive and high-performance system for student academic performance prediction," *Pattern Recognition*, vol. 137, 2023, doi: 10.1016/j.patcog.2023.109309.
[29] M. K. Assistant, N. Nidhi, S. Majithia, and N. Sharma, "Predictive model for students' academic performance using classification and feature selection techniques," in *2021 2nd International Conference on Computational Methods in Science and Technology (ICCMST)*, Dec. 2021, pp. 106–111, doi: 10.1109/ICCMST54943.2021.00032.
[30] C. F. Rodríguez-Hernández, M. Musso, E. Kyndt, and E. Cascallar, "Artificial neural networks in academic performance prediction: Systematic implementation and predictor evaluation," *Computers and Education: Artificial Intelligence*, vol. 2, 2021, doi: 10.1016/j.caeai.2021.100018.
[31] G. Latif, S. E. Abdelhamid, K. S. Fawagreh, G. Ben Brahim, and R. Alghazo, "Machine learning in higher education: students' performance assessment considering online activity logs," *IEEE Access*, vol. 11, pp. 69586–69600, 2023, doi: 10.1109/ACCESS.2023.3287972.
[32] M. Li, Y. Zhang, X. Li, L. Cai, and B. Yin, "Multi-view hypergraph neural networks for student academic performance prediction," *Engineering Applications of Artificial Intelligence*, vol. 114, 2022, doi: 10.1016/j.engappai.2022.105174.
[33] V. Matzavela and E. Alepis, "Decision tree learning through a predictive model for student academic performance in intelligent m-learning environments," *Computers and Education: Artificial Intelligence*, vol. 2, 2021, doi: 10.1016/j.caeai.2021.100035.
[34] A. Polyzou and G. Karypis, "Feature extraction for next-term prediction of poor student performance," *IEEE Transactions on Learning Technologies*, vol. 12, no. 2, pp. 237–248, 2019, doi: 10.1109/TLT.2019.2913358.
[35] F. Yang and F. W. B. Li, "Study on student performance estimation, student progress analysis, and student potential prediction based on data mining," *Computers and Education*, vol. 123, pp. 97–108, Aug. 2018, doi: 10.1016/j.compedu.2018.04.006.
[36] Dafid and Ermatita, "Filter-based feature selection method for predicting students' academic performance," in *2022 International Conference on Data Science and Its Applications (ICoDSA)*, Jul. 2022, pp. 309–314, doi: 10.1109/ICoDSA55874.2022.9862883.
[37] U. Imdad, W. Ahmad, M. Asif, and A. Ishtiaq, "Classification of students results using KNN and ANN," in *13th International Conference on Emerging Technologies*, 2017, pp. 1–6, doi: 10.1109/ICET.2017.8281651.
[38] W. Nuankaew and J. Thongkam, "Improving student academic performance prediction models using feature selection," in *2020 17th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON)*, Jun. 2020, pp. 392–395, doi: 10.1109/ECTI-CON49241.2020.9158286.
[39] S. D. A. Bujang *et al.*, "Multiclass prediction model for student grade prediction using machine learning," *IEEE Access*, vol. 9, pp. 95608–95621, 2021, doi: 10.1109/ACCESS.2021.3093563.
[40] K. Sixhaxa, A. Jadhav, and R. Ajoodha, "Predicting students performance in exams using machine learning techniques," in *2022 12th International Conference on Cloud Computing, Data Science & Engineering (Confluence)*, Jan. 2022, pp. 635–640, doi: 10.1109/Confluence52989.2022.9734218.

