IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received 5 July 2023, accepted 8 August 2023, date of publication 14 August 2023, date of current version 21 August 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3305248

RESEARCH ARTICLE

A Novel Student Achievement Prediction Method Based on Deep Learning and Attention Mechanism

YU LIU, YANCHUAN HUI, DONGXU HOU, AND XIAO LIU
School of Architectural and Civil Engineering, Shenyang University, Shenyang 110044, China
Key Laboratory of Environmental Geotechnical Engineering, Shenyang University, Shenyang 110044, China

Corresponding author: Xiao Liu (xiaoliu196202@163.com)

This work was supported in part by the National Natural Science Foundation of China under Grant 52078306; in part by the Doctoral Research Initiation of Foundation of Liaoning Province, China, under Grant 2023-BS-205; and in part by the Shenyang Science and Technology Planning Project under Grant 21-108-9-21.

ABSTRACT Student achievement is an important indicator for evaluating the quality of education. It can assess development potential of students and teaching level of lecturers. Predicting student achievement is an important aspect of education data mining, which can help teachers to guide learning process of students to improve student achievement and the quality of education. Existing methods for predicting achievement less focus on the correlation between influencing factors and student achievement, and ignore the influence of different factors on student achievement. Therefore, these models cannot achieve personalized analysis and guidance for students. To address these problems, this paper proposes a student achievement prediction model based on deep learning and attention mechanism (MCAG). The proposed model can study the correlation between various factors and highlight the influence of important factors. Firstly, the correlation between influencing factors and student achievement is analyzed using the maximum information coefficient and to determine the appropriate input parameter dimensions. Then, deep learning is used to extract high-dimensional and temporal features of the data, and the attention mechanism was used to effectively identify the importance of different attribute features for grades. Finally, the model predicts the final grades based on the fused features. The prediction performance of the proposed model has been validated through experiments, and compared with other baseline models, the accuracy of the proposed MCAG model is 94.22%, which indicates that the proposed model can predict student achievement more accurately.

INDEX TERMS Achievement prediction, data mining, deep learning, attention mechanism, characteristic selection.

I. INTRODUCTION

The quality of education is the focus of attention in the field of education, and improving the quality of education and teaching has always been one of the goals pursued by educators. In educational work, the hidden internal connections and patterns in the education data can be discovered by mining massive amounts of educational data, which helps improve education quality [1]. Student achievement is an important indicator of educational quality evaluation. It is also an important indicator for assessing students' developmental potential, level of development, and performance [2]. Student achievement prediction, as an important research branch in educational data mining, based on data such as course settings, student historical performance, and student behavior, helps teachers to timely and effectively intervene and guide students' learning processes by analyzing the potential effective information in teaching data, such as identifying at risk students and providing timely intervention measures [3]. Student achievement prediction has significant value for improving the quality of education and teaching. How to predict student achievement with high accuracy is an important problem during education.

The associate editor coordinating the review of this manuscript and approving it for publication was Zijian Zhang.

Previous achievement prediction methods mainly used statistical analysis methods to fit the data and predict achievement by collecting relevant data from the education management system, student ID cards, or questionnaire surveys [4]. These methods are simple and easy to operate. However, they have some disadvantages such as low computational efficiency, few factors are considered, and not applicable to large data sets. The accuracy of the prediction is often not satisfactory. With the development of machine learning, more and more machine learning methods are used to extract data features and predict student achievement. Existing achievement prediction methods can be divided into two categories: traditional machine learning and deep learning. Traditional machine learning models, such as decision trees [5], clustering [6], and matrix decomposition [7], are used to predict student achievement. Prediction methods based on traditional machine learning have some advantages such as small computational requirements and ease of implementation. However, these methods cannot fully explore the information contained in the data. The prediction accuracy of these methods can be further improved. With the development of machine learning technology, more and more scholars are trying to use deep learning technology to predict student achievements, such as recurrent neural networks [8], spike neural networks [9], and recursive neural networks [10]. The deep learning algorithms are not only applicable to large amounts of data but also capture the complex relationship between features. The prediction accuracy of the deep learning methods is higher than traditional machine learning algorithms.

Although the aforementioned work has achieved some achievements, there are still some problems that need to be addressed. Firstly, there is insufficient research on the correlation between various factors and student achievement. Many scholars only focus on the prediction methods and use all parameters in the collected dataset as inputs of the model. They ignore the selection of input features. Secondly, the current work assumes that all factors have an equal impact on student achievement. Different factors have different impacts on student achievement. Therefore, it is necessary to analyze the influence of different attribute features on the prediction accuracy. To address these challenges, this paper proposes a novel student achievement prediction method (MCAG) based on the maximum information coefficient (MIC), convolutional neural network (CNN), attention mechanism, and gate recurrent unit (GRU) model. The proposed method adopts the MIC to analyze the correlation between parameters and selects suitable parameters as inputs of the model. The CNN-Attention module is used to extract the dimensional features of the data and assign different attention weights to different attribute features. The GRU module captures the temporal features of data and fully extracts the internal features of data. The results of the two modules are fused to obtain student achievement. Finally, experiments are conducted on the surveyed dataset, and the results demonstrate the superiority of the proposed method.

The study is organized as follows: section II presents the related work about student achievement prediction. Section III introduces the method. Section IV explains the experiments. Section V discusses the results. The conclusion is drawn in Section VI.

II. RELATED WORK

A. DEEP LEARNING

Deep learning is an emerging field in machine learning. The CNN and RNN models are representative research achievements in this field. The CNN model mainly consists of convolutional layers, pooling layers, and fully connected layers, which can explore the high-dimensional spatial features of data [11]. The RNN is particularly effective for data with sequential characteristics. It can explore the temporal and semantic information of the data. Since the emergence of the RNN model, deep learning models have made breakthroughs in solving NLP problems such as speech recognition, language modeling, machine translation, and time series analysis. Its variants LSTM and GRU can avoid the gradient vanishing and exploding problems caused by long-term sequences [12], [13], [14]. The attention mechanism was first applied to image processing, aiming to enable the model to focus on the specified target during training [15]. The core idea of the attention mechanism in deep learning is to select the most important information among numerous pieces of information for the current task and assign different weights to them according to their importance [16]. In recent years, deep learning techniques and attention mechanisms have been widely used in various fields such as image processing [17], natural language processing [18], and speech recognition [19]. Their successful applications in these fields have also provided new ideas for their research in the field of educational data mining.

