International Journal of Assessment Tools in Education
2026, Vol. 13, No. 1, 186–204
https://doi.org/10.21449/ijate.1711610
journal homepage: https://dergipark.org.tr/en/pub/ijate
Research Article

Predicting middle school students' academic achievement in mathematics with a new hybrid approach

Gürkan Doğan [1*], **Demet Demiralp** [2]

[1]Munzur University, Faculty of Engineering, Department of Computer Engineering, Tunceli, Türkiye
[2]Munzur University, Tunceli Vocational School, Department of Child Care and Youth Services, Tunceli, Türkiye

ARTICLE HISTORY

Received: June 1, 2025

Accepted: Nov. 9, 2025

KEYWORDS

Academic achievement,
Student performance,
Artificial intelligence,
Convolutional neural networks,
Regression.

Abstract: The academic achievement of middle school students is one of the critical steps in their educational journey that affects their future academic and career prospects. During this transition period, students are faced with a variety of challenges related to their families, schools, and the individual themselves. Research shows that academic performance during these formative years is critical for future educational success and overall life success. Therefore, addressing students' academic achievement, which depends on various factors during this period, is of particular importance for both institutional and individual planning and orientation in the future. In this study, a new hybrid approach based on artificial neural networks that enables automatic analysis of data on family, school, and individual factors affecting middle school students' academic achievement in mathematics is proposed. A publicly available student performance dataset was used for training and testing the proposed hybrid approach and other models. This dataset consists of data such as mathematics grades, family information, residential information, and health status information for 395 students enrolled in two public schools in the Alentejo region of Portugal. The proposed approach achieved an R^2 score of 88.6% in experimental studies with this data set, providing approximately 3% higher accuracy than its closest competitor among other methods in the literature.

INTRODUCTION

The academic success of middle school students is one of the most important factors affecting their future academic and career prospects. In particular, these students are in a transitional period and face a variety of challenges, including physiological changes such as growth and development, social pressures, and high school entrance exams. However, students' development is also affected by many factors related to the family, school, and the individual (Huang & Yang, 2018). The family is the first contact an individual has with society after birth, and the first teachers of each individual are their parents. In fact, a survey conducted in 56 schools in Anhui Province, China, showed that factors such as the family's economic situation, the number of children, and parents' expectations of learning were significantly related to the children's high school entrance exam results (Bihua, 2013). Another factor that affects students' development is school. It has been discovered that students' self-sufficiency towards learning

*__Corresponding author at:__ Gürkan DOĞAN gurkandogan@munzur.edu.tr Munzur University, Faculty of Engineering, Department of Computer Engineering, Tunceli, Türkiye

e-ISSN: 2148-7456

plays a mediating role between teachers' behaviors and students' academic success (Lei & Shao, 2015). Another factor is that the individual himself/herself affects academic success. Taking this factor into consideration, researchers have shown in their study that students with higher academic control and ability to manage their emotions have better academic performance (Pekrun *et al*., 2002; Perry *et al*., 2001). Personal characteristics such as gender and self-confidence also play a role. For example, female students generally outperform male students in GPA scores (Karaman & Atar, 2019).

On the other hand, advances in Information Technologies (IT) have caused the databases of institutions and organizations to grow rapidly. This has increased the interest in Business Intelligence (BI) and Data Mining (DM) (Turban *et al*., 2008). Extracting meaningful information from all this data can only be done manually by experts. However, it is not advantageous for experts to analyze this data due to reasons such as the limited number of experts, the labor-intensive work required, excessive time loss and the possibility of missing important details. Alternatively, machine learning and artificial intelligence methods that have the ability to analyze big data can be used. Machine learning methods can play a role in improving students' academic success by analyzing their individual academic performance data (Rivas *et al*., 2021). Additionally, by analyzing big education data, such as student enrollments and demographics, it can identify patterns and insights that can help improve education policies and practices (Jin, 2021). Khasanah (2018) conducted a detailed study on the techniques and features used in educational data mining (EDM) to predict student performance. Accordingly, Decision Trees and Bayesian Networks were the most commonly used machine learning methods, while students' personal information, family information, pre-university characteristics, and university characteristics were the most commonly used features. Another study conducted by Sandra *et al*. (2021) focused on the classification algorithms of machine learning to predict students' learning success. Alongside conventional machine learning techniques, numerous studies have explored the integration and application of various facets of artificial intelligence in education. Zawacki-Richter *et al*. (2019) present a comprehensive overview of artificial intelligence applications within higher education. Certain systematic reviews concentrate exclusively on particular learning domains, such as mathematics (Hwang & Tu, 2021) and health (Sapci & Sapci, 2020). Similarly, some studies limit their research areas and only address a specific educational topic such as student assessment (González-Calatayud *et al*., 2021). In another study, Olabanjo *et al*. achieved high-accuracy predictions with a Radial Basis Function Neural Network (RBFNN) model incorporating cognitive and psychomotor data, emphasizing the importance of feature selection in model success (Olabanjo *et al*., 2022). Rajendran *et al*., on the other hand, achieved the best performance with a stacking algorithm by considering socio-demographic and lifestyle variables, demonstrating that stress and lifestyle have significant effects on academic achievement (Rajendran *et al*., 2022). Similarly, Matzavela and Alepis (2021) integrated decision tree learning into mobile learning environments, demonstrating that parental education, income level, and student employment status are related to academic achievement. A systematic review by Baashar *et al*. (2022) noted the widespread use of artificial neural networks (especially MLP) and their concentration at the higher education level, but highlighted their limitations in real-world applications. In a study conducted in Türkiye, Yağcı (2022) developed models that predicted final grades using only midterm and student grades; showed that the most successful results were obtained with RF and NN algorithms and that these models can be used for early intervention mechanisms. Finally, Hussain and Khan (2023) achieved high accuracy and low error rates in classification and regression tasks with decision trees optimized with genetic algorithms, demonstrating that this approach can identify at-risk students before the exam. However, similar studies in the literature have generally conducted one-way analyses based on students' past grades or online activity. Furthermore, student failure rates in Portugal are quite high compared to the European average, but local, data-driven models aimed at predicting or explaining this situation have previously been developed using simple machine learning models with low accuracy. These studies have typically applied a single algorithm (e.g., only decision trees or only Naive Bayes). Therefore, systematic comparisons of the performance of different data mining algorithms (DT, RF, ANN, SVM) have been lacking. In this context, the current study aims to answer the following research questions:

- To what extent can the academic achievement of secondary school students in Portugal be accurately predicted, specifically in Mathematics, using different data mining algorithms (e.g., SGD, RF, ANN, SVM)?
- Which data mining algorithm provides the highest accuracy and generalizability in predicting student achievement?

This study introduces a new methodology utilizing convolutional neural networks (CNN), a form of artificial intelligence, to forecast the academic performance of middle school students in mathematics. The proposed approach uses the CNN model as a feature extractor and features are selected from the extracted features using the F-regression technique. Finally, students' final grades are predicted using the selected features using the random forest regression method. Since the dataset includes personal information, family information, school information and information about the environment of the students, the effects of various factors affecting academic success are also taken into consideration in this study. In the experimental studies, the proposed methodology attained an R^2 accuracy around 3% superior to that of the nearest competing method.

METHOD

In this section, firstly, the student performance dataset is introduced to predict the performance of students in mathematics. Secondly, a newly developed neural network model is explained in detail. Thirdly, the optimization method used in training this ANN model and the learning rate planning are mentioned. Fourthly, feature selection and finally, the proposed approach are explained in detail.

Dataset

In this study, the student performance dataset (Cortez, 2014; Cortez & Silva, 2008) was used for training and testing the proposed approach and the state-of-the-art methods that were re-implemented. This dataset consists of data such as the success grades received in mathematics during the 2005-2006 academic year, information about their families, information about their environment, information about their health status, etc. of 395 students in two public schools in the Alentejo region of Portugal. The attributes, attribute types and descriptions of this data set are given in Table 1. As can be seen in this table, there are 33 attributes. Since “G3” among these attributes represents the final grade students received in mathematics, it was used as the estimated dependent variable in this study. The attributes other than the “G3” attribute were used as independent variables for the regression methods. Furthermore, since the dataset contained no missing data, missing data handling techniques were not applied. The training and testing of the regression methods were performed by randomly separating 276 (70%) of the data belonging to 395 students for training and 119 (30%) for testing in this dataset.