[41] W. Punlumjeak and N. Rachburee, “A comparative study of feature selection techniques for classify student performance,” in *2015 7th International Conference on Information Technology and Electrical Engineering: Envisioning the Trend of Computer, Information and Engineering*, 2015, pp. 425–429, doi: 10.1109/ICITEED.2015.7408984.
[42] M. Zaffar, M. A. Hashmani, and K. S. Savita, “Performance analysis of feature selection algorithm for educational data mining,” in *2017 IEEE Conference on Big Data and Analytics (ICBDA)*, Nov. 2017, pp. 7–12, doi: 10.1109/ICBDAA.2017.8284099.
[43] N. Nidhi, M. Kumar, and S. Agarwal, “Comparative analysis of heterogeneous ensemble learning using feature selection techniques for predicting academic performance of students,” in *2nd International Conference on Computational Methods in Science and Technology*, 2021, pp. 212–217, doi: 10.1109/ICCMST54943.2021.00052.
[44] Y. Shi, F. Sun, H. Zuo, and F. Peng, “Analysis of learning behavior characteristics and prediction of learning effect for improving college students’ information literacy based on machine learning,” *IEEE Access*, vol. 11, no. April, pp. 50447–50461, 2023, doi: 10.1109/ACCESS.2023.3278370.
[45] S. Sengupta, “Towards finding a minimal set of features for predicting students’ performance using educational data mining,” *International Journal of Modern Education and Computer Science*, vol. 15, no. 3, pp. 44–54, 2023, doi: 10.5815/ijmecs.2023.03.04.
[46] N. Rachburee and W. Punlumjeak, “A comparison of feature selection approach between greedy, IG-ratio, Chi-square, and mRMR in educational mining,” in *2015 7th International Conference on Information Technology and Electrical Engineering: Envisioning the Trend of Computer, Information and Engineering*, 2015, pp. 420–424, doi: 10.1109/ICITEED.2015.7408983.
[47] K. Sabaneh and R. Jayousi, “Prediction of students’ performance in e-learning courses,” in *2021 International Conference on Promising Electronic Technologies (ICPET)*, Nov. 2021, pp. 52–57, doi: 10.1109/ICPET53277.2021.00016.
[48] M. Garg and A. Goel, “Preserving integrity in online assessment using feature engineering and machine learning,” *Expert Systems with Applications*, vol. 225, 2023, doi: 10.1016/j.eswa.2023.120111.
[49] V. Shanmugarajeshwari and R. Lawrance, “Analysis of students’ performance evaluation using classification techniques,” *2016 International Conference on Computing Technologies and Intelligent Data Engineering (ICCTIDE'16)*, Kovilpatti, India, 2016, pp. 1-7, doi: 10.1109/ICCTIDE.2016.7725375.
[50] M. Q. Memon, S. Qu, Y. Lu, A. Memon, and A. R. Memon, “An ensemble classification approach using improvised attribute selection,” 2021, doi: 10.1109/ACIT53391.2021.9677093.
[51] A. Kasem, S. N. A. M. Shahrin, and A. T. Wan, “Learning analytics in Universiti Teknologi Brunei: predicting graduates performance,” in *2018 Fourth International Conference on Advances in Computing, Communication and Automation (ICACCA)*, Oct. 2018, vol. 2017, no. 4, pp. 1–5, doi: 10.1109/ICACCAF.2018.8776690.
[52] S. K. Trivedi, A. Sharma, P. Patra, and S. Dey, “Prediction of intention to use social media in online blended learning using two step hybrid feature selection and improved SVM stacked model,” *IEEE Transactions on Engineering Management*, pp. 1–16, 2022, doi: 10.1109/TEM.2022.3212901.
[53] X. Li, K. Jiang, H. Wang, X. Zhu, R. Shi, and H. Shi, “A novel K-means classification method with genetic algorithm,” *Proceedings of 2017 International Conference on Progress in Informatics and Computing*, pp. 40–44, 2017, doi: 10.1109/PIC.2017.8359511.
[54] H. Turabieh, “Hybrid machine learning classifiers to predict student performance,” *2019 2nd International Conference on New Trends in Computing Sciences, ICTCS 2019 - Proceedings*, 2019, doi: 10.1109/ICTCS.2019.8923093.
[55] S. Alraddadi, S. Alseady, and S. Almotiri, “Prediction of students academic performance utilizing hybrid teaching-learning based feature selection and machine learning models,” 2021, doi: 10.1109/WIDSTAIF52235.2021.9430248.
[56] S. S. M. Ajibade, N. B. Ahmad, and S. M. Shamsuddin, “An heuristic feature selection algorithm to evaluate academic performance of students,” *2019 IEEE 10th Control and System Graduate Research Colloquium*, pp. 110–114, 2019, doi: 10.1109/ICSGRC.2019.8837067.
[57] A. H. Nabizadeh, D. Goncalves, S. Gama, and J. Jorge, “Early prediction of students’ final grades in a gamified course,” *IEEE Transactions on Learning Technologies*, vol. 15, no. 3, pp. 311–325, 2022, doi: 10.1109/TLT.2022.3170494.
[58] V. B. Gladshiya and K. Sharmila, “An efficient approach of feature selection and metrics for analyzing the risk of the students using machine learning,” in *2021 International Conference on Advancements in Electrical, Electronics, Communication, Computing and Automation (ICAECA)*, Oct. 2021, pp. 1–6, doi: 10.1109/ICAECA52838.2021.9675507.
[59] F. Abbasi, M. Naderan, and S. E. Alavi, “Anomaly detection in internet of things using feature selection and classification based on logistic regression and artificial neural network on N-BaIoT dataset,” in *2021 5th International Conference on Internet of Things and Applications (IoT)*, May 2021, pp. 1–7, doi: 10.1109/IoT52625.2021.9469605.
[60] L. Rahman, N. A. Setiawan, and A. E. Permanasari, “Feature selection methods in improving accuracy of classifying students’ academic performance,” in *2nd International Conferences on Information Technology, Information Systems and Electrical Engineering*, 2017, no. 1, pp. 267–271, doi: 10.1109/ICITISEE.2017.8285509.
[61] I. Khan, A. Al Sadiri, A. R. Ahmad, and N. Jabeur, “Tracking student performance in introductory programming by means of machine learning,” in *2019 4th MEC International Conference on Big Data and Smart City (ICBDSC)*, Jan. 2019, pp. 1–6, doi: 10.1109/ICBDSC.2019.8645608.
[62] R. Alcaraz, A. Martinez-Rodrigo, R. Zangroniz, and J. J. Rieta, “Early prediction of students at risk of failing a face-to-face course in power electronic systems,” *IEEE Transactions on Learning Technologies*, vol. 14, no. 5, pp. 590–603, 2021, doi: 10.1109/TLT.2021.3118279.
[63] M. Skittou, M. Merrouchi, and T. Gadi, “Development of an early warning system to support educational planning process by identifying at-risk students,” *IEEE Access*, vol. 12, pp. 1–1, 2023, doi: 10.1109/access.2023.3348091.
[64] H. E. Abdelkader, A. G. Gad, A. A. Abohany, and S. E. Sorour, “An efficient data mining technique for assessing satisfaction level with online learning for higher education students during the COVID-19,” *IEEE Access*, vol. 10, pp. 6286–6303, 2022, doi: 10.1109/ACCESS.2022.3143035.
[65] C.-C. Kiu, “Data mining analysis on student’s academic performance through exploration of student’s background and social activities,” in *2018 Fourth International Conference on Advances in Computing, Communication and Automation (ICACCA)*, Oct. 2018, pp. 1–5, doi: 10.1109/ICACCAF.2018.8776809.
[66] A. Tarik, H. Aissa, and F. Yousef, “Artificial intelligence and machine learning to predict student performance during the COVID-19,” *Procedia Computer Science*, vol. 184, pp. 835–840, 2021, doi: 10.1016/j.procs.2021.03.104.
[67] M. Utari, B. Warsito, and R. Kusumaningrum, “Implementation of data mining for drop-out prediction using random forest method,” in *2020 8th International Conference on Information and Communication Technology (ICoICT)*, Jun. 2020, pp. 1–5, doi: 10.1109/ICoICT49345.2020.9166276.
[68] R. Patil, S. Salunke, M. Kalbhor, and R. Lomte, “Prediction system for student performance using data mining classification,” *Proceedings - 2018 4th International Conference on Computing, Communication Control and Automation, ICCUBEA 2018*, Jul. 2018, doi: 10.1109/ICCUBEA.2018.8697770.
[69] E. Al Nagi and N. Al-Madi, “Predicting students performance in online courses using classification techniques,” in *2020*

