RESEARCH ARTICLE

Academic achievement prediction in higher education through interpretable modeling

Sixuan Wang, Bin Luo *

School of Foreign Languages, Wuhan Business University, Wuhan, Hubei, People's Republic of China

* luobinw26@163.com

Abstract

Student academic achievement is an important indicator for evaluating the quality of education, especially, the achievement prediction empowers educators in tailoring their instructional approaches, thereby fostering advancements in both student performance and the overall educational quality. However, extracting valuable insights from vast educational data to develop effective strategies for evaluating student performance remains a significant challenge for higher education institutions. Traditional machine learning (ML) algorithms often struggle to clearly delineate the interplay between the factors that influence academic success and the resulting grades. To address these challenges, this paper introduces the XGB-SHAP model, a novel approach for predicting student achievement that combines Extreme Gradient Boosting (XGBoost) with SHapley Additive exPlanations (SHAP). The model was applied to a dataset from a public university in Wuhan, encompassing the academic records of 87 students who were enrolled in a Japanese course between September 2021 and June 2023. The findings indicate the model excels in accuracy, achieving a Mean absolute error (MAE) of approximately 6 and an R-squared value near 0.82, surpassing three other ML models. The model further uncovers how different instructional modes influence the factors that contribute to student achievement. This insight supports the need for a customized approach to feature selection that aligns with the specific characteristics of each teaching mode. Furthermore, the model highlights the importance of incorporating self-directed learning skills into student-related indicators when predicting academic performance.

Introduction

Context and motivation

Academic achievement is of paramount importance in educational contexts, serving as a key indicator of both learning ability and the effectiveness of school administration and teaching standards [1]. The prediction of academic achievement is a continuously evolving topic in educational management. The integration of predictive models in education empowers educators to make well-informed choices, offer specific support, and enhance teaching strategies, thereby improving student learning outcomes [2].

Previous research on achievement prediction primarily utilized statistical analysis methods to process data and forecast outcomes, with data mainly derived from educational management systems, student identification cards, or surveys [3]. ML techniques, known for their ability to tackle complex, nonlinear problems without presuppositions, are adept at identifying connections between various parameters [4]. The state-of-the-art ML techniques for prediction [5] include K-Nearest Neighbors (KNN), Decision Trees, Random Forests (RF), Support Vector Machines (SVM), Neural Networks, and Naive Bayes. Recent scholarly efforts, both domestically and internationally, have been geared towards increasing the precision of student achievement predictions through technological innovations in algorithms [6–8].

Despite these developments, challenges remain in the domain of achievement prediction. A primary issue is the limited alignment between the outcomes produced by ML algorithms and the foundational principles of education and instruction, leading to hesitancy among educators in relying on these models. Additionally, there is a gap in thorough data analysis, examination of relationships, and investigation into variables that impact student academic performance patterns.

Contribution of the study

In addressing these challenges, our study delivers distinctive contributions to the field of interpretable machine learning within the context of higher education. We delineate these contributions as follows:

- Theoretical contribution: this study introduces ML models coupled with game theory-based SHAP analysis which aims to develop and validate the XGB-SHAP model, a novel approach for interpreting machine learning-based predictions of student achievement, and explore its efficacy across various teaching modalities.

- Practical contributions: It evaluates the significance of different indicators and their positive or negative impacts on prediction outcomes, thus shedding light on the educational implications of achievement prediction models. The findings of this study provide empirical data support for teachers and educators, facilitating the refinement of their instructional strategies.

- Comparative analysis: It explores student achievement prediction models in three distinct educational settings: online, offline, and blended teachings. This exploration reveals variances in teaching patterns across these modes, yielding practical advice for educators in applying these prediction models.

Structure of the article

This paper is organized as follows: Section 'Literature review' presents a review of related literatures, providing a comprehensive review of the existing literature on student achievement prediction, examines the prevailing issues and identifies the gaps within the current body of research. Section 'Methodology' details the methodology employed in this study, introduces the interpretable performance prediction framework and the indicators system used in this paper and outlines the methodology used to conduct the data analysis for this paper. The findings and their implications are discussed in Sections 'Case study' and 'Results' respectively. The paper concludes with a summary of our key findings in the final Section 'Discussion and Conclusions'. Table 1 illustrates the list of abbreviations.

Table 1. List of abbreviations.

Abbreviation	Full form
ML	Machine Learning
SHAP	SHapley Additive exPlanations
XGBoost	Extreme Gradient Boosting
MAE	Mean Absolute Error
RF	Random Forest
SVM	Support Vector Machines
AI	Artificial Intelligence
BPNN	Backpropagation Neural Network

https://doi.org/10.1371/journal.pone.0309838.t001

Literature review

Previous research

Student achievement prediction indicators. Prediction accuracy largely depends on the careful selection of indicators. The initial and most critical step is the selection of appropriate input data. Previous research has identified three key groups of student-related features as pertinent input parameters: historical student performance, student engagement, and demographic data (Tomasevic et al., 2020).

Historical student performance has been consistently identified as a reliable predictor. For instance, DeBerard et al. [9] demonstrated that high school GPA is a strong predictor of college academic success. Similarly, Shaw et al. [10] found that combined SAT scores explain about 28% of the variance in first-year college GPA. Moreover, test scores have been used to predict future academic performance in various studies [11].

Regarding student engagement, a notable correlation with academic achievement has been observed [12]. Hussain et al. [13] identified a moderately strong positive correlation between student engagement and academic achievement. With evolving teaching formats like Massive Open Online Courses and the flipped classrooms, several studies have developed predictive models by analyzing student behaviors in learning management systems, such as video interactions, assignment submissions, and forum discussions [14]. With the innovation of modern educational technology tools, including artificial intelligence tools (such as ChatGPT) and virtual reality, significant roles have been played in enhancing student learning outcomes by integrating with educational theories like constructivism, experiential learning, and collaborative learning. These technologies, by offering immersive and interactive learning experiences, have increased student engagement, motivation, and critical thinking skills, thereby positively impacting academic performance [15, 16].

Studies have also considered demographic factors. Research indicates that demographic factors play a moderate role in predictive accuracy, with relevance around 60% in some studies, while others suggest that these variables have a limited impact on prediction precision [5, 17]. Additional indicators, such as student collaboration, teacher-student communication, and psychological factors like motivation and attitude, have also been explored. Recent studies emphasize the importance of considering learners' psychological well-being and cognitive processes in educational settings [18, 19]. These motivational and coping strategies remarkably influence students' learning approaches and overall educational outcomes [20].

The above discussion shows that student achievement is a composite of cognitive, behavioral, skill-based, and emotional outcomes derived from educational experiences [21]. Although there is a consensus on the selection of certain important indicators, the selection of the dataset for student achievement prediction varies from study to study. Selecting the most

suitable dataset depends largely on the specific goals and objectives of the researchers, with no universally accepted guidelines.

Student achievement prediction models. Originally, conventional statistical methods such as Discriminant Analysis and Multiple Linear Regression were the predominant approaches in the early stages of educational research [22]. Furthermore, Structural Equation Modeling (SEM) has been widely adopted in the social sciences. However, these traditional methods have often fallen short of delivering consistent and precise predictions or classifications [23].

Recently, an array of machine learning algorithms has been employed, including Multiple Regression, Probabilistic and Logistic Regression, Neural Networks, Decision Trees, Random Forests (RF), Genetic Algorithms, and Bayesian algorithms. These have shown varied levels of success in achieving high predictive accuracy [24]. Comparative studies of machine learning methods have been conducted, with Caruana et al. [25] exploring the performance evaluation of these models. Their research underscores a fundamental point: no single model or method universally excels across all problems and metrics. Tomasevic et al. [5] used the Open University Learning Analytics Dataset for a regression problem, finding that Artificial Neural Networks (ANN) and Decision Trees were the most effective, while KNN, SVM, and Bayesian linear regression were less successful.

While previous approaches using machine learning models for predicting student achievement have focused on model optimization [26], there are growing concerns regarding the opaque nature of complex models, which may hinder their broader application [27].