Artificial Neural Networks

Artificial neural networks (ANN) are a machine learning approach modeled after biological neural networks. This strategy simplifies the intricate architecture of the organic nervous system (e.g., neurons, dendrites, and axons) into a basic framework, hence facilitating the development of a mathematical processing model. The most unique characteristic of an artificial neural network (ANN) is its capacity to address intricate and unformulatable issues by optimizing the weights of connections among the interlinked neurons where information is processed. This procedure, which underpins deep learning techniques, is referred to as network training. Artificial Neural Network (ANN) models can be employed in several tasks, including data classification, regression, and pattern recognition (Pérez-Gomariz *et al*., 2023; Yüksel *et al*., 2023).

Table 1. *Attributes and descriptions in the student performance dataset.*

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

In this study, a new artificial neural network model based on convolutional neural networks was developed to predict students' academic achievement in mathematics. Figure 1 presents the block diagram of the constructed ANN model. This model fundamentally comprises four steps. In the first step, a data vector of size 58x1 is taken as input and passed through the batch normalization layer. In the second step, the normalized 58x1 data is passed through a one-dimensional convolution layer (Conv1D) with 11, 5, 3, 2 and 1-dimensional kernels. Then, by passing through the linear and parametric ReLU activation function and the batch normalization layer, new features of sizes 48x24, 54x24, 56x24, 57x24 and 58x24 are obtained, respectively. In addition to these processes, in this step, batch normalization was performed by ensuring that new information is learned with the fully connected layer when the data order is unimportant. Then, in this step, the features obtained from the batch normalization steps are concatenated. In the third step, the concatenated features are expressed as a single vector by passing through the serialization layer. In the fourth and last step, the final grade, which is the performance indicator of the students in the mathematics course, is predicted from the serialized features with the fully connected layer.

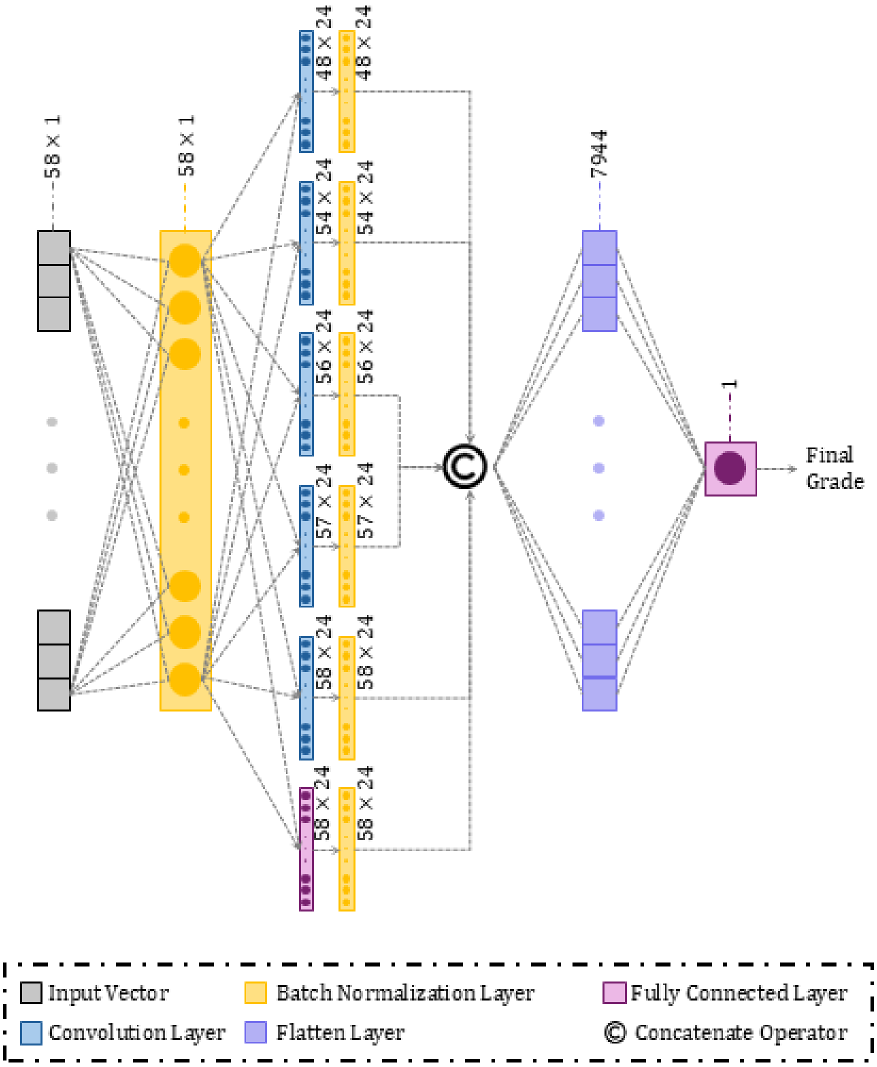

Figure 1. *Schematic of the artificial neural network (ANN) model.*

Learning Rate Scheduling

In this study, the "ReduceLROnPlateau" (RLRP) function in the Keras deep learning API is used to plan the learning rate. The RMSProp optimizer is employed to train the ANN model utilized as a feature extractor. This function reduces the learning rate by the factor parameter when the learning of the ANN model stabilizes. The determination of whether learning becomes stationary is based on the observation of the validation loss and the patience parameter value. This function enhances the efficiency of the ANN model's learning process. The parameters and their corresponding values utilized in the training of the created ANN model are presented in Table 2. The settings for configuring the learning rate are established as follows: patience is set to 50, factor is 0.1, the initial learning rate (Init_LR) is $1x10^{-2}$, and the minimal learning rate (Min_LR) is $1x10^{-9}$. The model's batch size is configured to 128, and the number of epochs is established at 1000. The hyperparameters used in the training process of the ANN model were optimized through an experimental iterative search process to maximize the accuracy and generalization success of the model.

Table 2. *Hyperparameters employed during the training of the ANN.*

Parameter	Epoch	Batch	Optimizer	LR Scheduling	Monitor	Patience	Factor	Init. LR	Min LR
Value	1000	128	RMSprop	RLRP	Val. Loss	50	0.1	$1e^{-2}$	$1e^{-9}$

Feature Selection

Feature selection is the act of eliminating or choosing specific features from a dataset to enhance the performance of models employed in tasks such as classification and regression. Unnecessary and irrelevant features in the dataset can increase the complexity of the model and lead to overfitting of the model or low generalization performance of the model. Therefore, making the right feature selection ensures that the model shows higher performance in terms of accuracy and speed in parallel with the feature dimension reduction (Doğan *et al*., 2024; Doğan & Ergen, 2023; İmak *et al*., 2023). There are many methods for feature selection such as filter, wrapper, embedded, intrinsic and univariate feature selection. In this study, many feature selection methods such as FPR test, mutual information, chi-squared, XGBRegressor, F-regression etc. were tried iteratively for feature selection and the best features were obtained with F-regression, which is one of the univariate feature selection methods. F-regression relies on the F-statistic, defined by the Equation 1 provided below:

$$F = \frac{r_{xy}^2}{1-r_{xy}^2} x(n-2) \tag{1}$$

In Equation 1, F represents the regression score, n denotes the number of samples, and r_{xy} signifies the correlation coefficient between the independent variable X and the dependent variable Y. r_{xy} can be expressed as follows:

$$r_{xy} = \frac{\sum_{i=1}^{n}(x_i-\vec{x})(y_i-\vec{y})}{\sqrt{\sum_{i=1}^{n}(x_i-\vec{x})^2}\sqrt{\sum_{i=1}^{n}(y_i-\vec{y})^2}} \tag{2}$$

In Equation 2, n represents the number of samples, x_i and y_i represent the ith independent and dependent variable values, respectively, $\vec{x}$ and $\vec{y}$ represent the mean of the x independent variable values and the mean of the y dependent variable values, respectively.