B. STUDENT ACHIEVEMENT PREDICTION METHOD

Student achievement prediction methods can be divided into three categories: statistical, traditional machine learning, and deep learning methods. For statistical methods, Liu et al. used linear regression to predict course achievement based on student classroom sign in time and seat selection that reflect student learning interests, achievement motivation, and personality traits, as well as psychological test data [4]. Statistical methods are easy to conduct for scholars, however, the prediction accuracy is limited due to these methods cannot extract the complex relationship between various factors and student achievement. For traditional machine learning methods, Zhang et al. adopted naive Bayes, decision trees, multilayer perceptions, and support vector machines to predict student achievement respectively based on student historical achievement and school behavior information. The results showed that the multilayer perceptron model had a better prediction effect [10]. Frances et al. proposed a classification and clustering method for predicting achievement. The results showed that the best prediction results were obtained when academic features, behavioral features, and additional features were comprehensively considered [20]. Emma Howard et al. studied the best time to apply a student achievement prediction system in a course using different prediction methods [21]. Backenköhler et al. proposed a data analysis method to determine the best choice of courses for the next semester. They explored the effects of different strategies on student achievement using the Markov decision process [22]. Wong et al. used a mixed dataset of real and simulated data to compare various decision tree-based algorithms and found that basic training data had no obvious effect on predicting student achievement [23]. Compared with statistical methods, traditional machine learning methods can extract the more plentiful information between education data and achieve more accurate results. However, with more and more education data are collected, traditional machine learning methods cannot deal with huge amount of data. Therefore, some scholars tried to use deep learning methods to predict student achievement.

For deep learning methods, S. Abubakari et al. established a student learning performance prediction model based on a neural network algorithm. This method extracted knowledge patterns from the student dataset, and selected accuracy as the evaluation metric of model performance. The results showed that the proposed neural network model had good prediction performance, especially in social science research [24]. Li et al. proposed a student achievement prediction model based on a dual-path attention mechanism, which considered the individual differences of students. The results showed that the proposed model could predict student achievement more accurately and had good interpretability [25]. Liu et al. proposed a student achievement prediction model based on an evolutionary spiking neural network. The experimental results showed that the proposed model could effectively improve the prediction accuracy of student achievement [9]. Zhang et al. proposed a dynamic key value memory network model based on the association between exercises and knowledge concepts, which improved the interpretability of the prediction results [26]. Chen et al. considered the relationship between knowledge structures when constructing the knowledge tracking model, which solved the problem of data sparsity in knowledge tracking. The proposed model also achieved a good prediction effect [27].

Existing research about deep learning methods to predict student achievement has already achieved some achievements. However, these methods also have some limitations. For example, they ignore the selection of input parameters. When applied in student achievement prediction, it involves many input parameters and is easy to reduce the feedback effectiveness due to the calculation complexity. Investigating the inter-relationship of various input parameters and selecting the most critical input parameters to control the dimension of the input parameter is important to enhance the effectiveness of machine learning. Additionally, the above research treated all factors equally in terms of their impact on student achievement, which affect the prediction accuracy.

C. CONTRIBUTIONS

The main contributions of this paper are as follows.

(1) The correlation between various factors and student achievement is discovered through data mining and analysis. The input parameters can be optimized by correlation coefficient.

(2) The proposed deep learning model can extract the complex nonlinear relationship between various attribute features and student achievement. The proposed method can effectively capture the importance of different attribute features to student achievement, and improve the prediction accuracy.

(3) The accuracy and effectiveness of the model are verified on the surveyed dataset. The experiment results showed that the proposed model has better prediction performance.

III. METHOD

A. FRAMEWORK OVERVIEW

Figure 1 shows the proposed MCAG model. Firstly, the input parameters are divided into different input modules based on their attributes. The correlation between the internal parameters of different input modules and the output parameters is studied using the MIC algorithm. The dimensions of the input parameters are reduced based on their correlation coefficient. The newly derived inputs are concatenated to obtain the new model input. The spatial and temporal features of data are explored by two sub-models, namely the CNN-Attention model and the GRU model. Finally, the prediction results of the two sub-models are fused through a fully connected layer to obtain the final prediction result. The details of the model are shown below.

B. CORRELATION COEFFICIENT

Student achievement is affected by many factors, such as family environment, attendance rates, and physical health. Based on previous research [28], this paper primarily categorizes the factors into three types: (1) student attribute characteristics, such as gender, family background, and parental education levels; (2) course attributes, such as the scores on regular tests and scores in related courses; (3) learning behavior characteristics, such as attendance rates and learning frequency.

Many factors influence student achievement. All of them are selected as inputs of the model is unrealistic. As we all know that the degree of influence of different parameters varies in predicting performance. For example, gender has little correlation with performance, but attendance has a significant influence [29]. Therefore, this paper uses the maximum information coefficient (MIC) method to study the correlation between each parameter and student achievement and reduces the dimensionality of input parameters based on their correlation coefficient. This approach preserves essential input features and avoids overfitting during the model training process by reducing input parameter dimensions.

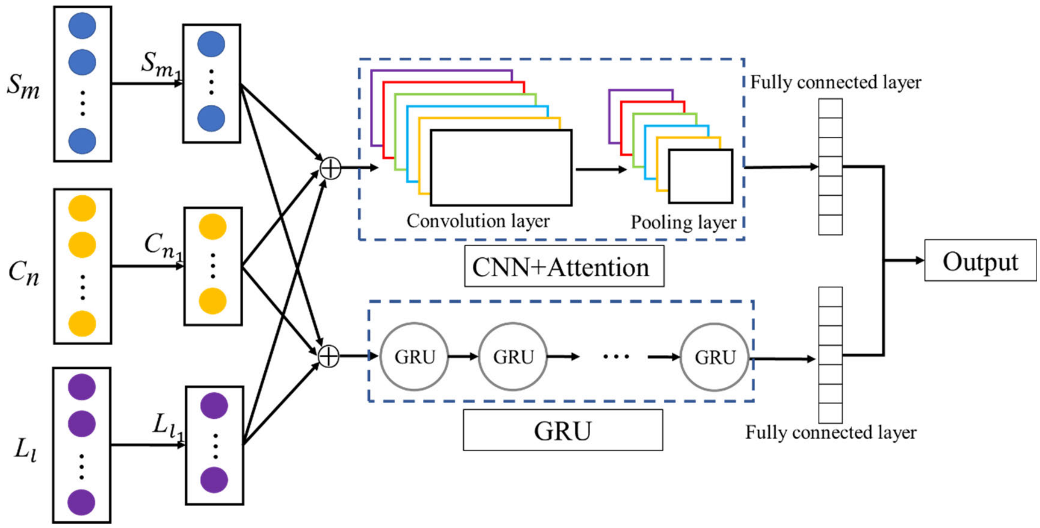


FIGURE 1. **MCAG model architecture.**

Assuming the parameter combination is $\{x_1, x_2, \ldots, x_n\}$, the correlation coefficient C is obtained using the MIC algorithm. Suitable input parameters $\{x_1, x_2, \ldots, x_t\}$, as shown in Figure 2, are selected using equation (1).

$$S = \begin{cases} 1 & C \geq \alpha \\ 0 & C < \alpha \end{cases} \tag{1}$$

where 1 represents the parameter selected as an input parameter. 0 represents the parameter that is not selected. can be determined through experiments.