Interpretable machine learning models. Nowadays, with the rapid development of artificial intelligence (AI) technology, ML models are being applied in many critical fields, such as education [28, 29], healthcare [30–32]. However, as the number of parameters soars, the 'black-box' nature of neural networks has raised concerns. Interpretable machine learning is a promising tool to alleviate concerns regarding the opacity of machine learning models. It equips ML models with the capability to articulate their processes in a manner comprehensible to humans [33].

Broadly, interpretable machine learning methods are divided into two categories: self-interpretation models and post-hoc interpretation methods [34]. Self-interpreting models typically have a simpler structure and include Linear models, Logistic Regression, and Decision Trees. Post-hoc interpretation methods involve either model-independent or model-specific techniques, applicable to various models but may require additional computational resources and analytical expertise.

Post-hoc or model-independent interpretation methods are extensively used in different scenarios. These include Partial Dependence Plot [35], Individual Conditional Expectation [36], Permutation Feature Importance [37], Local Interpretable Model-agnostic Explanations, and the SHAP method. The survey in the field of information resource management revealed that 83.7% of explainable ML applications utilize post-hoc explanation methods, with SHAP (51.2%) and feature importance analysis (34.1%) being the most common. Unlike traditional feature importance which indicates the significance of features without clarifying their impact on predictions, SHAP offers detailed explanations on both sample and feature levels through various visualizations like waterfall diagrams and feature dependency diagrams.

These interpretative approaches have been applied in diverse fields such as medicine, policymaking, and science, aiding in auditing predictions under circumstances like regulatory pressures and the pursuit of fairness [35]. However, the critical aspect of interpretability in machine learning models within the domain of educational management research remains underexplored.

Research gap

Given the aforementioned limitations, the interpretability of ML is a contentious issue. The various ML algorithms employed often fail to effectively elucidate the relationship between factors influencing students' academic performance and their grades. Additionally, they struggle to quantify the impact of each feature on the target value and to determine the positive or negative influence of each characteristic. To address these gaps in the literature, our study delves into the following areas:

- Feature Importance Analysis: Our research will quantify the influence of each feature on the prediction of student performance. This involves a detailed examination of the weight and significance of various factors in determining academic outcomes.

- Impact Assessment: We will assess the positive or negative impact of each feature on the target variable. This is crucial for understanding not only the magnitude of the influence but also its direction.

- Model Comparison: By comparing the interpretability and performance of different ML models, our study seeks to identify the most effective approaches for student achievement prediction.

- Practical Implications: We will discuss the practical implications of our findings, focusing on how increased interpretability can enhance educational practices and inform policy-making.

Through this comprehensive approach, our study seeks to bridge the gap in the current research by providing a clearer understanding of the mechanisms behind student achievement prediction models and their implications for educational stakeholders.

Methodology

Development of an interpretable performance prediction framework

As shown in Fig 1, we have developed an interpretable framework for performance prediction. The framework's core involves extracting five key features: academic factors, student engagement, demographic factors, psychological aspects, and self-directed learning abilities. These features form an input vector that accurately represents factors relevant to achievement prediction. The data for this study is sourced from three main systems: the Education Administration System (EAS), the Chaoxing Xuexitong System, and various questionnaires.

The methodology progresses in three phases. The initial phase involves creating an indicator system from these features. In the subsequent phase, we focus on constructing and elucidating performance prediction models. Four different ML algorithms are applied to our "learning" dataset. Their effectiveness is evaluated using two standard ML metrics: Mean Absolute Error (MAE) and R-squared (R^2). The optimal model is then selected based on these evaluations. The final stage of our methodology is the model interpretability phase, which accounts for the educational significance of the model by analyzing the importance and directional influence of the indicators. This phase aims to provide educators with insights to refine their teaching strategies.

Development of the indicator system

As mentioned in 'Literature review' section, prior research insights advocate categorizing student-related features into historical student performance, engagement, and demographic data [5]. To capture a holistic view of learner characteristics, we have expanded this system to include psychological factors and self-directed learning capabilities to form a student
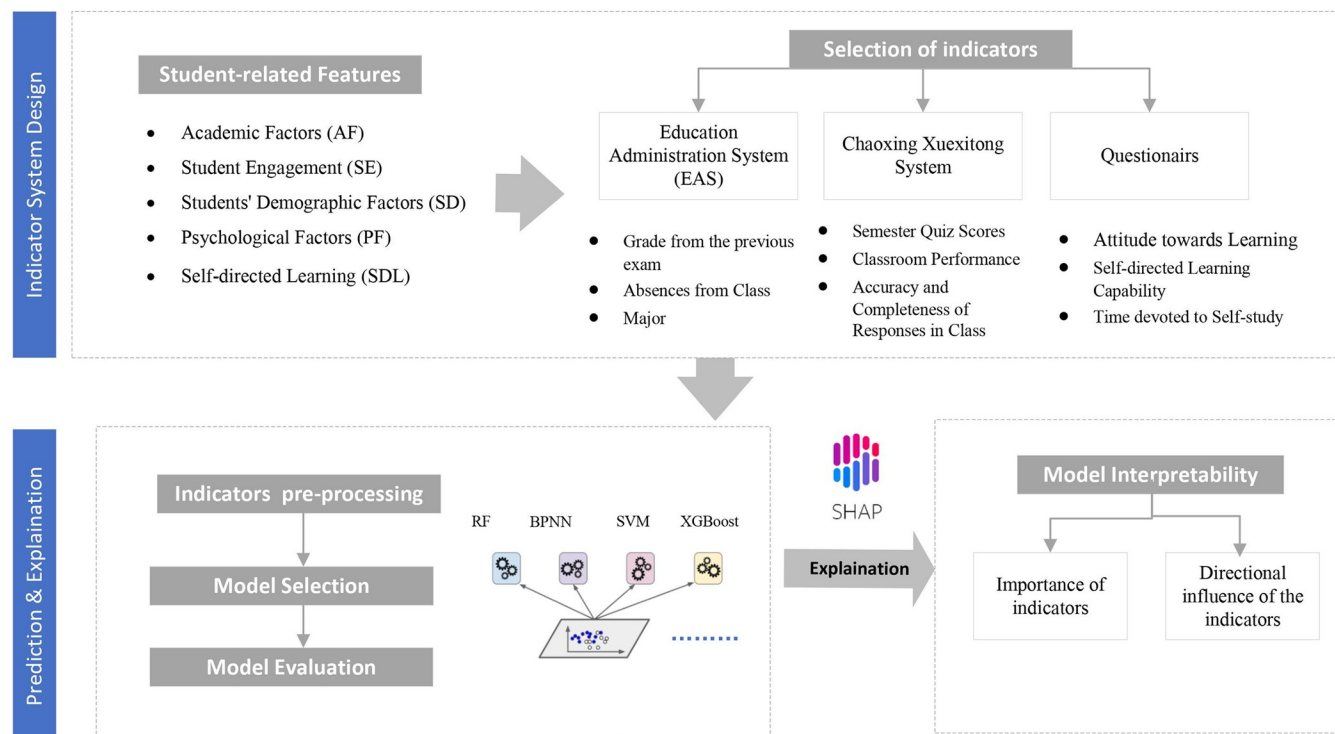
Fig 1. The framework of the interpretable academic achievement prediction model.

achievement prediction indicator system, as shown in Table 2. Considering the minimal variation in age, gender, and other demographic factors in our case study, we have chosen to focus solely on the major as the demographic data point.

Utilizing practical data from real-world scenarios, our model integrates these nine indicators to predict student achievement. The indicators collectively form an input vector matrix for predicting a student's academic performance, which can be defined as

$$x(i) = [x_{i,1}, x_{i,2}, x_{i,3}, \ldots \ldots, x_{i,k}]^T, \tag{1}$$

Table 2. Student achievement prediction indicator system.

