



Enhancing Student Success Prediction: A Comparative Analysis of Machine Learning Technique

Majdi Beseiso¹

Accepted: 17 January 2025 / Published online: 28 January 2025
© Association for Educational Communications & Technology 2025

Abstract

Predicting students' success is crucial in educational settings to improve academic performance and prevent dropouts. This study aimed to improve student performance prediction by combining advanced machine learning (ML) approaches. Convolutional Neural Networks (CNNs) and attention mechanisms were used for extracting relevant features from student data, such as grades, attendance, and demographic information. Then, these features were fed into different ML models, including Extreme Gradient Boosting (XGB), Random Forest (RF), K-Nearest Neighbors (KNN), Support Vector Machines (SVM), Boosted Decision Tree (BDT), and Naive Bayes (NB). The models were evaluated on multiple datasets, which enhanced the reliability and generalizability of the findings. Experimental results demonstrated that the XGB classifier, integrated with CNN and attention layer, consistently outperformed other models across two different datasets. The CNN + Attention + XGB model attained F1-score, recall, precision, and accuracy values of 1.0 for dataset 1 and 0.97 for dataset 2. This hybrid approach effectively captured complex patterns in student data and made accurate predictions. The study highlighted the potential of hybrid ML techniques for identifying struggling students and enabling targeted interventions for improved academic outcomes.

Keywords Convolutional neural network · Extreme gradient boosting classifier · Machine learning · Student performance · Support vector machine

Introduction

Student academic success is a paramount concern in educational institutions. Precise student performance prediction can empower educators to proactively intervene, offer tailored support, and significantly enhance the learning experience for students in need (Pasquini et al., 2004). Traditionally, student success has been gauged through examinations and assessments. However, these methods often provide a limited perspective of a student's potential and fail to capture the multifaceted nature of academic achievement (Alsariera et al., 2022).

The timely identification of potential academic problems is crucial. The use of readily accessible student data, such as school records and environmental factors, is essential to predict academic struggles. This timely detection can allow

the implementation of targeted support services, potentially preventing academic failure and redirecting students toward success (Hoffait & Schyns, 2017). The emergence of Machine Learning (ML) techniques presents a great opportunity to enhance the understanding of student success (Rastrollo-Guerrero et al., 2020). ML algorithms can analyze vast educational datasets, encompassing student demographics, prior academic performance, learning styles, and online activity. By identifying complex relationships within this data, ML models can more accurately predict academic performance than traditional methods (Romero & Ventura, 2007).

In pedagogical practice, the reliability and accuracy of ML models in forecasting student dropout are essential. It is because of a well-documented phenomenon known as the Pygmalion effect (Alyahyan & Düşteğör, 2020). In essence, students tend to fulfill the expectations placed upon them. If an ML model incorrectly predicts lower performance for a student, it could create a situation where the student receives less support or experiences lower expectations from educators, potentially leading to a self-fulfilling prophecy

✉ Majdi Beseiso
bsaiso@bau.edu.jo

¹ Department of Computer Science, Al-Balqa Applied, University, Al-Salt 19117, Jordan

and decreased motivation (Bognár & Fauszt, 2022). Therefore, ensuring the fairness and effectiveness of ML models in dropout prediction is essential to avoid negative biases impacting student success.

University dropout rates pose a significant challenge to achieving the cultural enrichment objectives within higher education. Public initiatives aimed at cultural development may be hampered if a substantial portion of the student population fails to complete their degrees. This concern is particularly pressing in Latin America, where dropout rates pose a serious challenge (Ma & Zhou, 2018). Studies have shown dropout rates exceeding 50% in some countries, with Colombia reporting a rate as high as 39% (Jiménez, 2022).

Therefore, it is imperative to develop intelligent decision-making tools to optimize student support and facilitate timely intervention. Such tools can help to identify and assist at-risk students on a larger scale, a task that would be both time-consuming and challenging if undertaken manually (Ojajuni et al., 2021). Moreover, it must be ensured that these intelligent systems are not designed as black boxes. Both institutional leadership and technical staff must have a comprehensive understanding of the functioning of these tools to ensure transparency, build trust, and refine the system's effectiveness over time (Rai, 2020).

Various ML methods including Support Vector Machine (SVM), K-Nearest Neighbor (KNN), Naive Bayes (NB), Artificial Neural Network (ANN), Decision Tree (DT), Logistic Regression (LR), and diverse ensemble methods, have been employed in past studies for predicting student performance (Marbouti et al., 2016; Ojajuni et al., 2021; Xu et al., 2019; Zeineddine et al., 2021). This study compared the performance of different ML models to determine the best model for predicting student academic success. Furthermore, this study employed a novel mix of CNN and attention layers with classic ML models such as SVM and XGB. This hybrid technique combined the capabilities of deep learning (DL) for feature extraction with the strength of conventional models for classification, resulting in higher prediction performance.

The main objectives of this study were to identify the key attributes that most significantly influence academic performance through correlation analysis, to predict students' academic success, and to compare the results of different algorithms for identifying the algorithm best suited for predicting student success and informing academic interventions. The key contributions include:

Integration of Hybrid Models A novel approach was proposed that integrated CNNs and attention mechanisms with standard ML models. This integration aimed to leverage the powerful feature extraction capabilities of DL while maintaining the interpretability and efficiency of classical methods (Al Mayahi & Al-Bahri, 2020; Yaacob et al., 2019).

Multi-Dataset Evaluation This study adopted a multi-dataset evaluation approach as in (Schmucker et al., 2022) to assess the methodology, emphasizing the importance of diverse datasets in ensuring a comprehensive assessment of the generalizability and robustness of the algorithms.

Feature Analysis and Engineering This study employed a thorough feature analysis and engineering approach to enhance the model's ability to predict student performance by considering diverse significant variables (Agyemang et al., 2024). This approach improved the model's accuracy by focusing on the relevant features influencing student outcomes.

Comprehensive Performance Metrics A comprehensive evaluation of the models was conducted using various metrics, including F1-score, recall, precision, and accuracy, to assess their predictive capabilities and ensure a robust understanding of model performance (Agyemang et al., 2024).

Overall, these contributions advance the field of educational data mining by enhancing the accuracy and effectiveness of student performance prediction models and facilitating timely intervention for weak learners. The study proposed a novel hybrid model, combining DL (CNN-attention) and traditional ML (such as SVM, XGB) for improved student performance prediction. The novel aspect of this study was its multimodal approach to forecasting student academic progress, which incorporated multiple advanced ML approaches. This study examined the performance of multiple algorithms across two independent datasets, providing a comprehensive validation of the conclusions through diverse and robust evaluation.

The structure of the remaining paper is as follows. Firstly, the literature review section thoroughly assesses the current research on ML-based models for predicting student performance. Then, the research methodology section elaborates on the data collection and preparation, feature extraction, classification, and hyperparameter tuning. The subsequent section comprehensively discusses the study results. Finally, the last section concludes the study and offers recommendations for further work.

Literature Review

A significant body of research, undertaken by both universities and independent teams, has explored student performance analysis (Alyahyan & Düşteğör, 2020; Haron et al., 2024; Hussain & Khan, 2023; Shirdastian et al., 2019). This field presents a multifaceted challenge. Not only is understanding student success essential, but developing appropriate methodologies for diverse situations adds another

layer of complexity. Numerous studies have demonstrated the potential of academic performance prediction methods to enhance administrative and teaching practices within educational institutions (Agyemang et al., 2024; Hussain & Khan, 2023; Yousafzai et al., 2020). Given the multifaceted nature of academic achievement, a comprehensive approach is necessary to predict student trajectories throughout their university journey (Martins et al., 2021).

ML offers a powerful approach to student performance analysis by leveraging historical data to predict future outcomes. This approach contrasts with traditional methods that rely solely on summative assessments like grades or points, which only provide a final performance picture. By incorporating ML techniques, educators and researchers can gain valuable insights to personalize learning experiences, generate predictive recommendations, and intervene proactively when support is most beneficial. This real-time responsiveness holds the potential to significantly improve educational outcomes (Al Mayahi & Al-Bahri, 2020).