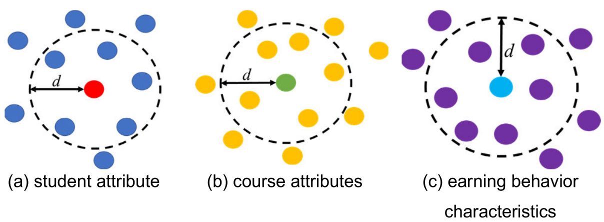


FIGURE 2. **Input parameter selection diagram.**

C. CNN-ATTENTION NETWORK

According to the above analysis, the factors can be primarily categorized into three types: student attribute characteristics, course attributes characteristics, and learning behavior characteristics. Each characteristic also includes many sub-factors. It is necessary to explore the features of multiple factors thoroughly. CNN is an important branch of neural networks and is composed of the input layer, convolutional layer, ReLU layer, pooling layer, and fully connected layer. The CNN structure contains channel domain and spatial domain, which can explore the spatial features of data through operations such as convolution and pooling. As shown in Figure 3, this paper constructs the input matrix for predicting student achievement based on the features of the CNN structure [30]. The three types of features that affect student achievement, namely student attribute characteristics, course attributes characteristics, and learning behavior characteristics, are constructed in the channel domain. In the spatial domain, it represents the student ID and sub-features of each characteristic. The advantage of constructing it this way is that it can extract the different importance of different factors.

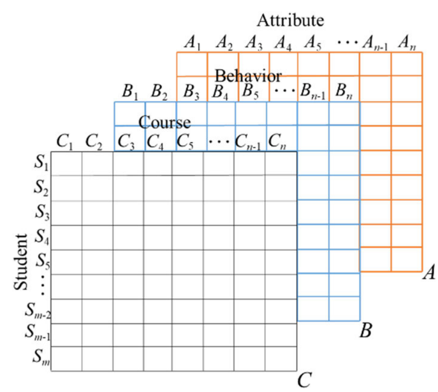


FIGURE 3. **Input matrix.**

The influence of the different characteristics on student achievement is different in the channel domain. Similarly, sub-features have different weights on student achievement in the spatial domain. To depict the effects of different factors on student achievement, this paper introduces an attention mechanism to construct a new input matrix, as shown in Figure 4. The original matrix X' is processed through a convolution operation (F_{tr}) to obtain matrix U. Then, matrix U is subjected to the maximum pooling operation (F_{sq} $(\cdot)$) to generate two compressed unit length matrices, which are $1 \times 1 \times$ C and $1 \times 1 \times$ W, respectively. Afterward, two fully connected layers (F_{ex} $(\cdot$, W$)$) transform U into two unit matrices with assigned weights. The mathematical expression

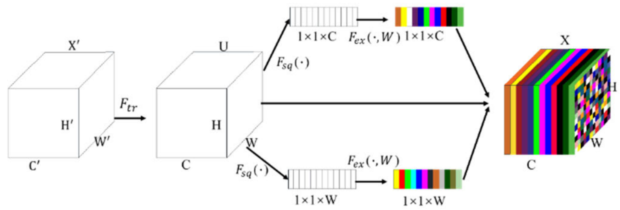


FIGURE 4. CNN-Attention structure.

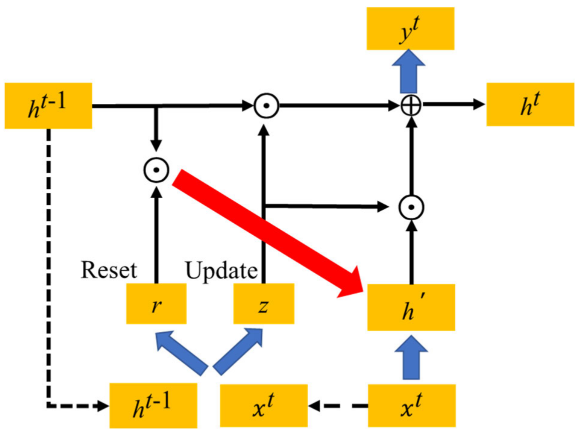


FIGURE 5. GRU structure.

of the channel domain attention mechanism is:

$$M_c(U) = \sigma(MLP(MaxPool(U))) = W_1\left(W_0\left(U^c_{\max}\right)\right) \tag{2}$$

where W_1 and W_0 are weight values.

The mathematical expression of the spatial domain attention mechanism is:

$$M_s(U) = \sigma(f(MaxPool(U))) \tag{3}$$

where $f(\bullet)$ is convolution operation.

The channel domain attention matrix and the spatial domain attention matrix are used to update the original matrix, respectively. The two updated matrices are then fused to obtain a new matrix X.

$$M(X) = M_c(U) + M_s(U) \tag{4}$$

The matrix X as the input of CNN. It is processed through fully connected layers and pooling layers to extract data characteristics.

$$Y = f(W_{CNN} * X + b_{CNN}) \tag{5}$$

where f is activation function; W_{CNN} is weight matrix; b_{CNN} is bias matrix.

D. GRU NETWORK

Student achievement prediction can be regarded as a temporal process. For example, a student's study time each day has a significant impact on their grade. A student who studies every day will typically gain more knowledge and perform better. Recurrent neural networks (RNNs) can effectively explore the temporal features of data and improve prediction accuracy. RNN is a common type of recurrent neural network, but it has problems of gradient vanishing and explosion due to its network structure. Subsequently, some variations of RNNs have been developed, such as LSTM and GRU. GRU is an improvement of the LSTM. GRU combines the forget gate and input gate into an update gate and combines the memory cell with the hidden layer into a reset gate, which simplifies the operation and enhances performance. A large number of experiments have shown that GRU has a simpler structure and fewer parameters than LSTM but has no significant difference in performance or even better performance in some tasks. Therefore, this paper chooses the GRU model to extract the temporal characteristic of data. Its primary structure is shown in the Figure 5 below.