The F-statistic adheres to the F-distribution with 1 and n-2 degrees of freedom and can be transformed into a p-value. Consequently, the features with the greatest ratings may be chosen.

Proposed Hybrid Approach

This study proposes a new hybrid approach utilizing convolutional neural networks to forecast students' performance in mathematics. The suggested approach comprises four distinct phases. Initially, the dataset is partitioned into 70% for training and 30% for testing, followed by the execution of one-hot encoding and standard scaling as data pretreatment measures. In the second stage, a new artificial neural network (ANN) model, derived from convolutional neural networks as seen in Figure 1, is constructed and utilized as a feature extractor. 7944 features are taken from the serialization layer of the created artificial neural network model. In the third stage, the most pertinent features are selected by excluding irrelevant and superfluous features from the extracted set using the F-regression feature selection approach. Consequently, the 66 most pertinent features are chosen from the 7944 features derived from the ANN model. In the last stage, the random forest regression technique is employed to forecast students' final mathematics grades utilizing the selected attributes. Figure 2 illustrates the design of the proposed hybrid approach.

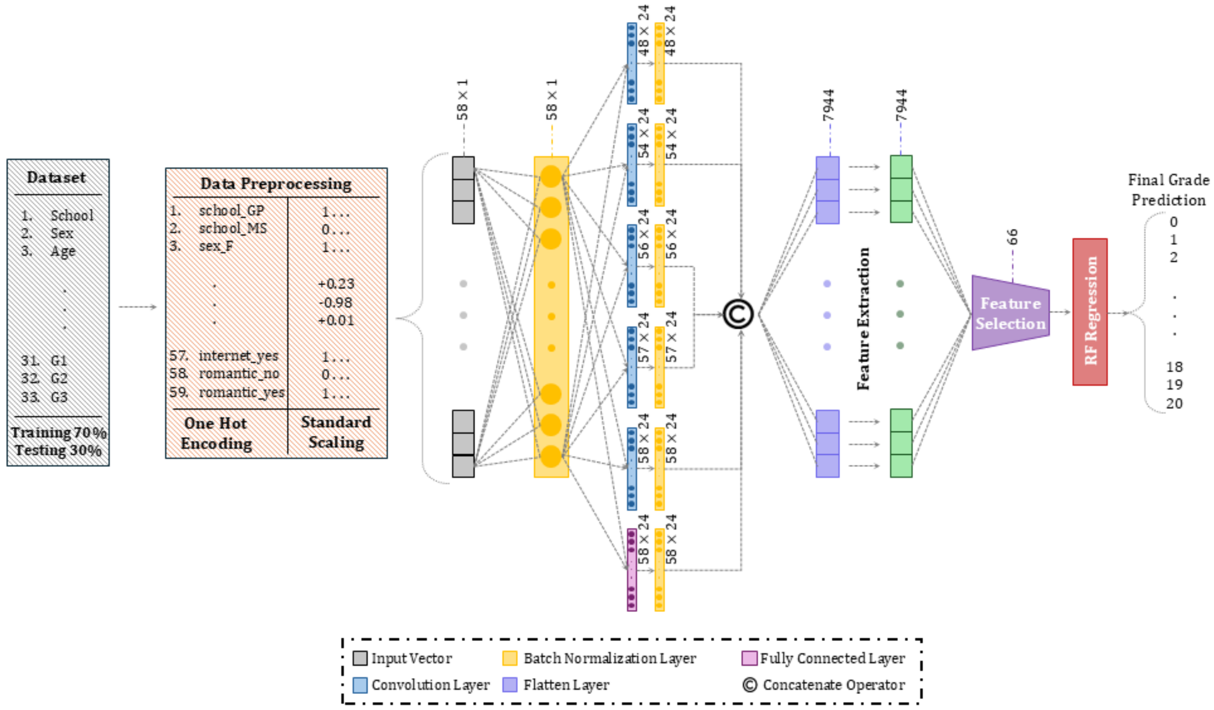

Figure 2. *Schematic of the proposed hybrid approach.*

Data Analysis and Evaluation Metrics

In this study, the prediction of students' performance in mathematics is considered as a regression problem. In order to overcome this problem, the proposed hybrid approach and the state-of-the-art methods such as Linear Reg., Robust Reg., Ridge Reg., Lasso Reg., Poly. Reg., SGD Reg., RF Reg., SVM Reg. and XGB Regression are used in the experiments. In order to compare the performances of all these approaches and methods, the mean absolute error (MAE), mean square error (MSE), root mean square error (RMSE) and R^2 metrics, which are widely available in the literature, are used. The equations of the MAE, MSE, RMSE and R^2 metrics are given in (3), (4), (5) and (6), respectively.

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|y_i - \hat{y}_i| \tag{3}$$

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2 \tag{4}$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2} \tag{5}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{N}(y_i - \bar{y})^2} \tag{6}$$

In the equations of these metrics, y_i denotes the true value, $\hat{y}_i$ signifies the predicted value, $\bar{y}$ indicates the mean of the true values, and N represents the sample size. MAE quantifies the discrepancy between expected and actual values by calculating the mean of the absolute differences between them. The Mean Squared Error (MSE) quantifies the deviation of projected values from actual values by calculating the average of the squared differences between them.

RMSE is derived by extracting the square root of the MSE data and serves the same function as MSE. R^2 quantifies the degree of fit between anticipated values and actual values. Given that $R^2 \in [0, 1]$, it can likewise be represented as a percentage.

The performance of the methods was analyzed using standard performance metrics such as MAE, MSE, RMSE and R^2 and 95% confidence interval (CI). To calculate the confidence intervals, the widely used bootstrap method, which does not rely on parametric assumptions, was used and 10,000 replications were performed. On the other hand, the Wilcoxon Signed-Rank Test was applied to evaluate the statistical significance of the results obtained by comparing the proposed approach with the most current models in terms of MAE, MSE, RMSE, and R^2. In this method, the percentage improvement provided by the proposed model compared to other models was calculated using the improvement criterion, and the statistical significance of the performance differences between the models was assessed using the *p*-value criterion. Furthermore, the significance criterion indicates the significance level of the difference based on the p-value. Accordingly, *** indicates a *p*-value of .001 or less, representing a very highly significant difference. ** denotes a *p*-value between .001 and .01, indicating a highly significant difference. * corresponds to a p-value between .01 and .05, indicating a significant difference, while ns signifies a *p*-value greater than .05, indicating that the difference is not statistically significant.

The model's losses and accuracies were measured using standard performance metrics such as MAE, MSE, RMSE, and R^2, and validated with independent test set results. This achieved balanced performance against both training data and unseen data. In regression problems, model loss values are measured using metrics such as MAE, MSE, and RMSE, while accuracy is measured using R^2. The closer a model's MAE, MSE, and RMSE values are to zero, the lower the error. Conversely, the closer a model's R^2 values are to one or 100%, the higher the model's accuracy.

Various techniques were used to increase and evaluate the model's generalizability. Firstly, ReduceLROnPlateau automatically reduces the learning rate when the validation loss fails to improve over a certain number of epochs, preventing the model from learning too quickly and getting stuck in local minima. This reduces the risk of overfitting and improves the model's performance on unseen data. Secondly, confidence intervals measure the uncertainty of model predictions and performance metrics; narrow and stable confidence intervals indicate consistent model performance and, therefore, high generalizability. Furthermore, the Bootstrap method (e.g., 10,000 resampling) was used to obtain the distribution of model performance across different data samples. This approach contributes to the assessment of generalizability by demonstrating that the model will perform similarly across different datasets. The combination of these techniques provides generalizability information by ensuring that the model not only achieves high accuracy but also produces reliable results on new and unseen data.