Kotsiantis et al. (2003) analyzed a large dataset of distance learning student performance and assessed the suitability of regression methods for predicting grades. Moreover, various regression algorithms were compared to determine the most appropriate option for accurately forecasting performance and providing analytical insights to support educators' decision-making processes. In (Marbouti et al., 2016), algorithms such as LR, SVM, DT, KNN, ANN, and NB were applied. Experiments were conducted utilizing the complete set of input variables and a subset selected through a feature selection process. The KNN algorithm demonstrated superior performance when all attributes were considered, while SVM and LR performed well on datasets that underwent a feature selection process.

By leveraging the information on current and projected student performance, teachers could proactively intervene and minimize student failures by offering targeted support and additional learning resources. In recent years, educational institutions have increasingly adopted learning management systems to deliver course content. These platforms have transformed traditional learning by incorporating interactive elements, such as gamified activities, simulations, and virtual reality experiences. This shift towards more engaging learning environments has also opened doors for researchers to gather and assess students' data in new and insightful ways (Bhutto et al., 2020; Sowunmi et al., 2016). Moreover, Xu et al. (2019) employed ML algorithms, including DT, neural networks, and SVM, to investigate the potential of internet usage data for predicting academic performance. Their findings indicated that student internet behavior patterns hold promise for predicting academic success, achieving an accuracy range of 71% to 76%.

Hasan et al. (2019) proposed a system to predict students' performance using ML algorithms. The system classified

students into two categories: bad or good performers. A dataset obtained from a university source was used for training the model. KNN and DT classifiers were employed for this purpose. Results showed that the DT classifier achieved maximum accuracy (94.44%). However, Hasan et al. (2019) only relied on accuracy as the performance metric, neglecting other potentially relevant metrics. Similarly, Ojajuni et al. (2021) proposed an ML-based system to classify student performance into binary categories: "good" or "bad." Their model, trained on a university dataset, employed KNN and DT algorithms. While the DT classifier achieved a high accuracy of 94.44%, the evaluation was limited to this single metric, potentially hindering the assessment of its overall effectiveness.

Bhutto et al. (2020) investigated the effectiveness of SVM and LR classifiers. Their model utilized features extracted from a preprocessed dataset acquired from an online learning platform. The aim was to categorize students' academic performance into three classifications: bad, average, and good. The results indicated that SVM achieved a higher accuracy (79%) than LR. To assess the system performance, Bhutto et al. (2020) used a combination of evaluation metrics derived from confusion matrix analysis, including F1-score, precision, accuracy, and recall.

Alsariera et al. (2022) explored the application of ML techniques for predicting student performance at the university level. Their research underscored the significance of student success in higher education institutes and the shortcomings of traditional assessment methods. Alsariera et al. (2022) also examined the existing research on ML applications in this area and noticed a considerable hike in the number of studies and a diverse range of algorithms being utilized. Their analysis suggested that artificial neural networks were the most effective in predicting academic performance, with factors like grades, demographics, and personal characteristics contributing significantly to the prediction accuracy.

Moreover, Yaacob et al. (2019) undertook a comparative study of five models to predict student success. Their findings indicated that the NB model demonstrated the best predictive performance, followed closely by the LR model. Similarly, Oyediji et al. (2020) reported that regression models often outperformed neural networks in terms of predictive accuracy. Alyahyan and Düşteğör (2020) identified student demographics and prior academic achievement were the predominant variables utilized in 69% of the reviewed studies. Moreover, Cumulative Grade Point Average (CGPA) was identified as the most commonly employed factor in predicting student performance.

Furthermore, Altabrawee et al. (2019) built a classifier for predicting student performance in a course at Muthanna University using different ML techniques. Altabrawee et al. (2019) focused on how the time spent on social media and using the Internet for learning impacted performance. Upon

evaluation, the neural network achieved the highest performance (0.807) and accuracy (77.04%), while five factors influencing student results were identified by the DT classifier. Kiu (2018) examined the correlation between students' participation in social activities and their final academic performance. While DT analysis was used to investigate this relationship, it revealed a weak correlation between these two factors.

Gil et al. (2018) highlighted the potential of large-scale data applications within educational contexts. Big data methods for learning analytics were employed in diverse applications, such as student grouping, course recommendation systems, fraud detection, risk identification, student skills assessments, data visualization, prediction of system performance, and collaborative learning opportunities. The predictive analysis focused on student progress, skills prediction, and behavior, which enhanced the significance of their work.

The literature supports the utility of ML algorithms as instrumental tools to devise models that predict student performance. Existing models demonstrate localized efficacy for predicting performance in individual courses. However, numerous challenges, such as model deployment, underfitting, overfitting, and prediction accuracy, require further attention. To address these limitations, this study proposed a novel hybrid approach combining CNN-attention and classic ML models (such as SVM, XGB) to improve student performance prediction. This approach aimed to enhance the accuracy and generalizability of predictions, making it a valuable tool for educators and institutions.

Research Methodology

In this study, two widely used classification algorithms, SVM and XGB, were first implemented and compared based on their predictive accuracy and other performance metrics. The research followed the data mining methodology illustrated in Fig. 1. The initial stage involved data preparation, which included cleaning, pre-processing, and feature engineering to eliminate redundancy, tackle missing values, and enhance the overall data quality and subsequent prediction accuracy. After that, the data was divided into testing and training sets. The training set was employed for training models, while the testing set was employed for evaluating performance. The model selection phase implemented the two classification algorithms, SVM and XGB, to generate predictive models. These models were then compared using various evaluation metrics to identify the most robust and efficient approach. Eventually, the selected model was further analyzed and interpreted to develop a recommendation system for predicting student success.

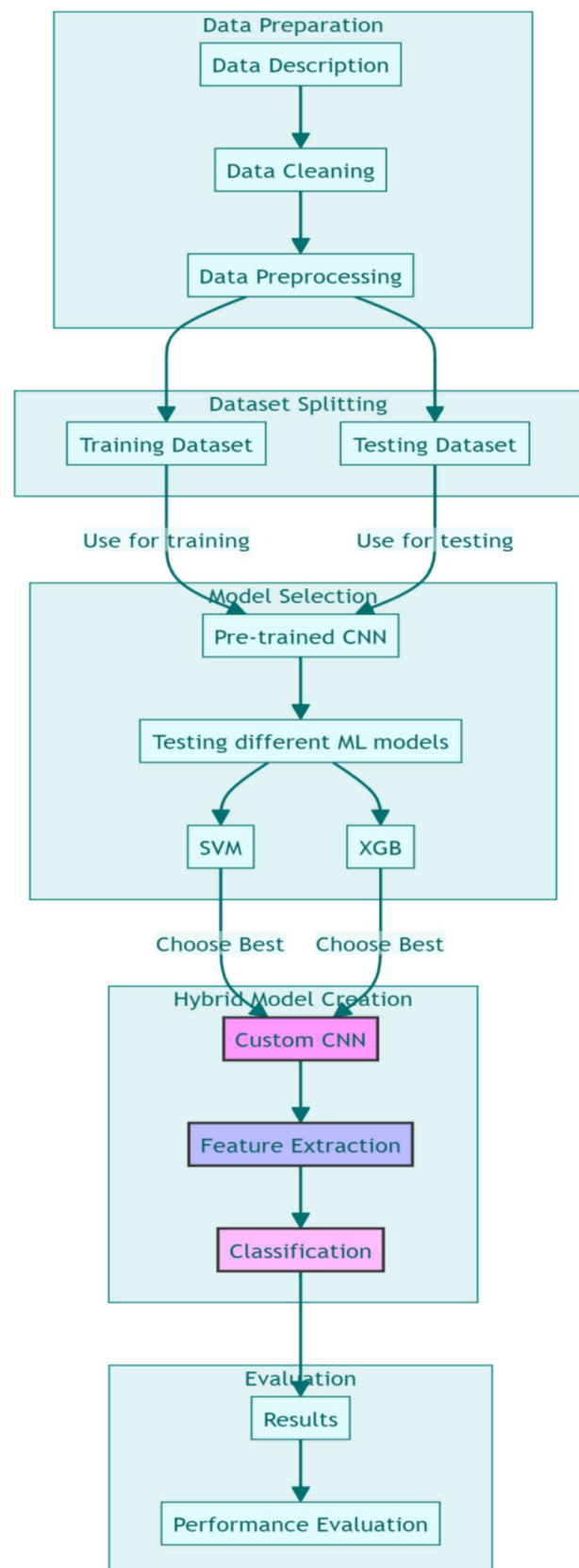


Fig. 1 The methodology used in this study

In the hybrid model creation step, the strengths of custom CNNs and traditional ML classifiers were combined to enhance performance. In this process, a custom CNN was first designed and trained for extracting meaningful features from input data, tailored specifically for the study problem. Once trained, this network was used to extract high-dimensional feature representations that effectively summarized the raw data, reducing its complexity while retaining critical information. These extracted features were then fed into a classifier (like SVM or XGB), to make the final predictions. This integration of CNN-based feature extraction with ML classifiers formed a hybrid model that leveraged the strengths of DL and the decision-making robustness of traditional classifiers, resulting in improved accuracy and efficiency.