International Conference on Intelligent Data Science Technologies and Applications (IDSTA), Oct. 2020, pp. 51–58, doi: 10.1109/IDSTA50958.2020.9264113.

[70] H. A. Mengash, "Using data mining techniques to predict student performance to support decision making in university admission systems," *IEEE Access*, vol. 8, pp. 55462–55470, 2020, doi: 10.1109/ACCESS.2020.2981905.

[71] Y. S. Alsalman, N. Khamees Abu Halemah, E. S. AlNagi, and W. Salameh, "Using decision tree and artificial neural network to predict students academic performance," in *2019 10th International Conference on Information and Communication Systems (ICICS)*, Jun. 2019, pp. 104–109, doi: 10.1109/IACS.2019.8809106.

[72] N. Tomasevic, N. Gvozdenovic, and S. Vranes, "An overview and comparison of supervised data mining techniques for student exam performance prediction," *Computers and Education*, vol. 143, Jan. 2020, doi: 10.1016/j.compedu.2019.103676.

[73] S. Alwarthan, N. Aslam, and I. U. Khan, "An explainable model for identifying at-risk student at higher education," *IEEE Access*, vol. 10, no. September, pp. 107649–107668, 2022, doi: 10.1109/ACCESS.2022.3211070.

[74] A. Bressane *et al.*, "Understanding the role of study strategies and learning disabilities on student academic performance to enhance educational approaches: A proposal using artificial intelligence," *Computers and Education: Artificial Intelligence*, vol. 6, 2024, doi: 10.1016/j.caeai.2023.100196.