Additionally, the effects of features such as school, age, sex, family, etc. in the dataset on the G3 output predicted by the proposed approach were investigated using the SHAP technique. Since the proposed approach is a complex and nested method that uses ANN for feature extraction, f_regression for feature selection, and Random Forest Regressor for prediction, the effect of the features on the output was performed using SHAP's KernelExplainer type.

The experimental training and testing of the proposed approach, together with the re-implemented state-of-the-art techniques, were performed on a workstation equipped with an Intel(R) Core(TM) i5 CPU @ 2.50GHz, 64 GB RAM, and a GeForce RTX 3090 @ 24GB graphics card. Additionally, the proposed approach was coded using software technologies including Python 3.10, Keras 2.10, and TensorFlow 2.10, while the Scikit-learn 1.2.2 machine learning package was employed to re-implement other advanced methodologies.

RESULTS

This study utilized a student performance dataset for the training and testing of the proposed methodology and contemporary methodologies. This dataset comprises the academic performance of 395 middle school study in Portugal, together with details regarding their familial backgrounds and environmental factors. 70% of this dataset was used for training the proposed approach and other state-of-the-art methods, and 30% for testing. The performance of the proposed approach was compared with state-of-the-art methods such as Linear Reg., Robust Reg., Ridge Reg., Lasso Reg., Poly. Reg., SGD Reg., RF Reg., SVM Reg., and XGB Regression, which are widely available in the literature and have high accuracy performance. In order to apply the experimental strategy applied to the proposed approach to these state-of-the-art methods in a healthy way, the state-of-the-art methods were re-implemented. In the training process of the proposed approach, the batch size was set to 128 and the number of iterations (epochs) was set to 1000.

The mean squared error (MSE) curve formed during the training process of the ANN model developed for the proposed approach is given in Figure 3a. Accordingly, the MSE, which was initially very high at approximately 429.95 and 101.63 for the training and test sets, respectively, decreased to approximately 1.40 and 4.07 starting from the 389th iteration. However, the curve of the learning rate that occurred over 1000 iterations is also given in Figure 3b. Accordingly, the learning rate (LR), which was 1×10^{-2} at the beginning, decreased to 1×10^{-3} in the 56th iteration, 1×10^{-4} in the 339th iteration and 1×10^{-9} in the 589th iteration. The lowest MSE was achieved with a learning rate of 1×10^{-5} for the training and test sets.

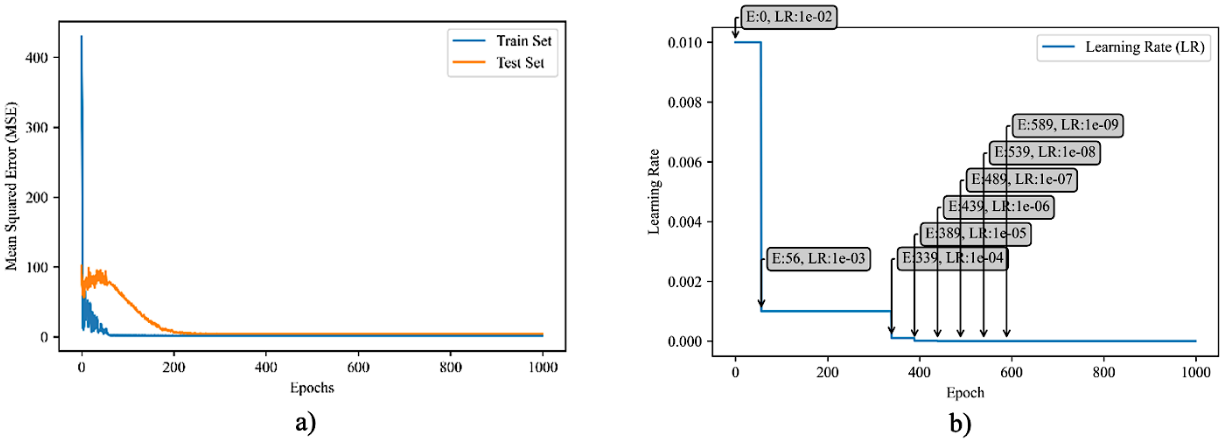

Figure 3. *Graphs of the training of the proposed ANN model: a) MSE curve and b) LR curve.*

In the second stage of the proposed approach, 7944 features were extracted from ANN. In the third stage, the most relevant features were selected using the F-regression technique. However, since the number of the most relevant features in feature selection is also the most important factor determining accuracy and error performance, the model's accuracy and error rates were checked by increasing the number from 1 to 7944 features and selecting the most relevant features. Figure 4 shows the minimum and maximum values of the MAE, RMSE and R^2 metrics graphically with the most relevant features selected according to the number of features. According to the mean absolute error graph given in Figure 4a, the highest MAE occurred with 1 feature with a score of 1.34, while the lowest MAE occurred with 67 features with a score of 1.05. According to the root mean square error graph given in Figure 4b, the highest RMSE was 2.06 with 1 feature, while the lowest RMSE was 1.58 with 66 features. According to the R^2 graph given in Figure 4c, the highest R^2 was 88.52% with 66 features, while the lowest R^2 was 80.57% with 1 feature. According to the graphs in Figure 4, RMSE and R^2 reached their best values with 66 features. For this reason, 66 features were used in the regression phase, which is the fourth stage of the proposed approach.

In the fourth stage of the proposed approach, the performance of the students was estimated with the random forest regression method. Student performance prediction was evaluated using the proposed model as well as state-of-the-art methods. Model performance was analyzed using standard performance metrics such as MAE, MSE, RMSE, and R^2, along with a 95% confidence interval (CI). Table 3 shows the performance comparison of the proposed approach and state-of-the-art methods.