Feature	Definition	Indicator
Academic Factors (AF)	Quantifies the academic performance of students, such as exam scores, GPA, or class rank.	1. Grade from the previous exam 2. Semester Quiz Scores
Student Engagement (SE)	Measures the level of student involvement and participation in educational activities, including class attendance, participation in discussions, completion of assignments, and extracurricular involvement.	3. Classroom Performance 4. Accuracy and Completeness of Responses in Class 5. Absences from Class
Psychological Factors (PF)	Focus on students' psychological characteristics and attributes that contribute to their academic success. They may include variables such as motivation, self-efficacy, self-regulation, learning styles, and attitudes toward learning.	6. Attitude towards Learning
Students' Demographic Factors (SD)	Recognizes the influence of students' backgrounds and contexts on their academic outcomes. This data includes variables such as age, gender, specialties or areas of study, and location.	7. Major
Self-directed Learning (SDL)	It refers to the process in which individuals take the initiative and responsibility for their own learning, without relying on external direction or instruction.	8. Self-directed Learning capability 9. Time devoted to self-study

where *i* indicates the student's serial number and *k* the *kth* indicator. The output set y_i represents the student's final exam score for the current period.

Model training

As SHAP is a model-agnostic interpretation framework, which enables it to be applied across a spectrum of common predictive models. This versatility allows SHAP to provide insights into the decision-making process of these models by quantifying the contribution of each feature to the prediction, thereby enhancing our understanding of the model's behavior regardless of its underlying structure or algorithmic approach. Commonly used ML models for academic achievement prediction include RF, BPNN, SVM, and XGboost. The rationale for selecting these four models is their proficiency as data-driven prediction methods. RF, an ensemble learning technique, amalgamates numerous decision trees, thereby reducing variance relative to individual trees. It is known for its superior average prediction performance. BPNN, a supervised learning algorithm, builds multi-layer neural networks inspired by biological neurons and employs a back-propagation algorithm for training, excelling in handling non-linear relationships and high-dimensional data. SVM has gained recognition for its effectiveness in classification, regression, and time-series prediction. XGBoost, enhancing the Gradient Boosting Decision Tree algorithm, stands out for its accuracy and flexibility.

To evaluate and select the most suitable model, we use MAE and R^2 as performance metrics, which can be defined as:

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|, \tag{2}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{3}$$

where *n* represents the number of samples. MAE measures the average absolute error between predicted and actual values, with a lower MAE indicating superior model performance. R^2 assesses the model's data fit, where a larger R^2 value generally signifies a better fit. An empirical R^2 value greater than 0.4 is considered indicative of a good fit [38].

In this research, a 5-fold cross-validation approach was implemented to fine-tune the hyperparameter to avoid overfit, optimizing them according to the mean value derived from each test set.

Model interpretability

Addressing the opaque nature of ML models, our research employs the SHAP method for interpretability. Developed by Lundberg and Lee in 2017 [39], SHAP merges various existing approaches to provide a reliable and intuitive explanation of model predictions. It does so by illustrating how predictions shift when certain variables are omitted. The Python SHAP package (https://github.com/slundberg/shap), enables the calculation of SHAP values for any selected model, and it is extensively utilized due to its versatility.

SHAP is characterized by three fundamental properties: local accuracy (the sum of feature attributions equals the model output), missingness (zero attribution for non-present features), and consistency (no decrease in feature attribution despite an increased marginal contribution). A notable advantage of SHAP is its model-agnostic nature, making it applicable to any machine learning model.

The principle of SHAP can be explained as follows: Assume the *ith* sample is x_i, with the *jth* feature of this sample being x_{ij}, and the model's predicted value for this sample as y_i. The

baseline value for the model (often the average of the target variable) is y_{base}. The SHAP value then follows the equation:

$$y_i = y_{base} + f(x_{i1}) + f(x_{i2}) + \ldots + f(x_{ik}) \tag{4}$$

where $f(x_{ik})$ is the SHAP value of x_{ij}. Intuitively, $f(x_{ij},1)$ indicates the contribution of the 1st feature in the *ith* sample to the final predicted value y_i. A value pf $f(x_i,1)$ greater than 1 implies that the feature enhances the predicted value, whereas a negative value suggests a diminishing effect.

Case study

Datasets

Data for this study was obtained from the EAS of a Wuhan-based public university. This system provided access to students' personal information, such as majors and academic grades. In addition, we gathered course-related learning data from the Chaoxing Xuexitong system, a widely used online education platform in China. To obtain data on self-study hours, learning attitudes, and self-directed learning indicators, we employed questionnaires as the methodological instrument. The learning attitude questionnaire adapted from the English-learning Motivation Scale developed by a Chinese scholar Meihua Liu [40] who is from Tsinghua University, a tool commonly utilized in in EFL teaching and learning in the Chinese context. For assessing self-directed learning capabilities, we used a questionnaire adapted from Jinfen Xu 's [41] self-directed learning capability scale. These questionnaires were administered in class under instructor supervision and lasted approximately 10 minutes each, aiming to evaluate students' learning attitudes and their aptitude for independent learning. The surveys were conducted midway through each semester. Our dataset encompasses data from 87 students enrolled in the Japanese course for the class of 2021, spanning three different learning modes. It includes nine indicators linked to student grades, amounting to a total of 2349 data entries. Table 3 shows the types of nine indicators.

While analyzing the datasets, an imbalanced data pattern was noted. To address this, we grouped students into three broad specialty categories: Arts, Science and Technology, and Arts and Sports. This categorization reduced data sparsity by assigning discrete values (1, 2, 3) to these groups.

Ethical considerations

The study was approved by the institutional review board, and the study runs from September 2021 to June 2023. All participants were not at risk if they chose or declined to participate.

Table 3. Description of indicator types.

Indicators	Indicator Type
Classroom Performance	Continuous
Accuracy and Completeness of Responses in Class	Continuous
Absences from Class	Continuous
Semester Quiz Scores	Continuous
Time devoted to Self-study	Continuous
Attitude towards Learning	Continuous
Self-directed Learning capability	Continuous
Major	Discrete
Grade from the previous exam	Continuous

https://doi.org/10.1371/journal.pone.0309838.t003

Parental consent is not required for undergraduate students participating in the study. Additionally, we explained the purpose of the study in the questionnaire, clarified that it was their right to participate or not to participate in the study, and informed all the participants that 'submitting answers' is considered informed consent for researchers to use their questionnaire responses and related data retrieved from EAS and Chaoxing platform in publications of the research.

Experimental setup

In this study, we conducted experiments employed PyCharm version 2022.3.3 as the compilation software, and implemented the algorithmic model using Python. The dataset was randomly partitioned into training and test sets in a 4:1 ratio for robust training and evaluation.

As state in the Methodology Section, we employ four classic ML models as our predictive model for academic performance. Table 4 presents the pseudo-code outlining the experimental procedures.

Results

Comparison of models

To obtain the optimal model parameters, the hyperparameters of the aforementioned four models were optimized separately. Table 5 displays the optimal hyperparameter combinations for the aforementioned four models.

Table 6 presents the comparison of the task performance of four models. Both BPNN and XGBoost show higher task performance compared to RF, while SVM lags in terms of task performance. The comparison indicates that XGBoost slightly surpasses BPNN, establishing XGBoost as the model with the best predictive performance. Therefore, this study selects the XGBoost model to fit all the data. SHAP values are used for interpretation.

Exploratory analysis utilizing XGBoost and SHAP

Given the effectiveness of the XGBoost model, it was selected for further analysis using SHAP to explore teaching patterns within the model across various teaching modes. SHAP offers insights into the influence of each indicator per sample, highlighting both positive and negative effects. In the associated figures, color coding is used to represent the magnitude of eigenvalues, with red indicating high values and blue representing low values.

Table 4. The pseudo-code for the XGBoost-Shap.

