Scientific Reports

https://doi.org/10.1038/s41598-026-40538-y

Article in Press

Explainable AI in education: integrating educational domain knowledge into the deep learning model for improved student performance prediction

Received: 20 October 2025

Accepted: 13 February 2026

Published online: 18 February 2026

Cite this article as: Qiang M., Liu Z. & Zhang R. Explainable AI in education: integrating educational domain knowledge into the deep learning model for improved student performance prediction. *Sci Rep* (2026). https://doi.org/10.1038/s41598-026-40538-y

Ming Qiang, Ziyang Liu & Ru Zhang

Explainable AI in Education: Integrating Educational Domain Knowledge into the Deep Learning Model for Improved Student Performance Prediction

Ming Qiang [1], Ziyang Liu[2], Ru Zhang [3]*
[1] Centre of International Education, Fuzhou Polytechnic, Fuzhou, 350108, Fujian, China
[2] Asia-Pacific AI Research Center, Asian Business Research Institute, Hongkong, China, 999077
[3] Department of Design, Hanyang University, Seoul 04763, Korea

* Corresponding authors: Ru Zhang (zhangru@hanyang.ac.kr)

Abstract

Although deep learning models, especially the Artificial Neural Network (ANN), are widely used for student performance prediction, their “black-box” nature often leads to unreliable learned relationships that contradict educational domain knowledge, limiting both trustworthiness and further performance improvement. Based on Shapley Additive Explanations (SHAP), this study developed an ANN using a public dataset containing 395 Portuguese high school students' mathematics performance records. The analysis identified key features influencing students' mathematics performance and revealed that the original correlations learned by the ANN were inconsistent with established educational domain knowledge. To address this issue, we proposed the Student Performance Prediction Explanation (SPPE) algorithm for optimizing ANN, which reassessed the contribution of 30 features under the guidance of educational domain knowledge. Both global and local interpretability analyses were conducted to examine the process of importance changes. Furthermore, this study found that after aligning the model with educational domain knowledge, the prediction accuracy of the proposed ANN achieved a 26.9% improvement compared with the original model. In addition, it outperformed some typical traditional machine learning algorithms. Additional experiments further confirmed that the proposed SPPE strategy is applicable to various ANN architectures, supporting its robustness across model structures within this dataset and reinforcing its generalizability and practical value. The findings of this study demonstrated that integrating educational domain knowledge can improve student performance prediction, contributing to the development of interpretable neural network frameworks and offering actionable insights for other educational applications.

Keywords: Student performance prediction; Explainable AI; Deep learning; SHAP; Educational domain knowledge; Educational technology

1. Introduction

Education plays a crucial role in transferring cultural legacy and values across generations. Undoubtedly, the academic performance of students is a vital benchmark to assess education quality. It plays an irreplaceable role in overseeing the quality of teaching and learning in schools. However, education is facing significant challenges such as rising dropout rates and persistent concerns about students' academic readiness for higher education. Given this situation, the study of students' academic achievement, especially predicting their performance, should be considered. Not only can it enable the administrative authority to put forward better policies in educational management [[1]] but it can also assist educators in improving their teaching methods [[2]]. In addition, students, especially those who are at risk of academic failure, can utilize feedback based on the predicted academic performance to adjust their learning strategies in a timely manner. However, accurate prediction requires comprehensive consideration of multiple determinants including socio-economic background, individualized learning styles, digital literacy, and environmental factors [[3]]. Limited by rudimentary analytical methods, early-stage research employed simplistic statistical models for performance prediction [[4]]. Yet, these investigations struggle to effectively explore the intricate interactions among variables within massive educational information and elucidate the results reasonably based on pedagogical theory and practice.

To overcome these limitations, Educational Data Mining (EDM) technologies have recently been introduced to predict student performance due to their superior performance [[5]]. As a data-driven pedagogical framework, EDM integrates Artificial Intelligence (AI), Knowledge Discovery in Databases (KDD), and data warehousing to analyze learning behaviors, validate educational strategies, and enhance educational outcomes [[6]]. Notably, Machine Learning (ML) algorithms such as decision trees [[7]], support vector machines [[8]], Bayesian networks [[9]], and ensemble methods [[10]] have been extensively applied[81]. However, the accuracy of these algorithms is often limited when dealing with noisy and complex educational datasets.

In order to seek more precise prediction, Artificial Neural Networks (ANNs) are increasingly explored due to their superior capability in detecting non-linear relationships [[11]-[15]]. For instance, Zacharis [[16]] developed an ANN model achieving 98.3% accuracy in predicting blended learning outcomes, while Saputra [[17]] attained 97.5% accuracy in e-learning performance prediction using activity logs. Although related studies on forecasting students' performance have achieved remarkable results, little effort has been made to explore issues related to the interpretability of ANN models in this field. The black-box nature of ANN increases the difficulty of understanding model predictions and restricts the discovery of useful information that is essential for creating customized interventions that improve student performance [[18]]. Based on such considerations, the establishment of interpretable ANN-based prediction models has received more attention recently. Özkurt [[19]] conducted a study to enhance the interpretability of ANN models by identifying the importance of features using XAI tools such as Local Interpretable Model-Agnostic Explanations (LIME) and SHapley Additive exPlanations (SHAP). However, as the interpretability of ANN-based models has improved in recent studies, new challenges have also emerged. In particular, the importance of some identified features does not match expectations based on established educational knowledge. For instance, contradicting prior research [[20], [21]], Özkurt's model [[19]] mistakenly identified the father's occupation as the least important factor and underestimated the influence of the mother's education [[22], [23]]. These discrepancies may be due to the models focusing too heavily on data patterns while neglecting the value of domain insights. This can reduce the trustworthiness of the prediction results and weaken the model's ability to uncover deeper and more meaningful correlations for further performance improvement.

To address this issue of misalignment between data-driven feature importance and domain knowledge, this study introduced an interpretable, domain-knowledge-aligned ANN model for student performance prediction. The model first employed the SHAP algorithm to reliably evaluate the contribution of 30 features to students' mathematics achievement. Then, this study proposed a novel Students' Performance Prediction Explanation (SPPE) algorithm, which is compatible with SHAP and designed to calculate weight scores for these features. Based on the importance of features represented by SHAP values, the features whose importance contradicts established educational domain knowledge are identified, and their weight scores are optimized in the loss function to guarantee the neural network can unearth reasonable correlation to improve the interpretability and prediction performance of the model.

To further clarify the focus of this study, three research questions are summarized as follows: 1) This study examines whether the feature importance patterns learned by ANN model-based student performance prediction align with established educational domain knowledge. 2) This study explores how educational domain knowledge can be systematically incorporated into ANN models to improve their interpretability. 3) This study investigates whether the integration of domain knowledge into ANN models can enhance the accuracy of student performance prediction. Therefore, the focus of this study is not on optimizing model architectures to pursue better predictive performance, but rather on examining how educational domain knowledge can be embedded into ANN models. This work is situated within the growing field of knowledge-infused AI, which seeks to combine data-driven machine learning with explicit human knowledge. To this end, we propose a novel framework, the Students' Performance Prediction Explanation (SPPE) algorithm. Distinct from conventional regularization methods that enforce generic model properties, SPPE provides a specific mechanism to inject context-dependent domain knowledge directly into the model's learning process. It achieves this by using SHAP values as a proxy for the model's learned feature importance hierarchy and optimizing this hierarchy to align with established educational theory. To the best of our knowledge, this is the first work in the field of student performance prediction that explores this particular approach to knowledge infusion, validating its role in influencing both interpretability and predictive accuracy.

The overall structure of the rest of this paper is outlined as follows. Section 2 reviews relevant research on traditional machine learning and neural network methods for student performance prediction. Section 3 introduces the methodology of the proposed framework, including dataset description, model construction, and SPPE optimization strategy. Section 4 reports the experimental results and discusses the interpretability and predictive performance in detail. Finally, Section 5 summarizes the main findings of the study, discusses limitations, and outlines future research directions.

2. Related works

2.1. Traditional machine learning approaches for student performance prediction

As the simplest algorithm, the Linear Regression (LR) method has been considered widely to predict students' performance. Sravani et al. and Dong et al. validated the feasibility of LR in this domain [[24], [25]]. Despite its widespread acceptance due to simplicity and explainability, LR is inferior to non-linear regression methods in capturing nonlinear relationships among variables. As one of the fundamental non-linear machine learning algorithms, Decision Trees (DTs) have been extensively studied. For instance, Hamoud et al. identified the J48 decision tree algorithm as more reliable than random tree and REPtree in the context of higher education [[26]].

However, the single machine learning algorithm such as decision tree has its limitations, such as overfitting which could reduce its performance in achieving more precise predictions. In this case, the ensemble method is preferred by some researchers to solve the overfitting problem through integrating prediction results from different machine learning algorithms [[27]]. For example, Kumar et al. [[28]] explored an ensemble approach combining ID3, J48, and tools in the Weka platform, achieving a prediction accuracy of 62.67%. Based on a students' dataset comprising 1,000 examples and 22 variables for assessing performance, Singh & Pal [[29]] proposed a reliable and valid prediction model ensembling bagging technique with Extra Tree (ET), K-nearest Neighbor (KNN), Naïve Bayes (NB), and DT [84].

Although traditional machine learning methods such as LR and DT can offer simplicity and interpretability, they lack the capacity to capture high-level nonlinear relationships between variables influencing students' performance [83]. In addition, they rely heavily on manual feature engineering, which limits their scalability.

2.2.Neural network approaches for student performance prediction

To further improve prediction performance, ANNs with superior feature extraction capabilities have been introduced to capture complex nonlinear relationships among variables. Song et al. [[30]] developed an Elastic Grey Wolf Optimization algorithm (EGWO) to optimize the weights and biases of Multilayer Perceptron (MLP), which is a basic form of ANN. As leading representatives of third-generation

ANNs, Recurrent Neural Networks (RNNs), Deep Neural Networks (DNNs), Convolutional Neural Networks (CNNs), and Long Short-Term Memory (LSTM) have demonstrated remarkable potential in students' performance prediction. RNNs were shown to achieve higher accuracy than multiple regression analysis in predicting students' final grades using log data from educational systems [[31]]. After conducting a thorough study on the dataset's features, Poudyal et al. [[32]] introduced the CNNs algorithm and optimized its prediction accuracy through transforming the uni-dimensional data into 2-dimensional image. Minn [[33]] proposed a Bayesian Knowledge Tracing and Long Short-Term Memory (BKT-LSTM) model to improve prediction accuracy by incorporating individual skill mastery, cross-skill learning transfer, and problem difficulty [82]. Comparative experiments by Sikder et al. [[34]] and Mohamed [[35]] further confirmed ANNs' superiority over algorithms like NB, SVM, and RF.

Despite their excellent prediction performance, the "black-box" nature of widely used ANNs hinders understanding and explaining corresponding prediction outcomes [[36]]. Maschler et al. [[37]] highlighted that more investigations should be made to strike an excellent balance between the accuracy and the trustworthiness of the students' performance prediction. Under these considerations, Özkurt [[19]] incorporated Explainable Artificial Intelligence (XAI) tools such as the SHAP with the developed ANN model to present insights on the input features influencing students' achievement. However, new concerns arise with the enhancement of the interpretability of ANN-based models. Some crucial features identified in the study are not in accordance with the established educational domain knowledge. For instance, unusually low weights were assigned to features previously proven to be significant, which impedes their intended role in predicting students' performance. The Özkurt's ANN identifies the father's job as the least crucial feature influencing students' study achievement [[19]], which is contrary to the established domain knowledge constructed by other studies [[20], [21]]. In contrast to the studies of Bassetto [[22]] and Da Silva et al. [[23]], this ANN model also indicated that the mother's education background has limited impact on students' academic performance. These contradictory phenomena mentioned above not only limit the further improvement but also reduce the reliability of the models.

Although neural networks have been demonstrated in many studies as more effective approaches for predicting student performance, their limited interpretability and tendency to capture variable importance that contradicts established educational knowledge restrict further improvements in model performance. Compared with the previous study on interpretability in student performance prediction [[19]], which mainly reported feature importance without addressing potential conflicts with domain knowledge, this study places greater emphasis on dealing with the discordance between model interpretability and educational domain knowledge, aiming to achieve improvements in both interpretability and prediction performance.
In order to clearly distinguish our approach from existing studies, Table 1 outlines the main methods, and limitations of representative works, together with the differences from this study.

