



OPEN ACCESS

EDITED BY

Giovanni Costa,
National Research Council (CNR), Italy

REVIEWED BY

Cristina Tripon,
Polytechnic University of Bucharest, Romania
Henri Tilga,
University of Tartu, Estonia
AHmed Kord,
Mansoura University, Egypt

*CORRESPONDENCE

Javier Linkolk Lopez-Gonzales
✉ javierlinkolk@gmail.com

RECEIVED 20 June 2025

ACCEPTED 02 October 2025

PUBLISHED 21 October 2025

CITATION

Gonzales W, Cordero Z, Abanto-Ramírez CD,
Susanibar Ramírez ET, Iftikhar H and
Lopez-Gonzales JL (2025) A hybrid AI
approach for predicting academic
performance in RBE students.
Front. Artif. Intell. 8:1651100.
doi: 10.3389/frai.2025.1651100

COPYRIGHT

© 2025 Gonzales, Cordero, Abanto-Ramírez,
Susanibar Ramírez, Iftikhar and
Lopez-Gonzales. This is an open-access
article distributed under the terms of the
[Creative Commons Attribution License \(CC BY\)](#). The use, distribution or reproduction in
other forums is permitted, provided the
original author(s) and the copyright owner(s)
are credited and that the original publication
in this journal is cited, in accordance with
accepted academic practice. No use,
distribution or reproduction is permitted
which does not comply with these terms.

A hybrid AI approach for predicting academic performance in RBE students

Willy Gonzales¹, Zindel Cordero¹, Carlos D. Abanto-Ramírez¹,
Edgar Tito Susanibar Ramírez², Hasnain Iftikhar³ and
Javier Linkolk Lopez-Gonzales^{1*}

¹Escuela de Posgrado, Universidad Peruana Unión, Lima, Peru, ²Facultad de Educación, Universidad Nacional José Fausto Sánchez Carrión, Huacho, Peru, ³Department of Statistics, University of Peshawar, Peshawar, Pakistan

Machine learning has advanced significantly in recent years and is being used in higher education to perform various types of data analysis. While the literature demonstrates the application of machine learning algorithms to predict performance in university education, no such applications are found in EBR, let alone in private institutions of a denominational nature, which presents an opportunity to study prediction in these institutions. To address this gap, this research aims to propose a predictive approach as a decision-support tool for regular basic education, using machine learning techniques. Among the techniques utilized, three machine learning models (Logistic Regression, Support Vector Machine, and Random Forest), along with deep learning models (AlexNet, Gated Recurrent Unit, and Bidirectional Gated Recurrent Unit), were analyzed, as well as ensemble models. Nonetheless, the Ensemble model, which combines deep learning and machine learning techniques, is preferred due to its superior accuracy, precision, and sensitivity performance metrics.

KEYWORDS

hybrid AI, artificial intelligence, predicting academic performance, RBE students, statistics

1 Introduction

Machine learning (ML) has advanced significantly in recent years and is increasingly applied in higher education (HE) to analyze student performance, predict dropout risk, and support decision-making (Fahd et al., 2022; Mashrur and Nonyelum, 2020; Singh and Kumar, 2020). Educational data mining (EDM) serves as a key methodology for extracting meaningful knowledge and patterns from academic databases, thereby enabling early detection of at-risk students and guiding timely interventions (Sultana and Khan, 2019; Lopez and Ramirez, 2021). A wide variety of ML models—including decision trees (DT), logistic regression (LR), support vector machines (SVM), random forests (RF), and artificial neural networks (ANN)—have been employed to predict academic performance with encouraging results (Ghosh and Sharma, 2022; Hussain and Khan, 2023; Reyes et al., 2025).

Recent studies have also explored deep learning (DL) architectures, such as convolutional neural networks (CNNs) and recurrent neural networks (RNNs), to capture complex temporal and non-linear dependencies in educational datasets (Mi and Yeung, 2019; Ahmed et al., 2022). While these approaches have demonstrated high predictive accuracy in various contexts, including student dropout and performance prediction, most have been developed either with ML or DL models in isolation.

In addition, other contributions in AI-driven educational analytics highlight the importance of integrating advanced computational approaches into teaching and learning practices. For example, [Tripon \(2022\)](#) examined how computational thinking skills can be promoted in STEM education through AI-enhanced teaching strategies, demonstrating the growing role of predictive and analytical tools in shaping educational outcomes. Incorporating such perspectives helps to situate our study within the broader landscape of AI applications in education while emphasizing the unique methodological contribution of our ML+DL ensemble approach in the RBE context. Furthermore, recent global research highlights that higher education students are actively forming perceptions about the role of AI tools such as ChatGPT in their learning processes ([Ravšelj et al., 2025](#)), offering valuable insights that can inform the design and contextualization of AI-driven predictive models in education. Moreover, applications in Regular Basic Education (RBE) remain limited, and virtually no work has addressed private denominational institutions, leaving a gap in context-specific predictive modeling.

[Table 1](#) summarizes representative studies from the literature, highlighting models used, datasets, and reported accuracies. As shown, ensemble approaches often outperform single models by combining complementary strengths ([Menchaca, 2024](#)), yet hybrid ML+DL ensembles remain underexplored in RBE contexts.

In contrast to prior studies that have applied ML or DL models independently (see [Chen and Ding, 2023](#); [Mohammadi et al., 2019](#); [Chavez et al., 2023](#); [Adefemi et al., 2025](#)), our work introduces a novel ensemble framework that hybridizes both paradigms. This framework integrates the interpretability and robustness of ML models (e.g., SVM, RF, LR; [Smith and Lee, 2017](#)) with the capacity of DL architectures (e.g., AlexNet, GRU, BiGRU) to capture non-linear feature interactions ([Brown, 2018](#); [Zhang and Chen, 2021](#)). By explicitly combining these complementary strengths, the ensemble is expected to outperform single-model baselines (cf. [Garcia and Alvarez, 2021](#)). Within the RBE institutional context—where datasets often contain both structured attributes and latent dynamic patterns—this hybrid strategy offers a unique methodological advantage that has not been previously investigated ([Nguyen and Do, 2019](#); [Castro and Paredes, 2022](#)).

TABLE 1 Summary of representative studies on student performance prediction.

Study	Models applied	Context	Reported accuracy
Ghosh and Sharma (2022)	SVM, RF	HE, India	SVM: 96.9%, RF: 81.3%
Reyes et al. (2025)	Multinomial LR	HE, Latin America	100% (AUC = 1)
Menchaca (2024)	Ensemble models	HE, Mexico	~90%
Baniata et al. (2024)	GRU	HE, Middle East	99.7%
Fanta and Alemu (2025)	RF	HE, Africa	99% (AUC = 100%)
Mi and Yeung (2019)	DL architectures	HE, Asia	92–96%

Motivated by the urgent need to strengthen the Peruvian education system, which is ranked 127th out of 137 countries in terms of quality ([Pachas, 2020](#)), and aligned with the National Education Project PEN 2036 ([Vargas, 2020](#)), this research proposes a predictive framework for private denominational RBE institutions. The approach follows a six-step methodology covering preprocessing, feature selection, normalization, model training, and evaluation. To the best of our knowledge, this is the first study in Peru to employ an ML+DL ensemble for academic performance prediction in RBE, providing an evidence-based decision-support tool for targeted interventions.