Algorithm 1: XGBoost-SHAP.
Input: D represents the dataset containing all samples, *test_per* is the percentage of test data to the total number of D. $\mathcal{M}$ represents the predictive model, and $\mathcal{N}$ represents the number of training epochs
Output: importance_scores_set refers to the collection of feature scores for each test sample, where these scores indicate the importance of the features.
importance_scores_set = ϕ
$\mathcal{D}\prime$ = feature_normal_and_mapping($\mathcal{D}$) // Continuous features are normalised
$D_{train}D_{test}$ = train_test_split($\mathcal{D}$, test_per)
for epoch = 0 to N do
$\mathcal{M}$ = model_update($\mathcal{M}, D_{train}$)
for $\in D_{test}$ do
importance_scores = SHAP($\mathcal{M}, D_{train}$)
importance_scores_set = importance_scores_set $\cup$ {importance_scores}

https://doi.org/10.1371/journal.pone.0309838.t004

Table 5. Optimal hyperparameter combinations of the four models.

Algorithms	Hyperparameter combinations	Optimal hyperparameter combination
SVM	C = (0.1,1,10,100) gamma = (0.0001,0.001,0.01,0.1,1)	C = 1 gamma = 0.001
BPNN	batch_size: = (8,16,32,64) lr = (0.00001,0.00004,by 0.000002) epoch = (500,800,1000,1500) min_epoch = (100,150,200) patience = (25,50,75) dropout_probability = (0.1,0.2,0.3)	batch_size: = 8 lr = 0.000038 epoch = 1000 min_epoch = 200 patience = 50 dropout_probability = 0.2 torch.optim.Adam
RF	n_estimators = (50,200, by = 10) max_depth = (20,50, by = 5) min_samples_leaf = (1,2,3,4,5,6) min_samples_split = (1,2,3,4)	n_estimators = 100 max_depth = 22 min_samples_leaf = 1 min_samples_split = 2
XGBoost	n_estimators = (100,1000,by = 100) max_depth = (3,20,by = 1) min_child_weight = (1,2,3,4,5) gamma = (0,0.1,0.2,0.3,0.4,0.5) subsample = (0.5,0.6,0.7,0.8,0.9,1) colsample_bytree = (0.5,0.6,0.7,0.8,0.9,1) reg_alpha = (0,0.1,1,2,3) reg_lambda = (0,0.1,1,2,3)	n_estimators = 100 max_depth = 6 min_child_weight = 1 gamma = 0 subsample = 1 colsample_bytree = 1 reg_alpha = 1 reg_lambda = 0

https://doi.org/10.1371/journal.pone.0309838.t005

Figs 2 and 3 shows the importance of indicators and a summary plot for offline teaching. The average SHAP value (horizontal axis) indicates the significance of each indicator, with their order of importance shown on the vertical axis in Fig 2. Key findings include classroom performance, previous exam grades, and student major as the most influential indicators. The impact of eigenvalues on each sample is depicted in Fig 3, where each row represents an indicator, each dot signifies a sample, and the SHAP value is plotted on the horizontal axis. Further analysis revealed a positive relationship between prior exam grades, self-directed learning ability, learning attitudes, and their effect on academic achievement predictions. Interestingly, occasional absences did not show a substantial negative influence on predicted grades, hinting at a divergence in the dynamics of college classrooms from high school settings. This might be attributed to the independent learning skills prevalent among college students. Moreover, it was noted that students majoring in Arts and Sports tend to have a slightly negative impact on predicted grades.

Analysis of online teaching using XGBoost and SHAP

Figs 4 and 5 presents the indicator importance and summary plot for online teaching. A key observation is the increased influence of previous exam grades on the predicted values in comparison to offline settings. This suggests that students with a strong academic foundation tend

Table 6. Comparison of the ability to predict and fit of the four models.

Arithmetic	Offline teaching		Online teaching		Blended teaching	
	MAE	R^2	MAE	R^2	MAE	R^2
RF	7.8333	0.6617	5.2094	0.7202	6.8144	0.7475
BPNN	7.6987	0.7148	5.3880	0.7042	6.1815	0.8085
SVM	10.9029	0.2170	7.5030	0.4454	10.0406	0.4536
XGBoost	7.5578	0.7321	5.3710	0.7146	5.9176	0.816

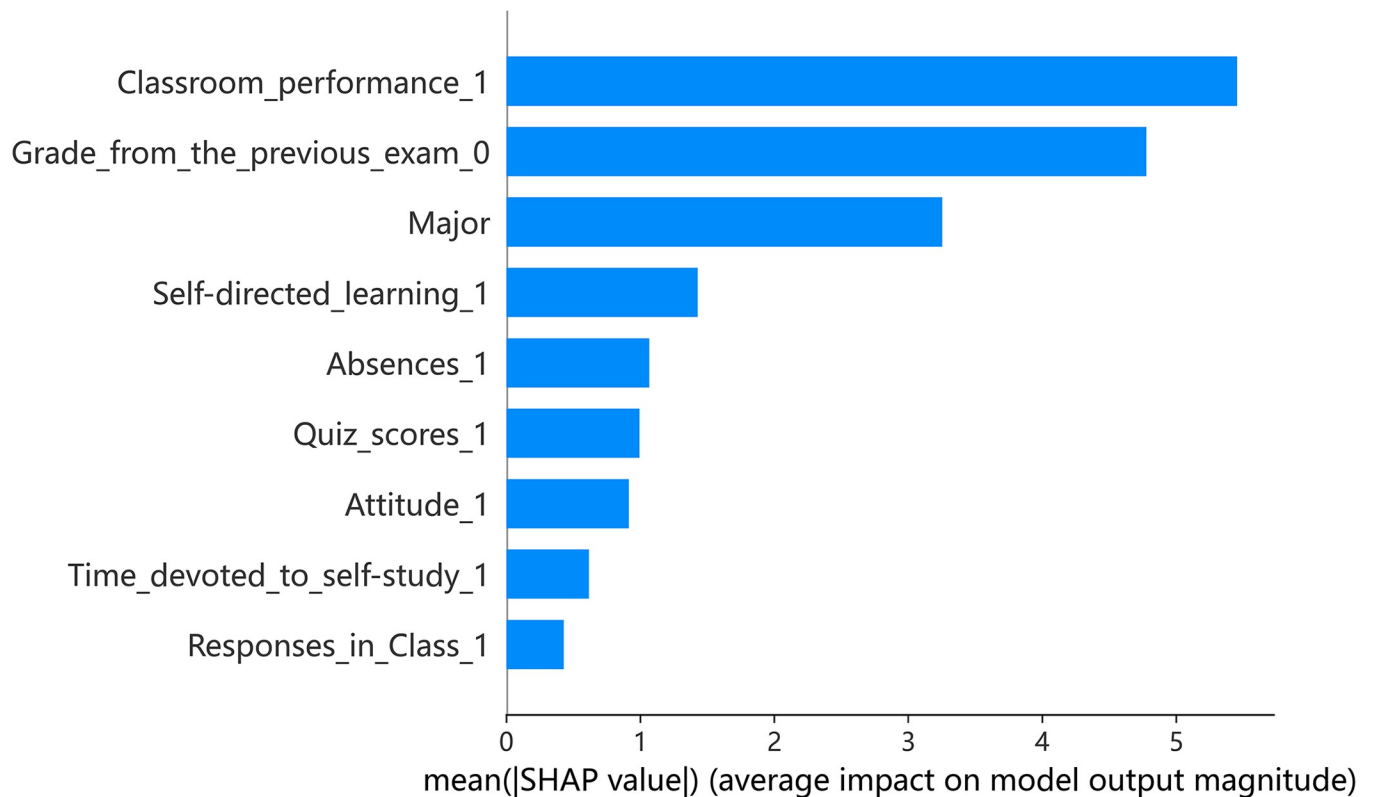https://doi.org/10.1371/journal.pone.0309838.t006

Fig 2. Indicator importance of offline teaching.

to be more self-directed, thereby enhancing their predicted performance more remarkably. The disparity in self-directed learning abilities is more evident in online courses, highlighting the detrimental effect of inadequate self-learning skills on performance. Students struggling with self-learning might not receive timely support, leading to poorer outcomes. In this context, classroom performance becomes a less critical predictor, and the influence of a student's major on predicted scores also diminishes. Interestingly, self-study time shows a positive correlation with predicted grades, while the relationship between quiz scores and performance prediction remains insignificant.