Data Description

To compare the robustness and effectiveness of the proposed algorithms, two different datasets, namely the Students' dataset for Grading (Khan et al., 2019) and Predict Students' Dropout and Academic Success (Martins et al., 2021), were used for the experimentations. The first dataset comprised student academic records for a course at Buraimi University College, Oman. This data was obtained from the college's registration department, adhering to all ethical guidelines. The training dataset encompassed data collected over three semesters, consisting of 152 instances and 11 attributes. Of these attributes, ten were considered predictive features, while the remaining one served as the target variable for classification. The features were categorized into two groups: categorical features and numerical features. Categorical features included Gen, Major, Year, Session, Sponsorship, Dorms, and Final_Grade. Numerical features included Attendance, PreReq_Grades, CGPA, and Grade 1_Cont. Final_Grade is the label of the data, indicating whether a student passed or failed.

The dataset comprised academic records of students enrolled in a face-to-face Phonetics and Phonology course. This course focused on the production of speech sounds and patterns by language learners without prior English language knowledge. Student assessment involved three examinations, each contributing 15, 15, and 50 marks to the final grade, and a 20-mark assignment. The first exam was administered in the sixth week of the semester, followed by the assignment. To identify struggling students early, the instructor needed to assess performance after the first exam. Despite the low weight of the first exam (15 marks), students still had to strive for high performance in the remaining 85 marks. As shown in Table 1, the dataset exhibited a well-distributed range of values for attendance, PreReq_Grades, CGPA, and Grade 1_Cont. The attendance ranged from 77 to 100%, reflecting consistent class participation. The minimum

Table 1 Descriptive analysis of numerical features in dataset 1

Feature	Mean	SD	Min	Max
Attendance	94.59	4.74	77.08	100.00
Grade 1_Cont	84.46	9.71	56.70	100.00
CGPA	2.86	0.54	1.51	3.93
PreReq_Grades	76.82	12.78	49.50	99.00

prerequisite grade was 50% (mandatory for course enrollment), and CGPA values ranged from 1.51 to 3.93.

Additionally, a second dataset sourced from the UC Irvine ML repository is incorporated into the analysis. The dataset 2 comprised 3630 instances, each characterized by 35 attributes and a binary classification label. It provided valuable insights into various factors influencing student success predictions.

Data Cleaning, Pre-processing, and Classification

To prepare the datasets for classification, a rigorous data cleaning and preparation process was implemented. This involved addressing missing values, removing outliers, and eliminating irrelevant features for ensuring the quality of data and model performance. Missing data was handled using appropriate imputation techniques. Exploratory Data Analysis (EDA) techniques, such as histograms and box plots, were employed to detect and remove outliers using the interquartile range (IQR) method. Additionally, correlation analysis was performed to identify and remove redundant or weakly correlated features that had minimal predictive power over the target variable.

The correlation matrix, showing the relationships between features and their impact on the predicted final grade, is presented in Fig. 2. Notably, Attendance, Grade 1_Cont, CGPA, and PreReq_Grades exhibited strong negative correlations with Final_Grade. Gender and Major were negatively associated with Dorm status, while Gender and Session were strongly negatively correlated. Moreover, Sessions were highly negatively correlated with CGPA. The 12 features with the lowest correlation to the target variable in the second dataset are highlighted in Fig. 3.

After cleaning and preprocessing the data, the data was split into a testing and training dataset. An 80:20 split was utilized, with 80% of the data allocated for training and the remaining 20% for testing. The training set was used to train the selected classification models, while the testing set served as an independent evaluation benchmark. The performance of these models was assessed using different metrics, which provided valuable insights into the efficacy of the employed classification techniques in predicting student performances on the given dataset.

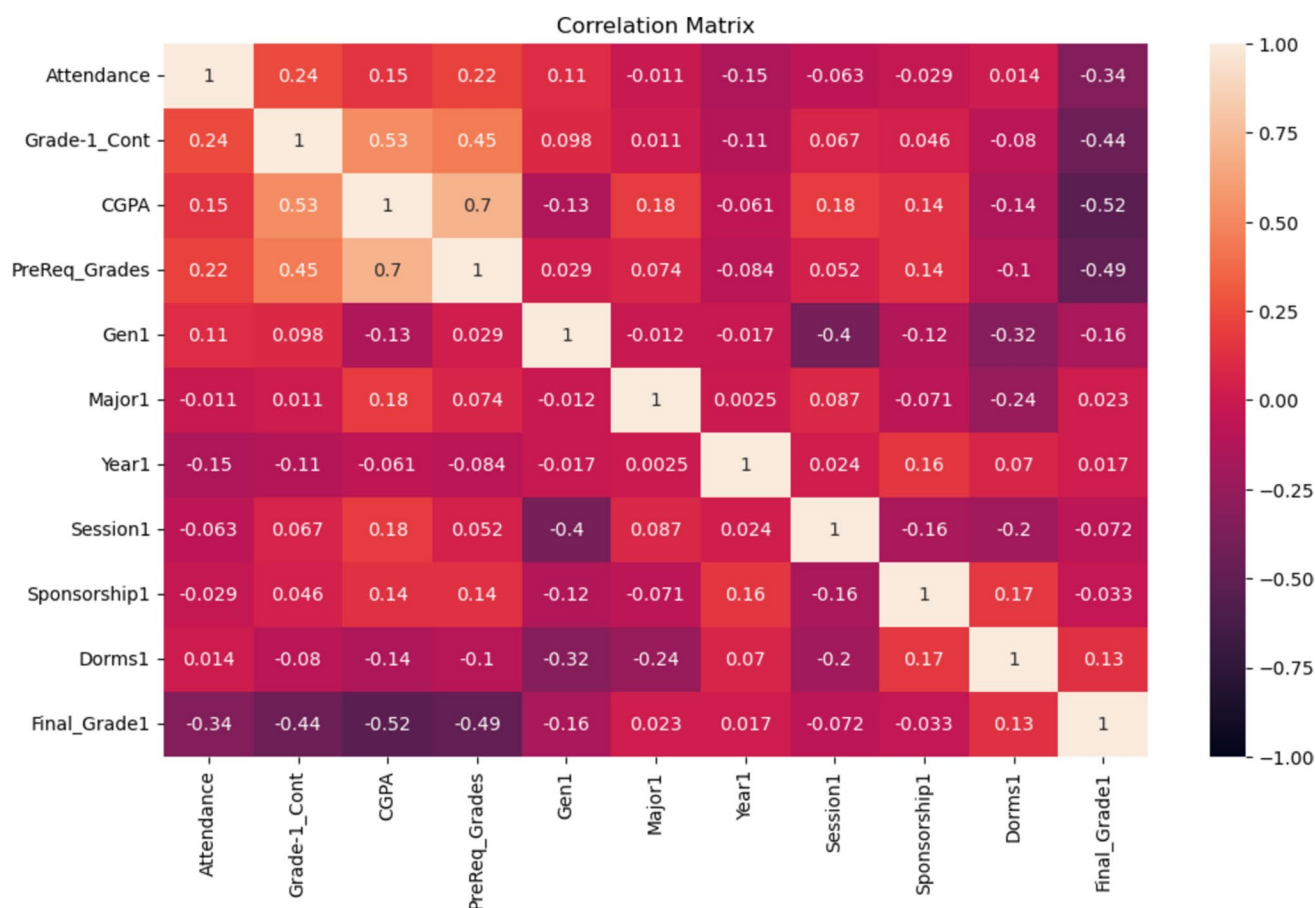


Fig. 2 Correlation of features

Feature Extraction Model

The output of the preceding feature selection stage served as the input to the feature extraction module. To effectively capture both collaborative and temporal relationships among the selected features, a CNN-based model, inspired by the work of Tavakolian et al. (2022), has been proposed. The proposed CNN architecture comprised two parallel convolutional paths, each consisting of two convolutional layers with kernel sizes of 3 and 5, respectively. These layers were designed to extract features from the feature sets, considering both recent and historical information. To further boost the model's capability to identify salient features, an attention layer, as described by Fu et al. (2023), was integrated. This attention mechanism assigned weights to different feature components, allowing the model to prioritize salient information. The convergence of the CNN and attention layers enabled the model to efficiently extract and select the most pertinent features from input data. To stabilize the training process and improve generalization, batch normalization was applied to the combined CNN-attention layers output, as suggested by De and Smith (2020).