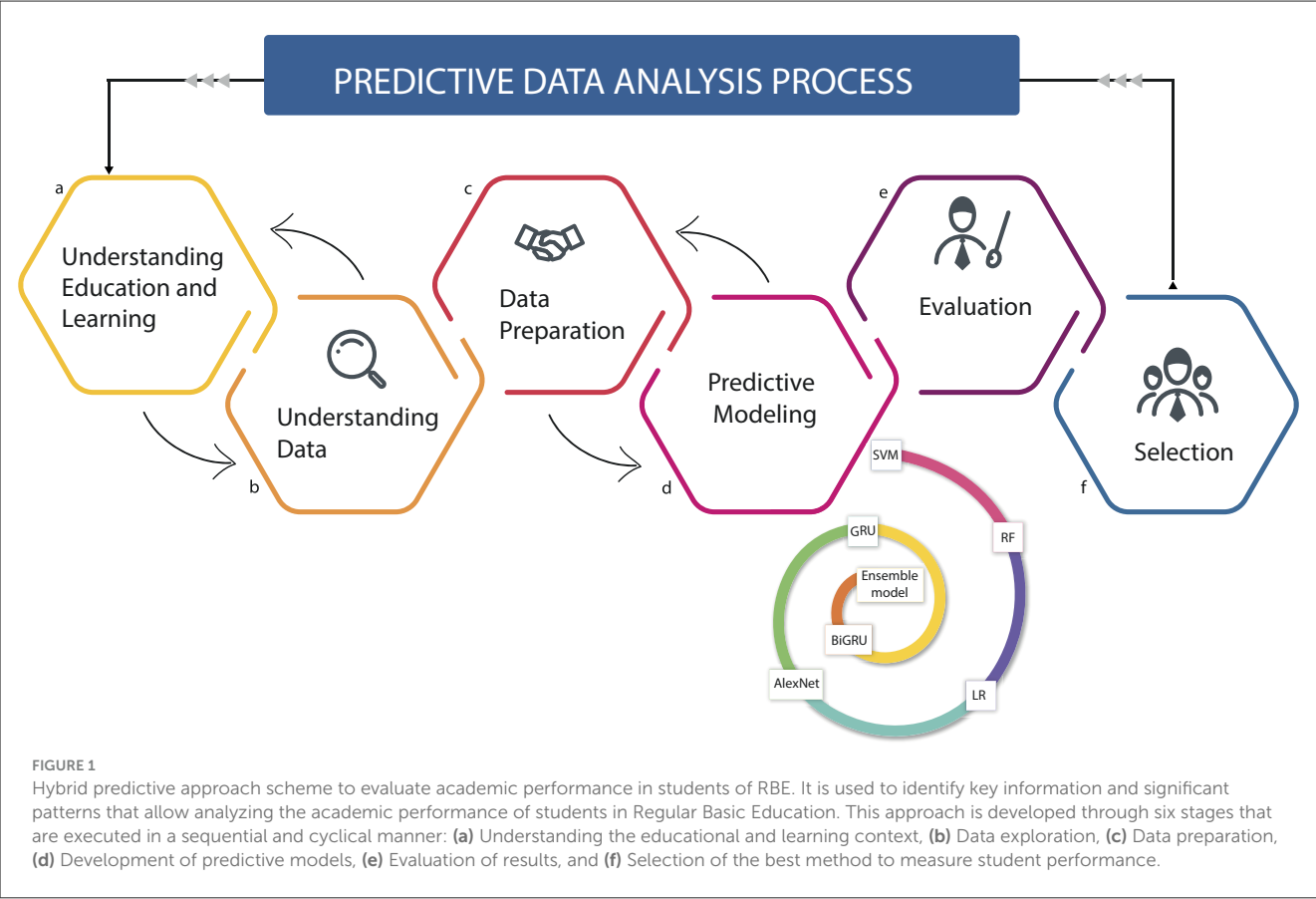
The remainder of the paper is structured as follows. Section 2 details the hybrid predictive approach used to evaluate the performance problem of RBE students. Section 3 presents the results and discussion, and Section 4 outlines the conclusions and directions for future research.

2 Materials and methods

This study uses a predictive approach to evaluate the comparative effectiveness of seven machine learning algorithms in predicting the academic performance of regular basic education students (see [Figure 1](#)). This approach allows identifying low-performing students by visualizing the results of predictive models, thus facilitating decision-making by educational institutions. The evaluated algorithms include: Support Vector Machine (SVM), Random Forest (RF), Logistic Regression (LR), AlexNet, Gated Recurrent Unit (GRU), Bidirectional GRU (BiGRU), and an Ensemble model.

2.1 Educational and learning comprehension

Academic performance is understood as a way of measuring how the teaching-learning process is developed. It is directly linked to the evaluation of acquired knowledge, so it is necessary to use various instruments that are properly structured and organized ([Pacheco et al., 2021](#)). Academic achievement as a complex phenomenon resulting from various personal and social variables, argued as an educational and evaluation element in most countries of the world. This performance is framed within the dynamics of interaction generated daily between students, teachers and the knowledge shared in the educational environment ([Santander, 2011](#)). Academic performance is determined by several factors, one of the most relevant of which is the understanding of the student's learning processes, which allows the development of appropriate methodological strategies ([Núñez et al., 2022](#)). Academic achievement defined as a complex process that could well be projected as an ascending property of an educational system, and where multiple variables are intertwined, with qualitative and quantitative aspects ([Ariza et al., 2018](#)). Other research indicates that academic achievement is linked to individual, psychological, cognitive and intellectual factors, as well as to variables related to educational processes, structural aspects and administrative elements. This perspective provides key concepts for understanding academic achievement as a pedagogical construct of a multicausal, multidimensional and multifactorial nature ([Murillo, 2024](#)). The



study of academic performance focuses on the grades obtained by students in the first through fifth grades of regular basic education, which represent the outcome of the educational process. In this context, academic performance is considered a fundamental indicator for decision-making at the institutional level, as it enables both the provision of incentives to high-performing students and the implementation of support measures for those with low results.

2.2 Understanding data

In this research, characteristics linked to students’ academic performance were selected, the most important being the grades obtained from their report cards throughout their time at school. Demographic variables such as gender, location by department and city, and courses completed were also included, as these data provide relevant information about students’ personal backgrounds and could have an impact on their academic performance (see Table 2). Meanwhile, the structure and presentation of the data followed the data organization presented in Kord et al. (2025).

2.3 Data preparation

2.3.1 Sample selection

The study sample consists of academic data from students at 11 schools belonging to the Adventist Educational Network in southern Peru, in the cities of Arequipa, Moquegua, Puno, Juliaca, Azángaro, Cusco, Espinar, Quillabamba, and Puerto Maldonado.

TABLE 2 Data description.

Attribute	Data classification	Possible values
Gender	Binary (symmetric)	F-M
Year	Categorical	2019–2021
Association	Categorical	MPS-MPLT-MSOP
Cities	Categorical	Cus-Esp-Pto Mald-Quill- Are-Mog-Aza-Jul-Pun
Courses	Categorical	Des Per-Cs Soc-Ed Rel-Ed Tra-Ed Fis-Com-Art Cul - Ing-Mat-Cie Tec-Ent Vir- Apr Aut
Qualification	Float	0–20

The database includes information collected between 2019 and 2021 (see Table 3). The students in the sample attend regular basic education, ensuring the homogeneity of the group in terms of educational level. No additional exclusion criteria were established, so data from all students at the selected institutions were included in the analysis.