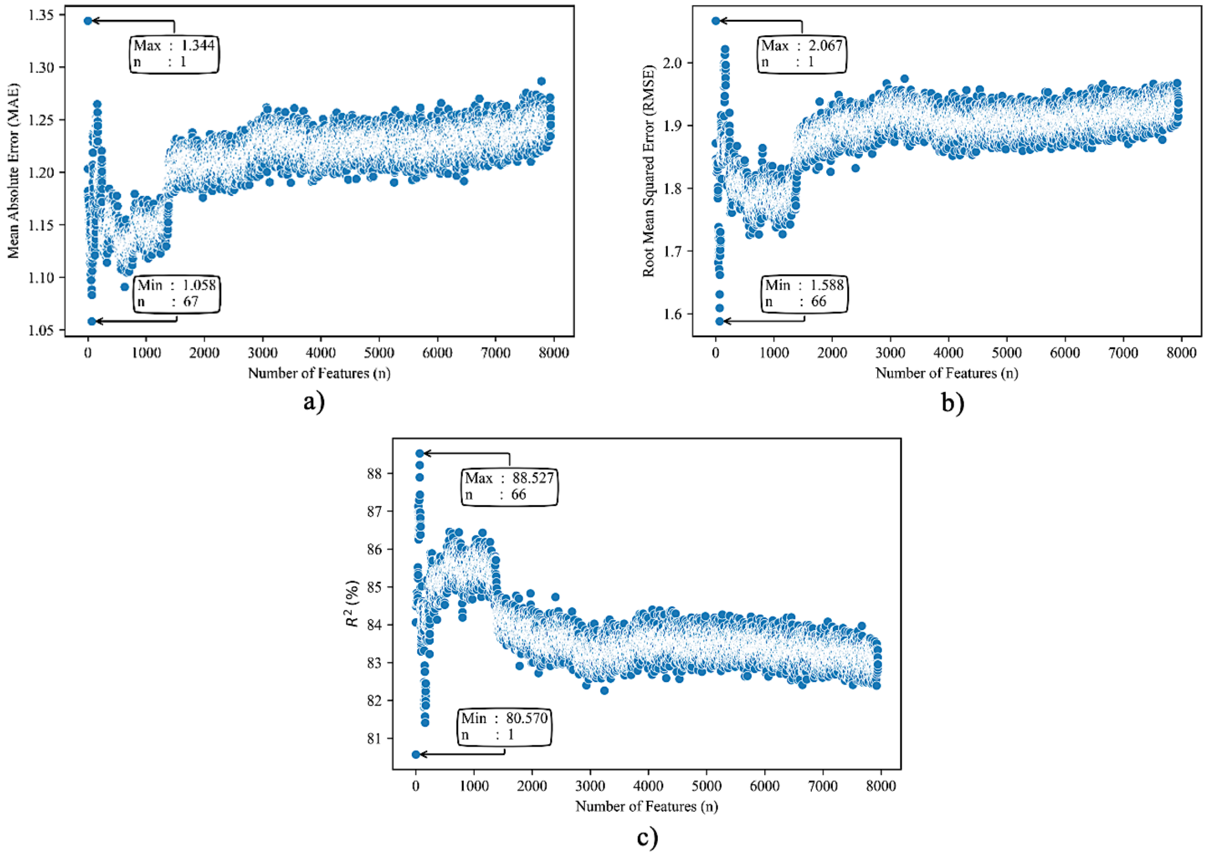

Figure 4. *Minimum (Min.) and maximum (Max.) error and accuracy rates according to the number of features (n); a) MAE, b) RMSE and c) R^2.*

Table 3. *Evaluation of the efficacy of the proposed approach vs leading techniques in forecasting student performance.*

Model	MAE ± CI	MSE ± CI	RMSE ± CI	R^2 ± CI
Linear Reg.	1.613 ± 0.327	5.781 ± 2.269	2.404 ± 0.472	.737 ± .088
Robust Reg.	1.117 ± 0.336	4.764 ± 2.739	2.182 ± 0.639	.783 ± .097
Ridge Reg.	1.437 ± 0.319	5.235 ± 2.319	2.288 ± 0.510	.761 ± .080
Lasso Reg.	1.285 ± 0.306	4.613 ± 2.315	2.147 ± 0.544	.790 ± .083
Poly. Reg.	3.327 ± 0.529	19.731 ± 6.122	4.442 ± 0.689	.102 ± .285
SGD Reg.	1.907 ± 0.281	6.058 ± 1.844	2.461 ± 0.374	.724 ± .082
RF Reg.	1.000 ± 0.256	3.100 ± 1.596	1.760 ± 0.457	.858 ± .070
SVM Reg.	1.831 ± 0.331	6.806 ± 2.428	2.609 ± 0.467	.690 ± .088
XGB Reg.	1.067 ± 0.273	3.470 ± 2.130	1.863 ± 0.575	.842 ± .091
ANN	1.453 ± 0.256	4.159 ± 1.281	2.039 ± 0.315	.810 ± .053
Prop. Approach	1.025 ± 0.210	2.487 ± 1.218	1.577 ± 0.382	.886 ± .052

CI: Confidence interval (%95)

Accordingly, the newly developed ANN model was more successful than the linear, SGD and SVM regression methods with a ratio of 1.45 in the MAE metric, while it became the third best method after the random forest (RF) and XGB regression methods with ratios of 4.15, 2.03 and 81.0% in the MSE, RMSE and R^2 metrics, respectively. The proposed approach was more successful than all methods with ratios of 1.02, 2.49, 1.58 and 88.6% in the MAE, MSE, RMSE and R^2 metrics, respectively. In the R^2 metric, it was approximately 3% more successful than its closest competitor, the RF regression method.

On the other hand, the comparative performance evaluation of the proposed approach with the state-of-the-art models was made in terms of error metrics (MAE, MSE, RMSE) and accuracy metric (R^2); the Wilcoxon Signed-Rank Test method was used to determine the statistical significance of this comparison.

Table 4 presents a statistical comparison of the MAE values of the proposed model and other regression models using the Wilcoxon Signed-Rank Test. This table shows that models such as Linear Regression, ANN, SGD, and SVM achieved significantly lower MAE values compared to the proposed model, and these differences were very highly statistically significant (***). In contrast, the differences observed for models such as Lasso, XGBoost, and Robust Regression were not statistically significant (ns).

Table 4. *Statistical comparison of the MAE values of the proposed model and other regression models using the Wilcoxon Signed-Rank Test.*

Model	MAE	Prop. MAE	Improve (%)	*p*-value	Significance
Poly. Reg.	3.3277	1.0252	69.19	.0000	***
SGD Reg.	1.9076	1.0252	46.26	.0000	***
SVM Reg.	1.8319	1.0252	44.04	.0000	***
Linear Reg.	1.6134	1.0252	36.46	.0001	***
ANN Reg.	1.4538	1.0252	29.48	.0004	***
Ridge Reg.	1.4370	1.0252	28.65	.0097	**
Lasso Reg.	1.2857	1.0252	20.26	.0833	ns
RF Reg.	1.0000	1.0252	-2.52	.5297	ns
Robust Reg.	1.1176	1.0252	8.27	.6118	ns
XGB Reg.	1.0672	1.0252	3.94	.7353	ns

ns: not significant, *: $p < .05$; **: $p < .01$; ***: $p < .001$.

Table 5 presents a statistical comparison of the MSE values of the proposed model and other regression models using the Wilcoxon Signed-Rank Test. Although Linear Regression, ANN, SGD, SVM, and Polynomial Regression models had significantly higher MSE values compared to the proposed model, these differences were found to be very highly statistically significant (***). Ridge Regression also showed a significant difference (*), but the significance level was lower. On the other hand, the differences between the Lasso, Random Forest, Robust Regression, and XGBoost models and the proposed model were not statistically significant (ns).

Table 6 shows the statistical comparison of the RMSE values of the proposed model and various regression models using the Wilcoxon Signed-Rank Test. Models such as Linear Regression, ANN, SGD, SVM, and Polynomial Regression have significantly higher RMSE values compared to the proposed model, and these differences are very highly statistically significant (***). The Ridge Regression model also presents a significant difference and is considered highly significant (**). In contrast, the differences observed between the Lasso, Random Forest, Robust Regression, and XGBoost models and the proposed model are not statistically significant (ns).

Table 5. *Statistical comparison of the MSE values of the proposed model and other regression models using the Wilcoxon Signed-Rank Test.*

Model	MSE	Prop. MSE	Improve (%)	*p*-value	Significance
Poly. Reg.	19.7311	2.4874	87.39	.0000	***
SGD Reg.	6.0588	2.4874	58.95	.0000	***
SVM Reg.	6.8067	2.4874	63.46	.0000	***
Linear Reg.	5.7815	2.4874	56.98	.0001	***
ANN Reg	4.1597	2.4874	40.20	.0004	***
Ridge Reg.	5.2353	2.4874	52.49	.0110	*
Lasso Reg.	4.6134	2.4874	46.08	.1165	ns
RF Reg.	3.1008	2.4874	19.78	.3979	ns
Robust Reg.	4.7647	2.4874	47.80	.5976	ns
XGB Reg.	3.4706	2.4874	28.33	.6609	ns

ns: not significant, *: $p < .05$; **: $p < .01$; ***: $p < .001$.

Table 6. *Statistical comparison of the RMSE values of the proposed model and other regression models using the Wilcoxon Signed-Rank Test.*

Model	RMSE	Prop. RMSE	Improve (%)	p-value	Significance
Poly. Reg.	4.4420	1.5771	64.49	.0000	***
SGD Reg.	2.4615	1.5771	35.93	.0000	***
SVM Reg.	2.6090	1.5771	39.55	.0000	***
Linear Reg.	2.4045	1.5771	34.41	.0001	***
ANN Reg.	2.0395	1.5771	22.67	.0004	***
Ridge Reg.	2.2881	1.5771	31.07	.0097	**
Lasso Reg.	2.1479	1.5771	26.57	.0833	ns
RF Reg.	1.7609	1.5771	10.44	.5297	ns
Robust Reg.	2.1828	1.5771	27.75	.6118	ns
XGB Reg.	1.8630	1.5771	15.34	.7353	ns

ns: not significant, *: $p < .05$; **: $p < .01$; ***: $p < .001$.