Calculation process of GRU is expressed as below:

$$z_u = \sigma\left(w_u * \left[h^{t-1}, x^t\right] + b_u\right) \tag{6}$$

$$z^r = \sigma\left(w_i * \left[h^{t-1}, x^t\right] + b_r\right) \tag{7}$$

$$z = \sigma\left(w * \left[h^{t-1} \odot z^r, x^t\right] + b\right) \tag{8}$$

$$h^t = z^u \odot c^{t-1} + \left(1 - z^u\right) \odot z \tag{9}$$

where h^{t-1} is the hidden at the previous moment; x^t is the input, w_u, w_i, w are the weight matrices of the reset gate z^r, updated gate z_u, and hidden layer, respectively; b_u, b_r and b are the bias vectors.

Output of the GRU:

$$y^t = \sigma\left(w' h^t\right) \tag{10}$$

Output of the MCAG model:

$$Output = f(FC(Y), FC(y)) \tag{11}$$

IV. EXPERIMENTS

A. DATASET ACQUISITION AND ANALYSIS

The dataset used in this paper is from a survey of sophomore, junior and senior students at a university in Nanjing, China. A total of 5678 students participated in the survey. The contents are divided into three categories: student attribute characteristics, course attributes characteristics, and student learning behavior characteristics. Each category contained six items, as shown in Table 1. Since the questionnaire contained non-numerical data, non-numerical data were converted to numerical data in advance to facilitate the experiment, as shown in Table 1. The questionnaires were sorted and screened to remove missing and abnormal questionnaires, and a total of 5420 valid data were obtained.

B. MODEL ESTABLISHMENT AND GYPERPARAMETER

This section mainly shows the process of constructing the model. Firstly, the maximum information coefficient (MIC) was used to study the correlation between each parameter and student achievement. The dimensions of input parameters were reduced based on their correlation coefficients. Then, the CNN-Attention module was used to explore the

TABLE 1. Statistical description of key parameters.

	Parameters	Features	Description	Digital expressing
Input	Student Attribute Characteristics	Gender	Man/Woman	1, 0
		Age	17~22	17, 18, 19, 20, 21, 22
		Parents' education level	High school and below/Undergraduate/Graduate	1, 2, 3
		Parents' marital status	Normal/Single-parent	1, 2
		Home address	Country/Urban	1, 2
		Physical health	Healthy/Ill	1, 2
	Course Attributes Features	Accompanying test results	0~100	0~100
		Number of class hours	One year/One semester/Other	1, 2, 3
		Time period of class	Morning/Afternoon/Evening	1, 2, 3
		Whether there is a midterm examination	Yes/No	1, 2
		Teaching language	Chinese/Other languages	1, 2
		Assessment method	Exam/Non-exam	1, 2
	Learning behavior characteristics	Attendance rate	0~100	0~100
		Learning frequency	High/Medium/Low	1, 2, 3
		Learning method	Self-study/Cooperation	1, 2
		Psychological quality	Strong/Medium/Weak	1, 2, 3
		Time management skills	Strong/Medium/Weak	1, 2, 3
		Classroom performance	Active/Quiet	1, 2
Output		Student achievement	Excellent/Good/Average/Poor	1, 0.8, 0.6, 0.4

high-dimensional features of the data and the different importance of these features, while the GRU model was used to explore the temporal features of the data. Finally, the results obtained from the two modules were fused to obtain the final prediction results.

Hyperparameters can directly affect the prediction performance of the neural network. Its optimal value may vary with specific networks and application areas and should be determined by experiments. Optimal hyperparameters of the proposed MCAG model are obtained by repeated experiments, as shown in Table 2.

Max-min normalization method is used to normalize the input data to facilitate the training process, as shown in Eq. (12).

$$x' = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \tag{12}$$

where x is the measured value, $x_{\max}$ and $x_{\min}$ are the maximum and minimum of measured values; x' is the normalized value.

TABLE 2. Optimal hyperparameters.

Model	Hyperparameters
MCAG	Convolution layer= 2； Pooling layer= 2, Kernel_size = 4, Dense = 1, Each batch = 128, Iteration = 200, Learning rate = 0.0001, Dropout rate = 0.5, Optimizer = Adam, Loss function= MSE, Activation function = ReLu

TABLE 3. Experiment configuration.

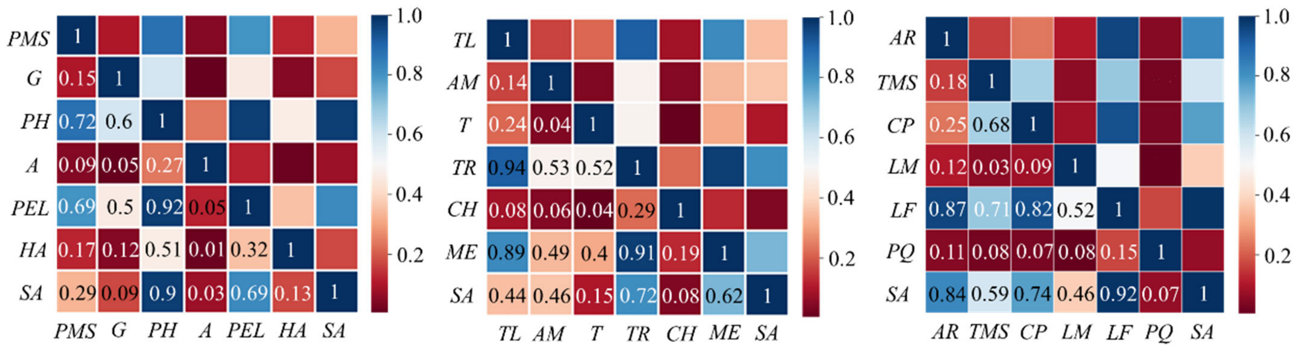
Parameters	Notes
Operation system	Windows 10
CPU	Intel(R) Core (TM) i7-10710U CPU @ 1.10GHz
GPU	GeForce MX350
Python	3.9
PyTorch	1.9.1

C. EXPERIMENT ENVIRONMENT

The hardware and software on the computer used in the experiment are listed in Table 3.

