All Element Selection Method in Classroom Social Networks and Analysis of Structural Characteristics

Zhaoyu Shou[1], Zhe Zhang[2], Jingquan Chen[3], Hua Yuan[4], Jianwen Mo[5]
School of Information and Communication, Guilin University of Electronic Technology, Guilin 541004, China[1, 2, 3, 4, 5]
Guangxi Wireless Broadband Communication and Signal Processing Key Laboratory[1]

***Abstract*—To deeply investigate the complex relationship between learners' structural characteristics in classroom social networks and the dynamics of learning emotions in smart teaching environments, an innovatively improved RP-GA. All Element Selection Method based on genetic algorithm is proposed. The method calculates the importance of factors based on the random forest model and guides the population initialization together with random numbers to achieve the differentiation and efficiency of factor selection; and utilized the Partial Least Squares regression model in conjunction with a cross-validation optimization model to enhance the accuracy of fitness evaluation, efficiently tackling the issues of premature convergence and low prediction accuracy inherent in traditional genetic algorithms for factor selection. Based on this method, the elements affecting learning emotions are precisely screened, and the intrinsic links between elemental changes and structural properties are deeply analyzed. Experiments show that RP-GA selects a small and efficient number of key elements on public datasets and significantly improves the prediction performance of classifiers such as SVM, NB, MLP, and RF. The proposed learning sentiment all-essential selection method provides effective conditions for classroom network structure characterization and future learning sentiment computation.**

Keywords*—*Genetic algorithms; element selection; random forest; partial least squares; classroom network

I. Introduction

With the integration and development of teaching practices and emerging information technologies, smart teaching has become an important driver of global education reform and has enabled interactive learning environments, data-driven decision support, and personalized learning experiences, among others, fundamentally reshaping the process of teaching and learning [1]. As an important carrier of information mapping the dimensions of student interactions, learning effectiveness, cognitive processes, and affective learning, the interaction mechanism between its structural properties and student performance has become a hot research topic [2]. Learning emotion, which refers to the learning-related emotional experiences that students produce in intelligent teaching and learning scenarios, covers the basic emotional states (e.g., concentration, confusion, anxiety, excitement, etc.) of the learning process. It is the emotional experience that accompanies students in the cognitive process related to learning content comprehension and knowledge construction, as well as the social-emotional responses generated in teaching activities such as teacher-student interaction and student-student interaction. These rich and diverse learning emotions do not exist in isolation, but are intertwined and influenced by each other with the structural characteristics of the classroom social network. Acquiring heterogeneous information data on students' academic performance, learning behaviors and social relationships through the smart teaching classroom provides a strong data foundation for analyzing the mutual influence relationship between learning emotions and the structural characteristics of classroom social networks, and lays a solid theoretical foundation and scientific basis for the computation and dissemination of learning emotions in the smart classroom scenario [3,4]. Classroom social networks with positivity can facilitate the transmission of positive emotions, while networks full of negativity may lead to an increase in negative emotions [5] The computation and transmission of learning emotions depend on the structural properties of classroom networks, and the structural properties of classroom networks are dynamically influenced by a variety of factors. Therefore, it is important to select the elements that have a significant effect on learning emotions to construct the classroom network, and then analyze the relationship between the changes in the elements and the structural characteristics of the network.

However, existing research on classroom social networks is constructed based on a series of established factors and provides only a preliminary analysis of their network structural characteristics [6-8]. But it failed to delve into how multiple factors affect the dynamics of classroom social networks, as well as how the computation and dissemination of learning emotions intertwine and interact with the structural properties of classroom social networks, a dynamic mechanism that is still under-researched in the existing research. In order to deeply explore the intrinsic connection between the structural characteristics of classroom social networks and learning emotions, this paper proposes a method called Random Partial-Genetic algorithm (RP-GA) for the selection of all elements of classroom social networks. In order to verify the reliability and validity, six public datasets were selected for systematic experiments, and the performance of the RP-GA algorithm on different datasets was comprehensively evaluated by using scientifically reasonable evaluation indexes and strictly controlling experimental variables. Meanwhile, in order to fit the actual needs of smart teaching scenarios, this study specifically collects the SCLE-Dataset of students' learning emotions under smart classrooms. This dataset covers multi-dimensional information of students in the learning process, including but not limited to academic performance, classroom interaction behavior and so on. The RP-GA algorithm is used to mine and analyze the SCLE-Dataset in depth, and the elements that have a significant impact on learning emotions are selected comprehensively and meticulously. Based on the selected elements, an all-element classroom network model is constructed, which fully considers the dynamic coupling mechanism between the elements and the structural characteristics of the classroom network. Using graph theory, complex networks and other theories and methods, we deeply analyze the dynamic change law between the elements and the network structure characteristics. The results of this research provide solid theoretical support for the future accurate calculation of the propagation of learning emotions in the classroom network, help educators better understand the process of the generation and development of students' learning emotions, and provide a scientific basis for optimizing teaching strategies and creating a positive learning atmosphere.

The main contributions of this study can be summarized in the following two points:

1) An innovative approach to full-factor selection is proposed in RP-GA. The RP-GA algorithm is an innovative extension of genetic algorithm (GA), based on the innovative improvements of factor importance and random number co-guided population initialization and R^2-based PLS regression fitness assessment, which significantly improves the search efficiency and predictive accuracy of the classifier compared with the existing algorithms. In classroom social network-specific scenarios, RP-GA is able to traverse and evaluate numerous factors affecting learning emotions more efficiently, providing a powerful tool for accurately identifying the elements affecting learning emotions.