2.3.2 Data collection

The total data set consisted of 155 transcripts from 3,247 Regular Basic Education students with 17 characteristics. These

TABLE 3 Detailed description of the dataset variables, including types, ranges, and educational context.

Number	Feature name	Type	Range	Feature description
1	DPCC	Number	0–20	Personal Development, Citizenship and Civics
2	CS	Number	0–20	Social Sciences
3	ER	Number	0–20	Religious education
4	EPT	Number	0–20	Education for Work
5	EF	Number	0–20	Physical education
6	C	Number	0–20	Communication
7	AC	Number	0–20	Art and Culture
8	I	Number	0–20	English
9	M	Number	0–20	Mathematics
10	CT	Number	0–20	Science and Technology
11	TICs	Number	0–20	It operates in virtual environments generated by ICTs
12	AA	Number	0–20	Manage learning autonomously
13	Grade	Categorical	1st–5th	Grade level
14	Year	Categorical	2019–2021	Year of data
15	Gender	Binary (symmetric)	F – M	Gender of student
16	Place of origin	Categorical	Cities	Place of origin
17	Educational Association	Categorical	Regions	Educational Association

TABLE 4 Table with processed data from grade transcripts 2019–2021.

Year	Assoc	School	Grad	Gd	DPCC	CS	ED	EPT	EF	C	AC	1	M	CT	TICs	AA
2019	MSOP	PARDO	Seg A	M	15	13	13	15	16	14	15	17	15	14	15	15
2019	MSOP	PARDO	Seg A	F	18	18	18	18	17	17	19	17	18	16	18	18
2019	MSOP	PARDO	Seg A	F	16	15	18	16	17	17	19	17	18	16	18	18
2019	MSOP	PARDO	Seg A	M	14	12	14	15	15	14	13	12	12	14	13	13

characteristics represent: year of study, place of origin, educational association, academic grade, gender, courses taken, and grades from the first to fifth years (see Table 4). They were classified as qualitative (nominal categorical and ordinal categorical) and quantitative. These variables were processed and analyzed using three machine learning models: Support Vector Machine (SVM), Random Forest (RF), and Logistic Regression (LR); as well as three deep learning models: AlexNet, Gated Recurrent Unit (GRU), and Bidirectional GRU (BiGRU); in addition to an ensemble model. The objective was to determine which of these models offers the best performance in predicting the academic performance of Regular Basic Education students, with a view to contributing to administrative decision-making within educational institutions. The analysis considered 12 subjects: Personal Development, Citizenship and Civics, Social Sciences, Religious Education, Education for Work, Physical Education, Communication, Art and Culture, English, Mathematics, Science and Technology, Develops in virtual environments generated by ICTs and Manages their learning independently. Likewise, five qualitative characteristics were incorporated: gender, place of origin, educational association, year of study and academic degree. Grades are expressed on a vigesimal scale (0 to 20), considering 11 as the minimum passing

grade. Likewise, it is important to emphasize that the characteristics of subjects and qualitative characteristics described are similar and coincide across institutions, thus achieving homogeneity in the data.

2.3.3 Dataset description and preprocessing

After preprocessing, the final dataset retained 17 features, encompassing academic indicators (such as grades in individual courses and the number of failed courses) along with demographic variables. The predictive task is formulated as a supervised **binary classification** problem. Specifically, the target label classifies students as either “Pass” or “Fail”, based on their academic results. This formulation allows for a straightforward assessment of academic success and failure, evaluating predictive models more directly aligned with educational decision-making.

Preprocessing was guided by three key criteria: completeness, consistency, and coherence. Records with missing values were imputed using the mean (for numerical attributes) or mode (for categorical attributes). Consistency was ensured by unifying categorical values (e.g., “Female” and “F” both mapped to “F”; “Male” and “M” mapped to “M”), while numerical attributes

(e.g., performance scores) were restricted to four decimal places. Coherence was verified by checking for outliers and distributional anomalies. Outlier detection revealed that some exchange students and Theology majors were older than the average student body. Still, these cases were retained since they aligned with the overall trends of the dataset.

To enable machine learning models to operate effectively, quantitative attributes were standardized to a mean of 0 and a standard deviation of 1. Each variable X_i was transformed into its standardized form Z_i using:

$$Z_i = \frac{(X_i - \bar{X})}{S_{n-1}}, \quad i = 1, \dots, n,$$

where

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i, \quad S_{n-1} = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2}.$$

2.3.4 Feature selection

To reduce dimensionality and enhance interpretability, a two-stage feature selection procedure was adopted:

1. **Correlation analysis.** Pairwise correlations between predictors and the target variable were computed. Features with negligible association to the outcome or those exhibiting strong multicollinearity (Pearson's $r > 0.85$) were flagged for removal.
2. **Recursive Feature Elimination (RFE).** Using logistic regression as a base estimator, RFE iteratively eliminated the least essential predictors based on coefficient weights. Cross-validation determined the optimal subset of features by maximizing predictive accuracy on the validation set.

This hybrid strategy combines the statistical efficiency of correlation analysis with the model-based refinement of RFE. As a result, the feature set was reduced from 22 to 12 attributes without loss of predictive power. In fact, models trained on the reduced set achieved slightly higher accuracy, F1-score, and AUC, highlighting both the computational efficiency and predictive robustness of the streamlined dataset.

2.4 Predictive modeling

This section details how the predictive models in this work were used to predict student performance. Hence, three machine-learning models, such as the Logistic regression model, Support Vector Machine model, and Random forest model; three deep learning models, including the Alexnet, the Gated Recurrent Unit, and Bidirectional Gated Recurrent Unit; and their proposed ensemble model. The details of each model are given below.

2.4.1 Logistic regression

The linear regression (LGR) is the basic method of binary classification, which estimates the probability of observation belonging to a particular category. The sigmoid function converts

linear combinations of features into probability values. The LGR model requires careful parameter tweaking and is best suited for binary classification problems. However, while using this strategy, it is critical to keep certain assumptions in mind, such as the assumption of feature independence. The LGR model is mathematically represented as follows: Let X represent an instance's feature vector, and y represent the binary class label (0 or 1). Given X , the logistic regression (LGR) model computes the probability that $y = 1$. This probability is denoted as:

$$P(y = 1/X) = \frac{1}{1 + e^{-(\alpha_0 + \alpha_1 X_1 + \alpha_2 X_2 + \dots + \alpha_P X_P)}} \quad (1)$$

where $P(y = 1/X)$ represents the probability that the class label y is one given the feature vector X , while e means the base of the natural logarithm (~ 2.71828), and $\alpha_0, \alpha_1, \alpha_2, \dots, \alpha_P$ are the coefficients (parameters) of the model to be learned during training. $X_1, X_2, \dots, X_P$ are the features of the instance X . The sigmoid function $\frac{1}{1+e^{-z}}$ maps the linear combination of features $\alpha_0 + \alpha_1 X_1 + \alpha_2 X_2 + \dots + \alpha_P X_P$ generates a value between 0 and 1, which indicates the probability that the observation is assigned to class 1. The model assigns the example to the class with the higher probability. Despite the effectiveness of logistic regression in binary classification, it is crucial to verify the validity of its assumptions, such as the assumption of linearity and feature independence, as these can impact its performance in real-world applications. Proper parameter tuning is also crucial for achieving optimal results with logistic regression.