[75] A. M. Olalekan, O. S. Egwuche, and S. O. Olatunji, "Performance evaluation of machine learning techniques for prediction of graduating students in tertiary institution," in *2020 International Conference in Mathematics, Computer Engineering and Computer Science (ICMCECS)*, Mar. 2020, pp. 1–7, doi: 10.1109/ICMCECS47690.2020.240888.

[76] V. L. Uskov, J. P. Bakken, A. Byerly, and A. Shah, "Machine learning-based predictive analytics of student academic performance in STEM education," *IEEE Global Engineering Education Conference, EDUCON*, pp. 1370–1376, Apr. 2019, doi: 10.1109/EDUCON.2019.8725237.

[77] G. S. Ramaswami, T. Susnjak, A. Mathrani, and R. Umer, "Predicting students final academic performance using feature selection approaches," in *2020 IEEE Asia-Pacific Conference on Computer Science and Data Engineering (CSDE)*, Dec. 2020, pp. 1–5, doi: 10.1109/CSDE50874.2020.9411605.

[78] G. Deeva, J. De Smedt, C. Saint-Pierre, R. Weber, and J. De Weerdt, "Predicting student performance using sequence classification with time-based windows," *Expert Systems with Applications*, vol. 209, 2022, doi: 10.1016/j.eswa.2022.118182.

[79] Q. Ni, Y. Zhu, L. Zhang, X. Lu, and L. Zhang, "Leverage learning behaviour data for students learning performance prediction and influence factor analysis," *IEEE Transactions on Artificial Intelligence*, pp. 1–12, 2023, doi: 10.1109/TAI.2023.3320118.

BIOGRAPHIES OF AUTHORS

Muhamad Aqif Hadi Alias received the B.Eng. degree in electronic engineering from Universiti Teknologi MARA (UiTM), Malaysia, in 2022. Currently, he is pursuing M.Sc. degree in electrical engineering at the School of Electrical Engineering, Universiti Teknologi MARA. His research interest includes the area of classification which specifically focuses on students' performance by obtaining their online learning activities. He can be contacted at email: aqifhadi7@gmail.com.

Najidah Hambali completed her Ph.D. in advanced process control at the School of Electrical Engineering, Universiti Teknologi MARA (UiTM). She received Bachelor of Engineering (B. Eng. Hons.) in electronics engineering from Universiti Sains Malaysia (USM) in 2004. In 2005, she joined Universiti Malaysia Pahang (UMP) as a tutor. After received her Master's Engineering Science (M.Eng.Sc.) in systems and control from The University of New South Wales (UNSW), Sydney, Australia in 2006, she continued working in UMP as a lecturer until 2010. In January 2011, she joined UiTM and later in October 2015, she started her Ph.D. study in UiTM. Currently, she is a senior lecturer in the Centre of System Engineering Studies (CSES), School of Electrical Engineering, College of Engineering UiTM, Shah Alam. Her current research interests are control system and nonlinear modelling for process control. She can be contacted at email: najidah@uitm.edu.my.

Mohd Azri Abdul Aziz is a senior lecturer at the School of Electrical Engineering, College of Engineering, Universiti Teknologi MARA (UiTM), Malaysia. He received the B.Eng. (Hons) degree in electronic engineering from University of Manchester Institute of Science and Technology (UMIST), UK. M.Sc. degree in electrical engineering and Ph.D. degree in electrical engineering, both from Universiti Teknologi MARA, Malaysia. He is also a fellow at Innovative Electromobility Research Lab (ITEM) UiTM, primarily researching on autonomous vehicles. His research area includes pattern classification, object detection, and IoT-based application. Currently, he is supervising and co-supervising 5 master's and 6 Ph.D. students. He has authored and co-authored 8 journals and 16 conference proceedings with 5 H-index and 73 citations. He can be contacted at email: azriaziz@uitm.edu.my.

Mohd Nasir Taib is a professor of control and instrumentation, Universiti Teknologi MARA, Selangor, Malaysia. His area of expertise are control system and signal processing. He is a former director of Malaysia Institute of Transport, Malaysia. He is also Head of the Advanced Signal Processing Research Group (ASPRG), UiTM Shah Alam. He has published more than 500 research and academic journals around the world and won many gold awards in innovations that he participated in. He can be contacted at email: dr.nasir@uitm.edu.my.

Rozita Jailani received her Ph.D. in automatic control and system engineering from Sheffield University, UK. She is currently an associate professor at the School of Electrical Engineering, College of Engineering, and a research fellow at the Integrative Pharmacogenomics Institute (iPROMISE), Universiti Teknologi MARA, Malaysia. Her research interests include intelligent control systems, rehabilitation engineering, assistive technology, instrumentation, artificial intelligence, and advanced signal and image processing techniques. She can be contacted at email: rozita@ieee.org.