D. METRICS

In this paper, in addition to the prediction accuracy, precision, recall, and F1-Measure are also selected for evaluating the performance of model classification prediction. Accuracy represents the proportion of correctly classified samples in the entire sample set. The higher the accuracy, the more accurate the prediction. Recall represents the proportion of correctly classified positive samples to the total number of predicted positive samples. Precision represents the proportion of correctly classified positive samples to the total number of positive samples predicted by the model. F1-Measure is a compromise between precision and recall.



(a) Student attribute (b) Course attribute (c) Learning behavior

FIGURE 6. Correlation coefficient.

The higher the F1-Measure value, the better the classification performance [31].

$$Accuracy = \frac{TP + TN}{TP + TN + FN + FP} \tag{13}$$

$$Precision = \frac{TP}{TP + FP} \tag{14}$$

$$Recall = \frac{TP}{TP + FN} \tag{15}$$

$$F1 - Measure = \frac{2 \times Precision \times Recall}{Precision + Recall} \tag{16}$$

where *TP* represents the samples that are truly positive and correctly identified as positive. *FP* represents the samples that are truly negative but mistakenly identified as positive. *FN* represents the samples that are truly positive but not correctly identified as positive. *TN* represents the samples that are truly negative and not identified as positive.

V. RESULTS AND DISCUSSIONS

In this section, the prediction performance of the proposed model is evaluated through experiments.

A. CORRELATION ANALYSIS OF PARAMETERS

In this section, the maximum information coefficient (MIC) method is used to study the correlation between each factor and student achievement. Dimension reduction is conducted based on the size of the correlation coefficient. The experiment is carried out from three aspects: student attribute characteristics, course attribute characteristics, and student learning behavior characteristics. The results are shown in Figure 6.Figure 6 shows the correlation between each factor and student achievement in each heatmap. It can be seen that there are differences in the correlation between each factor and student achievement on each heatmap. By comparing the three heatmaps, it is found that the correlation coefficient between student learning behavior features and student achievement is the largest, while that of student attribute features is the smallest. Therefore, the three factors that affect student achievement from largest to smallest are student learning behavior characteristics, course attribute characteristics, and student attribute characteristics.

For student attribute characteristics, the correlation coefficient between student physical health status and student achievement is the largest, indicating that student physical health status has the greatest impact on student achievement. The correlation coefficient between age, gender, and student achievement is the smallest, indicating that these two factors have the smallest impact on student achievement. For course attribute characteristics, the correlation coefficient between test scores and student achievement is the largest, indicating that regular in class tests are beneficial to improve student achievement. The correlation coefficient between course hours and student achievement is the smallest, indicating that the impact of course hours on student achievement can be ignored. For student learning behavior characteristics, the correlation coefficient between learning frequency and student achievement is the largest, indicating that maintaining a good learning frequency can help students acquire the knowledge and achieve good learning performance. The correlation coefficient between mental quality and student achievement is the smallest.

When the correlation coefficient is less than 0.4, the two parameters can be considered approximately unrelated [32]. Therefore, dimension reduction can be performed on the parameters, and the selected parameters are physical health status, parental education level, teaching language, assessment method, in class test scores, whether there is a midterm exam, attendance rate, time management ability, classroom performance, learning methods, and learning frequency.

where PMS is parents' marital status, G is gender, PH is physical health, A is age, PEL is parents' education level, HA is home address, TL is teaching language, AM is assessment method, T is time period of class, TR is accompanying test results, CH is number of class hours, ME is whether there is a midterm examination, AR is attendance rate, TMS is time management skills, CP is classroom performance, LM is learning method, LF is learning frequency, PQ is psychological quality, SA is student achievement, respectively.

Since the input parameters are not independent of each other, it can cause information redundancy and lead to model overfitting during training [33]. Therefore, this paper adopts the recursive feature elimination method to further simplify the input parameters and select appropriate parameter dimensions. The basic steps are as follows: first, all parameters as input of the model to obtain the accuracy of the model. Then, the feature with the lowest performance evaluation score is selected and removed from the feature set. Train the model again and calculate the accuracy of the model. Repeat steps 2 and 3 until no more features can be removed. The results are shown in Figure 7. It can be seen that the model performance reaches its best when the number of input features is 9. Therefore, the selected input parameters are physical health status, parental education level, assessment method, in class test scores, attendance rate, time management ability, classroom performance, learning methods, and learning frequency.

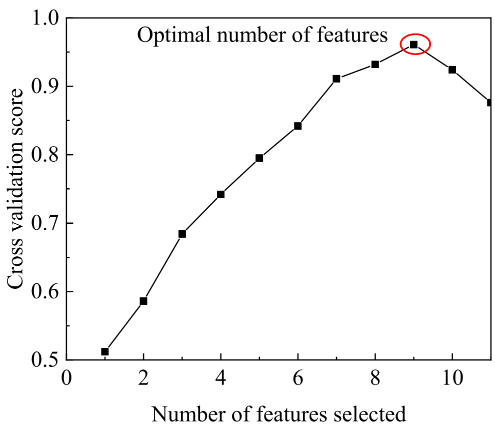


FIGURE 7. Cross validation scores of different features.

B. COMPARISON WITH OTHER MODELS

To verify the effectiveness of the proposed model, the proposed MCAG model is compared with four baseline models. They are Decision Tree [34], Support vector machine (SVM) [35], eXtreme Gradient Boosting (XGBoost) [36], and Back Propagation Neural Network (BPNN) [37]. Experiment results are presented in Table 4, while the histograms are in Figure 8.

As shown in Table 4, the accuracy, precision, recall, and F1-measure in the datasets are 94.22%, 92.85%, 95.53%, and 94.89% for the MCAG algorithm, 80.32%, 79.98%, 81.95% and 80.27% for BPNN, 87.38%, 89.56%, 91.71% and 89.61% for XGBoost, 82.01%, 86.33%, 84.62% and 87.02% for SVM, 85.43%, 87.91%, 90.51% and 88.69% for Decision Tree, respectively. Compared with the other four baseline models for achievement prediction, the proposed method based on deep learning in this paper achieved the best prediction results in four evaluation metrics. This is because baseline models did not extract more feature information specifically for the target grade data and treat the impact of each attribute feature on grades equally. However, the attention mechanism introduced in this paper can distinguish the different importance of each attribute feature and explore the impact of different features on grades. In addition, the proposed model in this paper also considers the temporal changes of each parameter and introduces the GRU model to capture the temporal features of the data. Therefore, the proposed model can explore more hidden information and greatly improve the prediction ability of the model. The experimental results also prove the effectiveness of the proposed method.