The architectural details of the proposed CNN-attention model are illustrated in Fig. 4. The subsequent stage following feature extraction was classification. As depicted in Fig. 4, the proposed architecture incorporated two dense layers, comprising 225 and 2 neurons, respectively. To avoid overfitting, a dropout layer with a 20% dropout rate was employed between the dense layers, as suggested by Hahn and Choi (2020). The total neurons in the initial dense layer were determined by the shape of the input dataset and the total features extracted from the CNN, resulting in 225 neurons. The final dense layer was dedicated to the classification task. The activation functions utilized for these layers were Swish and Softmax.

The model was initially trained as a pre-trained model for the first stage of classification. Subsequently, a two-stage classification approach was employed. The pre-trained model was utilized as a feature extractor, leveraging the CNN-attention layers for extracting 225 features from the first dense layer. These extracted features were then fed into the ML models, including SVM and XGB, for final classification. The model was trained over 500 epochs using a batch size of 64 for both training and validation. Adaptive

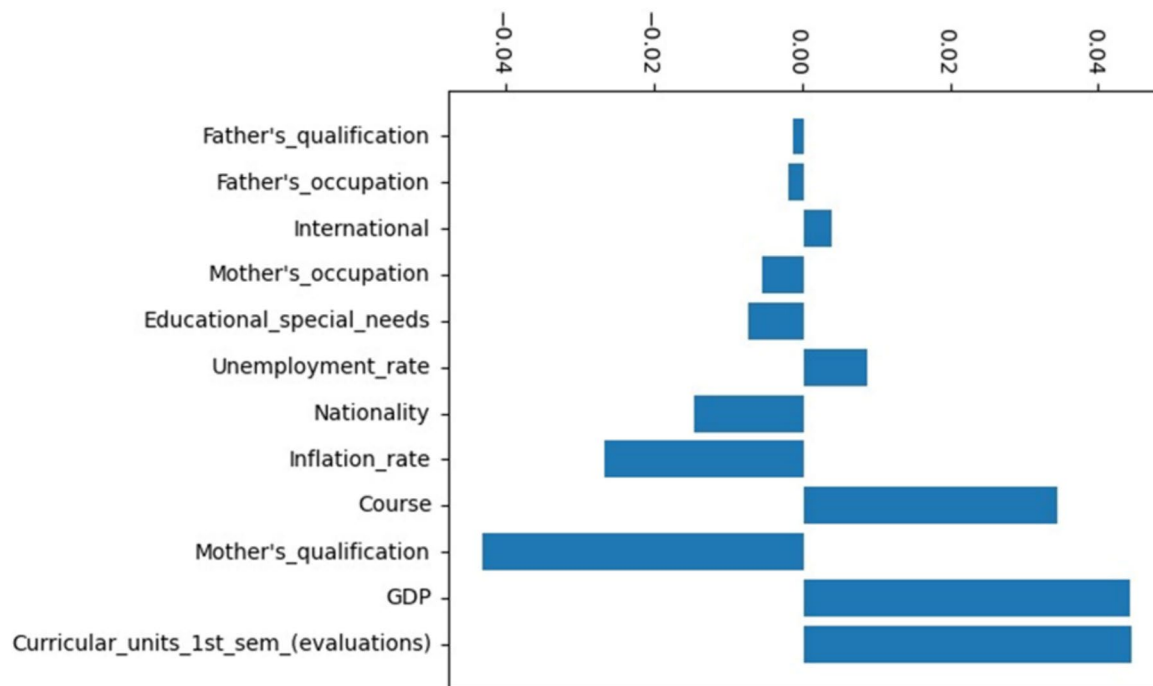


Fig. 3 Top 12 features with the lowest correlation to target

moment estimation (Adam) with focal loss was employed as the optimizer. Focal loss, as proposed by Mukhoti et al. (2020), was incorporated to mitigate class imbalance and prevent bias towards the majority class.

Classification Models

Data mining offers varied methods to predict students' success, such as classification, clustering, nearest neighbor, and logistic regression (Altabrawee et al., 2019; Haron et al., 2024). This study investigated SVM and XGB classification techniques to identify the most efficient model for predicting student final grades. SVM is a supervised ML algorithm employed for regression and classification tasks, while XGB is an ensemble learning technique that unifies the strength of multiple DTs to create an efficient predictive model. This exploration highlighted the breadth of options available within data mining for educational institutions seeking to improve student outcomes.

Support Vector Machine (SVM)

SVM aims to identify the optimal hyperplane that maximizes the margin between classes in a high-dimensional feature space. By utilizing support vectors, a subset of training data points, SVM is effective in high-dimensional spaces while maintaining computational efficiency.

To address the binary classification problem, the SVM algorithm was selected due to its effectiveness in handling imbalanced datasets. Following rigorous data preprocessing and partitioning into training and testing sets, the next step involved hyperparameter tuning to optimize the SVM's performance. This involved selecting the optimal kernel, a pivotal decision determining the shape of the classification boundary. Popular kernel functions, including Sigmoid, Radial Basis Function (RBF), Polynomial, and Linear, were considered, and their parameters were fine-tuned through techniques like grid search. The selected hyperparameters were then employed for training the SVM model on the training set, enabling it to discern the decision boundary that separates the two classes with maximum margin.

During the testing phase, the SVM effectively predicted class labels for new data points based on their positioning relative to the learned decision boundary. This streamlined process enhanced the SVM's ability to generalize to unseen data, enabling accurate binary classification predictions.

Extreme Gradient Boosting (XGB)

XGB, a powerful ML model that combines multiple simpler models to make accurate predictions, has gained significant attention in recent years (Amit, 2019; Zhao et al., 2020). XGB constructs a strong predictive model by sequentially training decision trees, where each subsequent tree focuses on correcting the errors made by its