2.4.2 Support vector machine

Support Vector Machine (SVM) model is one of the most commonly used ML algorithms to identify a hyperplane in N -dimensional space that classifies the data points. It helps find a plane that maximizes the margin. The diversity of N -dimensional space is based on the number of features; two features can be smoothly compared. However, it is more complex in the case of several features in classification. Maximizing the margin leads to a more accurate prediction. The margin refers to the distance between the decision hyperplane and the nearest instances, which is the number of that class. Most of the time, data is not linearly separable; in that case, SVM uses various types of functions called kernels to transform the data into the desired shape through input vectors. This study uses a linear kernel function. A linear kernel is one of the most straightforward functions for transforming data into the desired shape. These functions return the inner product between two suitable feature points in a dimensional space.

$$K(y_i, y_j) = (y_i \cdot y_{j+1}) \quad (2)$$

2.4.3 Random forest model

The ensemble learning technique known as Random Forest (RF) combines the predictive strength of many decision trees with the addition of randomization to reduce overfitting. RF produces multiple decision trees and uses bootstrapping to train each tree using a different data part. The final ranking is determined by combining the results from the individual tree, which can be

achieved by majority voting for classification tasks or by averaging for regression tasks. Mathematically, let's represent a Random Forest ensemble as a set of decision trees $(T_1, T_2, \dots, T_n)$, where each tree T_i is trained on a different bootstrapped sample from the original training dataset. The classification for a new sample X_{new} using the Random Forest ensemble can be expressed as follows:

For classification

$$\hat{y}_{\text{RF}}(X_{\text{new}}) = \text{MajorityVote}(T_1(X_{\text{new}}), T_2(X_{\text{new}}), \dots, T_n(X_{\text{new}})) \quad (3)$$

For regression

$$\hat{y}_{\text{RF}}(X_{\text{new}}) = \frac{1}{n} \sum_{i=1}^n T_i(X_{\text{new}}) \quad (4)$$

where: $\hat{y}_{\text{RF}}(X_{\text{new}})$ represents the final prediction or classification for the new sample X_{new} using the Random Forest ensemble. $T_i(X_{\text{new}})$ represents the prediction or classification made by the i^{th} decision tree in the ensemble for the new sample. Majority Vote (.) calculates the majority vote among the individual tree predictions for classification tasks, and s is the total number of decision trees in the RF ensemble. RF is powerful. After all, it may average or combine the outputs of several trees, each trained on a separate subset of the data, because it can decrease overfitting. The model's resilience and capacity for generalization are improved by this ensemble technique, making it a popular option for various machine-learning applications.

2.4.4 The AlexNet model

One of the significant developments in computer vision and deep learning, AlexNet, is a large-scale convolutional neural network (CNN) architecture that has significantly increased picture classification precision in Image Network datasets. Three entirely interconnected layers, one final softmax layer for classification, and five convolutional layers make up an AlexNet architecture. A number of new methods have also been introduced, such as dropout regularization to prevent over-adjustment, data improvement by picture reflection and cropping, and the use of rectified linear units (ReLU) as activation functions. AlexNet's primary contribution was paving the way for further developments in the field of image identification by demonstrating the efficacy of deep neural networks for such tasks. Currently, a lot of the most sophisticated CNN architectures are built on top of AlexNet's foundations and keep pushing the limits of computer vision applications like image recognition.

2.4.5 The gated recurrent unit model

The RNN design known as the Gated Recurrent Unit (GRU) solves the fading gradient issue and makes it possible to identify persistent dependencies in the order of data. It is presented as a more straightforward and effective design substitute for conventional long-term memory units (LSTMs). GRU units are made up of update and reset gates to regulate the information flow within the network. The updated door determines what exactly

should be added to the new input and what should be kept from the prior hidden state. By pressing the reset button, the network may calculate the number of hidden states from the past that are important to the input at hand. GRU can record both short- and long-term dependencies with these gates by selectively updating and resetting its hidden state in response to the input sequence. GRU's streamlined architecture, which requires fewer parameters and expedites training, is one of its advantages over LSTM. This is especially helpful when working with huge data sets or constrained computational resources. Additionally, GRU and LSTM demonstrated comparable performance in a number of tasks, including machine translation, sentiment analysis, language modeling, and language recognition. Because GRU technology can manage both short- and long-term dependencies, it has shown effectiveness in modeling sequence data. It is extensively utilized in many different domains, such as sequence data production, time series analysis, and natural language processing. To increase efficiency and accuracy, scientists and industry professionals are still investigating and fine-tuning GRU-based models and their modifications and combinations with other methods.

2.4.6 The bidirectional gated recurrent unit model

A general-purpose deep learning architecture called the Bidirectional Gated Recurrent Unit (BiGRU) is utilized for sequential data modeling, including time series, text, and speech data. It is a development of the conventional GRU design that adds two-way processing to let networks remember past, present, and future input patterns. Two contemporaneous GRU layers comprise the architecture; one processes the order of inputs forward and the other backward. The outputs from these two layers are concatenated when going through dense layers for regression and classification. For applications involving natural language processing, including machine translation, sentiment analysis, and named entity recognition, BiGRU is especially helpful when information about previous and upcoming input sequences is available. BiGRU is frequently utilized as a benchmark model to assess more intricate designs against, as it has demonstrated state-of-the-art performance for an extensive array of workloads. All things considered, sequential data modeling has been using BiGRU, a strong and adaptable deep-learning method, more and more in recent years. Time sequence data, speech, text, and other sequential data types are frequently processed using RNNs, a deep learning model. RNNs are made to work with variable-length sequences, in contrast to classic neural networks, which accept fixed inputs and generate fixed outputs. This is accomplished by adding loops to the network that allow information to endure over time. Because of this, they are especially good at processing stimuli that are sequential or have a temporal component.

2.4.7 The proposed ensemble model

Ensemble learning is an approach for improving model accuracy and efficacy. It is a powerful meta-learning strategy that combines weak and strong learners to boost the effectiveness of the weak learner. This article uses the ensemble approach to increase the accuracy of several models for BC illness prediction.

Combining several models tries to improve performance over individual models. This work creates an ensemble using six models: LGR, SVM, RF, Alexnet, GRU, and BiGRU. The ensemble model is built using a weighting approach. The weighted average ensemble approach allows several models to contribute to a forecast based on their confidence level or anticipated performance. Each member's contribution to the final forecast is weighed against the model's performance in a weighted ensemble. The model weights are tiny positive numbers totaling one, reflecting each model's degree of trust or predicted performance. This work assigns the Ensemble model weights: LGR: 0.1511, SVM: 0.2151, RF: 0.1179, Alexnet: 0.1912, GRU: 0.1329, and BiGRU: 0.1818.

2.5 Evaluation

The proposed intelligence hybrid system has been evaluated by six different performance metrics (accuracy, sensitivity, specificity, F1 score, Brier score, and error rate), an equal prediction test statistical test (the Diebold-Mariano test), and a visual analysis (bar plot, line plot, and level plot; Iftikhar et al., 2023; Cuba et al., 2024; Gonzales et al., 2024; Alshanbari et al., 2023). The specifications are as follows:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{TN} + \text{FN}}, \quad (5)$$

The following formulas show that TP is a true positive number, TN is a false negative number, FP is a false positive number, and FN is a false negative number.