Table 1. Comparison of closely related studies and this work.

Study	Method	Limitation	Differences from this study
Sravani et al. [24]; Dong et al. [25]	LR	Cannot capture nonlinear relationships; limited predictive accuracy	Our work moves beyond LR by incorporating nonlinear ANN models while addressing interpretability through domain knowledge
Hamoud et al. [26]	Decision Tree (J48)	Overfitting; low generalizability	We focus on ANN-based framework that integrates domain knowledge to reduce overfitting and improve trustworthiness
Kumar et al. [28]; Singh & Pal [29]	Ensemble methods (ID3, J48, KNN, NB, ET)	Still require heavy manual feature engineering; limited interpretability	Our framework reduces reliance on feature engineering by using ANN feature extraction and interpretable techniques
Song et al. [30]; Sikder et al. [34]; Mohamed [35]	ANN (MLP, CNN, LSTM, DNN, RNN)	“Black-box” nature; lack of interpretability	We introduce SPPE to enhance interpretability while maintaining predictive performance
Özkurt [19]	ANN + SHAP	Some feature importance contradicted established educational knowledge	Our work explicitly aligns interpretability results with domain knowledge to avoid contradictions

3. Methodology

3.1. The Overview of the Proposed Interpretable Neural Network-Based Framework

A novel interpretable approach was proposed in this study to predict student performance based on the ANN model to explore a balance between model interpretability and accuracy. Figure 1 visualizes the proposed framework with the comprehensive, multi-stage process. The dataset analyzed was sourced from the UCI Machine Learning Repository, containing 395 instances and 30 features related to student performance in Mathematics [[38]]. Although the sample size is relatively limited, this dataset is among the most widely used benchmarks in student performance prediction research, offering a rich set of features that capture demographic, social, and school-related aspects. Its extensive application in previous studies has also demonstrated its reliability. Therefore, we selected it as the basis for our analysis. Data pre-processing was initially conducted to prepare the data for machine learning models, including converting string variables, standardization and splitting the original dataset into training and test subsets. Then an ANN model was built to be trained and tested based on the pre-processed data. In this study, the ANN was trained using a regression strategy, where the model learned to predict continuous academic performance scores. This regression-based training choice affects optimisation and interpretation in two ways. First, predicting the continuous $G3$ score provides a smoother gradient signal than directly optimising a three-class objective, which stabilises learning on this small tabular dataset and allows the SPPE regularisation term to adjust feature contributions in a gradual manner. Second, SHAP explanations are computed on the trained model outputs; a continuous output avoids threshold-driven discontinuities and makes global/local contribution trends more interpretable. However, for evaluation purposes, the focus was not on minor numerical variations but on the model’s ability to classify students into practical performance categories (e.g., fail, good, excellent). Therefore, to align our assessment with the practical decision-making needs of educators, we employed classification-oriented metrics such as accuracy, precision, recall, F1-score, and confusion matrix analysis to assess the practical effectiveness of the model in student performance prediction. This hybrid evaluation strategy is also consistent with established practices in the field using this benchmark dataset [38]. Distinguished from the research carried out by Özkurt [19], which only employed SHAP to determine the significance of features affecting student performance, this study not only provided the importance of input features but also identified the features whose importance contradicts the domain knowledge and optimized them based on the proposed SPPE algorithm. We have added a summary table to clarify the meanings and roles of SHAP and SPPE in the proposed framework, as shown in Table 2.

Table 2. Summary of key interpretability concepts used in this study.

Term	Definition	Role in this study
SHAP	A model-agnostic method that attributes a prediction to each input feature based on Shapley values, measuring its marginal contribution.	Used to quantify both global and local feature importance and to detect inconsistencies between data-driven feature importance and educational domain knowledge.
SPPE	The proposed Students' Performance Prediction Explanation algorithm that adjusts feature weight scores during training according to domain knowledge and SHAP feedback.	Integrated into the ANN loss function to increase the weight scores of domain-relevant features and decrease those of less relevant features, thereby aligning learned correlations with educational knowledge

		while preserving predictive performance.

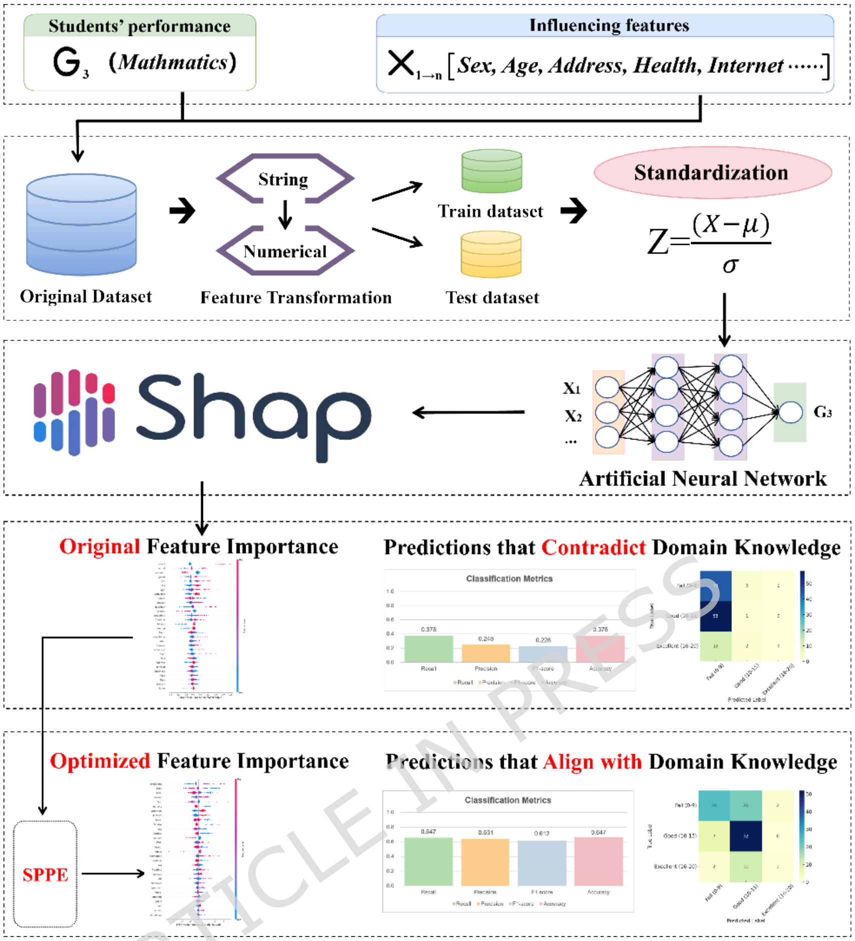

Students' performance
G_3 (Mathmatics)
Influencing features
$X_{1\to n}$ [Sex, Age, Address, Health, Internet ······]
Original Dataset
String
Numerical
Feature Transformation
Train dataset
Test dataset
Standardization
$Z=\frac{(X-\mu)}{\sigma}$
Shap
X1
X2
...
G3
Artificial Neural Network
Original Feature Importance
Predictions that Contradict Domain Knowledge
Classification Metrics
Optimized Feature Importance
Predictions that Align with Domain Knowledge
SPPE

Figure 1. The framework of the proposed interpretable neural network aligned with domain knowledge for student performance prediction.

3.2. Data Understanding and Pre-processing

This study selected mathematics performance as the target for student performance prediction, considering mathematics as a foundational discipline that not only cultivates logical reasoning and problem-solving skills but also serves as a critical determinant of students' readiness for higher education and STEM careers. The dataset used in this study contains information on secondary school students in Portugal and includes 30 features, as shown in Table 3, including school-related, social/emotional, and demographic factors hypothesized to influence student performance.

In Portugal, mathematics is a core subject integrated into multiple areas of the secondary education curriculum. Student achievement in this subject is assessed across three grading periods: G1, G2, and G3, with scores assigned on a 20-point scale. Among these, G3, the final grade, is considered the most representative measure of students' overall academic achievement and was used to evaluate the prediction accuracy of the proposed algorithm, consistent with the approach taken by Cortez et al. [[38]].

Table 3. The details of features in the dataset.

Category	Feature	Description
Demographic features	sex	student’s sex (binary: 'F' - female or 'M' - male)
	age	student’s age (numeric: from 15 to 22)
	address	student’s home address type (binary: 'U' - urban or 'R' - rural)
	Pstatus	parent’s cohabitation status (binary: 'T' - living together or 'A' - apart)
	Medu	mother’s education (categorical: from 0 to 4) 0 – None, 1 – Primary education (4th grade), 2 – 5th to 9th grade, 3 – Secondary education, 4 – Higher education
	Mjob	mother’s job (categorical: 'teacher', 'health' care related, civil 'services', 'at_home' or 'other')
	Fedu	father’s education (categorical: from 0 to 4) 0 – None, 1 – Primary education (4th grade), 2 – 5th to 9th grade, 3 – Secondary education, 4 – Higher education
	Fjob	Father’s job (categorical: 'teacher', 'health' care related, civil 'services', 'at_home' or 'other')
	guardian	student's guardian (categorical: 'mother', 'father' or 'other’)
	famsize	family size (binary: 'LE3' - less or equal to 3 or 'GT3' - greater than 3)
	health	current health status (numeric: from 1 - very bad to 5 - very good)
	internet	Internet access at home (binary: yes or no)
Social/Emotional features	famrel	quality of family relationships (ordinal: from 1 - very bad to 5 - excellent)

	romantic	with a romantic relationship (binary: yes or no)
	Dalc	workday alcohol consumption (ordinal: from 1 - very low to 5 - very high
	Walc	weekend alcohol consumption (ordinal: from 1 - very low to 5 - very high)
	freetime	free time after school (ordinal: from 1- very low to 5- very high)
	goout	going out with friends (ordinal: from 1- very low to 5- very high)
School-related features	school	student's school (binary: 'GP' - Gabriel Pereira or 'MS' - Mousinho da Silveira)
	failures	number of past class failures (ordinal: n if 1 ≤ n < 3, else 4)
	reason	reason to choose this school (categorical: close to 'home', school 'reputation', 'course' preference or 'other')
	traveltime	home to school travel time (ordinal: 1- < 15 mins., 2- 15 to 30 mins, 3- 30 mins to 1 hour or 4 - > 1hour)
	studytime	weekly study time (ordinal: 1- < 2 hours, 2- 2 to 5 hours, 3- 5 to 10 hours or 4 - > 10 hours)
	schoolsup	extra educational school support (binary: yes or no)
	famsup	family educational support (binary: yes or no)
	activities	extra-curricular activities (binary: yes or no)
	paid	extra paid classes within the course subject (Math) (binary: yes or no)
	nursery	attended nursery school (binary: yes or no)
	higher	wants to take higher education (binary: yes or no)
	absences	number of school absences (numeric: from 0 to 93)
Students’ grades	G1	first period grade (numeric: from 0 to 20)
	G2	second period grade (numeric: from 0 to 20)

G3	final period grade (numeric: from 0 to 20)

Before training the proposed model, data pre-processing is indispensable to provide cleaned data for subsequent model training and testing. The pre-processing procedures carried out in this study comprise three steps: 1) Transformation. Since string features such as Sex, Mjob, Reason cannot be handled by the model directly, the initial step of pre-processing stage is to convert these features into a numerical format. We distinguished between ordinal variables (e.g., Medu, Fedu, traveltime) which have an inherent order, and nominal variables (e.g., Mjob, Fjob, reason) which do not. In this study, label encoding was applied to transform each categorical value into an integer representation (e.g., 0, 1, 2, 3) to maintain computational efficiency. To ensure this encoding strategy did not introduce bias, a sensitivity analysis comparing it with one-hot encoding was conducted and is reported in Section 4.4.2. This approach was chosen because it efficiently converts categorical data without causing a substantial increase in feature dimensionality and allows for a more straightforward interpretation of feature importance in subsequent analysis. 2) Dataset partitioning. The original dataset is divided into training and testing dataset (70%/30% ratio). 3) Standardization. This study employed standardization to reduce the effect of magnitude between different features through mapping the value of data to a specific range (a mean of 0 and a standard deviation of 1). This step is explained by the equation (1) below:

$$Z = \frac{(X-\mu)}{\sigma} \quad (1)$$

where Z stands for the standardized value of data and X means the value of original data related to the student's feature. μ and σ represent the mean and the standard deviation of the value in the original dataset.