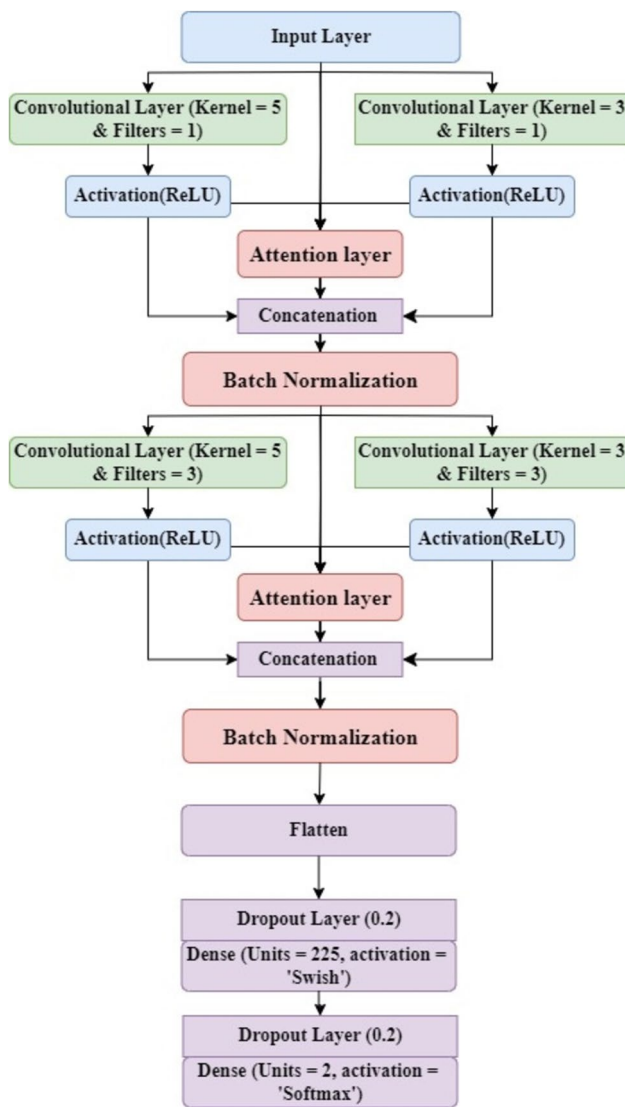


Fig. 4 The architecture of proposed CNN with attention

predecessors. This iterative process, known as gradient boosting, allows the model to improve its predictions gradually over time. The "gradient" part of the name refers to the optimization technique employed to minimize the model errors. By combining the predictions of many decision trees, XGB can achieve high accuracy, even on complex datasets. It is a popular choice for tasks like classification and regression, where it often outperforms other ML methods.

XGB is employed in this study to optimize the training process, which resulted in faster convergence and improved performance. It incorporated regularization techniques to prevent overfitting and employed advanced algorithms for efficient computation. This approach enabled the handling of complex student data and the identification of subtle patterns, avoiding oversimplification of the problem.

Hyperparameter Tuning

To optimize model performance, a grid search cross-validation approach was employed to tune hyperparameters. This technique is essential in ML since it enhances model accuracy and generalizability. A computational grid was defined to explore potential values for key hyperparameters. Hyperparameters, such as the maximum tree depth and number of trees, were considered for XGB. For SVM, the kernel type and regularization parameter were tuned. Additionally, the learning rate, a parameter controlling the step size during optimization, was carefully selected to balance convergence speed and model performance. The optimal hyperparameter values, determined through the grid search process, are presented in Table 2 and Table 3 for XGB and SVM, respectively.

Results

The predictive model developed for this research followed a pipeline encompassing data collection, cleaning, preprocessing, classification, model training, and evaluation. The primary objectives of this study were twofold: predicting academic success and evaluating the performance of the selected ML methods. To assess the effectiveness of these models, the study employed a rigorous evaluation framework utilizing F1-score, recall (sensitivity), precision, and accuracy as performance metrics. The formulas to compute accuracy, precision, recall, and F-measure are provided in Eqs. 1, 2, 3, and 4, respectively.

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \quad (1)$$

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

$$Recall = \frac{TP}{TP + FN} \quad (3)$$

Table 2 Hyperparameter values for XGB

Parameter name	No. of trees	Learning rate	Maximum depth
Value	100	0.01	1

Table 3 Hyperparameter values for SVM

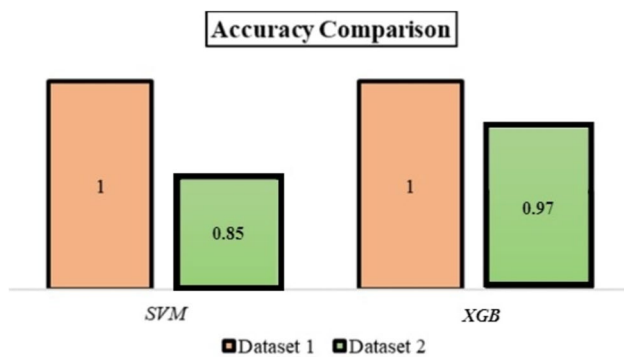
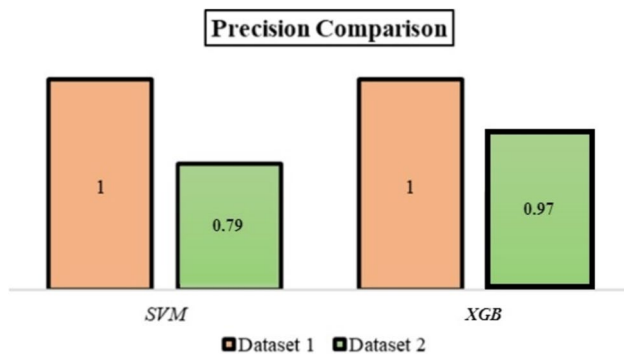
Parameters name	Kernel	Regularization parameter
Value	Linear	1.0

Table 4 Performance evaluation for dataset 1

Classifiers	Accuracy	Precision	Recall	F1Score
CNN + Attention + SVM	1.0	1.0	1.0	1.0
CNN + Attention + XGB	1.0	1.0	1.0	1.0

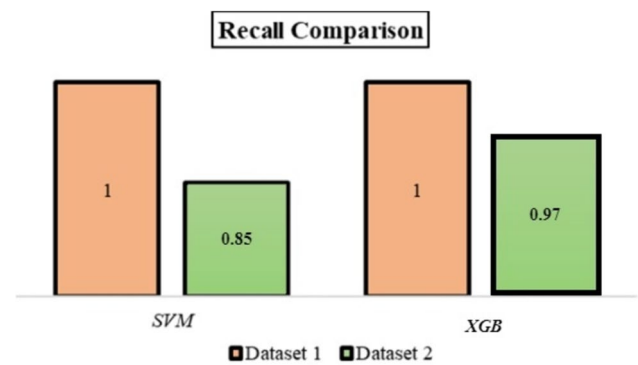
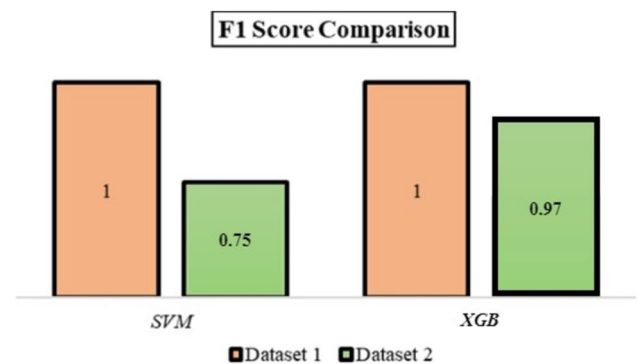
Table 5 Performance evaluation for dataset 2

Classifiers	Accuracy	Precision	Recall	F1Score
CNN + Attention + SVM	0.85	0.79	0.85	0.75
CNN + Attention + XGB	0.97	0.97	0.97	0.97

**Fig. 5** Accuracy comparison of prediction models**Fig. 6** Precision comparison of prediction models

$$F1Score = \frac{2 * precision * recall}{precision + recall} \quad (4)$$

The evaluation results of the prediction models with SVM and XGB are given in Tables 4 and 5 for Dataset 1 and Dataset 2, respectively. The results of this comparison are visually presented in Figs. 5, 6, 7 and 8. Specifically, Fig. 5 illustrates a comparison of accuracy, Fig. 6 presents a comparison of precision, Fig. 7 depicts a comparison of recall (sensitivity), and Fig. 8 shows a comparison of F1 Score for both models and datasets. Despite their distinct

**Fig. 7** Recall/ sensitivity comparison of prediction models**Fig. 8** F1 score comparison of prediction models

algorithmic approaches, both SVM and XGB models demonstrated robust performance metrics, including high precision, accuracy, and sensitivity, highlighting their efficacy in the given classification task. Notably, both algorithms achieved 100% performance on the first dataset, as shown in Table 4. However, their performance on the second dataset was slightly lower, with XGB exhibiting better performance than the other, as shown in Table 5.

Moreover, to gain deeper insights into the relative performance of the proposed classifier, a comparative analysis was conducted involving additional classification algorithms, namely Random Forest (RF), KNN, NB, and Boosted Decision Tree (BDT). These models were trained and evaluated on both datasets, and the results are summarized in Table 6 for Dataset 1 and Table 7 for Dataset 2. The main objective of this research was to determine and assess the most effective model for predicting students' dropout and academic performance. In accordance with the results presented in Tables 6 and 7, NB was determined to be the least accurate model. In contrast, RF, SVM, BDT, KNN, and XGB demonstrated superior performance across the selected evaluation metrics, making them promising candidates for predicting student performance.