Blended teaching: Insights from XGBoost and SHAP

Figs 6 and 7 examines the indicator importance and summary plot for blended teaching. In this teaching mode, the impact of self-directed learning skills is more notable compared to other teaching methods, possibly due to the adoption of flipped classroom techniques. Self-directed learning shows a stronger positive correlation with both previous exam grades and quiz scores. Furthermore, the relevance of attitude towards learning is accentuated, suggesting its growing importance in blended learning environments where independent study is emphasized.

Discussion and conclusions

The prediction of academic achievement in higher education has become an increasingly prominent topic within the field of education [42]. In today's information age, the tremendous growth of educational institutions' electronic data "…can be utilized for discovering unknown
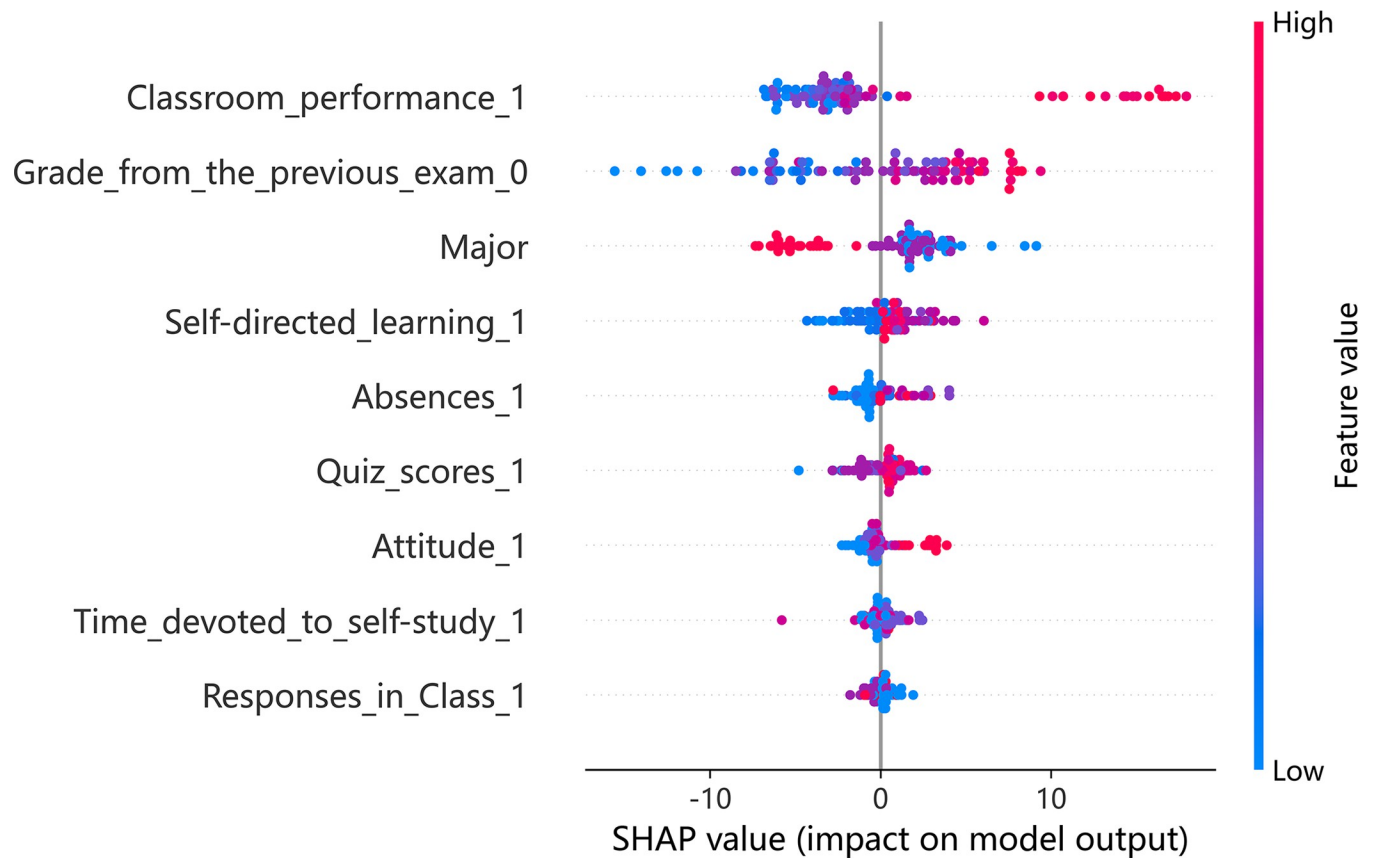
Fig 3. Summary plot of offline teaching.

patterns and trends" [43].Recent researches on predicting student performance are frequently spearheaded by educators identifying as "AI" educators to identify features that can be used to make predictions [44], to identify algorithms that can improve predictions [45], and to quantify aspects of student performance. However, analyzing performance, providing high-quality education strategies for evaluating the students' performance from these abundant resources are among the prevailing challenges universities face [46].

In this research, we have developed the XGB-SHAP model, integrating XGBoost with SHAP, to systematically explore the relationship between grade prediction and diverse indicators across various teaching methods. Focused on university Japanese language classes, our study demonstrated XGBoost's superior performance over other models, as evidenced by R^2 and MAE metrics. The integration of SHAP offered a clear visual representation, highlighting the mode and directional influence of each indicator and sheds light on the educational implications of ML structures in pedagogy. The study also supported that the XGB-SHAP model can be effectively used in the field of educational management research.