3.3. The Structure of the Developed Artificial Neural Network

This study experimentally constructed a 6-layer neural network containing 1,705 trainable parameters. As shown in Figure 2, the first layer, referred to as the input layer, contains 32 neurons, while the last layer, a linear output layer, consists of a single neuron without an activation function, used to predict G3. To capture refined information from high-level feature representations for accurate prediction, the number of neurons per layer gradually decreases with increasing network depth. Additionally, the Rectified Linear Unit (ReLU) was chosen as the activation function.

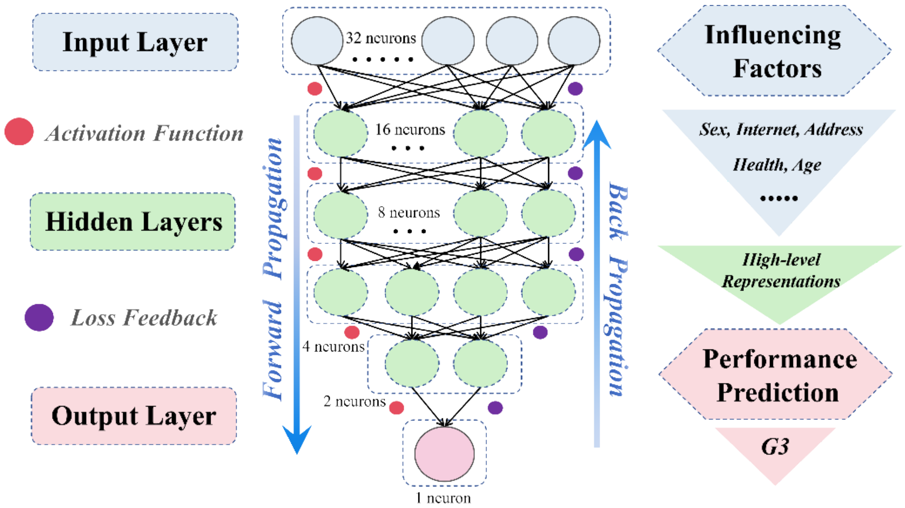


Figure 2. The structure of the developed artificial neural network.

3.4. SHAP-based Interpretability Optimization

The works of Özkurt have demonstrated that using SHAP values for predictions offers a feasible and effective method to elucidate feature contributions to students' performance [[19]]. However, the analysis of SHAP values often does not align with established educational domain knowledge. To further enhance both the interpretability and accuracy of ANN-based student performance prediction models, this study proposed a students' performance prediction explanation algorithm (Figure 3), inspired by the contextual decomposition algorithm, proposed in a previous study on the interpretability of AI models [[39]], to align the learned correlations in the ANN model with domain knowledge.

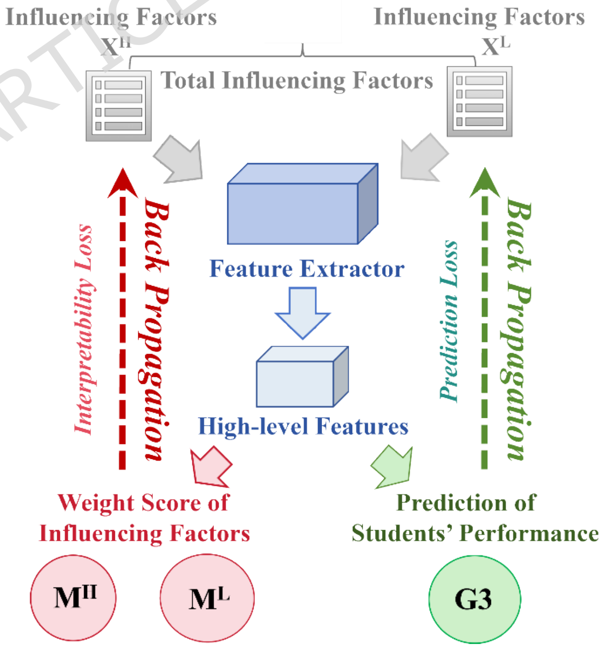


Figure 3. SHAP-based interpretability optimization process for the ANN model with the SPPE algorithm.

This study first made an extensive investigation on the related works of features affecting students' performance and identified the features that have a strong or weak correlation with student performance respectively. This process involved a systematic, multi-stage approach. Initially, a broad literature review identified factors with a strong consensus on their impact on academic achievement. This was followed by a more focused analysis of studies relevant to the dataset's specific educational context to account for cultural and systemic variability. Finally, we considered the distributional properties of the dataset itself to identify features prone to statistical bias.

Based on this synthesis, the features are classified into two categories, X^H and X^L, to align with the established educational domain knowledge. X^H represents the influencing features whose significance the model should be guided to increase during training, while the rest of the features whose importance should be reduced are regarded as X^L. A feature was included in X^H if the literature consistently showed a significant correlation with academic performance. A feature was included in X^L if it met one of several criteria: 1) literature suggested a negligible impact; 2) findings across studies were highly contradictory; or 3) dataset-specific issues, such as severe class imbalance, indicated a high risk of learning spurious correlations. We explicitly acknowledge that this categorization is context-dependent and would require re-evaluation for datasets from different socioeconomic or cultural settings.

Equation (2) describes the process of obtaining the prediction of students' performance (G3) in the forward propagation. To output the G3, the ANN-based model $g(x)$ with Z layers receives data x as input and employs a series of $v(x)$ called logits including linear operation and activation function in each layer to learn high-level representations.

$$g(x) = \text{Linear}_{\text{output}}(v(x)) = \text{Linear}_{\text{output}}(v_Z(v_{Z-1}(...(v_2(v_1(x)))))) \qquad (2)$$

In this equation, x denotes the input feature vector of a student. $v_j(\cdot)$ denotes the transformation at the j-th consisting of a linear operation (weights and biases) and an activation function. □ is the total number of layers in the ANN and $g(x)$ is the final predicted performance score (G3).

To quantify the weight score of features in the model training process, the Contextual Decomposition (CD) score [[39]] M^H and M^L are introduced to represent the weight score of X^H and X^L for the subsequent optimization. Through affecting the weight score of $X^H$and X^L, the corresponding SHAP value of features can be optimized.

Equation (3) elucidates the method of attaining the M^H and M^L in forward propagation. On the foundation of X^H and X^L, the logits $v(x)$ was disintegrated into corresponding weight scores M^H and M^L by the CD algorithm $v^{CD}(x)$. The CD algorithm employed a series of layer-specific disintegration, $\square_j^{CD}(x)$, to each corresponding layer $v_j(x)$ and the operation was conducted repeatedly across every layer until the logits were disintegrated completely.

$$M^{H}, M^{L} = v^{CD}(x) = v_{Z}^{CD}(v_{Z-1}^{CD}(...(v_{2}^{CD}(v_{1}^{CD}(x)))))$$

(3)

Here, M^{H} represents the aggregated weight score of features classified as high-importance (X^{H}), while M^{L} represents the aggregated weight score of low-importance features (X^{L}). $v_{j}^{CD}(\cdot)$ denotes the contextual decomposition operation applied at the j-th layer to separate the contributions of X^{H} and X^{L}.

After obtaining the predicted G3, M^{H} and M^{L}, they were optimized in the loss function at the back propagation. This study utilized Mean-Square Error (MSE) as the loss function to minimize the difference between the model's prediction value and actual value. As shown in equation (4), L is made up of two parts, one is the loss function of the students' performance prediction, and the other is the loss function of the weight scores M^{H} and M^{L}. n stands for the number of training samples while □(□) represents the prediction value of students' performance and y is the actual students' performance. The user-defined coefficient κ_1 and κ_2 were regarded as regularization terms to pursue balance in the loss function of optimizing G3 prediction and weight scores. κ_1 and κ_2 were experimentally assigned values of 1 and 0.1 to achieve a higher performance. During training, seeking the minimum value of the loss function for the weight scores, M^{L} was decreased and M^{H} was increased in a balanced proportion to optimize the importance of X^{H} and X^{L}.

$$L = \kappa_1 \frac{1}{n}\sum_j |g(x_j) - y_j| + \kappa_2 \frac{1}{n}\sum_j |M_j^{L} - M_j^{H}|$$

(4)

Conceptually, the SPPE algorithm can be viewed as a form of knowledge-infused regularization. However, it is distinct from conventional SHAP-informed regularization methods that typically enforce domain-agnostic properties such as sparsity or monotonicity. The novelty of SPPE lies in its specific function as a knowledge infusion mechanism. It uses SHAP values not merely as a diagnostic tool, but as an active and differentiable bridge to inject external, context-specific domain knowledge directly into the model's training process. While many knowledge-infused AI systems require architectural modifications or structured knowledge graphs, SPPE offers a flexible approach that regularizes a model's learned feature importance hierarchy against a “soft” target derived from empirical literature. This process compels the model to reconcile purely data-driven patterns with established theoretical knowledge, thereby enhancing its alignment with the domain it seeks to model.

It is important to note that both SHAP and the proposed SPPE algorithm operate at the level of associations between input features and model predictions rather than causal relationships. Throughout this study, terms such as “impact”, “effect”, or “influence” are used in a descriptive, correlational sense to indicate how changes in a feature are associated with changes in the model's output, and not to claim causal effects on students' actual achievement.

3.5. Implementation Details

This study utilized pytorch to construct the neural network. Grid-search method was employed to determine the optimal structure of the proposed ANN. In addition, the model has been trained using the final hyperparameter configuration selected by grid search (Sec.~3.6). Unless otherwise specified, the number of epochs has been set to 100, with batch size 32 and learning rate 0.001. Due to its ability to efficiently adjust learning rates and handle sparse gradients, Adam was selected as the optimizer to minimize the training loss and facilitate the parameter optimization process. To ensure the reproducibility of the experiments, all models were implemented using Python 3.8.10 and PyTorch 1.10.2. A fixed random seed (set to 42) was applied to all stochastic processes, including data splitting and weight initialization. Additionally, as the dataset contained no missing values, no imputation techniques were applied prior to the standardization described in Section 3.2. Moreover, a comprehensive validation framework that includes standard metrics composed of accuracy, precision, recall, F1-score, and confusion matrix analysis was applied to assess the performance of the original and optimized ANN.

3.6. Model Training and Hyperparameter Tuning

The ANN architecture was designed following a series of preliminary experiments to find a suitable balance between model complexity and performance for this specific tabular dataset. The final architecture consists of six fully connected layers with 128, 64, 32, 16, 8, and 1 neuron(s) respectively. This structure was found to be sufficiently deep to capture complex non-linear relationships without being overly susceptible to overfitting. The Rectified Linear Unit (ReLU) was used as the activation function for all hidden layers due to its effectiveness in preventing the vanishing gradient problem. The model was trained using the Adam optimizer, chosen for its adaptive learning rate capabilities and robust performance, with Mean-Square Error (MSE) serving as the loss function, consistent with our regression-based training approach.

To determine the optimal hyperparameters, we employed a systematic grid search methodology. The search space for the learning rate was [0.001, 0.005, 0.01], for the batch size was [16, 32, 64], and for the epochs was [50, 100, 150]. For the key parameters of our proposed SPPE algorithm, the search space for both κ1 (for X^H) and κ2 (for X^L) was [0.1, 0.5, 1.0, 2.0]. The final selected hyperparameters, which yielded the best performance on the validation sets, were a learning rate of 0.001, a batch size of 32, 100 epochs, and κ1 = 1.0, κ2 = 1.0. All results in Sec.~4 have been reported under this selected configuration.

4. Experimental Results

4.1. Analysis of the Interpretability Measures

Both global and local interpretability were utilized to explain the rationality of the optimized ANN model. The global interpretation concentrates on the overall influence of 30 features on the model, while the local interpretation was designed to analyze the contribution of these features for individual samples, revealing how features influence specific prediction results of the model.