Table 6 Comparative performance evaluation of prediction models for dataset 1

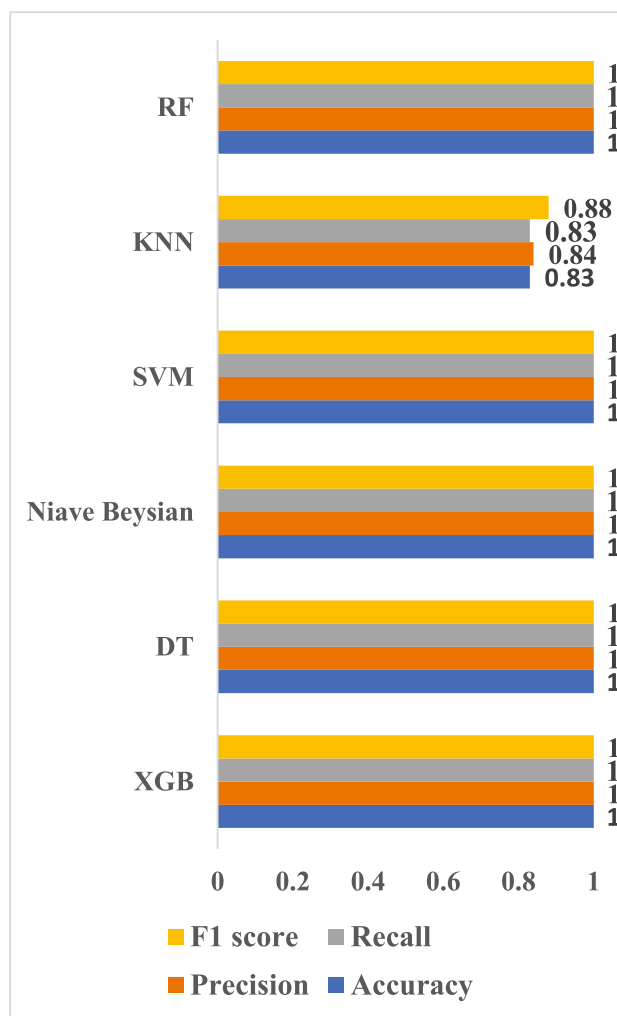
Classifiers	Accuracy	Precision	Recall	F1 Score
RF	1.0	1.0	1.0	1.0
KNN	0.83	0.84	0.83	0.88
SVM	1.0	1.0	1.0	1.0
NB	1.0	1.0	1.0	1.0
DT	1.0	1.0	1.0	1.0
XGB	1.0	1.0	1.0	1.0

Table 7 Comparative performance evaluation of prediction models for dataset 2

Classifiers	Accuracy	Precision	Recall	F1 Score
RF	0.86	0.88	0.86	0.80
KNN	0.90	0.90	0.90	0.88
SVM	0.85	0.79	0.85	0.79
NB	0.54	0.81	0.54	0.60
DT	0.85	0.72	0.85	0.78
XGB	0.97	0.97	0.97	0.97

Figures 9 and 10 present a comparative analysis of the performance metrics of different classification algorithms for Dataset 1 and Dataset 2, respectively. Based on these evaluation metrics, the XGB model consistently outperformed other models in predicting academic success and dropout of students. The superior performance of the CNN-attention-XGB combination can be attributed to the effective feature extraction of the pre-trained CNN, which provided a robust foundation for subsequent XGB classification. Additionally, the augmented feature set derived from the CNN output enhanced the overall predictive capability of the XGB model. The comparative analysis offered a clear and insightful evaluation of the performance of the six algorithms across different datasets, enabling a rapid assessment of their weaknesses and strengths. Overall, the findings showed that XGB is the most competitive classifier compared to others and that student grading is the most significant factor in predicting student performance.

A significant contribution of this work lies in its emphasis on feature importance analysis. The proposed CNN-attention-XGB model leveraged Shapley values to quantify the influence of each feature on the overall predictions. Figure 11 illustrates the feature importance ranking. As evident from Fig. 11, the model's performance is notably enhanced by incorporating features from educational specializations. Additionally, age, occupying a central position within the dataset, proved to be a valuable predictor. The CNN extracted 225 features, with features 127, 213, and 100 emerging as the top three most influential features. Notably,

**Fig. 9** Comparing evaluation metrics of the prediction models for dataset 1

the middle-tier features within the dataset exhibited a notable impact on the final prediction.

Discussion

This study highlights the significant role of machine learning techniques in accurately predicting student success, particularly through the CNN-attention-XGB model. The model achieved superior performance metrics by integrating advanced feature extraction with robust classification, outperforming other algorithms like SVM, RF, and KNN. The importance of key predictors, such as educational specialization and age, underscores the value of understanding student-specific factors. These insights enable educators to implement timely interventions, such as tailored support programs, to improve academic outcomes and reduce dropout rates.

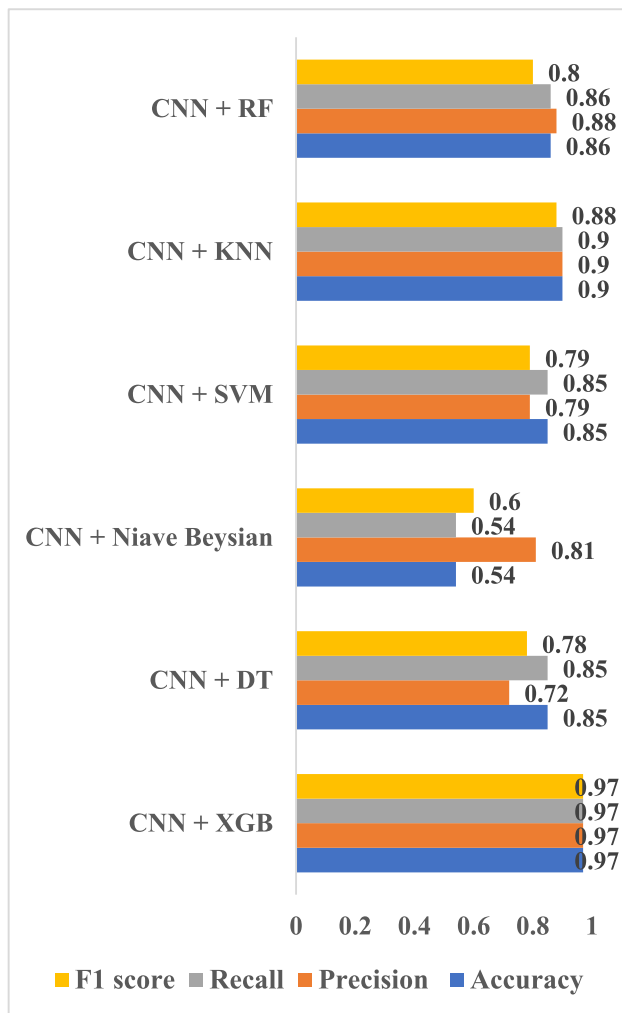


Fig. 10 Comparing evaluation metrics of the prediction models for dataset 2

The comparative analysis of classification models also provided valuable insights into the strengths and weaknesses of different approaches. While simpler models like Naïve Bayes and KNN struggled with accuracy, more sophisticated algorithms, such as XGB, demonstrated superior performance due to their ability to handle complex patterns and feature interactions. These findings highlight the versatility of ML methods and reinforce the importance of careful model selection and feature engineering in achieving optimal results.

Figures 5, 6, 7 and 8 provide a comprehensive view of the model's performance across several evaluation metrics. Figure 5 presents a comparison of accuracy between the SVM and XGB models, showing that both algorithms achieved perfect accuracy (100%) on Dataset 1, while XGB slightly outperformed SVM on Dataset 2. This highlights the ability of XGB to handle more complex data effectively. In terms of Fig. 6 (precision) and Fig. 7 (recall), XGB continued to

demonstrate superior performance, particularly in correctly identifying positive cases and detecting at-risk students. The model's high recall suggests that it is particularly adept at identifying students who may need early intervention. Moreover, Fig. 8 (F1-score) reinforces this point, as XGB maintained a balanced performance in terms of both precision and recall, making it a highly reliable classifier.