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}}. \quad (6)$$

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}} \quad (7)$$

$$\text{F1-Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (8)$$

The F1 score ranges from zero to one. A score near to one implies more incredible model performance, whereas a value close to zero indicates poor performance.

$$\text{Brier Score} = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2, \quad (9)$$

where n is the number of instances, y_i is the observed binary output, and $\hat{y}_i$ is the predicted binary output for instance i . Brier Score ranges between 0 and 1, where a value closer to zero indicates better performance, while a value close to 1 shows poor performance of the predictive models.

$$\text{Error} = (y_i - \hat{y}_i), \quad (10)$$

where y_i represents actual value and $\hat{y}_i$ denotes the predicted value in specific data points.

Furthermore, to the performance metrics, the Diebold-Mariano (DM) test has been performed to examine the significance of the variations in the prediction performance of the predictive models (Diebold and Mariano, 2002). This equal prediction test is commonly used to compare predictions from various models (Qureshi et al., 2024; Iftikhar et al., 2024b; Qureshi et al., 2025; Iftikhar et al., 2024a). The DM statistic may be calculated using the following equation:

$$\text{DM}_s = \frac{\bar{y}}{\sqrt{\text{Var}(\bar{y})}}, \quad (11)$$

where,

$$\bar{y} = \frac{1}{H} \sum_{h=1}^H y_h, \quad y_h = (c_h - \tilde{c}_{1h})^2 - (c_h - \tilde{c}_{2h})^2, \quad (12)$$

$$\text{Var}(\bar{y}) = \frac{1}{H} (2 \sum_{j=1}^{h-1} r_j + r_0), \quad \text{and } r_j = \text{cov}(y_h - y_{h-j}). \quad (13)$$

the estimated value of the first predicting model is $\tilde{c}_{1h}$, whereas the predicted value of the second predicting model at time h is $\tilde{c}_{2h}$.

2.6 Selection

The results obtained after applying various algorithms to data from regular basic education students show that the ensemble model performs best in predicting academic performance, standing out for its high precision, accuracy, and sensitivity. This suggests that its implementation could be useful for educational institutions to design more effective policies and strategies aimed at strengthening and supporting students based on their estimated performance.

3 Results and discussion

Table 5 provides the performance metrics of various machine learning, deep learning, and ensemble models assessed through a cross-validation method over 500 iterations. It is categorized into three distinct scenarios, each representing a different split of training and testing data: 50%–50% (Scenario 1), 75%–25% (Scenario 2), and 90%–10% (Scenario 3). The models evaluated include Support Vector Machine (SVM), Random Forest (RF), Logistic Regression (LR), AlexNet, Gated Recurrent Unit (GRU), Bidirectional GRU (BiGRU), and an Ensemble model.

Each model is assessed based on metrics such as Accuracy, Sensitivity, Specificity, F1 Score, Brier Score (BS), and Error Rate. In all three scenarios, the Ensemble model consistently delivers the highest scores in terms of accuracy, sensitivity, specificity, and F1 score, while also achieving the lowest error rates and Brier scores. This indicates that integrating multiple models leads to enhanced performance and robustness.

When analyzing individual models, deep learning-based frameworks like AlexNet, GRU, and BiGRU generally outperform conventional machine learning models such as SVM, RF, and LR in most instances. AlexNet and BiGRU show strong sensitivity

TABLE 5 Performance measures: results of the evaluation of the deep learning, machine learning, and ensemble models with a cross-validation approach of 500 executions.

Models	Accuracy	Sensitivity	Specificity	FS	BS	Error
1st scenario (50%, 50%)						
SVM	0.9055	0.9538	0.9538	0.9599	0.0242	0.0945
RF	0.8826	0.9217	0.9217	0.9281	0.0359	0.1174
LR	0.8854	0.9296	0.9296	0.9388	0.0325	0.1146
Alexnet	0.9089	0.9582	0.9582	0.9641	0.0208	0.0911
GRU	0.8953	0.9518	0.9518	0.9467	0.0258	0.1047
BiGRU	0.8994	0.9522	0.9522	0.9515	0.0272	0.1006
Ensemble	0.9244	0.9715	0.9715	0.9774	0.0187	0.0756
Models	Accuracy	Sensitivity	Specificity	F1 score	Brier score	Error
2nd scenario (75%, 25%)						
SVM	0.9111	0.9594	0.9594	0.9655	0.0216	0.0889
RF	0.8882	0.9273	0.9273	0.9337	0.0333	0.1118
LR	0.8910	0.9352	0.9352	0.9444	0.0299	0.1090
Alexnet	0.9145	0.9638	0.9638	0.9697	0.0182	0.0855
GRU	0.9009	0.9574	0.9574	0.9523	0.0232	0.0991
BiGRU	0.9050	0.9578	0.9578	0.9571	0.0246	0.0950
Ensemble	0.9244	0.9715	0.9715	0.9774	0.0187	0.0756
3rd scenario (90%, 10%)						
SVM	0.9137	0.9637	0.9637	0.9698	0.0196	0.0864
RF	0.8908	0.9316	0.9316	0.9380	0.0313	0.1093
LR	0.8936	0.9395	0.9395	0.9487	0.0279	0.1065
Alexnet	0.9171	0.9681	0.9681	0.9740	0.0162	0.0830
GRU	0.9035	0.9617	0.9617	0.9566	0.0212	0.0966
BiGRU	0.9076	0.9621	0.9621	0.9614	0.0226	0.0925
Ensemble	0.9244	0.9715	0.9715	0.9774	0.0187	0.0756

and specificity values, highlighting their effectiveness in accurately identifying both positive and negative instances. Among the machine learning models, SVM ranks highest, followed closely by LR and RF.

As the proportion of training data increases across the scenarios (from 50% in Scenario 1 to 90% in Scenario 3), the accuracy, sensitivity, specificity, and F1 score tend to rise, while the error rate and Brier score tend to decrease. This illustrates that a larger training dataset aids in enhancing model generalization and performance. In summary, the findings highlight the advantages of deep learning techniques compared to traditional machine learning models in this classification challenge. Furthermore, the ensemble model emerges as the top performer across all evaluated metrics, making it the most dependable option for this particular issue.

On the other hand, after evaluating the model performance with accuracy metrics (accuracy, sensitivity, specificity, F1 score, Brier score, and error), the same prediction statistical tests (DM tests) were used to explore the significance of different prediction abilities between models. Table 6 shows the DM test of each

model pair to determine the outcome quality (mean performance measurement). The results of the DM test (*p*-values) are shown in Table 6. For clarity, the DM test examines whether the forecasting errors of two models are statistically different. A *p*-value below 0.05 indicates that the difference in predictive performance between two models is statistically significant at the 5% level, while higher *p*-values suggest that the models perform similarly. In this study, the table confirms that the ensemble model consistently achieves statistically significant improvements compared to the other six prediction models across all data-splitting scenarios.