The results reveal that, the study of student achievement prediction, using student-related features, such as student historical achievement, student engagement and demographic data, which have been used as important input features in the previous literature, is not sufficient. With the development of society and the diversification of teaching and learning modes, the importance of self-directed learning skills in the prediction of university students' performance has been demonstrated in this study. Psychological factors such as attitude towards learning should also be taken into account. The impact of a student's major on foreign
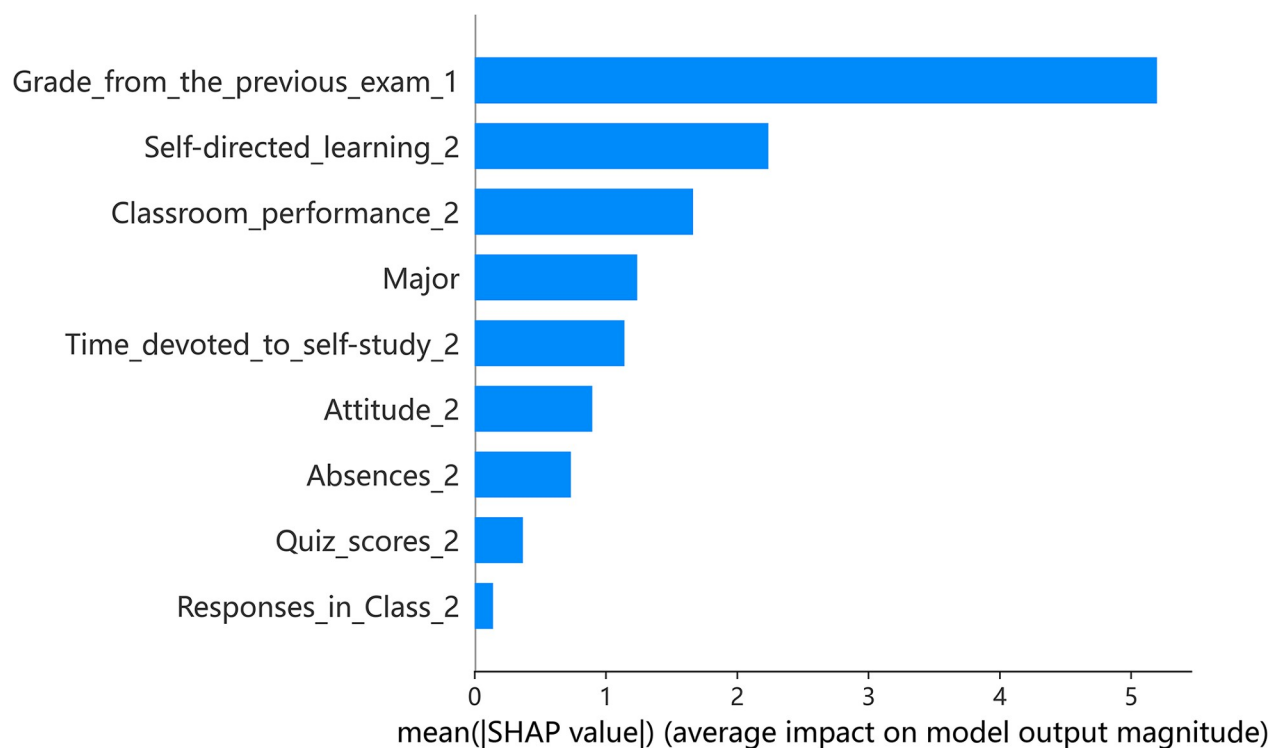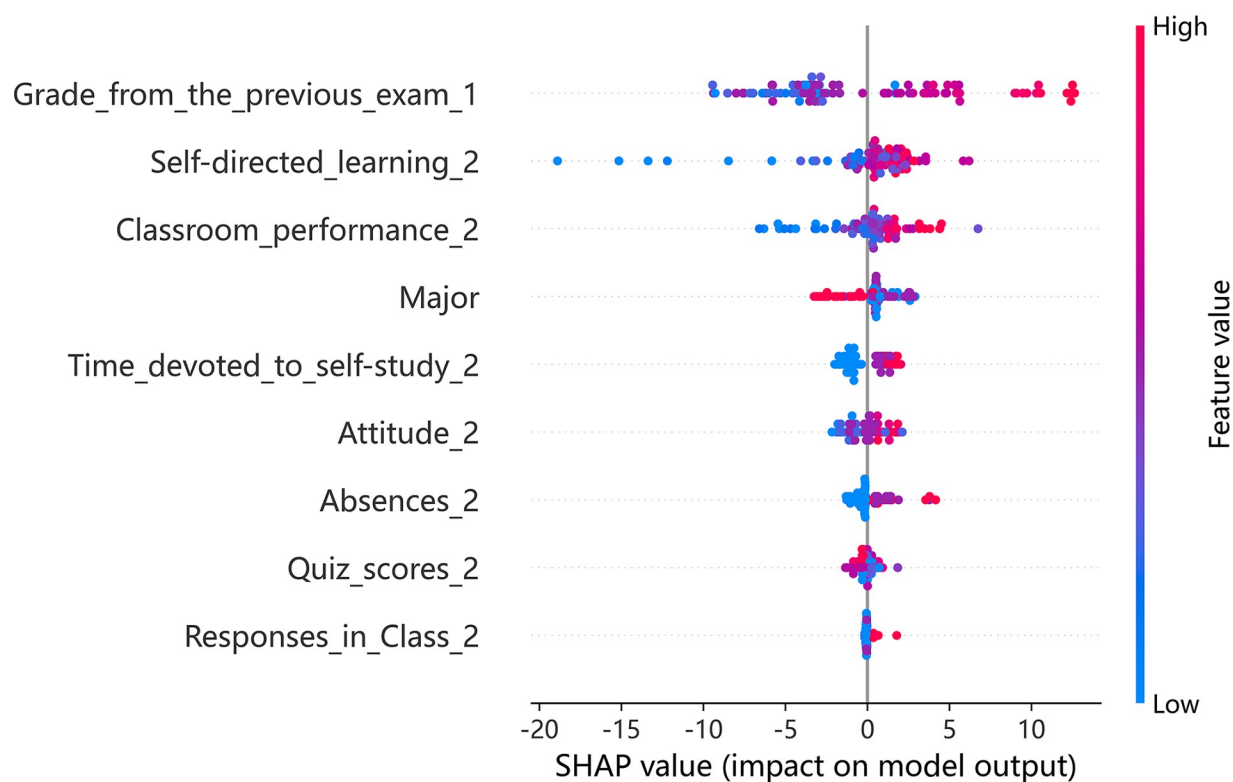
Fig 4. Indicator importance of online teaching.
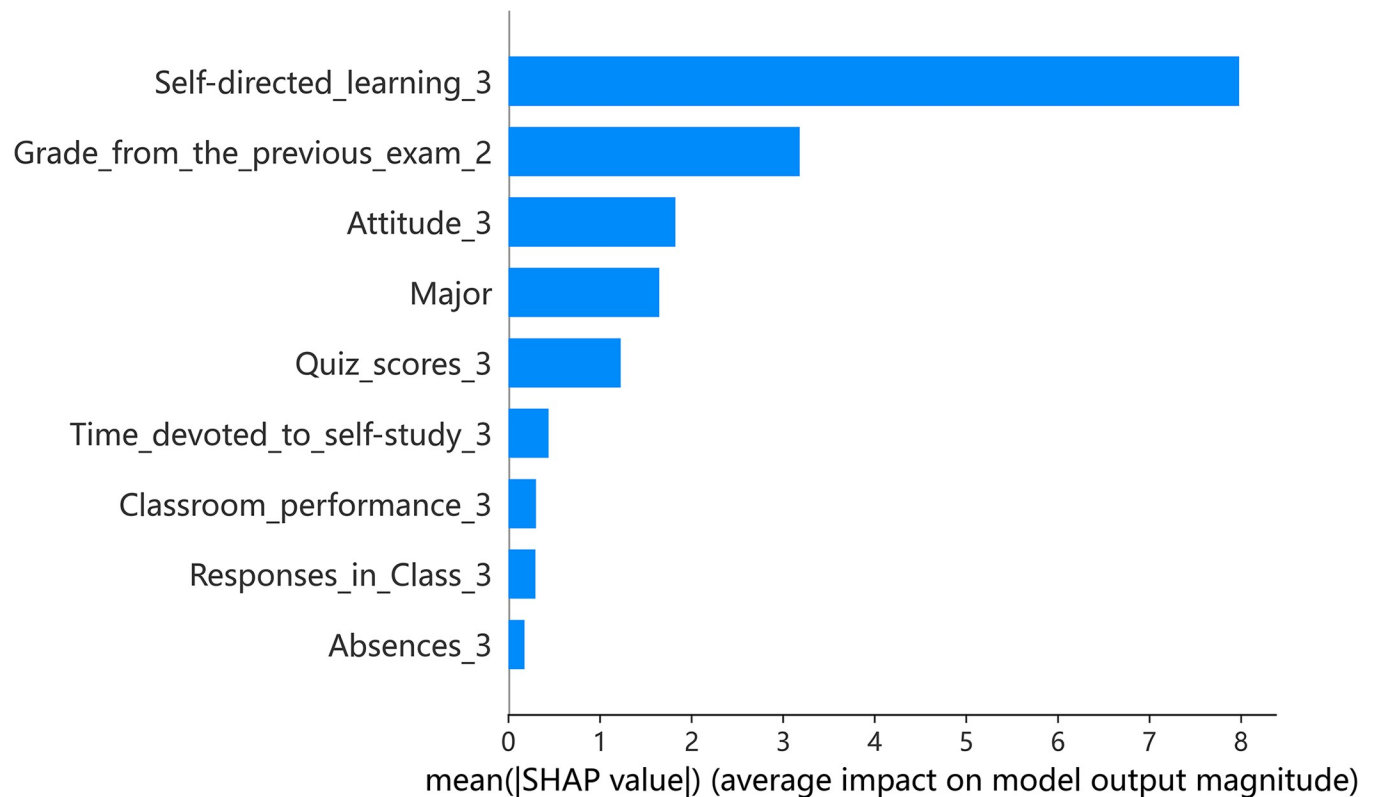
Fig 5. Summary plot of online teaching.

Fig 6. Indicator importance of blended teaching.

https://doi.org/10.1371/journal.pone.0309838.g006

language learning is considerable, which indicate differences in learning environments, cultural factors, motivation to learn foreign languages. While classroom response accuracy and attendance appeared less critical. This suggests a potential shift in focus within higher education classrooms, advocating for a tailored approach to characteristic selection based on teaching modes. This methodology provides educators with a quantitative view of how educational processes affect student achievement.