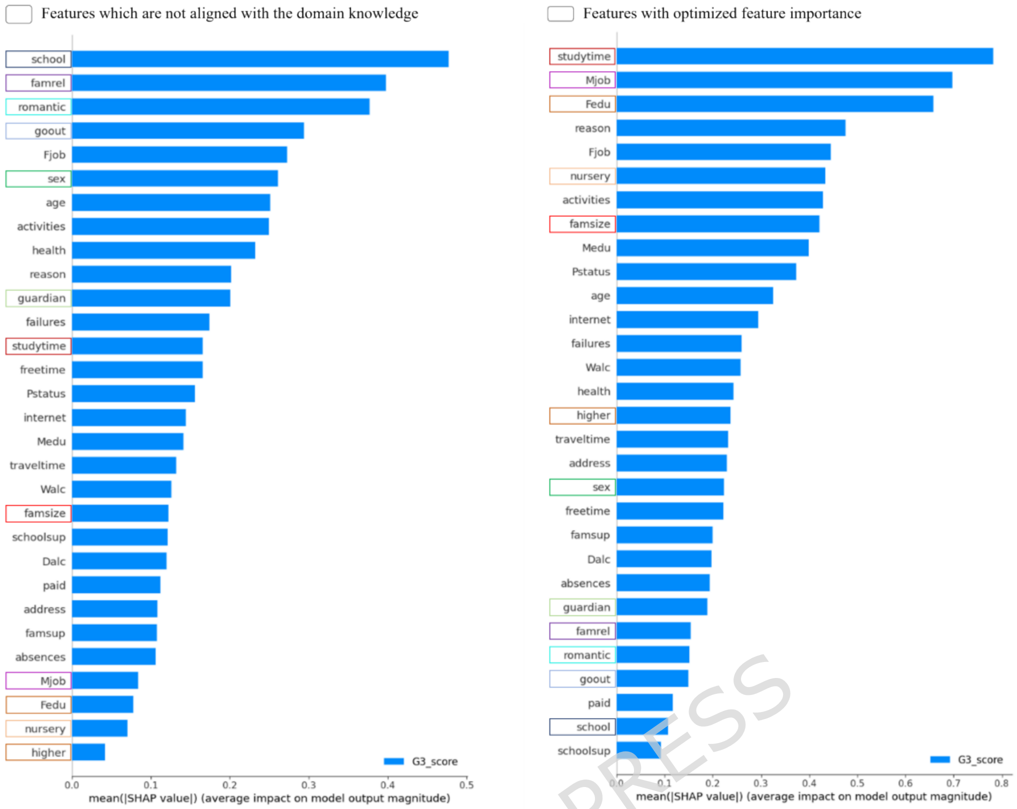
For the purposes of evaluation, the continuous grade predictions generated by the regression model were converted into three discrete performance categories: ``Fail'' (grades 0-9), ``Good'' (grades 10-15), and 'Excellent' (grades 16-20). These thresholds are not arbitrary; they are directly based on the official grading standards of the Portuguese education system, from which the dataset was collected. This ensures that our evaluation aligns with the practical and pedagogical context of the data. The performance of the models was assessed using the independent test set (30% of the original dataset) to ensure the robustness of the results on unseen data. The metrics used for evaluation include accuracy, precision, recall, and F1-score, which were chosen to provide a comprehensive assessment of the model's practical utility in correctly identifying students across these meaningful performance categories.

4.1.1. Global Explanation of the ANN

The global explanation provides an overview of the relative significance of all features in the ANN's prediction process. The feature importance plot (Figure 4) displays all features and ranks their contributions to the original and optimized ANN models in descending order. By comparing the two plots in Figure 4 from the perspective of educational domain knowledge, some representative features with the most significant differences were identified and highlighted.

According to educational domain knowledge, some features such as School, Famrel, Romantic, Goout, Sex, and Guardian were assigned unusually high importance in the original ANN model shown in Figure 4 (a). Meanwhile, the importance of other relevant features including Higher, Nursery, Fedu, Mjob, Famsize and Studytime was partially overlooked to some extent. To address these misalignments, the importance of these features was systematically adjusted during model training through the proposed SPPE method to enable them to play a more appropriate and reasonable role in prediction shown in Figure 4 (b).

The designed mechanism could optimize the weight scores of identified features by affecting the shift tendency of corresponding SHAP values. The high compatibility between SHAP and SPPE primarily stems from their mutual reliance on feature contribution decomposition principles, with both enabling quantification of nonlinear feature interactions and providing measurable indicators for feature contributions. SPPE's optimization of identified feature weight scores is based on re-calibrating neural network internal representations during the training phase, thereby systematically influencing the prioritization of feature importance in the model. Given that SHAP values derive from the model’s final learned parameters, any weight adjustments made by SPPE inherently propagate to subsequent SHAP computations during the model training phase. Consequently, alterations in weight scores of the identified features directly impact corresponding SHAP values through modifications in the feature marginal contributions to predictive results.



(a) Feature importance of the original ANN

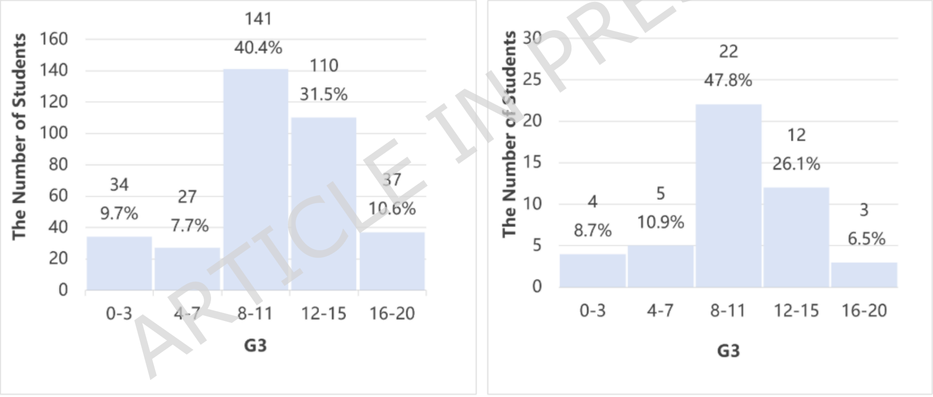
(b) Feature importance of the optimized ANN

Figure 4. The comparison of feature importance plot based on the original and optimized ANN.

A global explanation is firstly provided for features that require decreased importance. A detailed discussion of the identified features and the underlying reasons for their reduced contribution is presented below.

Sex. A common stereotype deems that male students outperform their female counterparts, particularly in science, mathematics and engineering. However, this assumption is unreliable [[40]]. Several scholars have demonstrated that gender differences in students' academic achievement are minimal. A meta-analysis of 1, 286, 350 individuals across 242 studies published between 1990 and 2007 [[41]], indicated male and female students shared a similar academic performance (effect size d = 0.05 and variance ratio = 1.08). To verify the widespread applicability of these findings, analysis of large datasets based on probability sampling of U.S. adolescents spanning 20 years had also been conducted. The observed effect sizes (-0.15 to 0.22) and variance ratios (0.88 to 1.32) further confirmed that gender is not a decisive feature in predicting academic performance. Else et al. [[40]] reinforced this conclusion through a meta-analysis of two prominent international datasets. Their comprehensive study of 493,495 students aged 14-16 from 69 countries reported a mean mathematics achievement effect size of d < 0.15, indicating no significant gender-based performance disparity.

School. Although the feature (School) is often considered important for predicting students' academic achievement, the original model's assignment of top importance to this feature may be misleading. This could be attributed to the imbalanced distribution of students across the two schools represented in the dataset. The feature (School) consisted of two categories: Gabriel Pereira (GP) with 349 students and Mousinho da Silveira (MS) with only 46 students. Such a large disparity likely caused the model to overfit information from GP while underrepresenting MS, which increased the model’s bias. Rather than using data re-sampling techniques, which might artificially inflate the feature's importance against the guidance of domain knowledge, our proposed SPPE algorithm directly addresses this by reducing the weight of the “school” feature during training. This choice is further supported by studies suggesting that the feature (School) has little direct effect on students' academic performance, despite the common assumption that it plays a significant role [42, 43]. Furthermore, the uniformity of local education quality further reduces the influence of the feature (School) in forecasting students’ performance. The government of Portugal has implemented a series of measures to minimize disparities in education quality across the country, such as the introduction of the Programme for International Student Assessment (PISA) and the standardization of the teacher certification system [44]. Figure 5 visualizes the distribution of student performance between the two schools and shows that there is little observable difference in mathematics achievement. Therefore, it is reasonable for the optimized model to appropriately reduce the importance assigned to this feature.



(a) GP distribution in G3

(b) MS distribution in G3

Figure 5. Distribution of students' G3 scores by school.

Romantic. Students’ romantic relationships have not been identified as major factors influencing academic performance [[45], [46]]. For instance, Pietrulewicz [[47]] conducted a study involving 901 adolescent students who had been in at least one romantic relationship, and found that romantic involvement had limited predictive value for students’ academic performance. Awe et al. [[46]] confirmed this finding through an empirical study of 200 students (124 male and 76 female) from eight departments at the Federal University Oye-Ekiti, Nigeria. They applied chi-square tests, ANOVA, and t-tests for independent samples. The results indicated no significant impact of romantic relationships on academic performance (F = 0.721, df = 199, p > 0.05).

Goout. The original ANN model may have misinterpreted the relationship between going out and students’ academic performance, leading to an overestimation of the importance of this feature. The model may have assumed that going out reduces study time and negatively affects academic outcomes. However, several studies have indicated that the effect of going out depends heavily on its purpose, which is not specified in the original dataset [[48], [49]]. For instance, Li et al. [[48]] pointed out that going out for recreational activities unrelated to academic goals does not contribute to improved performance, but positive and organized activities such as sports or academic competition are beneficial to promote students' academic performance [[49]].

Guardian. The confusion between guardians and parents may have misled the original model to assign excessive importance to the feature (Guardian) in predicting students' academic performance. This role does not necessarily have a direct impact on students' academic performance. Cahn [[50]] notes that while guardians are authorized to make legal decisions such as those related to medical care or property management, they are often excluded from making educational decisions. Recent research further suggested that guardians show stronger correlations with academic achievement only when they are the students' parents [[51]].

Famrel (Family relationships). Family relationships generally consist of the relationships between students and their relatives, including cousins, grandparents and parents. But some researches emphasize that these relationships may have different impacts on students' performance [[52], [53], [54], [55]]. Specifically, Ginther & Pollak [[52]] and Tanskanen & Danielsbacka [[53]] suggest insignificant correlations between extended family members (cousins and grandparents) and academic outcomes, whereas Peng et al. [[54]] and Rathee & Kumari [[55]] identify parental relationships as pivotal predictors. Lacking further identification of the family relationship in the original dataset may confuse the original ANN to assign the misappropriate importance to this feature. Furthermore, Famrel was assessed using five subjective rating levels, heavily reliant on participants' self-reported perceptions. But the subjective judgment may not accurately present participants' true feelings, potentially leading to biased statistical data.

A global explanation is then provided for features that require increased importance. A detailed discussion of the identified features and the underlying reasons for their increased contribution is presented below.

Study time. The importance of study time should be emphasized in predicting students' performance. This aligns with research findings, such as Liu [[56]], which reported a strong correlation between study time and academic performance. Cassidy [[57]] deems that plenty of study time could enable the students to expand their knowledge storage and exercising their self-discipline competence to ensure their impressive academic performance. By conducting a regression analysis on a dataset from California elementary schools, Jez et al. [[58]] quantified this correlation, showing that each 15-minute increase in daily study time can lead to an approximate 1% improvement in academic achievement. Especially, students with poor financial conditions experienced a more significant augmentation of up to 1.5%.