Table 7. *Statistical comparison of the R^2 values of the proposed model and other regression models using the Wilcoxon Signed-Rank Test.*

Model	R^2	Prop. R^2	Improve (%)	p-value	Significance
Poly. Reg.	0.1024	0.8868	766.06	.0000	***
SGD Reg.	0.7244	0.8868	22.43	.0000	***
SVM Reg.	0.6904	0.8868	28.46	.0000	***
Linear Reg.	0.7370	0.8868	20.33	.0001	***
ANN Reg.	0.8108	0.8868	9.38	.0004	***
Ridge Reg.	0.7618	0.8868	16.41	.0110	*
Lasso Reg.	0.7901	0.8868	12.24	.1165	ns
RF Reg.	0.8589	0.8868	3.25	.3979	ns
Robust Reg.	0.7832	0.8868	13.23	.5976	ns
XGB Reg.	0.8421	0.8868	5.31	.6609	ns

ns: not significant, *: $p < .05$; **: $p < .01$; ***: $p < .001$.

Table 7 shows the R^2 values of the proposed model compared to those of different regression models using the Wilcoxon Signed-Rank Test. The R^2 performances of Linear Regression, ANN, SGD, SVM, and especially Polynomial Regression models show very highly statistically significant differences compared to the proposed model (***). The Ridge Regression model, on the other hand, presents a significant difference, meeting the significance threshold at p = .0110 (*). On the other hand, the differences between the Lasso, Random Forest, Robust Regression, and XGBoost models and the proposed model are not statistically significant (ns).

The impact of the features in the dataset on the predicted output (G3) of the proposed approach is illustrated by the summary SHAP plot presented in Figure 5. The effect of the features in the dataset on the predicted output (G3) of the proposed approach is shown by the summary SHAP plot presented in Figure 5. Accordingly, the features that have the greatest impact on the G3 output are, respectively, second period grade (G2), first period grade (G1), number of absences, male sex, female sex, school type (GP: Gabriel Pereira, MS: Mousinho da Silveira), current health status, weekend alcohol consumption (Walc), student's home address (U: Urban, R: Rural), workday alcohol consumption, family size (LE3: less or equal to 3), parent's cohabitation status (T: cohabiting), etc. In other words, it can be said that characteristics such as high G1 and G2 grades, low school absenteeism, gender status, school type, good health status and low alcohol consumption positively affect the student's success.

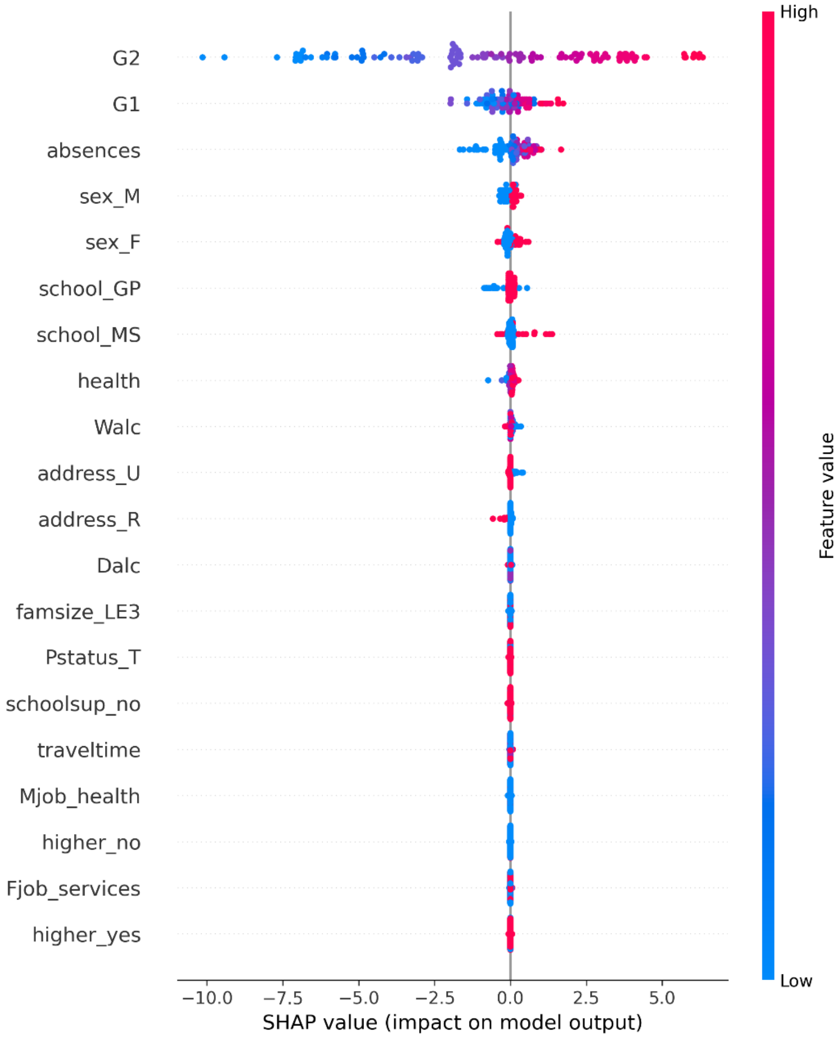

Figure 5. *SHAP summary plot demonstrating the contribution of each feature to the predicted output (G3), as determined by the proposed approach.*

In summary, the proposed model outperformed all regression models when evaluated in terms of error metrics (MSE and RMSE). Improvements of up to 87% in MSE and 64% in RMSE were achieved, and most of these differences were found to be statistically significant ($p < .05$). Significant differences were observed, particularly when compared to traditional methods such as Polynomial, SGD, SVM, ANN and Linear regression. In terms of the accuracy metric R^2, the proposed model also had the highest explanatory power (88.68%), and the differences between it and the classical models were statistically significant. However, when compared to robust models such as RF, XGB, Lasso and Robust, the R^2 difference was lower (3-10%), and these differences were not found to be statistically significant. However, considering the decrease in error rates, the proposed model offers a more balanced structure in terms of both accuracy and generalizability. The Wilcoxon Signed-Rank Test results also support this superiority. As a result, the proposed method stands out as the most successful model by offering lower error and higher explanatory power against both traditional and modern regression models.

DISCUSSION

The proposed regression approach demonstrated statistically superior performance metrics such as MAE, MSE, RMSE, and R^2 compared to other commonly used regression models. Specifically, because the proposed model was designed by combining approaches such as ANN, f_regression, and RF regressor, it achieved approximately 7% and 3% higher accuracy than the plain ANN and RF regressor models, respectively. The proposed approach's high accuracy performance, as seen in these ablation results, is believed to be due to three reasons. First, the use of ANN as a feature extractor. Because ANN can learn and therefore capture complex data patterns better than traditional machine learning models, it can extract features with complex context information. Second, the f_regression feature selection technique was used to select the 66 most important features from the 7944 features extracted by the ANN. Third, the RF regressor was used for G3 estimation from these 66 selected features. While many regressors exist, such as Linear, Ridge, Lasso, and SVR, the RF regressor was preferred due to its high performance on small and nonlinear datasets and its robustness against overfitting. Indeed, when the error and accuracy rates in Table 3 are examined, it can be seen that the RF regressor exhibits the best performance, excluding the proposed approach.

This superior performance of the proposed approach over other commonly used regression models is supported by the Wilcoxon Signed-Rank Test results, which revealed statistically significant improvements in performance metrics such as MAE, MSE, RMSE, and R^2. According to the Wilcoxon Signed-Rank test results presented in Tables 4–7, the proposed model performed significantly better in MAE, MSE, and RMSE, particularly compared to Polynomial Regression, SGD Regression, SVM Regression, and Linear Regression models ($p < .001$). These results demonstrate that the proposed model significantly reduces error rates and increases prediction accuracy. However, no statistically significant differences were observed in comparisons with Lasso, Random Forest, Robust, and XGBoost Regression models ($p > .05$). An examination of R^2 values reveals that the proposed model significantly increases its explanatory power, with the largest increase occurring approximately 766% compared to Polynomial Regression. Overall, the findings indicate that the proposed approach produces more stable and reliable estimates compared to classical and some advanced regression methods, particularly with significant improvements in accuracy and error rates.