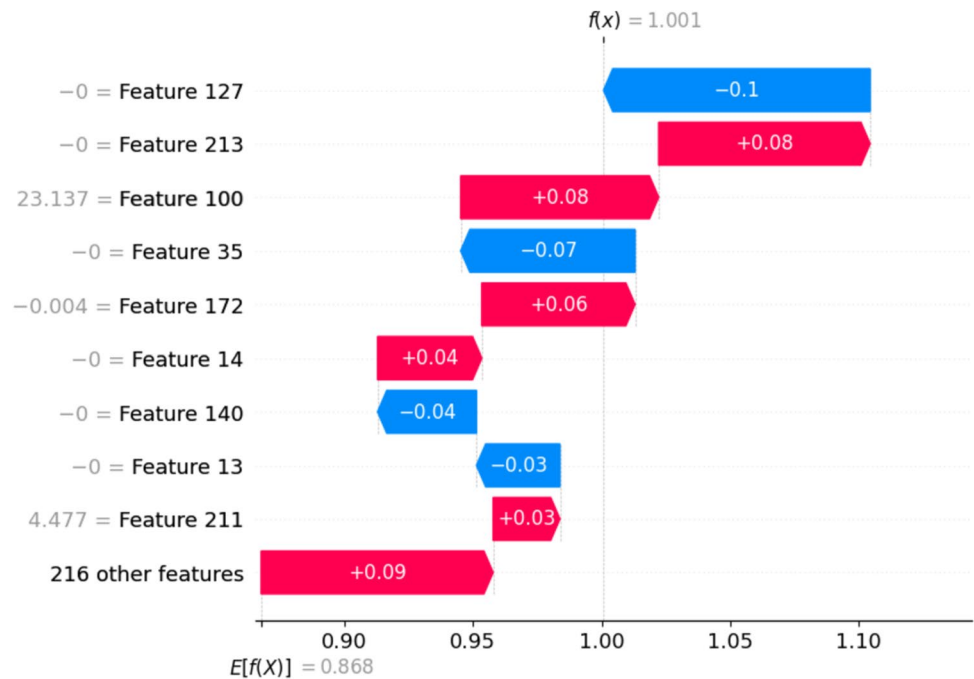
The feature importance analysis, as shown in Fig. 11, further emphasizes the value of educational specialization and age in predicting student success. The CNN component extracted 225 features, which were then processed and prioritized by XGB, highlighting specific data points that are most influential. These insights are vital for developing targeted educational strategies that take into account the individual characteristics of students. Additionally, Figs. 9 and 10, which compare the performance metrics of various models, demonstrate that XGB consistently outperformed other algorithms such as RF, KNN, and NB, reinforcing its superiority in predicting student success and dropout. The results from these figures suggest that XGB, with its capacity to handle complex feature interactions, remains the most effective model for real-world educational data, which is often dynamic and multifaceted.

Despite these promising findings, there are some practical challenges to consider. The complexity of the CNN-attention-XGB model might limit its interpretability for educators and administrators. Simplifying model outputs and focusing on explainable techniques could make these tools more accessible. Additionally, incorporating diverse datasets and exploring other ML methods could further improve predictions. This research paves the way for future innovations, encouraging educational institutions to adopt data-driven strategies to enhance student success.

Conclusion and Future Recommendations

Predicting student academic performance and identifying potential dropouts are crucial challenges in higher education. This study investigated the effectiveness of ML algorithms in addressing these issues. Two datasets from a higher education institution, comprising 152 and 3630 instances, were acquired and preprocessed. A diverse range of ML algorithms, encompassing RF, KNN, SVM, NB, BDT, and XGB, were employed and evaluated to predict student academic performance and dropout rates. The results demonstrated that RF, SVM, BDT, and XGB achieved 100% sensitivity, precision, accuracy, and F1-score in predicting student grading. However, for dropout prediction, these models achieved over 90% accuracy. The combination of XGB with CNN and attention outperformed other classifiers in predicting academic performance, achieving the highest accuracy for both grading and dropout prediction. This suggests that grading

Fig. 11 Importance of each feature in final classification



data plays a more significant role in predicting student performance compared to dropout data.

Moreover, the key features influencing student performance, such as attendance, Grade-1_Cont, CGPA, and Pre-Req_Grades, highlight the importance of academic engagement and prior achievement. To optimize student outcomes, instructors should prioritize support for students with lower CPGA and encourage regular class attendance. Future research can explore the generalization of these findings by applying the models to datasets from different institutions, faculties, and departments. Additionally, visualizing and interpreting decision tree results can provide valuable insights to support low-performing students.

Data Availability Data available on request from the author.

References

- Agyemang, E. F., Mensah, J. A., Ampomah, O. A., Agyekum, L., Akuoko-Frimpong, J., Quansah, A., & Akinlosotu, O. M. (2024). Predicting Students' Academic Performance Via Machine Learning Algorithms: An Empirical Review and Practical Application. *Computer Engineering and Intelligent Systems*, 15(1), 86.
- Al Mayahi, K., & Al-Bahri, D. M. (2020). Machine Learning Based Predicting Student Academic Success. In *International Congress on Ultra Modern Telecommunications and Control Systems and Workshops* (Vols. 2020-October). IEEE. <https://doi.org/10.1109/ICUMTS1630.2020.9222435>
- Alsariera, Y. A., Baashar, Y., Alkawsi, G., Mustafa, A., Alkahtani, A. A., & Ali, N. (2022). Assessment and Evaluation of Different Machine Learning Algorithms for Predicting Student Performance. *Computational Intelligence and Neuroscience*, 2022, 4151487. <https://doi.org/10.1155/2022/4151487>
- Altabrawee, H., Ali, O. A. J., & Ajmi, S. Q. (2019). Predicting Students' Performance Using Machine Learning Techniques. *Journal of University of Babylon for Pure and Applied Sciences*, 27(1), 194–205. <https://doi.org/10.29196/jubpas.v27i1.2108>
- Alyahyan, E., & Düşteğör, D. (2020). Predicting academic success in higher education: literature review and best practices. *International Journal of Educational Technology in Higher Education*, 17(1). <https://doi.org/10.1186/s41239-020-0177-7>
- Amit, K. (2019). Machine Learning Techniques. In *Artificial Intelligence and Soft Computing*. <https://doi.org/10.1201/9781315219738-13>
- Bhutto, S., Siddiqui, I. F., Arain, Q. A., & Anwar, M. (2020). Predicting Students' Academic Performance Through Supervised Machine Learning. In *ICISCT 2020 - 2nd International Conference on Information Science and Communication Technology*. IEEE. <https://doi.org/10.1109/ICISCT49550.2020.9080033>
- Bognár, L., & Fauszt, T. (2022). Factors and conditions that affect the goodness of machine learning models for predicting the success of learning. *Computers and Education: Artificial Intelligence*, 3, 100100. <https://doi.org/10.1016/j.caeai.2022.100100>
- De, S., & Smith, S. L. (2020). Batch normalization biases residual blocks towards the identity function in deep networks. *Advances in Neural Information Processing Systems*, 2020-December, 19964–19975.
- Fu, H., Bai, Y., Guo, T., & Mei, S. (2023). What Can a Single Attention Layer Learn? A Study Through the Random Features Lens. *Advances in Neural Information Processing Systems*, 36.
- Gil, D., Fernández-Alemán, J. L., Trujillo, J., García-Mateos, G., Luján-Mora, S., & Toval, A. (2018). The effect of green software: A study of impact factors on the correctness of software. *Sustainability (Switzerland)*, 10(10), 3471. <https://doi.org/10.3390/su10103471>
- Hahn, S., & Choi, H. (2020). Understanding dropout as an optimization trick. *Neurocomputing*, 398, 64–70. <https://doi.org/10.1016/j.neucom.2020.02.067>
- Haron, N. H., Mahmood, R., Amin, N. M., Ahmad, A., & Jantan, S. R. (2024). An Artificial Intelligence Approach to Monitor and Predict Student Academic Performance. *Journal of Advanced Research in Applied Sciences and Engineering Technology*, 44(1), 105–119. <https://doi.org/10.37934/araset.44.1.105119>