Famsize (Family size). Family size demonstrates significant predictive power for academic performance, which warrants further emphasis on its importance. Marks [[59]] discovered that family size is a major contributor to affect students' reading and mathematics performance through analyzing data from 30 countries. This study has inspired further research to explore the importance of this feature in students' performance. AlSaleh et al. [[60]] made an extensive research on the students in Arabic area and revealed that family size can account for the considerable parts of the effects on students' academic achievement. A Chinese study [[61]] specified these effects that students from a family size greater than 3 are approximately 17% less likely to complete secondary education than the peers from a family size less or equal to 3. The larger family means educational resources must be distributed among more children, limiting access to individual academic support, which in turn hinders students' educational achievement [[21]].

Mjob (Mother's job). Mother's job has been underestimated in its importance as a predictor of academic performance. Several studies have demonstrated that a mother's occupation serves as a robust academic predictor. Azizah et al. [[62]] employed Ordinary Least Squares (OLS) regression to investigate participants from 24 Indonesian provinces, revealing that a mother's occupation has both short-term and long-term impacts on students' academic performance. Wagner [[63]] found that mothers in prestigious occupations are more likely to raise children with exceptional academic achievement, based on data from 7,716 mother-daughter pairs in the Educational Longitudinal Study.

Fedu (father's education). Various studies employing diverse methodologies demonstrate a significant correlation between father's education and students' academic achievement. For instance, Ossai et al. [[64]] applied two-way ANOVA to 3,214 students' exam scores guided by the theory of planned behavior, revealing that fathers' education significantly impacts students' academic outcomes, particularly for those with master's degrees. Wamala et al. [[65]] employed a multiple linear regression and summary statistics to survey 5,148 records of sixth grade students enrolled in Ugandan primary schools. This study concluded that fathers' education is a pivotal determinant for students' academic achievement and children born to fathers with primary, secondary, and post-secondary education performed better in all disciplines than those with non-educated fathers. It may be attributed to that father, especially those with inferior education background, know that excellent academic performance could bring superior educational background and vocational career to their children [[66]]. As a result, they tend to offer emotional support, set academic goals, and foster a study-oriented home environment, all of which contribute significantly to students' academic success.

Nursery. Attending nursery school (i.e., preschool experience) can have a lasting and significant influence on students' academic achievement. Ulferts et al. [[67]], based on data from 17 longitudinal studies across 9 European countries involving 16,461 students aged 3 to 15, discovered that sustained interdependence exists between students' educational outcomes and their nursery school experiences. This long-term effect extends across different academic stages. This finding aligns with a comprehensive study by Bagudo [[68]], who led an empirical investigation of 600 students' academic records (300 boys and 300 girls) from Sokoto State primary schools. Furthermore, Bustamante et al. [[69]] confirmed the impact is also effective to the students from different races. Based on Piaget's cognitive development theory and Brunner's discovery learning theory, Tande [[70]] conducted a comparative study involving 159 level one pupils in 4 primary schools, to reveal that significant disparities stay in the academic achievement of students who received nursery education and those did not.

Higher (Higher education expectation). Several studies confirmed that students’ expectations of higher education can be a crucial predictor of their academic performance, deserving higher importance. Trinidad [[71]] conducted longitudinal research on 15,244 tenth-grade US students and discovered that improving students’ education expectation could reinforce their academic achievement in college entering examinations. Similar findings have been observed in developing countries. Latikal [[72]] conducted a quantitative study on 142 seventh-grade students in Indonesia, applying multiple linear regression analysis to reveal that students' higher education expectations play an important role in shaping their academic success. This effect can be largely attributed to self-efficacy and study motivation. Self-efficacy means the students’ confidence in completing tasks in study. It is the source to maintain students’ active participation in the study so as to fight against the challenge of the study persistently. In addition, students with identified study motivation could make the most of their study time and resources to attain ideal academic results, which consolidates their commitment to higher education in the future.

4.1.2. Local Explanation of the ANN

Local interpretation focuses on assessing the contribution of features to individual samples, revealing how features influence specific prediction outcomes of the model. By calculating SHAP values and visualizing the summary plot of SHAP values, this method offers a detailed account of each feature’s directional impact (positive/negative) on student performance. In detail, the summary plot ranks features vertically based on their accumulated average absolute SHAP values, with color encoding the magnitude of feature values. For instance, red denotes higher feature values, whereas blue represents lower ones. When the SHAP values of a feature cluster densely in regions distant from zero, this indicates the feature exerts substantial influence on the model’s predictions.
A local explanation is firstly discussed for features that require decreased importance, as presented below.

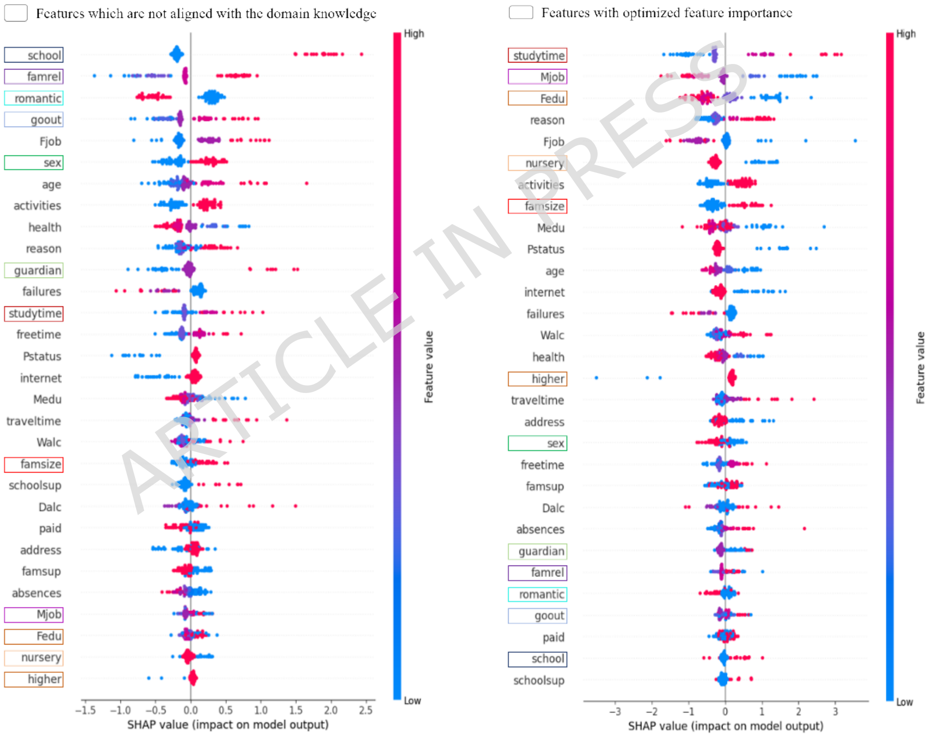
Categorical features. Categorical features (School, Sex, Guardian and Romantic) experience a similar optimization in Figure 6. Taking feature Romantic as an example, Figure 6 (a) depicts that the original ANN assigns an entirely negative contribution (red dots) to students’ performance in the study while attributing a completely positive impact to those not in a relationship (blue dots). This situation may not align with educational reality and overlooks the special cases. While the common belief is that romantic relationships negatively affect academic performance, there are situations where they can actually be beneficial. In contrast, the optimized ANN shown in Figure 6 (b) corrects the original ANN's rigid and absolute interpretations which failed to account for the variability in how romantic relationship impacts individual students. Figure 6 (b) exhibits that the distribution of red dots (romantic relationship) and blue dots (no romantic relationship) overlap and are located in the area of negative and positive contribution. The narrower distribution range of these 2 variables indicates the importance of the feature Romantic is decreased, aligning more closely with domain knowledge. The clustering of blue dots (no romantic relationship) in the positive contribution area suggests that not being involved in a romantic relationship could be beneficial for students' academic performance, and vice versa. Some studies further revealed the correlation between romantic relationships and students' performance. Arbinaga et al. [[73]] emphasized that emotional dependency within a romantic relationship can lead to a lack of confidence and motivation when students face academic challenges. In addition, the emotional fluctuations and stress associated with romantic relationships may hinder students' learning efficiency. Harahap et al. [[74]] discovered being in a romantic relationship may narrow students' social circles and reduce interactions with peers, weakening their social support networks and, in turn, potentially impairing their academic performance.

Ordinal features. Ordinal features (Goout and Famrel) contain five ordinal values, ranging from very low (1) to very high (5). Taking Goout as an example, the original ANN in Figure 6 (a) assigns unreasonably high importance to this feature, with different feature values dispersing independently into negative or positive contributions. Specifically, blue and bluish-violet dots (representing low values of Goout) cluster in the negative contribution area, while purple and red dots (representing high values of Goout) cluster in the positive contribution area. However, this prediction is overly rigid and contradicts domain expectations. The optimized ANN in Figure 6 (b) reduced the importance of Goout to an appropriate level and overlapped different feature values with both negative and positive contribution areas. For instance, pink and red dots (representing high Goout values) predominantly align with negative impacts on academic performance. Gupta & Gueneau [[75]] validated that students' frequent social outings compromise their time management and attention, while fostering detrimental habits such as irregular study and rest patterns. These behaviors undermine their physical and mental health, indirectly impairing academic performance.

A local explanation is then discussed for features that require increased importance, as presented below.

Categorical features. As a representative of categorical features (Famsize, Mjob, Higher, and Nursery), the Nursery feature is selected to illustrate the local explanation of features with increased importance. Figure 6 (a) shows that both red and blue dots align with the zero SHAP value line in the original ANN, implying no impact of students' nursery school attendance on academic performance. However, this contradicts established educational domain knowledge, which suggests a correlation between nursery school experience and academic performance [[76], [77]]. The optimized ANN, as shown in Figure 6 (b), uncovers this relationship, revealing that students who attended nursery school tend to underperform academically, whereas those without such experience achieve better results. A widely held opinion suggests that parents may be key to explaining this phenomenon, which goes against common perceptions. Due to heavy workloads, parents often have no choice but to entrust nursery schools with both the education and care of their children. However, students raised in environments lacking parental involvement are more likely to face behavioral and emotional challenges, which can hinder their academic progress [[78]].

Ordinal features. In the original ANN Figure 6 (a), Studytime ranks 13th in feature importance. This plot provides more clues to confirm that the importance of this feature should be enhanced. i.e. part of blue dots (representing low values of Studytime) cluster in the positive contribution area implies that deficient study time could be beneficial to students' academic performance. This is inconsistent with established educational research. After optimizing, the ANN shifted the feature contribution to align with these expectations. Figure 6 (b) shows that feature Studytime has a remarkable influence on students' mathematics grade (the importance of this feature is increased to the 1 place by the optimized ANN). This prediction is also supported by existing educational domain knowledge. Baliyan & Khama [[79]] established a strong correlation between study time and mathematics performance through a quantitative analysis of 168 Botswana secondary students, identifying insufficient study time as a key detriment. Spitzer's [[80]] cross-national study of 6,000 students further reinforced this finding, demonstrating that increased study time significantly boosts mathematics scores, particularly among low-performing students.



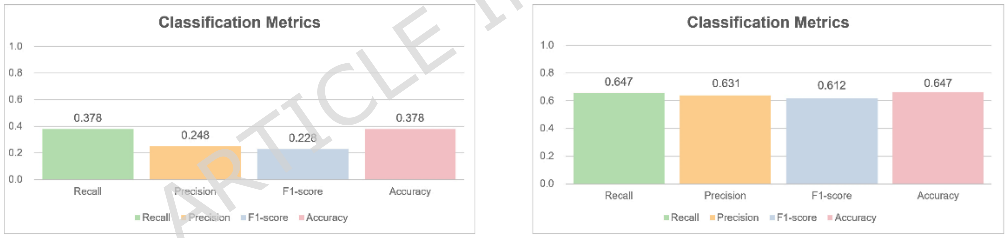
(a) Summary plot of SHAP values based on the original ANN

(b) Summary plot of SHAP values based on the optimized ANN

Figure 6. Comparison of SHAP values' summary plot between the original and optimized ANN.