Our study also shows that the factors influencing student performance vary: offline teaching values classroom performance, while online teaching and blended teaching emphasize independent learning. In blended teaching, quiz scores have a remarkable positive impact, differing from the trends in other modes. This could be attributed to quizzes acting as formative assessments in blended learning, enhancing student participation and providing continual feedback. Consequently, teaching strategies and support systems should be adapted to meet the distinct needs of each teaching mode to optimize learning outcomes.

Acknowledging the formidable technical challenges associated with interpretable machine learning models in practical educational contexts, it is imperative to recognize their substantial contributions in enhancing our comprehension and utility of achievement prediction models. Additionally, they play a pivotal role in mitigating the skepticism harbored by educators towards machine learning models deployed for achievement prediction. Moving forward, there exist several promising avenues for exploration within the realm of interpretable machine models that merit thorough investigation: first, expand the dataset to cover more academic areas, different institutions, and varied student groups. This will test the model's effectiveness in diverse settings. Second, the refinement and augmentation of existing interpretable
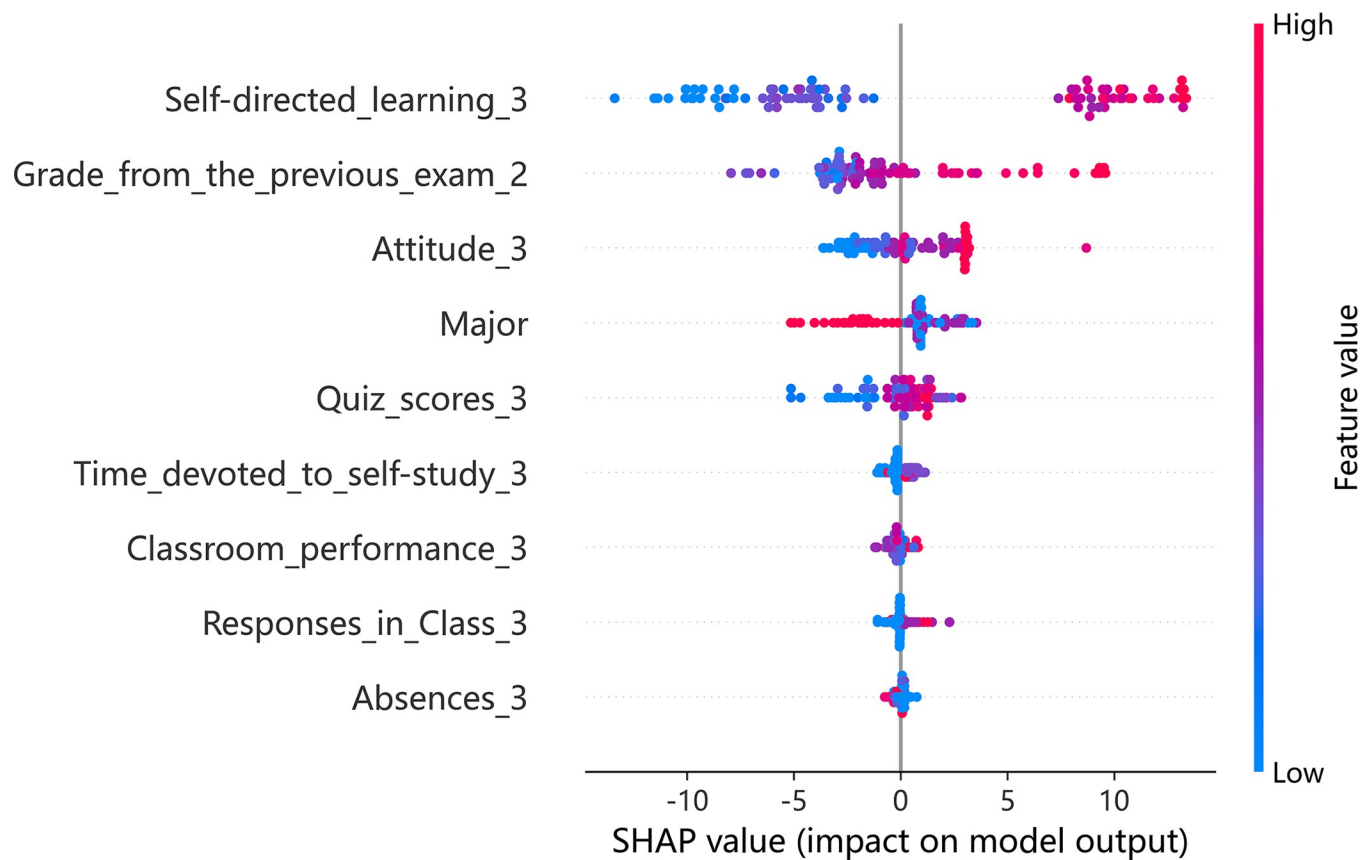
Fig 7. Summary plot of blended teaching.

models to enhance their accuracy and utility. These directions offer promising avenues for furthering the application and acceptance of interpretable machine learning in educational settings.

Supporting information

S1 File. Original data.
(XLSX)

Author Contributions

Conceptualization: Sixuan Wang.

Data curation: Sixuan Wang.

Formal analysis: Sixuan Wang.

Funding acquisition: Sixuan Wang.

Investigation: Sixuan Wang, Bin Luo.

Methodology: Sixuan Wang.

Software: Sixuan Wang, Bin Luo.

Supervision: Sixuan Wang.

Validation: Sixuan Wang.

Writing – original draft: Sixuan Wang.

Writing – review & editing: Sixuan Wang, Bin Luo.

References

1. You J W. Identifying significant indicators using LMS data to predict course achievement in online learning. The Internet and Higher Education, 2016, 29: 23–30. https://doi.org/10.1016/j.iheduc.2015.11.003

2. Musso M, Kyndt E, Cascallar E, et al. Predicting Mathematical Performance: The Effect of Cognitive Processes and Self-Regulation Factors. Education Research International, 2012, 2012(1): 250719. https://doi.org/10.1155/2012/250719

3. Namoun A, Alshanqiti A. Predicting student performance using data mining and learning analytics techniques: A systematic literature review. Applied Sciences, 2020, 11(1): 237. https://doi.org/10.3390/app11010237

4. Ma T, Wu L, Zhu S, et al. Multiclassification prediction of clay sensitivity using extreme gradient boosting based on imbalanced dataset. Applied Sciences, 2022, 12(3): 1143. https://doi.org/10.3390/app12031143

5. Tomasevic N, Gvozdenovic N, Vranes S. An overview and comparison of supervised data mining techniques for student exam performance prediction. Computers & education, 2020, 143: 103676. https://doi.org/10.1016/j.compedu.2019.103676

6. Liu C, Wang H, Du Y, et al. A predictive model for student achievement using spiking neural networks based on educational data. Applied Sciences, 2022, 12(8): 3841. https://doi.org/10.3390/app12083841

7. Liu C, Wang H, Yuan Z. A method for predicting the academic performances of college students based on education system data. Mathematics, 2022, 10(20): 3737. https://doi.org/10.3390/math10203737

8. Baashar Y, Alkawsi G, Mustafa A, et al. Toward predicting student's academic performance using artificial neural networks (ANNs). Applied Sciences, 2022, 12(3): 1289. https://doi.org/10.3390/app12031289

9. DeBerard M S, Spielmans G I, Julka D L. Predictors of academic achievement and retention among college freshmen: A longitudinal study. College student journal, 2004, 38(1): 66–81.

10. Shaw E J, Marini J P, Beard J, et al. The Redesigned SAT® Pilot Predictive Validity Study: A First Look. Research Report 2016–1. College Board, 2016.