Furthermore, according to the analysis results of the SHAP summary graph presented in Figure 5, the G2 and G1 variables appear to be the factors that most strongly influence the prediction performance of the proposed approach. This suggests that previous semester grades (first and second semesters) play a decisive role in students' final academic success. Furthermore, an increase in the absences variable appears to negatively impact the outcome of the proposed approach, confirming that high absences reduce academic achievement. The gender variables (sex_M and sex_F) have a relatively small but noticeable impact on the recommended approach. The school type (school_GP and school_MS) and health status (health) variables also exert partial influence on the decision to use the recommended approach, suggesting that students' individual and environmental circumstances may have indirect effects on achievement. The other variables have limited impact. Overall, the SHAP analysis indicates that the model's decisions are meaningful and explainable, with students' past academic performance (G1 and G2) being the strongest predictor.

As noted in the literature, the academic success of middle school students is influenced by many factors, including family, school, and individual characteristics (Huang & Yang, 2018). The findings of this study also revealed that this relationship is in this direction, as seen in the SHAP summary graph in Figure 5. Bihua's (2013) study in China reported that the family's economic status, number of children, and parents' learning expectations were significantly related to students' success on the high school entrance exam. Similarly, this study determined that the variables Fjob, Mjob, Pstatus, and schoolsup, which represent the family's economic status, famsize, which indicates family size, and higher and famsup, which indirectly reflect parental expectations, exhibited significant relationships with academic success. The findings suggest that parents' occupational level and economic conditions contribute to students' academic success as they improve; family support and higher education expectations strengthen success. Conversely, the number of family members plays a significant role in students' academic success in families with three or fewer members. Lei and Shao's (2015) study emphasized the mediating role of students' self-efficacy in the relationship between teacher behaviors and academic achievement. Similarly, this study found that school type (school), school support (schoolsup), and travel time to school (traveltime) had significant effects on achievement. Furthermore, when evaluated in terms of individual characteristics, students with lower alcohol consumption (Dalc, Walc) were found to exhibit higher academic performance. These results are consistent with Pekrun *et al.* (2002) and Perry *et al.* (2001), who stated that high self-control and academic control increase achievement. Also, as stated in the study by Karaman and Atar (2019), this research found that the average achievement of male students (sex_M) was higher than that of female students (sex_F). In the literature, machine learning methods are widely used in the analysis of training data (Jin, 2021; Khasanah, 2018; Rivas *et al*., 2021). Previous studies have mostly focused on classification and prediction methods using Decision Trees, Bayesian Networks, and artificial neural network-based models on large-scale datasets (Baashar *et al*., 2022; Sandra *et al*., 2021). In this study, a hybrid model combining artificial neural network (ANN)-based feature extraction and random forest (RF) regressor was developed using a small-scale dataset. Using the 66 most significant variables selected with the f_regression method from 7944 features extracted by the ANN, students' academic success in mathematics was predicted with an R^2 accuracy of 88%. In this respect, the study supports model integration and feature selection approaches suggested in the literature and provides an innovative example for course-based applications at the secondary education level. However, the small size of the data set and the interpretability of the model may be limited.

Developed using the Student Performance dataset, the proposed approach contributes to data-based decision-making in education by predicting students' final academic success with high accuracy. The proposed approach highlights first-term (G1) and second-term (G2) grades as the strongest predictors, highlighting the importance of closely monitoring students' intra-term performance. For example, if a student's G1 grade is low, the proposed approach can identify early on the likelihood of this student failing at the end of the term, allowing teachers to implement timely individual support, additional homework, or mentoring programs. Similarly, the negative impact of the absences variable demonstrates the critical role of regular attendance in academic performance. This finding allows for the establishment of early warning systems in schools to closely monitor absence trends. Furthermore, SHAP analysis increases the model's explainability, allowing teachers and administrators to understand the variables underlying the proposed approach's decisions. Consequently, this approach not only predicts student success but also provides a powerful tool for early identification of at-risk students, developing targeted intervention strategies, and creating personalized learning plans.

Implications

The findings of this study offer important implications for both educational practitioners and researchers. For practitioners, the proposed model's high-accuracy prediction of student achievement can be directly applied to the development of early warning and intervention systems. For example, teachers can identify students at risk of failure by monitoring students' G1 and G2 grades and attendance records at the beginning of a semester and create additional tutoring support, mentoring, or individualized learning plans for these students. Educational administrators can use the outputs from this model to data-driven restructuring of school policies and plan targeted interventions, such as reducing absenteeism rates and improving health and social conditions.

From a research perspective, the results demonstrate that explainable artificial intelligence (XAI)-based models can be effectively used in educational data. Future studies could focus on testing the generalizability of the model on datasets collected from different regions or educational systems, improving predictive performance by adding psychosocial variables (motivation, parental involvement, emotional state, etc.), and dynamically modeling student achievement trends using time-series approaches. Furthermore, making model outputs interpretable by teachers, parents, and students using explainability methods like SHAP will support the ethical and reliable use of AI in education. Ultimately, this study provides a strong foundation not only for predicting student achievement but also for developing data-driven, human-centered educational strategies.

Limitations

While the findings of this study reveal important conclusions, several limitations must be considered. First, the study used only the UCI Student Performance dataset, which represents a limited sample size obtained from two Portuguese schools (Gabriel Pereira and Mousinho da Silveira). Therefore, the generalizability of the model to different educational systems, cultural contexts, or socioeconomic conditions is limited. Second, the variables included in the dataset are largely limited to numerical and categorical demographic information; important variables that could influence achievement, such as students' motivation level, psychological factors, family dynamics, or teaching quality, are not included in the model. This partially limits the model's explanatory power and ignores the influence of behavioral or emotional factors in predicting achievement.

Third, the model used in the study adopts a static approach; that is, it does not dynamically track students' developmental changes or learning curves over time. Therefore, the model is limited in its ability to reflect changes in student performance over subsequent weeks or the effects of instructional interventions. Fourth, although the SHAP analysis provides an explanation of the model's decision-making process, these explanations are based solely on variables in the existing dataset. Therefore, the model's interpretability depends on the scope of the data, and the level of explanation may change when new or different types of variables are added.

Finally, the validation process for the proposed approach was conducted with a single dataset with a relatively small sample size of 649 students. This creates uncertainty about whether the proposed approach will perform equally well across different student groups or datasets. A limited sample size, in particular, can limit the learning capacity of complex models and prevent the full statistical emergence of the effects of some variables. Therefore, future research using larger, multicenter datasets encompassing diverse demographic groups will further enhance the generalizability of the model.

CONCLUSION

This study proposes a new hybrid methodology utilizing convolutional neural networks to forecast the academic performance of middle school children. The initial phase of the proposed methodology involves the development of a neural network utilizing one-dimensional convolutions for automated feature extraction. In the second stage, f-regression technique is used to select important features from the features extracted by this neural network. In the final stage, the chosen features are input into the random forest regression model, which is then trained to produce the students' final grade predictions as output. As a result of the experimental tests, the proposed hybrid approach has achieved a better success than all other methods by reaching 88.6% R^2 accuracy, 1.0 MAE, 2.5 MSE score.

This methodology employs one-dimensional convolutional neural networks, which are extensively utilized across several domains, including speech recognition, signal processing, natural language processing, financial time series analysis, and big datasets. This study was undertaken due to the absence of literature on the application of one-dimensional convolutional neural networks for predicting students' performance in a course. Furthermore, despite the limited size of the student performance dataset included in this work, our proposed hybrid methodology has demonstrated superior efficacy relative to other leading methodologies. Future research will aim to evaluate the overall academic achievement of students by examining all their courses rather than concentrating on a single course using a larger data set. Thus, the effect of the factors that negatively affect the academic achievement of the student can be observed more clearly.