4.2. The Prediction Performance Comparison of the Original and Optimized ANN

To investigate whether aligning the learned correlations of the ANN with educational domain knowledge improves prediction performance, the optimized ANN model was compared to the original ANN model across different metrics. Considering that these models were originally trained in a regression manner, we first reported RMSE and MAE to evaluate their performance. The original ANN achieved an RMSE of 10.44 and an MAE of 9.49, while the optimized ANN reduced them to 4.27 and 3.35, respectively. In addition, we used four key metrics including accuracy, precision, recall, and F1-score to assess the results after converting the task into a classification problem, as this approach better fits the educational setting. It is worth noting that in the results, recall, precision, and F1-score are calculated using the weighted average method. In this case, the recall value equals accuracy because weighted recall reflects the overall proportion of correctly predicted samples. The results, shown in Figure 7, demonstrated that the optimized model outperforms the original model on all metrics. Specifically, accuracy increased from 37.8% to 64.7%, reflecting a 26.9% improvement in the model’s ability to correctly predict student performance. Similarly, precision rose from 24.8% to 63.1%, indicating more accurate identification of true positives. The recall metric improved from 37.8% to 64.7%, suggesting that the optimized model is better at identifying students at risk of underperformance. Finally, the F1-score, which balances precision and recall, improved from 22.8% to 61.2%, further validating the overall robustness of the optimized model. These findings indicate that incorporating educational domain knowledge into the ANN model through the proposed SPPE algorithm is beneficial for improving student performance predictions. Moreover, McNemar’s test confirmed that the improvement is statistically significant (p = 0.00072, p < 0.001). This statistical test was computed on the paired predictions of the held-out test set (N=119). The analysis of discordant pairs revealed that the optimized model correctly classified a significant number of student cases that the original model failed to identify, further supporting the reliability of the performance gain.

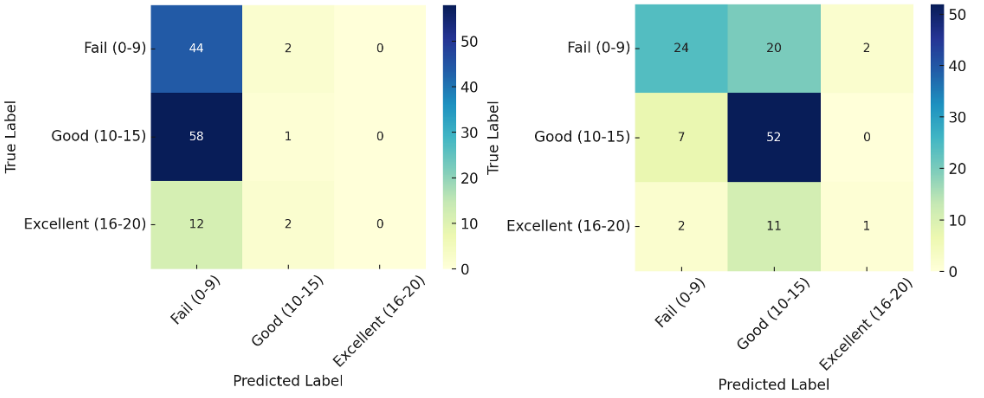


(a) Classification metrics based on the original ANN model

(b) Classification metrics based on the optimized ANN model

Figure 7. Performance comparison between the original and optimized ANN models.

In addition to the numerical metrics, the confusion matrices of the original and optimized ANN models were analyzed to gain deeper insights into their performance. Students' academic performance was classified into three categories: fail (0–9), good (10–15), and excellent (16–20) for analysis. As shown in Figure 8 (a), the diagonal line from the top left to the bottom right represents the number of correctly classified samples. For the original ANN, 44 samples were correctly classified as fail, while only 1 sample was correctly classified as good, and no samples were correctly classified as excellent. In contrast, for the optimized ANN in Figure 8 (b), 24 samples were correctly classified as fail, 52 samples as good, and 1 sample as excellent. The original and optimized models misclassified 74 instances and 43 instances, respectively. The discrepancy suggests that the original ANN model struggles to differentiate between students in these three categories, which could lead to underestimating their academic performance and misclassifications between these groups. In contrast, the optimized ANN model is better at distinguishing between students who are likely to perform well and those who are not.



(a) Confusion matrix based on the original ANN

(b) Confusion matrix based on the optimized ANN

Figure 8. Confusion matrices of original and optimized ANN models for students' academic performance classification (Fail: 0–9; Good: 10–15; Excellent: 16–20)

In addition to comparing the original and optimized ANN models, this study further evaluated the performance of the optimized ANN against several representative traditional machine learning algorithms that are widely regarded as strong baselines for tabular data, including Random Forest (RF), Support Vector Machine (SVM), k-Nearest Neighbors (KNN), Decision Tree (DT), XGBoost, and LightGBM. As shown in Table 4, all traditional models achieved higher accuracy than the original ANN model, with RF obtaining the best overall accuracy among them. This finding is consistent with previous research, such as Shwartz-Ziv's study [85], which demonstrated that classical machine learning methods like GBDT outperform neural networks on small to medium-sized tabular datasets. One plausible explanation is that these models have fewer parameters and are easier to optimize, whereas neural networks with more parameters typically require larger datasets or specialized architectures (e.g., TabNet) to achieve comparable performance. However, once educational domain knowledge was integrated, the optimized ANN outperformed all the traditional baselines across all metrics shown in Table 4. This result suggests that embedding domain knowledge enables the ANN to capture more reasonable and effective variable relationships, which can provide a viable alternative pathway to improving neural network performance beyond architectural optimization.

Table 4. The performance comparison between optimized ANN and other traditional machine learning algorithms.

Model Name	**Accuracy**	**Precision**	**Recall**	**F1-score**
RF	0.563	0.679	0.563	0.485
SVM	0.504	0.479	0.504	0.372
KNN	0.554	0.599	0.554	0.527
DT	0.436	0.455	0.436	0.443
XGBoost	0.563	0.576	0.563	0.542
LightGBM	0.554	0.535	0.554	0.479

Optimized ANN	**0.647**	**0.631**	**0.647**	**0.612**

4.3. Investigating the Influence of Loss Weight κ_2 on Model Performance

The loss weight parameter κ_2 plays a crucial role in guiding the model to balance classification performance and interpretability in the context of student performance prediction. To explore its impact, we varied κ_2 and evaluated its effect on prediction metrics (accuracy and F1-score) and the weight score M^H of identified important features required increased importance.

As illustrated in Figure 9, the weight score M^H of identified important features steadily rises with increasing κ_2, suggesting that the model increasingly prioritizes these features during training to reflect their domain-aligned importance. Meanwhile, both accuracy and F1-score improve as κ_2 increases from 0.001 to 0.1, reaching their peaks at κ_2=0.1. This demonstrates that introducing a moderate emphasis on domain-aligned feature importance helps guide the ANN model toward more meaningful and generalizable patterns in student performance prediction.

However, when κ_2 becomes too large (κ_2=1), a slight decline in accuracy and F1-score is observed. This implies that excessive regularization causes the model to over-prioritize interpretability at the cost of classification effectiveness, potentially suppressing important data-driven patterns that are not explicitly captured by prior knowledge. Conversely, when κ_2 is too small (i.e. κ_2=0.001), the model tends to ignore domain constraints, focusing solely on data correlations and potentially reinforcing spurious or biased relationships. Therefore, setting κ_2=0.1 offers an optimal balance between prediction accuracy and interpretability in this study for student performance prediction.

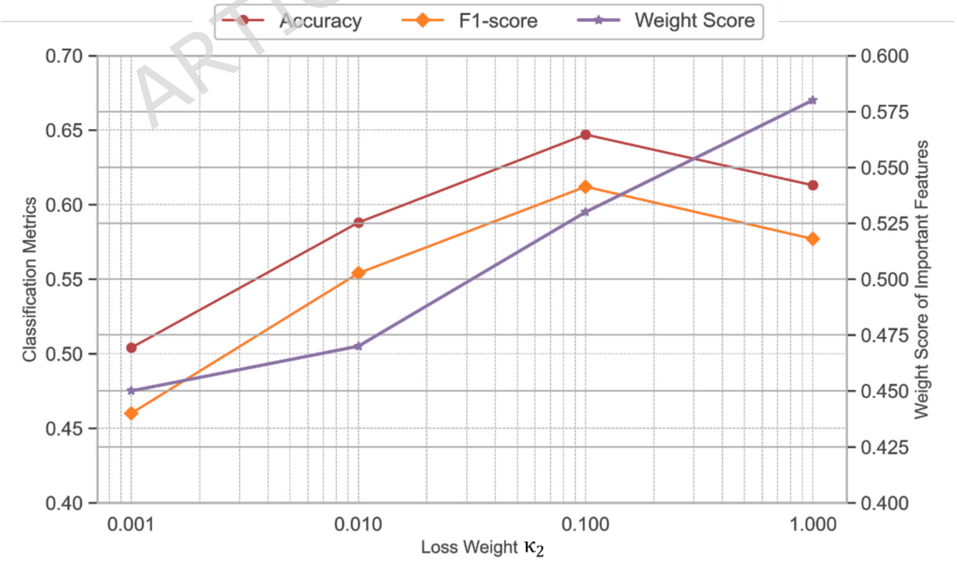


Figure 9. The influence of loss weight κ_2 on model performance.

4.4. The Applicability Analysis of the SPPE

4.4.1. Applicability of SPPE across different ANN architectures

To further evaluate the generalizability of the proposed SPPE algorithm, we extended our experiments by applying SPPE to several neural network architectures previously proposed in other studies, all trained on the same Portuguese student dataset. In these replications, the set of features identified for importance enhancement and reduction was kept consistent with that defined in this study, for the sake of consistency, comparability, and to validate the rationality of the domain-informed feature selection strategy. Specifically, four representative ANN models proposed by Aslam et al. [86], Huang et al. [87], Zhao et al. [88], and Yang et al. [90] were re-implemented and evaluated before and after applying our optimization strategy. As shown in Table 5, the weight score M^H of the identified important features increased across all models, indicating that SPPE effectively guided the networks to prioritize domain-aligned information. In addition, performance metrics including accuracy, precision, recall, and F1-score improved significantly after applying SPPE. For example, in the model from Aslam et al. [86], accuracy improved from 38.7% to 50.4%, and F1-score increased from 21.6% to 44.5%. Similar improvements were observed in the models by Huang et al. [87], Zhao et al. [88], and Yang et al.[90]. These results can effectively demonstrate the broad applicability and effectiveness of the proposed SPPE strategy.

Table 5. Performance comparison across different ANN model structures before and after optimization using SPPE

	Original ANN					**Optimized ANN aligned with domain knowledge**				
	Accuracy	**Precision**	**Recall**	**F1-score**	M^H	**Accuracy**	**Precision**	**Recall**	**F1-score**	M^H
Aslam et al.[86]	0.387	0.149	0.387	0.216	0.122	0.504	0.539	0.504	0.445	0.497
Huang et al. [87]	0.378	0.394	0.378	0.353	0.419	0.546	0.534	0.546	0.519	0.475
Zhao et al. [88]	0.429	0.453	0.429	0.435	0.456	0.563	0.497	0.563	0.528	0.501
Yang et al. [90]	0.395	0.480	0.395	0.247	0.407	0.538	0.570	0.538	0.514	0.504

However, it should be noted that the criteria used in this study to determine discrepancies between SHAP-derived feature importance and educational domain knowledge were mainly informed by findings of prior literature, the distributional characteristics of the current dataset, and the local educational context in which the data were collected. The experimental results demonstrate that the proposed strategy effectively guides ANNs to learn feature relationships that are more consistent with established educational knowledge. However, when applying the proposed approach to other datasets or educational settings, the importance of variables identified for improvement or reduction may differ, and it is suggested to take into account the specific local context, cultural factors, and data characteristics to ensure the validity of the proposed SSPE algorithm.

4.4.2. Sensitivity analysis of feature encoding schemes

To examine the influence of feature encoding on both SHAP values and the SPPE mechanism, we have compared the optimized ANN under two encoding schemes for all categorical variables: label encoding (our main setting) and one-hot encoding. Label encoding has been initially adopted to avoid an excessive increase in input dimensionality and the corresponding risk of overfitting on this relatively small dataset. In a supplementary experiment, we re-trained the same ANN architecture with SPPE using one-hot encoded features while keeping all other training configurations unchanged. As summarized in Table 6, the optimized ANN with one-hot encoding achieved accuracy, precision, recall, and F1-score of 0.642, 0.626, 0.642, and 0.605, respectively, which are very close to the results obtained with label encoding (0.647, 0.631, 0.647, and 0.612). In addition, the global SHAP-based feature importance ranking and the qualitative local explanations remained highly similar across the two settings, and the set of features adjusted by SPPE did not change. These findings suggest that the arbitrary ordering induced by label encoding has not materially affected either the predictive performance or the interpretability behavior of the proposed framework.