Figure 8 shows the improvement of the proposed model compared with each baseline model. It can be seen intuitively from the figure that the proposed model obtained significant improvements in four indicators, especially in accuracy. Therefore, it can be proved that the proposed model can effectively predict student achievement.

TABLE 4. Prediction results of different models (%).

Models	Accuracy	Precision	Recall	F1-Measure
Decision Tree	85.43	87.91	90.51	88.69
SVM	82.01	86.33	84.62	87.02
XGBoost	87.38	89.56	91.71	89.61
BPNN	80.32	79.98	81.95	80.27
MCAG	94.22	92.85	95.53	94.89

C. ABLATION EXPERIMENTS

The proposed algorithm mainly consists of CNN and GRU. Ablation experiments are conducted to verify the prediction performance of different parts. We choose three sub-models for the experiments.

- CNN-attention: The proposed method only comprises attention mechanism and CNN algorithms.
- GRU: The proposed method only comprises GRU algorithms.
- CNN-GRU: The proposed method comprises CNN and GRU algorithms.

The experimental results are shown in Table 5 and Figure 9. It can be seen that the proposed model achieved the highest values in accuracy, precision, recall, and F1-measure, indicating that the proposed model has the best prediction performance. Compared with the CNN-GRU model, the proposed model improved by 1.67%, 2.16%, 1.89%, and 2.45% in the four metrics of accuracy, precision, recall, and F1-measure respectively. The reason is that the attention mechanism in the proposed model can consider the different importance of different parameters. The MCAG model can more fully explore the internal features of the data and improve the prediction accuracy of the model. Compared with the CNN-Attention model, the proposed model improved by 4.68% in accuracy, 4.13% in precision, 5.14% in recall, and 5.63% in F1-measure. The reason is that the GRU module in the proposed model considers the temporal features of the data, such as student achievement being closely related to the

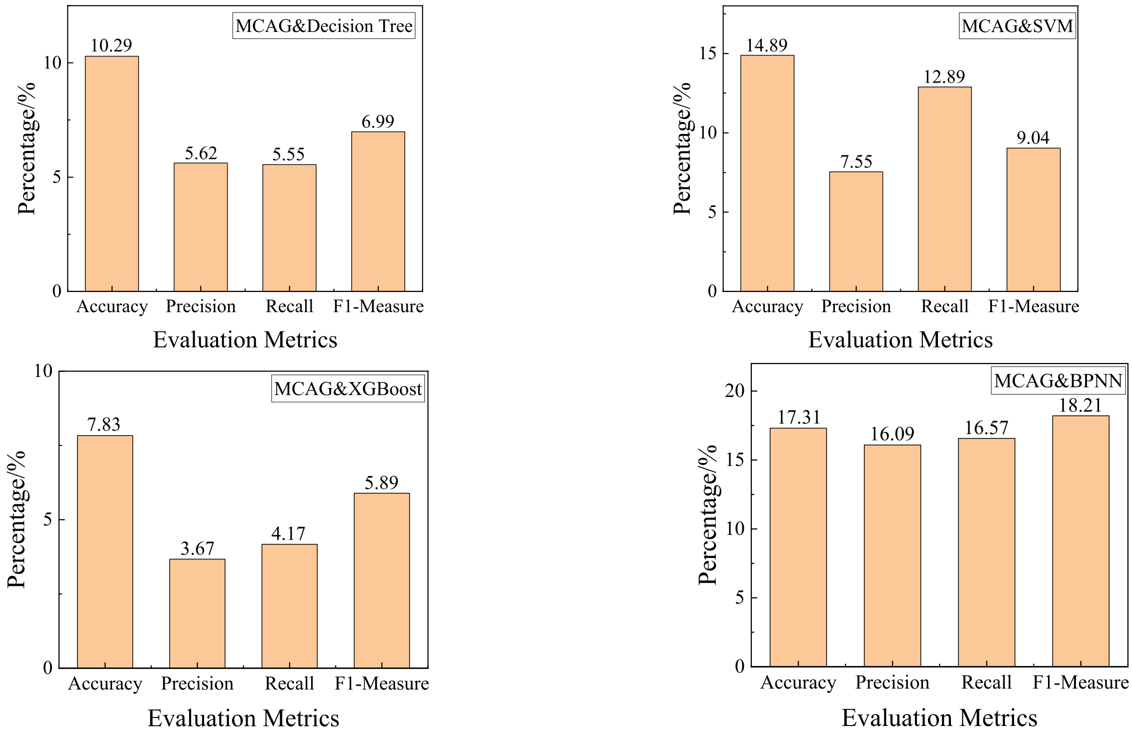


FIGURE 8. Comparison results of models.

TABLE 5. Prediction results of ablation experiments.

Models	Accuracy	Precision	Recall	F1-Measure
CNN-Attention	90.01	89.17	90.86	89.83
GRU	87.83	88.69	89.91	88.66
CNN-GRU	92.68	90.89	93.76	92.62
MCAG	94.22	92.85	95.53	94.89

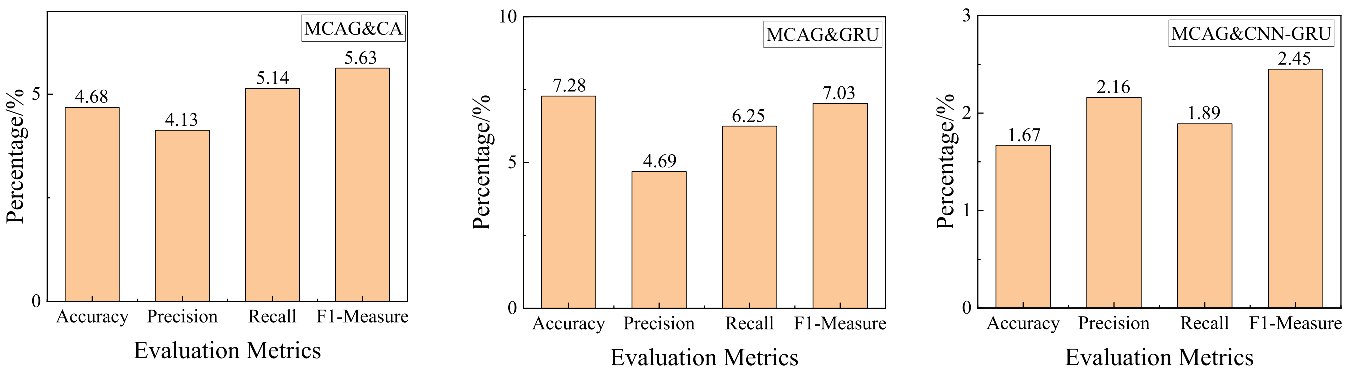


FIGURE 9. Comparison results of models.

accumulation of their study time. Therefore, the MCAG model captures data features more comprehensively and the prediction results are better. Compared with the GRU model, the proposed model improved by 7.28% in accuracy, 4.69% in precision, 6.25% in recall, and 7.03% in F1-measure. The reason is that the CNN-Attention module in the proposed model can extract the dimensional features of the data and distinguish the different importance, which can improve the prediction performance of the model. In conclusion, the proposed MCAG model can more accurately capture the changing features of the data and effectively predict student achievement.