11. Lei Z, Tong D, Zhuoping W. The prediction of academic achievement and analysis of group characteristics for mooc learners based on data mining. Chongqing Higher Educ. Res, 2021, 2: 1–13. https://doi.org/10.15998/j.cnki.issn1673-8012.2021.06.009

12. Li X, Zhang Y, Cheng H, et al. Student achievement prediction using deep neural network from multi-source campus data. Complex & Intelligent Systems, 8, 5143–5156. 2022. https://doi.org/10.1007/s40747-022-00731-8

13. Hussain M, Zhu W, Zhang W, et al. Student Engagement Predictions in an e-Learning System and Their Impact on Student Course Assessment Scores. Computational intelligence and neuroscience, 2018, 2018(1): 6347186. https://doi.org/10.1155/2018/6347186 PMID: 30369946

14. Riestra-González M, del Puerto Paule-Ruíz M, Ortin F. Massive LMS log data analysis for the early prediction of course-agnostic student performance. Computers & Education, 2021, 163: 104108. https://doi.org/10.1016/j.compedu.2020.104108

15. Al Shloul T, Mazhar T, Iqbal M, et al. Role of activity-based learning and ChatGPT on students' performance in education. Computers and Education: Artificial Intelligence, 2024: 100219. https://doi.org/10.1016/j.caeai.2024.100219

16. Mallek F, Mazhar T, Shah S F A, et al. A review on cultivating effective learning: synthesizing educational theories and virtual reality for enhanced educational experiences. PeerJ Computer Science, 2024, 10: e2000. https://doi.org/10.7717/peerj-cs.2000 PMID: 38855256

17. Kovacic Z. Early prediction of student success: Mining students' enrolment data. 2010.

18. Ahmad S, El-Affendi M A, Anwar M S, et al. Potential future directions in optimization of students' performance prediction system. Computational Intelligence and Neuroscience, 2022, 2022(1): 6864955. https://doi.org/10.1155/2022/6864955 PMID: 35619762

19. Kukkar A, Mohana R, Sharma A, et al. Prediction of student academic performance based on their emotional wellbeing and interaction on various e-learning platforms. Education and Information Technologies, 2023, 28(8): 9655–9684. https://doi.org/10.1007/s10639-022-11573-9

20. Boekaerts M, Niemivirta M. Self-regulated learning: Finding a balance between learning goals and ego-protective goals. Handbook of self-regulation. Academic Press, 2000.

21. Cogliano M C, Bernacki M L, Hilpert J C, et al. A self-regulated learning analytics prediction-and-intervention design: Detecting and supporting struggling biology students. Journal of educational psychology, 2022, 114(8): 1801. https://doi.org/10.1037/edu0000745

22. Vandamme J P, Meskens N, Superby J F. Predicting academic performance by data mining methods. Education Economics, 2007, 15(4): 405.

23. Kyndt E, Musso M, Cascallar E, et al. Predicting academic performance: The role of cognition, motivation and learning approaches. A neural network analysis.Methodological challenges in research on student learning. Antwerp: Garant, 2015.

24. Yağcı M. Educational data mining: prediction of students' academic performance using machine learning algorithms. Smart Learning Environments, 2022, 9(1): 11. https://doi.org/10.1186/s40561-022-00192-z

25. Caruana R, Niculescu-Mizil A. An empirical comparison of supervised learning algorithms. Proceedings of the 23rd international conference on Machine learning. 2006: 161–168. https://doi.org/10.1145/1143844.1143865

26. Shalev-Shwartz S, Ben-David S. Understanding machine learning: From theory to algorithms. Cambridge university press, 2014.

27. Du M, Liu N, Hu X. Techniques for interpretable machine learning. Communications of the ACM, 2019, 63(1): 68–77.

28. Munir H, Vogel B, Jacobsson A. Artificial intelligence and machine learning approaches in digital education: A systematic revision. Information, 2022, 13(4): 203. https://doi.org/10.3390/info13040203

29. Sanusi I T, Oyelere S S, Vartiainen H, et al. A systematic review of teaching and learning machine learning in K-12 education. Education and Information Technologies, 2023, 28(5): 5967–5997. https://doi.org/10.1007/s10639-022-11416-7

30. Raza A, Uddin J, Almuhaimeed A, et al. AIPs-SnTCN: Predicting anti-inflammatory peptides using fastText and transformer encoder-based hybrid word embedding with self-normalized temporal convolutional networks. Journal of chemical information and modeling, 2023, 63(21): 6537–6554. https://doi.org/10.1021/acs.jcim.3c01563 PMID: 37905969

31. Akbar S, Raza A, Al Shloul T, et al. pAtbP-EnC: identifying anti-tubercular peptides using multi-feature representation and genetic algorithm based deep ensemble model. IEEE Access, 2023. https://doi.org/10.1109/ACCESS.2023.3321100

32. Ullah M, Akbar S, Raza A, et al. DeepAVP-TPPred: identification of antiviral peptides using transformed image-based localized descriptors and binary tree growth algorithm. Bioinformatics, 2024, 40(5): btae305. https://doi.org/10.1093/bioinformatics/btae305 PMID: 38710482

33. Doshi-Velez F, Kim B. Towards a rigorous science of interpretable machine learning. arXiv preprint arXiv:1702. 08608, 2017. https://doi.org/10.48550/arXiv.1702.08608

34. Carvalho D V, Pereira E M, Cardoso J S. Machine learning interpretability: A survey on methods and metrics. Electronics, 2019, 8(8): 832. https://doi.org/10.3390/electronics8080832

35. Murdoch W J, Singh C, Kumbier K, et al. Definitions, methods, and applications in interpretable machine learning. Proceedings of the National Academy of Sciences, 2019, 116(44): 22071–22080. https://doi.org/10.1073/pnas.1900654116 PMID: 31619572

36. Goldstein A, Kapelner A, Bleich J, et al. Peeking inside the black box: Visualizing statistical learning with plots of individual conditional expectation. journal of Computational and Graphical Statistics, 2015, 24(1): 44–65. https://doi.org/10.1080/10618600.2014.907095

37. Smith G, Mansilla R, Goulding J. Model class reliance for random forests. Advances in Neural Information Processing Systems, 2020, 33: 22305–22315.

38. Plonsky L, Ghanbar H. Multiple regression in L2 research: A methodological synthesis and guide to interpreting R2 values. The Modern Language Journal, 2018, 102(4): 713–731. https://doi.org/10.1111/modl.12509

39. Lundberg S M, Erion G, Chen H, et al. From local explanations to global understanding with explainable AI for trees. Nature machine intelligence, 2020, 2(1): 56–67. https://doi.org/10.1038/s42256-019-0138-9 PMID: 32607472

40. Liu M. Chinese students' motivation to learn English at the tertiary level. Asian EFL Journal, 2007, 9(1): 126–146.

41. Jinfen Xu, Renzhong Peng, Weiping Wu. Survey and Analysis of Non-English Major College Students' Autonomous English Learning Ability. Foreign Language Teaching and Research, 2004 (01),64–68.

42. Hellas A, Ihantola P, Petersen A, et al. Predicting academic performance: a systematic literature review [C]//Proceedings companion of the 23rd annual ACM conference on innovation and technology in computer science education. 2018: 175-199.doi:10.1145/3293881.3295783.

43. Alyahyan E, Düştegör D. Predicting academic success in higher education: literature review and best practices. International Journal of Educational Technology in Higher Education, 2020, 17(1): 3. https://doi.org/10.1186/s41239-020-0177-7

44. Molnár G, Kocsis Á. Cognitive and non-cognitive predictors of academic success in higher education: a large-scale longitudinal study. Studies in Higher Education, 2023: 1–15. https://doi.org/10.1080/03075079.2023.2271513

45. Yakubu M. N., & Abubakar A. M. (2022). Yakubu M N, Abubakar A M. Applying machine learning approach to predict students' performance in higher educational institutions. Kybernetes, 2022, 51(2): 916–934. https://doi.org/10.1108/K-12-2020-0865

46. Albreiki B, Zaki N, Alashwal H. A systematic literature review of student'performance prediction using machine learning techniques. Education Sciences, 2021, 11(9): 552. https://doi.org/10.3390/educsci11090552