Table 6. Performance of the optimized ANN under different feature encoding schemes.

Encoding scheme	Accuracy	Precision	Recall	F1-score
Label encoding (proposed)	**0.647**	**0.631**	**0.647**	**0.612**
One-hot encoding	0.642	0.626	0.642	0.605

5. Discussion

This study proposed an interpretable neural network framework that integrates educational domain knowledge to improve both the explainability and predictive performance of student achievement models. The experimental results demonstrate that while traditional ANN models can achieve strong predictive accuracy, their interpretability often conflicts with established educational theories. By embedding domain-informed constraints through the SPPE algorithm, the optimized ANN achieved significant improvements in RMSE, MAE, and classification metrics, while aligning feature importance rankings with educational expectations. These findings indicate that in educational prediction tasks, model reliability and rationality depend not merely on data volume or architectural sophistication but on whether the learning process incorporates human knowledge to bridge data correlations with pedagogical causality.

Mechanistically, the SPPE algorithm introduces SHAP-informed weight regulation into the loss function, guiding the model to adjust feature importance dynamically during backpropagation. This design establishes a balance between data-driven learning and knowledge-driven regularization, preventing the model from overfitting spurious patterns. As a result, the framework achieves dual gains in interpretability and performance, providing a more trustworthy AI paradigm for educational analytics and policy decision-making. The robustness of this methodological approach is further supported by the findings in Section 4.4, where the SPPE strategy demonstrated consistent performance improvements across several different ANN architectures.

Compared with prior works that mainly used XAI tools for post-hoc interpretation, this study shifts from “explanation after learning” to “explanation-guided learning.” Instead of interpreting model behavior retrospectively, the model learns under interpretability constraints, achieving a unified optimization of accuracy and transparency. This approach is generalizable to other educational data mining tasks, such as learning behavior modeling and intervention evaluation, where both predictive precision and interpretability are crucial.

These findings indicate that in educational prediction tasks, model reliability and rationality depend not merely on data volume or architectural sophistication, but on whether the learning process incorporates human knowledge to bridge data correlations with patterns that are consistent with pedagogical theories about factors associated with student achievement, rather than with strong causal claims.

Nonetheless, several limitations should be acknowledged. The dataset employed in this study is relatively small and context-specific, which may limit the generalizability of findings. Future research should validate the proposed framework using larger and more diverse educational datasets. Moreover, the quantitative encoding of domain knowledge relies mainly on literature synthesis and statistical reasoning, necessitating further refinement through expert input. We have clarified our systematic process for this, but acknowledge its inherent context-dependency and the need for careful adaptation when applying the framework to new settings. Lastly, the current SPPE framework focuses on static features, and future extensions could incorporate temporal or behavioral variables to capture dynamic aspects of student learning [89].

In conclusion, this study confirms the feasibility and effectiveness of embedding educational domain knowledge into neural networks for student performance prediction. The SPPE algorithm enhances both interpretability and predictive validity, providing a promising pathway toward the deeper integration of explainable AI and educational intelligence. Extending these conclusions to different educational systems, cultures, or datasets requires additional empirical validation, which we have left as future work.

6. Conclusion

This study proposed an interpretable artificial neural network (ANN) model that integrates educational domain knowledge to predict students’ mathematics performance. By combining the SHAP algorithm for feature importance analysis with the novel Students’ Performance Prediction Explanation (SPPE) mechanism, the model dynamically rectifies feature importance inconsistencies during training. To the best of our knowledge, this study is among the first attempts in student performance prediction to systematically incorporate educational domain knowledge into an ANN training process to improve interpretability and performance. Experiments on the UCI public student dataset demonstrated that the proposed approach effectively enhanced both interpretability and predictive performance, yielding feature importance distributions more aligned with educational theory and achieving significantly higher prediction accuracy and reliability.

Unlike previous studies that focused primarily on performance optimization or post-hoc explainability, this work emphasizes the paradigm of knowledge-constrained interpretable learning, where domain expertise is embedded into the model training process. This approach enables the neural network to move beyond purely data-driven correlations toward educationally meaningful association patterns that are aligned with existing theories, without asserting causal relationships.

In conclusion, the findings confirm the feasibility and effectiveness of integrating domain knowledge into deep learning models for educational prediction tasks, contributing to the advancement of explainable AI in education. Future research may extend this framework to larger and cross-cultural datasets and explore enhanced architectures incorporating attention mechanisms or adaptive optimization strategies to further improve interpretability and generalization.

Author Contributions: Conceptualization, M.Q. and R.Z.; methodology, M.Q.; validation, R.Z., investigation, M.Q.; resources, ZY.L.; writing—original draft preparation, M.Q.; writing—review and editing, M.Q., ZY.L. and R.Z.; project administration, ZY.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The original dataset used in the study is openly available at https://archive.ics.uci.edu/dataset/320/student+performance. The code used in this study is available from the corresponding author upon request.

Conflicts of Interest: The authors declare no conflicts of interest.

References

[1] Buenaño-Fernández, D.; Gil, D.; Luján-Mora, S. Application of machine learning in predicting performance for computer engineering students: A case study. Sustainability 2019, 11(10), 2833.

[2] Belachew, E.B.; Gobena, F.A. Student performance prediction model using machine learning approach: the case of Wolkite university. Int. J. Adv. Res. Comput. Sci. Softw. Eng. 2017, 7(2), 46–50.

[3] White, G.L. Adaptive learning technology relationship with student learning outcomes. J. Inf. Technol. Educ. 2020, 19, 113–130.

[4] Lau, E.T.; Sun, L.; Yang, Q. Modelling, prediction and classification of student academic performance using artificial neural networks. SN Appl. Sci. 2019, 1(9), 982.

[5] Khan, A.; Ghosh, S.K. Student performance analysis and prediction in classroom learning: A review of educational data mining studies. Educ. Inf. Technol. 2021, 26(1), 205–240.

[6] Romero, C.; Ventura, S. Educational data mining and learning analytics: An updated survey. Wiley Interdiscip. Rev. Data Min. Knowl. Discov. 2020, 10(3), e1355.

[7] Li, G.Y.; Han, G. The behavior analysis and achievement prediction research of college students based on XGBoost gradient lifting decision tree algorithm. In Proceedings of the 2019 7th International Conference on Information and Education Technology, 2019; pp. 289–294.

[8] Bhutto, E.S.; Siddiqui, I.F.; Arain, Q.A.; Anwar, M. Predicting students' academic performance through supervised machine learning. In Proceedings of the 2020 International Conference on Information Science and Communication Technology (ICISCT), 2020; pp. 1–6.

[9] Xing, W.; Li, C.; Chen, G.; Huang, X.; Chao, J.; Massicotte, J.; et al. Automatic assessment of students' engineering design performance using a Bayesian network model. J. Educ. Comput. Res. 2021, 59(2), 230–256.

[10] Marwaha, A.; Singla, A. A study of factors to predict at-risk students based on machine learning techniques. In Intelligent Communication, Control and Devices; Springer: Singapore, 2020; pp. 133–141.

[11] Drywień, M., Górnicki, K., & Górnicka, M. Application of artificial neural network to somatotype determination. *Applied Sciences*, 2021, **11**(4): 1365.

[12] Santiago, A.M.; Rodríguez, J.I.B.; Torres, J.A.O.; Rabasa, J.A.G.; Izaguirre, J.M.V.; Alejandro G.F. Detecting methotrexate in pediatric patients using artificial neural networks. *Applied Sciences*, 2024, **15**(1): 306.

[13] Qiu, Y.; Hui, Y.; Zhao, P.; Dou, J.; Bhattacharya, S.; Dai, B.; Yu, J. The novel adaptive graph neural network-based coke quality prediction for coal samples with missing properties in sustainable smart cokemaking applications. Fuel 2025, 397, 135377.

[14] Liang, X.; Yao J.; Zhang, W.; Wang, Y. A novel fault diagnosis of a rolling bearing method based on variational mode decomposition and an artificial neural network. *Applied Sciences*, 2023, **13**(6): 3413.

[15] Qiu, Y.; Hui, Y.; Zhao, P.; Cai, C.H.; Dai, B.; Dou, J.; Bhattacharya, S.; Yu, J. A novel image expression-driven modeling strategy for coke quality prediction in the smart cokemaking process. Energy 2024, 294, 130866.

[16] Zacharis, N.Z. Predicting student academic performance in blended learning using artificial neural networks. Int. J. Artif. Intell. Appl. 2016, 7(5), 17–29.

[17] Saputra, E.P. Prediction of evaluation result of e-learning success based on student activity logs with selection of neural network attributes based on PSO. J. Phys. Conf. Ser. 2020, 1641(1), 012074.

[18] Ribeiro, M.T.; Singh, S.; Guestrin, C. Anchors: High-precision model-agnostic explanations. In Proceedings of the AAAI Conference on Artificial Intelligence, 2018; 32(1).

[19] Özkurt, C. Assessing student success: The impact of machine learning and XAI-BBO approach. J. Smart Syst. Res. 2024, 5(1), 40–54.

[20] Mirashrafi, S.B.; Bol, G.; Nakhaeizadeh, G. The effect of individual factors, family background and socioeconomic status on university admission in Iran in 2007. Lit. Inf. Comput. Educ. J. 2013, 4(3), 1042–1048.

[21] García, M.V.E.; García, M.J. Educational level and parents' occupation: Their influence on the academic performance of university students. RIDE Ibero-Am. J. Res. Educ. Dev. 2019, 10(19).

[22] Bassetto, C.F. Family background and school performance: An approach with binary variables from SARESP results. Braz. J. Popul. Stud. 2019, 36, e0077.

[23] Da Silva, I.V.; da Silva, M.T.; da Silva Martins, S.A. Predictive success factors in school performance: An analysis of the large-scale assessment in Brazil. In Proceedings of the 2019 International Conference on ICT, Society, and Human Beings, 2019; pp. 161–168.
[24] Sravani, B.; Bala, M.M. Prediction of student performance using linear regression. In Proceedings of the 2020 International Conference on Emerging Technologies (INCET), 2020; pp. 1–5.
[25] Dong, C.; Guo, Y. Prediction of university students' academic level based on linear regression model. Int. J. Contin. Eng. Educ. Life Long Learn. 2020, 30(2), 204–218.
[26] Hamoud, A.; Hashim, A. S.; & Awadh, W. A. Predicting student performance in higher education institutions using decision tree analysis. International Journal of Interactive Multimedia and Artificial Intelligence, 2018, 5, 26-31.
[27] Salman, S.; Liu, X. Overfitting mechanism and avoidance in deep neural networks. arXiv Prepr. 2019, arXiv:1901.06566.
[28] Kumar, A.D.; Selvam, R.P.; Palanisamy, V. Prediction of student performance using hybrid classification. Int. J. Recent Technol. Eng. 2019, 8(4), 6566–6570.
[29] Singh, R.; Pal, S. Machine learning algorithms and ensemble technique to improve prediction of students performance. Int. J. Adv. Trends Comput. Sci. Eng. 2020, 9(3).
[30] Song, Y.; Meng, X.; Jiang, J. Multi-layer perception model with elastic grey wolf optimization to predict student achievement. PLoS ONE 2022, 17(12), e0276943.
[31] Okubo, F.; Yamashita, T.; Shimada, A.; Ogata, H. A neural network approach for students' performance prediction. In Proceedings of the 7th International Learning Analytics & Knowledge Conference, 2017; pp. 598–599.
[32] Poudyal, S.; Mohammadi-Aragh, M.J.; Ball, J.E. Prediction of student academic performance using a hybrid 2D CNN model. Electronics 2022, 11(7), 1005.
[33] Minn, S. BKT-LSTM: Efficient student modeling for knowledge tracing and student performance prediction. arXiv Prepr. 2020, arXiv:2012.12218.
[34] Sikder, M.F.; Uddin, M.J.; Halder, S. Predicting students yearly performance using neural network: A case study of BSMRSTU. In Proceedings of the 2016 5th International Conference on Informatics, Electronics & Vision (ICIEV), 2016; pp. 524–529.
[35] Kerkeb, O.S.K. Feature engineering, mining for predicting student success based on interaction with the virtual learning environment using artificial neural network. Ann. Rom. Soc. Cell Biol. 2021, 25(6), 12734–12746.
[36] Geman, S.; Bienenstock, E.; Doursat, R. Neural networks and the bias/variance dilemma. Neural Comput. 1992, 4(1), 1–58.
[37] Maschler, M.; Zamir, S.; Solan, E. Game theory. Cambridge: Cambridge University Press, 2020.
[38] Cortez, P.; Silva, A.M.G. Using data mining to predict secondary school student performance. 2008; Available online: https://repositorium.sdum.uminho.pt/handle/1822/8024 (accessed on Day Month Year).
[39] Singh, C.; Murdoch, W.J.; Yu, B. Hierarchical interpretations for neural network predictions. arXiv Prepr. 2018, arXiv:1806.05337.
[40] Else-Quest, N.M.; Hyde, J.S.; Linn, M.C. Cross-national patterns of gender differences in mathematics: a meta-analysis. Psychol. Bull. 2010, 136(1), 103.
[41] Lindberg, S.M.; Hyde, J.S.; Petersen, J.L.; Linn, M.C. New trends in gender and mathematics performance: a meta-analysis. Psychol. Bull. 2010, 136(6), 1123.