Figure 2 illustrate the performance metrics of various predictive models evaluated under three distinct scenarios: (a) 50% training data and 50% testing data, (b) 75% training data and 25% testing data and (c) 90% training data and 10% testing data. The models considered include Support Vector Machine (SVM), Random Forest (RF), Logistic Regression (LR), AlexNet, Gated Recurrent Unit (GRU), Bidirectional GRU (BiGRU), and an Ensemble model. The metrics employed for evaluation consist of Accuracy, Sensitivity, Specificity, F1 Score

TABLE 6 All predictive models (p-values) of DM test results are considered in all training and test dataset scenarios.

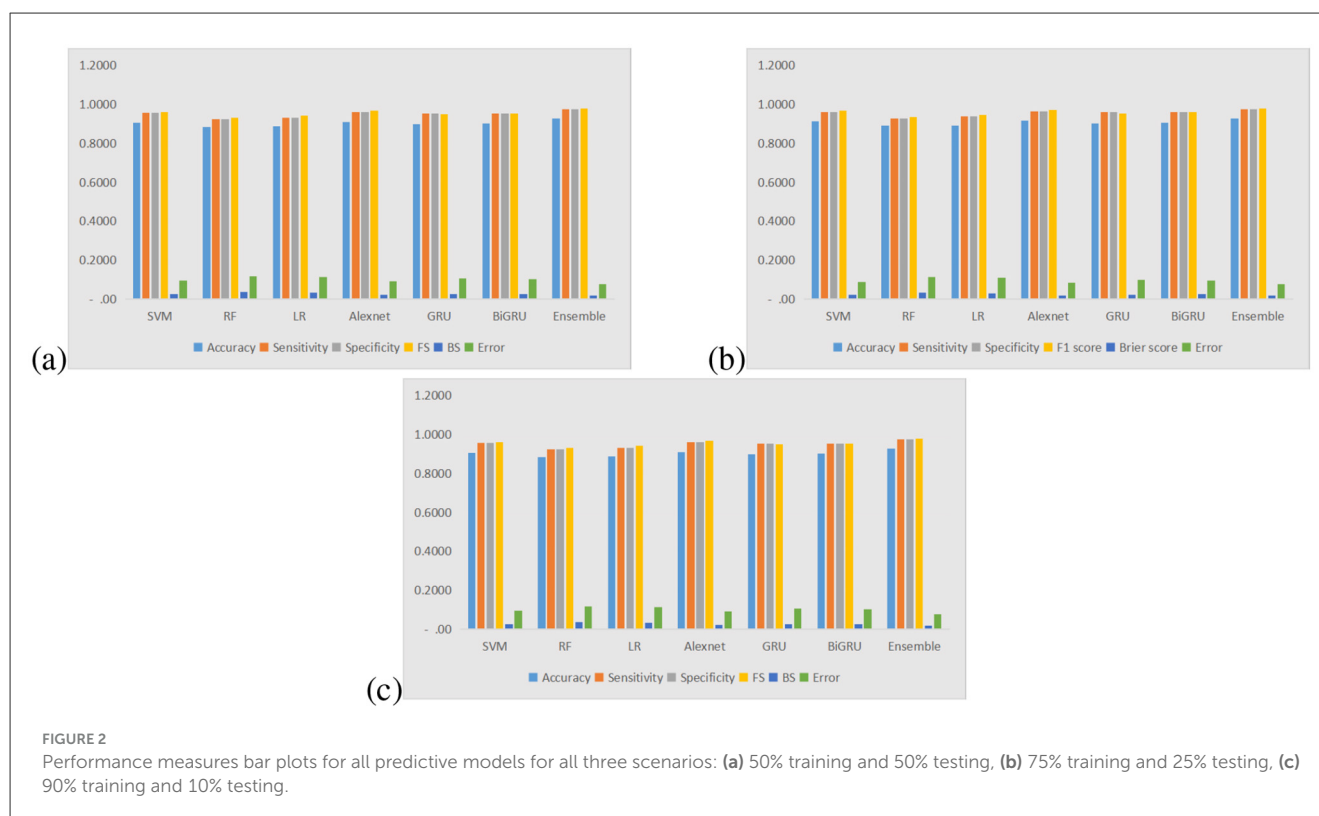
Model	SVML	RF	LGR	Alexnet	GRU	BiGRU	Ensemble
1st scenario (50%, 50%)							
SVML	0.00	0.02	0.01	0.99	0.06	0.05	0.99
RF	0.98	0.00	0.98	0.98	0.98	0.98	0.99
LGR	0.99	0.03	0.00	0.99	0.97	0.98	0.99
Alexnet	0.01	0.02	0.01	0.00	0.03	0.02	0.99
GRU	0.94	0.02	0.03	0.97	0.00	0.93	0.98
BiGRU	0.95	0.02	0.02	0.98	0.07	0.00	0.98
Ensemble	0.01	0.01	0.01	0.01	0.02	0.02	0.00
2st scenario (75%, 25%)							
SVML	0.00	0.02	0.01	0.99	0.06	0.05	0.99
RF	0.99	0.00	0.98	0.99	0.98	0.98	0.99
LGR	0.99	0.02	0.00	0.99	0.97	0.98	0.99
Alexnet	0.01	0.02	0.01	0.00	0.03	0.02	0.99
GRU	0.94	0.02	0.03	0.97	0.00	0.93	0.98
BiGRU	0.95	0.02	0.02	0.98	0.07	0.00	0.99
Ensemble	0.01	0.01	0.01	0.01	0.02	0.01	0.00
3st scenario (90%, 10%)							
SVML	0.00	0.01	0.01	0.99	0.06	0.05	0.99
RF	0.99	0.00	0.98	0.99	0.98	0.98	0.99
LGR	0.99	0.02	0.00	0.99	0.97	0.98	0.99
Alexnet	0.01	0.01	0.01	0.00	0.03	0.02	0.99
GRU	0.94	0.02	0.03	0.97	0.00	0.93	0.98
BiGRU	0.95	0.02	0.02	0.98	0.07	0.00	0.99
Ensemble	0.01	0.01	0.01	0.01	0.02	0.01	0.00

(FS), Brier Score (BS), and Error. Throughout all scenarios, the models demonstrate consistently high levels of accuracy, sensitivity, specificity, and F1 scores, indicating their effectiveness in classification tasks. The Ensemble model consistently achieves the best results, exhibiting minimal error and advanced classification abilities.

- In scenario (a) **50% training–50% testing**, the models show somewhat lower accuracy relative to scenarios with a greater amount of training data. This indicates that a smaller training dataset causes a slight decrease in the models' generalization capabilities.
- In scenario (b) **75% training–25% testing**, the performance across all models improves, with higher accuracy and decreased error. This underscores the notion that more training data positively influences model performance.
- In scenario (c) **90% training–10% testing**, the models achieve optimal performance, as evidenced by the highest accuracy and the lowest error rates. The Ensemble model particularly distinguishes itself as the most dependable model, demonstrating strong generalization capabilities.

- **SVM, RF, and LR:** these traditional machine learning models display competitive results but tend to fall short compared to deep learning models (AlexNet, GRU, and BiGRU) regarding sensitivity and specificity.
- **AlexNet, GRU, and BiGRU:** these deep learning models illustrate enhanced classification capabilities, particularly in sensitivity and F1 score. Their effectiveness in learning intricate patterns within data contributes to their outstanding performance.
- **Ensemble model:** this model consistently surpasses all other models, suggesting that integrating multiple classifiers enhances predictive accuracy and robustness.