VI. CONCLUSION AND FUTURE WORK

Student achievement prediction is a research hotspot in the field of educational data mining in recent years. It is also an important goal for learning analytics. Because current related research has not considered that different factors have different effects on student achievement, and that student achievement can be regarded as temporal data, this paper proposes a student achievement prediction model based on MIC, CNN, attention mechanism, and GRU. The following conclusions can be drawn.

(1) The order of correlation of the parameters with student achievement from largest to smallest is student learning behavior characteristics, course attribute characteristics, and student attribute characteristics. It indicates that we should focus on guiding students to develop good learning behavior.

(2) Compared with the existing baseline models, the proposed MCAG model not only explores temporal and dimension characteristics of data but also further portrays the influence of different parameters on the prediction results.

(3) The experimental results demonstrate that the values of the evaluation metric of the proposed MCCB model are larger than the four baseline models and three sub-models, indicating that the proposed model outperforms the four baseline models and two sub-models, the proposed model can predict student achievement more accurately.

In future work, we plan to consider more related factors to further demonstrate the prediction performance of the proposed MCAG model. Additionally, we aim to explore more effective deep learning architectures to improve the accuracy of student achievement prediction.

DECLARATION OF COMPETING INTEREST

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this article.

REFERENCES

[1] P. Gou, "Teaching English using mobile applications to improve academic performance and language proficiency of college students," *Educ. Inf. Technol.*, pp. 1–15, May 2023, doi: 10.1007/s10639-023-11864-9.

[2] J. Chaparro-Peláez, S. Iglesias-Pradas, F. J. Rodríguez-Sedano, and E. Acquila-Natale, "Extraction, processing and visualization of peer assessment data in moodle," *Appl. Sci.*, vol. 10, no. 1, p. 163, Dec. 2019, doi: 10.3390/app10010163.

[3] H. Li, H. Li, S. Zhang, Z. Zhong, and J. Cheng, "Intelligent learning system based on personalized recommendation technology," *Neural Comput. Appl.*, vol. 31, no. 9, pp. 4455–4462, Sep. 2019, doi: 10.1007/s00521-018-3510-5.

[4] A. Namoun and A. Alshanqiti, "Predicting student performance using data mining and learning analytics techniques: A systematic literature review," *Appl. Sci.*, vol. 11, no. 1, p. 237, Dec. 2020, doi: 10.3390/app11010237.

[5] M. Pandey and V. Kumar Sharma, "A decision tree algorithm pertaining to the student performance analysis and prediction," *Int. J. Comput. Appl.*, vol. 61, no. 13, pp. 1–5, Jan. 2013, doi: 10.5120/9985-4822.

[6] X. Zhang, R. Xue, B. Liu, W. Lu, and Y. Zhang, "Grade prediction of student academic performance with multiple classification models," in *Proc. 14th Int. Conf. Natural Comput., Fuzzy Syst. Knowl. Discovery (ICNC-FSKD)*, Huangshan, China, Jul. 2018, pp. 1086–1090, doi: 10.1109/FSKD.2018.8687286.

[7] Z. Ren, X. Ning, and H. Rangwala, "Grade prediction with temporal course-wise influence," 2017, *arXiv:1709.05433*.

[8] B.-H. Kim, E. Vizitei, and V. Ganapathi, "GritNet: Student performance prediction with deep learning," 2018, *arXiv:1804.07405*.

[9] C. Liu, H. Wang, Y. Du, and Z. Yuan, "A predictive model for student achievement using spiking neural networks based on educational data," *Appl. Sci.*, vol. 12, no. 8, p. 3841, Apr. 2022, doi: 10.3390/app12083841.

[10] L. Zhang, X. Xiong, S. Zhao, A. Botelho, and N. T. Heffernan, "Incorporating rich features into deep knowledge tracing," in *Proc. 4th ACM Conf. Learn. Scale*, Apr. 2017, pp. 169–172, doi: 10.1145/3051457.3053976.

[11] C. Zhou, H. Xu, L. Ding, L. Wei, and Y. Zhou, "Dynamic prediction for attitude and position in shield tunneling: A deep learning method," *Autom. Construction*, vol. 105, Sep. 2019, Art. no. 102840, doi: 10.1016/j.autcon.2019.102840.

[12] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, Nov. 1997, doi: 10.1162/neco.1997.9.8.1735.

[13] L. Yu, J. Qu, F. Gao, and Y. Tian, "A novel hierarchical algorithm for bearing fault diagnosis based on stacked LSTM," *Shock Vibrat.*, vol. 2019, pp. 1–10, Jan. 2019, doi: 10.1155/2019/2756284.

[14] J. Zhang, P. Wang, R. Yan, and R. X. Gao, "Long short-term memory for machine remaining life prediction," *J. Manuf. Syst.*, vol. 48, pp. 78–86, Jul. 2018, doi: 10.1016/j.jmsy.2018.05.011.

[15] V. Mnih, N. Heess, and A. Graves, "Recurrent models of visual attention," in *Advances in Neural Information Processing Systems*. Cambridge, MA, USA: MIT Press, 2014, pp. 2204–2212, doi: 10.48550/arXiv.1406.6247.

[16] K. Xu, J. Ba, R. Kiros, K. Cho, A. Courville, R. Salakhutdinov, R. Zemel, and Y. Bengio, "Show, attend and tell: Neural image caption generation with visual attention," 2015, *arXiv:1502.03044*.

[17] L. Chen, H. W. Zhang, J. Xiao, L. Q. Nie, J. Shao, W. Liu, and T. S. Chua, "SCA-CNN: Spatial and channel-wise attention in convolutional networks for image captioning," in *Proc. 30th IEEE Conf Comput. Vis. Pattern Recognit. (CVPR)*. Piscataway, NJ, USA, Jul. 2017, pp. 5659–5667, doi: 10.48550/arXiv.1611.05594.