[42] Abdulkadiroglu, A.; Angrist, J.; Pathak, P. The elite illusion: Achievement effects at Boston and New York exam schools. Econometrica 2014, 82(1), 137–96.
[43] Anderson, K.; Gong, X.; Hong, K.; Zhang, X. Do selective high schools improve student achievement? Effects of exam schools in China. China Econ. Rev. 2016, 40, 121–134.
[44] Marôco, J. International large-scale assessments: Trends and effects on the Portuguese public education system. In Monitoring Student Achievement in the 21st Century: European Policy Perspectives and Assessment Strategies, 2020; pp. 207–222.
[45] Schmidt, J.; Lockwood, B. Love and other grades: A study of the effects of romantic relationship status on the academic performance of university students. J. Coll. Stud. Retent.: Res. Theory Pract. 2017, 19(1), 81–97.
[46] Awe, P.B.; Akinyemi, E.O.; Omolayo, B.O.; Balogun, M.O. Dating status and sociability as predictors of academic performance among university students. Arch. Bus. Res. 2018, 6(4), 18–30.
[47] Pietrulewicz, B. Significance of relationships and psychosocial adaptation during adolescence. Sci. Educ. South Sci. Cent. Natl Acad. Pedagog. Sci. Ukr. 2016, 5, 7–29.
[48] Li, Y.; Bebiroglu, N.; Phelps, E.; Lerner, R.M.; Lerner, J.V. Out-of-school time activity participation, school engagement and positive youth development: Findings from the 4-H study of positive youth development. J. Youth Dev. 2008, 3(3), 22–27.
[49] Poulin, F.; Denault, A.S. Friendships with co-participants in organized activities: Prevalence, quality, friends' characteristics, and associations with adolescents' adjustment. New Dir. Child Adolesc. Dev. 2013, 2013(140), 19–35.
[50] Cahn, N. Planning options for the daily care of a minor in the event of an adult's incapacity or death. In Tax, Estate, and Lifetime Planning for Minors, 2006; pp. 125.
[51] Wibowo, B.Y.; Nurmala, M.D. The influence of parental involvement on students' achievement at SMK Al-Insan Kerotek Cilegon. J. Guid. Couns. Res. 2024, 9(1), 1–?
[52] Ginther, D.K.; Pollak, R.A. Family structure and children's educational outcomes: Blended families, stylized facts, and descriptive regressions. Demography 2004, 41(4), 671–696.
[53] Tanskanen, A.O.; Danielsbacka, M. Intergenerational family relations: An evolutionary social science approach. London: Taylor & Francis, 2018; p. 182.
[54] Peng, X.; Sun, X.; He, Z. Influence mechanism of teacher support and parent support on the academic achievement of secondary vocational students. Front. Psychol. 2022, 13, 863740.
[55] Rathee, N.; Kumari, P. Parent-child relationship and academic achievement: An exploratory study on secondary school students. Int. J. Health Sci. 2022, 6(S3), 6267–6275.
[56] Liu, M. The relationship between students' study time and academic performance and its practical significance. BCP Educ. Psychol. 2022, 7, 412–415.
[57] Cassidy, S. Exploring individual differences as determining factors in student academic achievement in higher education. Stud. High Educ. 2012, 37(7), 793–810.
[58] Jez, S.J.; Wassmer, R.W. The impact of learning time on academic achievement. Educ. Urban Soc. 2015, 47(3), 284–306.
[59] Marks, G.N. Family size, family type and student achievement: Cross-national differences and the role of socioeconomic and school factors. J. Comp. Fam. Stud. 2006, 37(1), 1–46.

[60] AlSaleh, A.; Alabbasi, A.A.; Ayoub, A.E.; Hafsyan, A. The effects of birth order and family size on academic achievement, divergent thinking, and problem finding among gifted students. J. Educ. Gift. Young Sci. 2021, 9(1), 67–73.
[61] Shen, Y. The effect of family size on children's education: Evidence from the fertility control policy in China. Front. Econ. China 2017, 12(1).
[62] Azizah, S.N.; Saleh, S.; Sulistyaningrum, E. The effect of working mother status on children’s education attainment: Evidence from longitudinal data. Economies, 2022, 10(2): 54.
[63] Wagner, M. V. Modeling my mother? An exploration of the relationship between a mother's occupational status and her daughter's career aspirations [dissertation]. Boston College, 2013.
[64] Ossai, M.C.; Ethe, N.; Edougha, D.E.; Okeh, O.D. Parental educational levels and occupations as determinants of their children's examination integrity and academic performance. International Journal of Educational Reform, 2023: 10567879231213066.
[65] Wamala, R.; Kizito, O.S.; Jjemba, E. Academic achievement of Ugandan sixth grade students: Influence of parents' education levels. Contemporary Issues in Education Research, 2013, 6(1): 133–142.
[66] Vadivel, B.; Alam, S.; Nikpoo, I.; Ajanil, B. The impact of low socioeconomic background on a child’s educational achievements. Education Research International, 2023, 2023(1): 6565088.
[67] Ulferts, H.; Wolf, K.M.; Anders, Y. Impact of process quality in early childhood education and care on academic outcomes: Longitudinal meta-analysis. Child Development, 2019, 90(5): 1474–1489.
[68] Bagudo, A.A. Influence of nursery education on cognitive competence among pupils in public primary schools in Sokoto State: Does it persist or fade? Sokoto Educational Review, 2013, 14(2): 9.
[69] Bustamante, A.S.; Dearing, E.; Zachrisson, H.D.; Vandell, D.L. Adult outcomes of sustained high-quality early child care and education: Do they vary by family income? Child Development, 2022, 93(2): 502–523.
[70] Tande, K. Early childhood education and its influence on academic performance of level one pupils in selected primary schools. Journal of Education Policy and Entrepreneurial Research, 2016, 3(7).
[71] Trinidad, J.E. Stable, unstable, and later self-expectations’ influence on educational outcomes. Educational Research and Evaluation, 2019, 25(3–4): 163–178.
[72] Latikal, D. G. Motivation and learning achievement as determinants of interest in continuing higher education among senior high school students. Journal of Economic Education and Entrepreneurship Studies, 2024, 5(3): 383–392.
[73] Arbinaga, F.; Mendoza-Sierra, M.I.; Caraballo-Aguilar, B.M.; Buiza-Calzadilla, I.; Torres-Rosado, L.; Bernal-López, M.; et al. Jealousy, violence, and sexual ambivalence in adolescent students according to emotional dependency in the couple relationship. Children, 2021, 8(11): 993.
[74] Harahap, A.Z.; Harahap, A.C.P.; Putri, M.A.; Hasibuan, H.; Suryani, S.; Hajariansyah, P. Depiction of moral and spiritual development of students at MTS Cerdas Murni. Tarbiatuna: Journal of Islamic Education Studies, 2023, 3(2): 237–242.
[75] Gupta, R.; Gueneau, C. Feature correlation with student education performance. Journal of Student Research, 2021, 10(2).
[76] Mikas, D. The impact of emotional and behavioural problems on school achievement of pupils. Pedagogika (Pedagogical Research), 2012, 9(1–2): 83–99.

[77] Krieg, S.; Curtis, D.; Hall, L.; Westenberg, L. Access, quality and equity in early childhood education and care: A South Australian study. Australian Journal of Education, 2015, 59(2): 119–132.
[78] Garon-Carrier, G.; Bégin, V. The (limited) contribution of early childcare arrangements to social and academic development among Canadian children. Developmental Psychology, 2021, 57(11): 1855.
[79] Baliyan, S.P.; Khama, D. How distance to school and study hours after school influence students' performance in mathematics and English: A comparative analysis. Journal of Education and e-Learning Research, 2020, 7(2): 209–217.
[80] Spitzer, M. W. H. Just do it! Study time increases mathematical achievement scores for grade 4–10 students in a large longitudinal cross-country study. European Journal of Psychology of Education, 2022, 37(1): 39–53.
[81] Sun, X.; Li, B.; Sutcliffe, R.; Gao, Z.; Kang, W.; Feng, J. Wse-MF: A weighting-based student exercise matrix factorization model. Pattern Recognition, 2023.
[82] Shen, S.; Liu, Q.; Huang, Z.; Zheng, Y.; Yin, M.; Wang, M.; Chen, E. A survey of knowledge tracing: Models, variants, and applications. IEEE Transactions on Learning Technologies, 2024.
[83] Gao, Z.; Zhang, Y.; Zhang, R.; Sun, X.; Feng, J. Do gender or major influence the performance in programming learning? Teaching mode decision based on exercise series analysis. Computational Intelligence and Neuroscience, 2022.
[84] Shen, G.; Yang, S.; Huang, Z.; Yu, Y.; Li, X. The prediction of programming performance using student profiles. Education and Information Technologies, 2023.
[85] Shwartz-Ziv, R.; and Armon, A. Tabular data: Deep learning is not all you need. Information Fusion, 2022, 81, pp.84-90.
[86] Aslam, N.; Khan, I.; Alamri, L.; Almuslim, R. An improved early student's academic performance prediction using deep learning. International Journal of Emerging Technologies in Learning, 2021, 16(12): 108–122.
[87] Huang, C.; Zhou, J.; Chen, J.; Yang, J.; Clawson, K.; Peng, Y. A feature weighted support vector machine and artificial neural network algorithm for academic course performance prediction. Neural Computing and Applications, 2023, 35(16): 11517–11529.
[88] Zhao, Y.; Wang, H.; Wang, P. The investigation of influence related to feature dimensionality processing on deep learning-based student performance prediction. In: 2024 3rd International Conference on Robotics, Artificial Intelligence and Intelligent Control (RAIIC), 2024: 321–326.
[89] Gao, Z.; Yan, H.; Liu, J.; Zhang, X.; Lin, Y.; Zhang, Y.; Feng, J. Tracing distinct learning trajectories in introductory programming course: a sequence analysis of score, engagement, and code metrics for novice computer science vs. math cohorts. International Journal of STEM Education, 2025.
[90] Yang, X.; Zhang, H.; Chen, R.; Li, S.; Zhang, N.; Wang, B.; Wang, X. Research on forecasting of student grade based on adaptive K-Means and deep neural network. Wireless Communications and Mobile Computing, 2022, 2022(1): 5454158.

Acknowledgements

The authors received no specific funding for this work.

Competing interests

The authors declare no competing interests.

Data and Code Availability

The dataset analyzed during the current study is available in the UCI Machine Learning Repository. The source code and processed data used to support the findings of this study are available from the corresponding author upon reasonable request.