The Brier score and error metrics decrease with an increase in the training data proportion, signifying that more training data results in better-calibrated predictions and fewer misclassification incidents. The lowest error rates are recorded in the **90% training–10% testing** scenario, reinforcing the advantages of augmented training data. This analysis reveals that enhancing the amount of training data substantially boosts model performance across all evaluated metrics. Deep learning models (AlexNet,



GRU, BiGRU) and the Ensemble model consistently outperform conventional machine learning methods. Notably, the Ensemble model exhibits superior classification accuracy, sensitivity, and specificity, positioning it as the most trustworthy predictive model in all tested scenarios. These insights highlight the value of employing ensemble and deep learning approaches for critical classification tasks, particularly when ample training data is available.

4 Conclusions

The findings allow us to conclude that the chosen deep learning, machine learning, and ensemble methods demonstrated a strong predictive ability, primarily due to the linear correlation between academic characteristics and student outcomes. Among the techniques utilized, three machine learning models (Logistic Regression, Support Vector Machine, and Random Forest), along with deep learning models (AlexNet, Gated Recurrent Unit, and Bidirectional Gated Recurrent Unit), were analyzed, as well as ensemble models. Nonetheless, the Ensemble model, which combines deep learning and machine learning techniques, is preferred due to its superior accuracy, precision, and sensitivity performance metrics. The critical factors for assessing a student's academic performance at the Peruvian university include the number of Failed Courses and the grades received in the first 2 years, as these are crucial determinants for the student's academic success. To ensure that a student maintains good academic performance in subsequent years, proactive measures should be taken in alignment with the model's predictions during the

initial two years. Furthermore, the consistency of the proposed predictive system was validated by dividing the entire dataset into three training and testing scenarios [(90%, 10%), (75%, 25%), and (50%, 50%)], conducting a comparative assessment of the models using six performance metrics, graphical analyses, and statistical testing through five hundred simulation runs. According to the evaluation results, the ensemble model consistently outperformed the other models across all three training and testing scenarios.

Despite the promising results, several limitations must be acknowledged. First, the dataset is restricted to 11 schools of the Adventist Educational Network in the macro-south of Peru, which raises questions about external validity. The predictive power of the model may be context-dependent, as socio-economic, cultural, and institutional factors not captured in the dataset could significantly influence academic performance in other settings. Therefore, extrapolation of results beyond this specific institutional and geographical context should be approached with caution. Second, although rigorous data cleaning and verification procedures were performed, reliance on institutional records may still introduce biases or gaps in data collection that could affect the outcomes. Finally, while the models achieve high predictive performance, the absence of explicit interpretability analysis (e.g., SHAP or LIME) limits their immediate practical applicability for educators and decision-makers.

Future research should address these limitations by (i) testing the model on external datasets from diverse institutional and cultural contexts to assess generalizability, (ii) incorporating broader socio-economic and demographic variables to capture a wider range of determinants of student success, (iii) evaluating

longitudinal datasets to study dynamic academic trajectories, and (iv) integrating explainability frameworks to provide actionable insights into the role of specific features. By pursuing these directions, subsequent studies can enhance the robustness, fairness, transparency, and practical utility of AI-driven predictive systems in education. Likewise, can also be extended to other scenarios with different datasets (Ifthikhar et al., 2024; Cabello-Torres et al., 2022; Cruz et al., 2020).

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Author contributions

WG: Conceptualization, Writing – original draft, Writing – review & editing, Investigation. ZC: Investigation, Writing – review & editing, Conceptualization, Writing – original draft. CA-R: Writing – review & editing, Investigation, Supervision, Formal analysis, Writing – original draft, Data curation. ES: Writing – original draft, Writing – review & editing, Methodology, Supervision, Project administration, Validation, Funding acquisition, Software. HI: Conceptualization, Writing – review & editing, Supervision, Formal analysis, Software, Data curation, Writing – original draft. JL-G: Visualization, Funding acquisition, Resources, Validation, Formal analysis, Project administration, Writing – original draft, Conceptualization, Investigation, Data curation, Supervision, Writing – review & editing, Methodology, Software.

References

- Adefemi, K. O., Mutanga, M. B., and Jugoo, V. (2025). Hybrid deep learning models for predicting student academic performance. *Math. Comput. Appl.* 30:59. doi: 10.3390/mca30030059
- Ahmed, S., Helmy, Y., and Ouf, S. (2022). “A deep learning framework for predicting the student’s performance in the virtual learning environment,” in *2022 5th International Conference on Computing and Informatics (ICCI)* (Cairo), 240–250. doi: 10.1109/ICCI54321.2022.9756058
- Alshanbari, H. M., Ifthikhar, H., Khan, F., Rind, M., Ahmad, Z., and El-Bagoury, A. A. A. H. (2023). On the implementation of the artificial neural network approach for forecasting different healthcare events. *Diagnostics* 13:1310. doi: 10.3390/diagnostics13071310
- Ariza, C. P., Toncel, L., Á. R., and Blanchar, J. S. (2018). El rendimiento académico: una problemática compleja. *Rev. Bol. Redipe* 7, 137–141.
- Baniata, L. H., Kang, S., Alsharaiah, M. A., and Baniata, M. H. (2024). Advanced deep learning model for predicting the academic performances of students in educational institutions. *Appl. Sci.* 14:1963. doi: 10.3390/app14051963
- Brown, M. (2018). Deep learning in education: opportunities and challenges. *J. Artif. Intell. Educ.*
- Cabello-Torres, R. J., Estela, M. A. P., Sánchez-Ccoyllo, O., Romero-Cabello, E. A., Ávila, F. F. G., Castañeda-Olivera, C. A., et al. (2022). Statistical modeling approach for pm10 prediction before and during confinement by COVID-19 in south lima, peru. 12:16737. doi: 10.1038/s41598-022-20904-2
- Castro, M., and Paredes, J. (2022). Unique challenges in predictive modeling for regular basic education. *J. Educ. Res.*
- Chavez, H., Chavez-Arias, B., Contreras-Rosas, S., Alvarez-Rodríguez, J. M., and Raymundo, C. (2023). Artificial neural network model to predict student performance using nonpersonal information. *Front. Educ.* 8:1106679. doi: 10.3389/feduc.2023.1106679
- Chen, S., and Ding, Y. (2023). A machine learning approach to predicting academic performance in Pennsylvania’s schools. *Soc. Sci.* 12:118. doi: 10.3390/socsci12030118
- Cruz, A. R. H. D. L., Ayuque, R. F. O., Cruz, R. W. H. D. L., Lopez-Gonzales, J. L., and Gioda, A. (2020). Air quality biomonitoring of trace elements in the metropolitan area of huancayo, peru using transplanted tillandsia capillaris as a biomonitor. *An. Acad. Bras. Cienc.* 92:e20180813. doi: 10.1590/0001-3765202020180813
- Cuba, W. M., Huaman Alfaro, J. C., Ifthikhar, H., and López-Gonzales, J. L. (2024). Modeling and analysis of monkeypox outbreak using a new time series ensemble technique. *Axioms* 13:554. doi: 10.3390/axioms13080554
- Diebold, F. X., and Mariano, R. S. (2002). Comparing predictive accuracy. *J. Bus. Econ. Stat.* 20, 134–144. doi: 10.1198/073500102753410444
- Fahd, K., Venkatraman, S., Miah, S. J., and Ahmed, K. (2022). Application of machine learning in higher education to assess student academic performance, at-risk, and attrition: a meta-analysis of literature. *Educ. Inf. Technol.* 27, 1–33. doi: 10.1007/s10639-021-10741-7
- Fanta, S., and Alemu, D. (2025). Student academic performance prediction using random forest. *Afr. J. Inf. Syst.*
- Garcia, M., and Alvarez, J. (2021). Hybrid machine and deep learning models for predictive analytics. *Expert Syst. Appl.*
- Ghosh, S., and Sharma, R. (2022). Application of machine learning algorithms for student performance prediction. *Educ. Inf. Technol.*