- Hasan, H. M. R., Rabby, A. K. M. S. A., Islam, M. T., & Hossain, S. A. (2019). Machine Learning Algorithm for Student's Performance Prediction. In *2019 10th International Conference on Computing, Communication and Networking Technologies, ICCCNT 2019*. IEEE. <https://doi.org/10.1109/ICCCNT45670.2019.8944629>
- Hoffait, A. S., & Schyns, M. (2017). Early detection of university students with potential difficulties. *Decision Support Systems*, 101, 1–11. <https://doi.org/10.1016/j.dss.2017.05.003>
- Hussain, S., & Khan, M. Q. (2023). Student-Performer: Predicting Students' Academic Performance at Secondary and Intermediate Level Using Machine Learning. *Annals of Data Science*, 10(3), 637–655. <https://doi.org/10.1007/s40745-021-00341-0>
- Jiménez, M. C. (2022). Abandono universitario tarea en la que Iberoamérica se sigue rajando. In *Periódico UNAL*.
- Khan, I., Al Sadiri, A., Ahmad, A. R., & Jabeur, N. (2019). Tracking student performance in introductory programming by means of machine learning. In *2019 4th MEC International Conference on Big Data and Smart City, ICBDS 2019*. IEEE. <https://doi.org/10.1109/ICBDSC.2019.8645608>
- Kiu, C. C. (2018). Data Mining Analysis on Student's Academic Performance through Exploration of Student's Background and Social Activities. In *Proceedings - 2018 4th International Conference on Advances in Computing, Communication and Automation, ICACCA 2018*. IEEE. <https://doi.org/10.1109/ICACCAF.2018.8776809>
- Kotsiantis, S. B., Pierrakeas, C. J., & Pintelas, P. E. (2003). Preventing student dropout in distance learning using machine learning techniques. In *Lecture Notes in Artificial Intelligence (Subseries of Lecture Notes in Computer Science): Vol. 2774 PART 2* (pp. 267–274). https://doi.org/10.1007/978-3-540-45226-3_37
- Koutina, M., & Kermanidis, K. L. (2011). Predicting postgraduate students' performance using machine learning techniques. *IFIP Advances in Information and Communication Technology*, 364 AICT(PART 2), 159–168. https://doi.org/10.1007/978-3-642-23960-1_20
- Ma, X., & Zhou, Z. (2018). Student pass rates prediction using optimized support vector machine and decision tree. In *2018 IEEE 8th Annual Computing and Communication Workshop and Conference, CCWC 2018* (Vols. 2018-January). IEEE. <https://doi.org/10.1109/CCWC.2018.8301756>
- Marbouti, F., Diefes-Dux, H. A., & Madhavan, K. (2016). Models for early prediction of at-risk students in a course using standards-based grading. *Computers and Education*, 103, 1–15. <https://doi.org/10.1016/j.compedu.2016.09.005>
- Martins, M. V., Tollo, D., Machado, J., Baptista, L. M. T., & Realinho, V. (2021). Early Prediction of student's Performance in Higher Education: A Case Study. In *Advances in Intelligent Systems and Computing: Vol. 1365 AIST*. Springer International Publishing. https://doi.org/10.1007/978-3-030-72657-7_16
- Moser, G., Tier, B., Crump, R., Khatkar, M., & Raadsma, H. (2009). A comparison of five methods to predict genomic breeding values of dairy bulls from genome-wide SNP markers. *Genetics Selection Evolution*, 41(1), 1–16. <https://doi.org/10.1186/1297-9686-41-56>
- Mukhoti, J., Kulharia, V., Sanyal, A., Golodetz, S., Torr, P. H. S., & Dokania, P. K. (2020). Calibrating deep neural networks using focal loss. *Advances in Neural Information Processing Systems*, 2020-December, 15288–15299.
- Ojajuni, O., Ayeni, F., Akodu, O., Ekanoye, F., Adewole, S., Ayo, T., Misra, S., & Mbarika, V. (2021). Predicting Student Academic Performance Using Machine Learning. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics): Vol. 12957 LNCS*. Springer International Publishing. https://doi.org/10.1007/978-3-030-87013-3_36
- Oyediji, A. O., Salami, A. M., Folorunsho, O., & Abolade, O. R. (2020). Analysis and Prediction of Student Academic Performance Using Machine Learning. *JITCE (Journal of Information Technology and Computer Engineering)*, 4(01), 10–15. <https://doi.org/10.25077/jitce.4.01.10-15.2020>
- Pasquini, M., Picardi, A., Biondi, M., Gaetano, P., & Morosini, P. (2004). Relevance of anger and irritability in outpatients with major depressive disorder. *Psychopathology*, 37(4), 155–160. <https://doi.org/10.1159/000079418>
- Rai, A. (2020). Explainable AI: From black box to glass box. *Journal of the Academy of Marketing Science*, 48, 137–141.
- Rastrollo-Guerrero, J. L., Gómez-Pulido, J. A., & Durán-Domínguez, A. (2020). Analyzing and predicting students' performance by means of machine learning: A review. *Applied Sciences*, 10(3), 1042.
- Romero, C., & Ventura, S. (2007). Educational data mining: A survey from 1995 to 2005. *Expert Systems with Applications*, 33(1), 135–146. <https://doi.org/10.1016/j.eswa.2006.04.005>
- Schmucker, R., Wang, J., Hu, S., & Mitchell, T. (2022). Assessing the Performance of Online Students - New Data, New Approaches, Improved Accuracy. *Journal of Educational Data Mining*, 14(1), 1–45.
- Shirdastian, H., Laroche, M., & Richard, M. O. (2019). Using big data analytics to study brand authenticity sentiments: The case of Starbucks on Twitter. *International Journal of Information Management*, 48, 291–307. <https://doi.org/10.1016/j.ijinfomgt.2017.09.007>
- Sowunmi, O. Y., Misra, S., Fernandez-Sanz, L., & Medina-Merodio, J. A. (2016). Framework for Academic Advice through Mobile Applications. In *25th International Conference on Information Systems Development, ISD 2016* (pp. 332–344).
- Tavakolian, A., Hajati, F., Rezaee, A., Fasakhodi, A. O., & Uddin, S. (2022). Fast COVID-19 versus H1N1 screening using Optimized Parallel Inception[Formula presented]. *Expert Systems with Applications*, 204, 117551. <https://doi.org/10.1016/j.eswa.2022.117551>
- Varoquaux, G., & Colliot, O. (2023). Evaluating Machine Learning Models and Their Diagnostic Value. In *Neuromethods* (Vol. 197, pp. 601–630). https://doi.org/10.1007/978-1-0716-3195-9_20
- Xu, X., Wang, J., Peng, H., & Wu, R. (2019). Prediction of academic performance associated with internet usage behaviors using machine learning algorithms. *Computers in Human Behavior*, 98, 166–173. <https://doi.org/10.1016/j.chb.2019.04.015>
- Yaacob, W. F. W., Nasir, S. A. M., Yaacob, W. F. W., & Sobri, N. M. (2019). Supervised data mining approach for predicting student performance. *Indonesian Journal of Electrical Engineering and Computer Science*, 16(3), 1584–1592. <https://doi.org/10.11591/ijeecs.v16.i3.pp1584-1592>
- Yousafzai, B. K., Hayat, M., & Afzal, S. (2020). Application of machine learning and data mining in predicting the performance of intermediate and secondary education level student. *Education and Information Technologies*, 25(6), 4677–4697.
- Zeineddine, H., Braendle, U., & Farah, A. (2021). Enhancing prediction of student success: Automated machine learning approach. *Computers and Electrical Engineering*, 89, 106903. <https://doi.org/10.1016/j.compeleceng.2020.106903>
- Zhao, W., Li, J., Zhao, J., Zhao, D., Lu, J., & Wang, X. (2020). XGB model: Research on evaporation duct height prediction based on XGBoost algorithm. *Radioengineering*, 29(1), 81–93. <https://doi.org/10.13164/re.2020.0081>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.