[18] A. M. Rush, S. Chopra, and J. Weston, "A neural attention model for abstractive sentence summarization," 2015, *arXiv:1509.00685*.

[19] D. Bahdanau, J. Chorowski, D. Serdyuk, P. Brakel, and Y. Bengio, "End-to-end attention-based large vocabulary speech recognition," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Piscataway, NJ, USA, Mar. 2016, pp. 4945–4949, doi: 10.48550/arXiv.1508.04395.

[20] B. K. Francis and S. S. Babu, "Predicting academic performance of students using a hybrid data mining approach," *J. Med. Syst.*, vol. 43, no. 6, p. 162, Jun. 2019, doi: 10.1007/s10916-019-1295-4.

[21] E. Howard, M. Meehan, and A. Parnell, "Contrasting prediction methods for early warning systems at undergraduate level," *Internet Higher Educ.*, vol. 37, pp. 66–75, Apr. 2018, doi: 10.1016/j.iheduc.2018.02.001.

[22] M. Backenkohler and V. Wolf, "Student performance prediction and optimal course selection: An MDP approach," in *Software Engineering and Formal Methods*. Cham, Switzerland: Springer, 2018, pp. 40–47, doi: 10.1007/978-3-319-74781-1_3.

[23] J. C. F. Wong and T. C. Y. Yip, "Measuring students' academic performance through educational data mining," *Int. J. Inf. Educ. Technol.*, vol. 10, no. 11, pp. 797–804, 2020, doi: 10.18178/ijiet.2020.10.11.1461.

[24] M. S. Abubakari, F. Arifin, and G. G. Hungilo, "Predicting students' academic performance in educational data mining based on deep learning using TensorFlow," *Int. J. Educ. Manage. Eng.*, vol. 10, no. 6, pp. 27–33, Dec. 2020, doi: 10.5815/ijeme.2020.06.04.

[25] M. Y. Li, X. D. Wang, S. L. Ruan, K. Zhang, and Q. Liu, "Student performance prediction model based on two-way attention mechanism," *J. Comput. Res. Develop.*, vol. 57, no. 8, pp. 1729–1740, 2020. [Online]. Available: https://kns.cnki.net/kcms/detail/detail.aspx?FileName=JFYZ202008020&DbName=CJFQ2020

[26] J. N. Zhang, X. J. Shi, I. King, and D. Y. Yeung, "Dynamic key-value memory networks for knowledge tracing," in *Proc. 26th Int. Conf. World Wide Web*, 2017, pp. 765–774, doi: 10.48550/arXiv.1611.08108.

[27] P. Chen, Y. Lu, V. W. Zheng, and Y. Pian, "Prerequisite-driven deep knowledge tracing," in *Proc. IEEE Int. Conf. Data Mining (ICDM)*, Nov. 2018, pp. 39–48, doi: 10.1109/ICDM.2018.00019.

[28] Y. Jin, "Predicting university students' academic performance based on educational data," Southeast Univ., Dhaka, Bangladesh, Tech. Rep., 2021, doi: 10.27014/d.cnki.gdnau.2021.002318.

[29] T. Fang, "Research on student achievement prediction model based on ensemble learning algorithm," Guilin Univ. Electron. Technol., Guilin, China, Tech. Rep., 2022, doi: 10.27049/d.cnki.ggldc.2022.000645.

[30] C. Chen, B. Wu, P. Jia, and Z. Wang, "A novel hybrid deep neural network prediction model for shield tunneling machine thrust," *IEEE Access*, vol. 10, pp. 123858–123873, 2022, doi: 10.1109/ACCESS.2022.3224184.

[31] B. Wu, "Research and application of grade prediction model based on decision tree algorithm," Xi'an Univ. Technol., Xi'an, China, Tech. Rep., 2019. [Online]. Available: https://kns.cnki.net/KCMS/detail/detail.aspx?dbname=CMFD201902&filename=1019859395.nh

[32] T. K. Koo and M. Y. Li, "A guideline of selecting and reporting intraclass correlation coefficients for reliability research," *J. Chiropractic Med.*, vol. 15, no. 2, pp. 155–163, Jun. 2016, doi: 10.1016/j.jcm.2016.02.012.

[33] L. Yu and H. Liu, "Efficient feature selection via analysis of relevance and redundancy," *J. Mach. Learn. Res.*, vol. 5, pp. 1205–1224, Apr. 2016, doi: 10.1023/B:JODS.0000045365.56394.b4.

[34] J. L. Zhu, Y. L. Kang, R. B. Zhu, and S. B. Liu, "College academic achievement early warning prediction based on decision tree model," in *Proc. 7th CCF Conf. Big Data* Berlin, Germany: Springer, 2019, pp. 351–365, doi: 10.1007/978-981-15-1899-7_25.

[35] I. Burman and S. Som, "Predicting students academic performance using support vector machine," in *Proc. Amity Int. Conf. Artif. Intell. (AICAI)*, Piscataway, NJ, USA, Feb. 2019, pp. 756–759, doi: 10.1109/AICAI.2019.8701260.

[36] T. Q. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*. New York, NY, USA: ACM, 2016, pp. 785–794, doi: 10.48550/arXiv.1603.02754.

[37] A. Mueen, B. Zafar, and U. Manzoor, "Modeling and predicting students' academic performance using data mining techniques," *Int. J. Modern Educ. Comput. Sci.*, vol. 8, no. 11, pp. 36–42, Nov. 2016, doi: 10.5815/ijmecs.2016.11.05.

YU LIU received the M.Sc. and Ph.D. degrees in geotechnical engineering from Northeastern University, China, in 2012 and 2016, respectively. He is currently an Associate Professor with the School of Architectural and Civil Engineering, Shenyang University, China. His current research interests include underground engineering, machine learning, and education.

YANCHUAN HUI received the Ph.D. degree in mechanics from Wuhan University, China, in 2019. He is currently an Associate Professor with the School of Architectural and Civil Engineering, Shenyang University, China. His current research interests include artificial intelligence and machine learning.

DONGXU HOU received the M.Sc. degree in constructional engineering from Shenyang Jianzhu University, China, in 2009, and the Ph.D. degree in constructional engineering from the Dalian University of Technology, China, in 2015. He is currently an Associate Professor with the School of Architectural and Civil Engineering, Shenyang University, China. His current research interests include underground engineering and education.

XIAO LIU received the Ph.D. degree in constructional engineering from Northeastern University, China, in 2008. She is currently a Professor with the School of Architectural and Civil Engineering, Shenyang University, China. Her current research interests include construction engineering and machine learning.