Funding

The author(s) declare that no financial support was received for the research and/or publication of this article.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declare that no Gen AI was used in the creation of this manuscript.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

Publisher’s note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

- Gonzales, S. M., Iftikhar, H., and López-Gonzales, J. L. (2024). Analysis and forecasting of electricity prices using an improved time series ensemble approach: An application to the peruvian electricity market. *AIMS Math* 9, 21952–21971. doi: 10.3934/math.20241067
- Hussain, S., and Khan, M. Q. (2023). Student-performulator: predicting students' academic performance at secondary and intermediate level using machine learning. *Ann. Data Sci.* 10, 637–655. doi: 10.1007/s40745-021-00341-0
- Iftikhar, H., Gonzales, S. M., Zywiolok, J., and López-Gonzales, J. L. (2024a). Electricity demand forecasting using a novel time series ensemble technique. *IEEE Access* 12, 88963–88975. doi: 10.1109/ACCESS.2024.3419551
- Iftikhar, H., Khan, M., Żywiolek, J., Khan, M., and López-Gonzales, J. L. (2024c). Modeling and forecasting carbon dioxide emission in pakistan using a hybrid combination of regression and time series models. *Heliyon* 10:e33148. doi: 10.1016/j.heliyon.2024.e33148
- Iftikhar, H., Khan, M., Khan, Z., Khan, F., Alshanbari, H. M., and Ahmad, Z. (2023). A comparative analysis of machine learning models: a case study in predicting chronic kidney disease. *Sustainability* 15:2754. doi: 10.3390/su15032754
- Iftikhar, H., Khan, M., Turpo-Chaparro, J. E., Rodrigues, P. C., and López-Gonzales, J. L. (2024b). Forecasting stock prices using a novel filtering-combination technique: application to the pakistan stock exchange. *AIMS Math* 9, 3264–3288. doi: 10.3934/math.2024159
- Kord, A., Aboufetouh, A., and Shohieb, S. M. (2025). Academic course planning recommendation and students' performance prediction multi-modal based on educational data mining techniques. *J. Comput. High. Educ.* 1–39. doi: 10.1007/s12528-024-09426-0
- Lopez, J., and Ramirez, M. (2021). Early warning system for at-risk students using data mining. *Comput. Educ.*
- Mashrur, A., and Nonyelum, I. (2020). Machine learning for education: a review. *Int. J. Educ. Technol. High. Educ.*
- Menchaca, D. P. (2024). Modelos predictivos de rendimiento académico universitario mediante aprendizaje automático. *Rev. Cient. Salud Desarro. Hum.* 5, 974–991. doi: 10.61368/r.s.d.h.v5i2.204
- Mi, F., and Yeung, D.-Y. (2019). "Deep learning models for predicting student performance," in *Proceedings of the AAAI Conference on Artificial Intelligence* [Washington, DC: Association for the Advancement of Artificial Intelligence (AAAI)].
- Mohammadi, M., Dawodi, M., Tomohisa, W., and Ahmadi, N. (2019). "Comparative study of supervised learning algorithms for student performance prediction," in *2019 International Conference on Artificial Intelligence in Information and Communication (ICAIIIC)* (Okinawa), 124–127. doi: 10.1109/ICAIIIC.2019.8669085
- Murillo, N. A. (2024). "Factores asociados al rendimiento académico en educación secundaria: una revisión sistemática," in *Asociación de Psicología y Educación Facultad de Educación-cfp Universidad Complutense de Madrid* (Madrid: Universidad Complutense de Madrid), 73.
- Nguyen, P., and Do, M. (2019). Predictive modeling in education: a systematic review. *Educ. Technol. Soc.*
- Núñez, J. C., Tuero, E., Fernández, E., Añón, F. J., Manalo, E., and Rosário, P. (2022). Efecto de una intervención en estrategias de autorregulación en el rendimiento académico en primaria: estudio del efecto mediador de la actividad autorregulatoria. *Rev. Psicodidáctica* 27, 9–20. doi: 10.1016/j.psicod.2021.09.001
- Pachas, R. (2020). *Gestión de la Educación Básica en el Perú*. Ministerio de Educación del Perú.
- Pacheco, R. J. P., Valdez, M. J. P., Viteri, S. P., and Duque, M. F. R. (2021). Gestión educativa: factor clave en la implementación del currículo de educación física. *Rev. Venez. Gerenc.* 26, 232–247. doi: 10.52080/rvgluz.26.e5.16
- Qureshi, M., Iftikhar, H., Rodrigues, P. C., Rehman, M. Z., and Salar, S. A. (2024). Statistical modeling to improve time series forecasting using machine learning, time series, and hybrid models: a case study of bitcoin price forecasting. *Mathematics* 12:3666. doi: 10.3390/math12233666
- Qureshi, M., Ishaq, K., Daniyal, M., Iftikhar, H., Rehman, M. Z., and Salar, S. A. (2025). Forecasting cardiovascular disease mortality using artificial neural networks in Sindh, Pakistan. *BMC Public Health* 25:34. doi: 10.1186/s12889-024-21187-0
- Ravšelj, D., Keržič, D., Tomaževič, N., Umek, L., Brezovar, N., Iahad, N. A., et al. (2025). Higher education students' perceptions of chatgpt: a global study of early reactions. *PLoS ONE* 20:e0315011. doi: 10.1371/journal.pone.0315011
- Reyes, J. E. A., Peñafiel, E. F. M., Barrionuevo, T. P. M., and Castelo, G. V. (2025). Estudio del rendimiento académico mediante la comparación de modelos de regresión y árboles de clasificación. *Telos: Rev. Estud. Interdiscip. Cienc. Soc.* 27, 94–115. doi: 10.36390/telos271.08
- Santander, O. A. E. (2011). El rendimiento académico, un fenómeno de múltiples relaciones y complejidades. *Rev. Vanguardia Psicol. Clín. Teó. Práct.* 2, 144–173.
- Singh, A., and Kumar, P. (2020). Machine learning for student performance prediction: a survey. *J. Educ. Learn.*
- Smith, J., and Lee, K. (2017). Classical machine learning algorithms in education: review and applications. *Int. J. Mach. Learn. Educ.*
- Sultana, N., and Khan, S. (2019). Student performance prediction using machine learning. *Int. J. Adv. Comput. Sci. Appl.*
- Tripon, C. (2022). Supporting future teachers to promote computational thinking skills in teaching stem—a case study. *Sustainability* 14:12663. doi: 10.3390/su141912663
- Vargas, C. (2020). *El reto de la educación en el Perú: Proyecto Educativo Nacional al 2036*. Consejo Nacional de Educación.
- Zhang, W., and Chen, L. (2021). Gru and bigru models for educational sequence prediction. *Knowl.-Based Syst.*