Declaration of Conflicting Interests and Ethics

The authors declare no conflict of interest. This research study complies with research publishing ethics. The scientific and legal responsibility for manuscripts published in IJATE belongs to the authors.

Data Availability Statement

The research data supporting this study are publicly available at the UCI Machine Learning Repository: https://archive.ics.uci.edu/dataset/320/student+performance.

Informed Consent

Written informed consent was obtained from the participants who participated in this study.

Contribution of Authors

Gürkan Doğan: Investigation, Conceptualization, Methodology, Visualization, Software, and Writing-original draft. **Demet Demiralp**: Investigation, Formal analysis, Supervision, Writing review & Editing, Validation.

Orcid

Gürkan Doğan https://orcid.org/0000-0003-2497-8348
Demet Demiralp https://orcid.org/0000-0003-3271-8511

REFERENCES

Baashar, Y., Alkawsi, G., Mustafa, A., Alkahtani, A.A., Alsariera, Y.A., Ali, A.Q., Hashim, W., & Tiong, S.K. (2022). Toward predicting student's academic performance using artificial neural networks (ANNs). *Applied Sciences*, *12*(3). https://doi.org/10.3390/app12031289

Bihua, Z. (2013). Analysis of family and school factors influencing students' academic performance. *Education Study*, *34*(3), 88-97.

Cortez, P. (2014). *Student performance*. Machine Learning Repository. https://doi.org/10.24432/C5TG7T

Cortez, P., & Silva, A. (2008) Using data mining to predict secondary school student performance. In A. Brito, J. Teixeira (Eds.), *Proceedings of 5th future business technology conference, FUBUTEC 2008* (pp. 5-12). EUROSIS. https://repositorium.uminho.pt/server/api/core/bitstreams/991a0e2b-249d-466d-afef-937d975ff7fc/content

Doğan, G., & Ergen, B. (2023). A new approach based on convolutional neural network and feature selection for recognizing vehicle types. *Iran Journal of Computer Science*, *6*(2), 95-105. https://doi.org/10.1007/s42044-022-00125-6

Doğan, G., Imak, A., Ergen, B., & Sengur, A. (2024). A new hybrid approach for grapevine leaves recognition based on ESRGAN data augmentation and GASVM feature selection. *Neural Computing and Applications, 36,* 7669-7683. https://doi.org/10.1007/s00521-024-09488-2

González-Calatayud, V., Prendes-Espinosa, P., & Roig-Vila, R. (2021). Artificial intelligence for student assessment: A systematic review. *Applied Sciences*, *11*(12), Article 5467. https://doi.org/10.3390/app11125467

Huang, T., & Yang, Y. (2018). An empirical analysis of the influencing factors of middle school students' academic achievement in China: Based on the following survey data of CEPS (2014-2015). In *Proceedings of the 2018 2nd International Conference on Education Science and Economic Management (ICESEM 2018)*. Atlantis Press. https://doi.org/10.2991/icesem-18.2018.208

Hussain, S., & Khan, M.Q. (2023). Student-Performulator: 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

Hwang, G.J., & Tu, Y.F. (2021). Roles and research trends of artificial intelligence in mathematics education: A bibliometric mapping analysis and systematic review. *Mathematics*, *9*(6), Article 584. https://doi.org/10.3390/math9060584

İmak, A., Doğan, G., Şengür, A., & Ergen, B. (2023). Asma yaprağı türünün sınıflandırılması için doğal ve sentetik verilerden derin öznitelikler çıkarma, birleştirme ve seçmeye dayalı yeni bir yöntem [A new method based on extracting, combining and selecting deep features from natural and synthetic data for classification of grapevine leaf species]. *International Journal of Pure and Applied Sciences*, *9*(1), 46-55. https://doi.org/10.29132/ijpas.1144615

Jin, Y. (2021). *Research on key technologies of student's learning performance prediction based on big learning data* [Unpublished master's thesis]. Southeast University.

Karaman, P., & Atar, B. (2019). The effects of student and school level characteristics on academic achievement of middle school students in Turkey. *Journal of Measurement and Evaluation in Education and Psychology*, *10*(4), 391-405. https://doi.org/10.21031/epod.564819

Khasanah, A.U. (2018). A review of student's performance prediction using educational data mining techniques. *Journal of Engineering and Applied Sciences*, *13*(6), 5302-5307.

Lei, H., & Shao, C. (2015). The relationship between teacher's caring behavior and students' academic performance: The mediating role of learning effectiveness. *Psychological Development and Education*, *31*(2), 188-197. https://doi.org/10.16187/j.cnki.issn1001-4918.2015.02.08

Matzavela, V., & Alepis, E. (2021). Decision tree learning through a predictive model for student academic performance in intelligent M-Learning environments. *Computers and Education: Artificial Intelligence*, *2*, Article 100035. https://doi.org/10.1016/j.caeai.2021.100035

Olabanjo, O.A., Wusu, A.S., & Manuel, M. (2022). A machine learning prediction of academic performance of secondary school students using radial basis function neural network. *Trends in Neuroscience and Education*, *29*, Article 100190. https://doi.org/10.1016/j.tine.2022.100190

Pekrun, R., Goetz, T., Titz, W., & Perry, R.P. (2002). Academic emotions in students' self-regulated learning and achievement: A program of qualitative and quantitative research. *Educational Psychologist*, *37*(2), 91-105. https://doi.org/10.1207/S15326985EP3702_4

Pérez-Gomariz, M., López-Gómez, A., & Cerdán-Cartagena, F. (2023). Artificial neural networks as artificial intelligence technique for energy saving in refrigeration systems-A review. *Clean Technologies*, *5*(1), 116-136. https://doi.org/10.3390/cleantechnol5010007

Perry, R.P., Hladkyj, S., Pekrun, R.H., & Pelletier, S.T. (2001). Academic control and action control in the achievement of college students: A longitudinal field study. *Journal of Educational Psychology*, *93*(4), 776-789. https://doi.org/10.1037/0022-0663.93.4.776

Rajendran, S., Chamundeswari, S., & Sinha, A.A. (2022). Predicting the academic performance of middle- and high-school students using machine learning algorithms. *Social Sciences and Humanities Open*, *6*(1), *Article 100357*. https://doi.org/10.1016/j.ssaho.2022.100357

Rivas, A., González-Briones, A., Hernández, G., Prieto, J., & Chamoso, P. (2021). Artificial neural network analysis of the academic performance of students in virtual learning environments. *Neurocomputing*, *423*, 713-720. https://doi.org/10.1016/j.neucom.2020.02.125

Sandra, L., Lumbangaol, F., & Matsuo, T. (2021). Machine learning algorithm to predict student's performance: A systematic literature review. *TEM Journal*, 1919-1927. https://doi.org/10.18421/TEM104-56

Sapci, A.H., & Sapci, H.A. (2020). Artificial intelligence education and tools for medical and health informatics students: Systematic review. *JMIR Medical Education*, *6*(1), Article e19285. https://doi.org/10.2196/19285

Turban, E., Sharda, R., Aronson, J.E., & King, D. (2008). *Business intelligence: A managerial approach*. Prentice Hall.

Yağcı, M. (2022). Educational data mining: prediction of students' academic performance using machine learning algorithms. *Smart Learning Environments*, *9*, Article 11. https://doi.org/10.1186/s40561-022-00192-z

Yüksel, N., Börklü, H.R., Sezer, H.K., & Canyurt, O.E. (2023). Review of artificial intelligence applications in engineering design perspective. *Engineering Applications of Artificial Intelligence*, *118*, Article 105697. https://doi.org/10.1016/j.engappai.2022.105697

Zawacki-Richter, O., Marín, V.I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education – where are the educators? *International Journal of Educational Technology in Higher Education*, *16*(1), Article 39. https://doi.org/10.1186/s41239-019-0171-0