2) Construct a multi-layer heterogeneous classroom network model based on all elements, while analyzing key structural properties such as network diameter, average degree, clustering coefficient and so on. In-depth exploration of the role of the mechanism that influences the dynamic impact of changes in the emotional elements of learning on the structural characteristics of the network. The construction and analysis of the all-element classroom network model not only provides a solid theoretical foundation for the effective use of network structural characteristics to guide and regulate the learning affective state in the field of educational technology, but also provides a scientific basis for future research on learning affective computation and dissemination.

II. Related Work

In smart teaching scenarios, educational researchers are faced with the challenge of accurately screening out the elements affecting learning emotions from massive data [9]. As a powerful global optimization search algorithm [10], genetic algorithm shows great potential in model optimization and parameter tuning with its robustness, adaptivity and parallel processing capability. In recent years, a variety of improved algorithms based on genetic algorithms have shown remarkable achievements in the field of factor selection research. Izabela et al. [11] proposed a Genetic Algorithm with Aggressive Mutation and Reduced Factors (GAAMmf), which scales down the number of factors while performing the factor selection. Aram et al. [12] proposed an Alternating Sorting Method Genetic Algorithm (ASMGA), which combines genetic algorithm and maximum bounded factor selection in a hybrid packing-filtering algorithm. Deng et al. [13] proposed a Factor Threshold Guided Genetic Algorithm (FTGGA), which first applies the ReliefF algorithm to filter the redundant factors, and then further evaluates the retained subset of factors by FTGGA, which exhibits higher classification accuracy and a smaller subset of factors on a 12-gene microarray dataset. However, these methods mainly focus on the optimization of the algorithm itself, ignoring the important impact of factor quality on algorithm performance. Studies have shown that factor selection and factor importance assessment are critical to the performance of genetic algorithms, and researchers have begun to explore factor importance assessment methods in depth, providing an effective means of assessing the contribution of factor importance to the predictive power of a model. Razmjoo et al. [14] proposed two incremental ranking methods for factors for classification tasks with the aim of developing a factor importance ranking strategy for effective removal of irrelevant factors from the classification model. Kaneko et al. [15] proposed a new Cross-Validated Ranking Factor Importance (CVPFI) method, which achieves stable computation even with a small number of samples and is capable of assessing the importance of strongly correlated factors. Du et al. [16] computed the importance of influencing factors and improved the prediction accuracy of the risk of conflict by constructing a Random Forest model that contains multiple decision trees. Although these methods perform well in specific situations, they fail to effectively utilize factor importance to guide the initial population construction of the genetic algorithm, resulting in inefficient convergence of the algorithm. In order to break through the limitations of existing research, this paper proposes the RP-GA full-factor selection method. The RP-GA algorithm is an innovative extension of the Genetic Algorithm (GA) with two key innovative improvements. Firstly, the algorithm is based on factor importance and random numbers jointly guiding population initialization. Previous studies have failed to fully integrate factor importance into the initial population construction of genetic algorithms, while the RP-GA algorithm, by combining factor importance and random numbers, can increase the proportion of high-quality solutions in the population, accelerate the convergence speed of the algorithm, and thus improve the quality of the final solution. Secondly, R^2-based PLS regression is used for fitness assessment. This innovative fitness assessment can more accurately measure the fitness of individuals compared to traditional methods, providing more effective guidance for the evolution of the algorithm. With these two innovations, the RP-GA algorithm is expected to provide a more efficient and accurate solution for accurately screening the elements affecting learning emotions in smart teaching scenarios.

In the field of classroom social network research, researchers have found that network structural properties [17] (e.g., network diameter, mean degree, clustering coefficient, etc.) are closely related to the exchange of information or the propagation of learning emotions in classroom networks. Tang et al. [18] established a database based on multidimensional data such as student identity, seating relationship, and social relationship, constructed a classroom social network through the seating similarity between learners, and used the CRITIC algorithm and the CRITIC algorithm and entropy weight method to obtain the combination weights, and proposed GRA-TOPSIS multi-decision fusion algorithm to mine the key student nodes with negative influence. The algorithm can objectively evaluate learners based on the classroom social network and provide a theoretical basis for learning emotion calculation and dissemination. Xie et al. [19] showed that the network density, average degree, and clustering coefficients of the classroom social network have a significant impact on the learning emotion calculation and dissemination of students. Classroom social networks with smaller network diameters and higher network densities are conducive to the dissemination of learning emotions among students; networks with higher clustering coefficients, although conducive to the rapid dissemination of learning emotions within a small group, may form information silos and impede cross-group dissemination. The current field of classroom network research has yielded some results, but most of them follow an established model. Without selecting the factors affecting learning emotions, researchers often consider all the attributes of students, including academic performance, classroom interaction behavior and emotional traits, as potential factors affecting learning emotions, and construct a classroom network model based on them. Subsequently, we focus on analyzing the dynamic changes of students' learning emotions over time within the framework of the network model, in an attempt to reveal the evolution of learning emotions at different stages of teaching and learning. However, this research paradigm has significant methodological flaws, the core of which lies in the failure to analyze the factors of the classroom network in depth and identify the key elements that affect learning emotions. Theoretically, the influence of various student attributes on affective learning is not uniformly distributed, but has a complex hierarchical and causal relationship. Conducting research without distinguishing between critical and non-critical elements will introduce a large amount of redundant information, increase the complexity and computational cost of the research model, and the interfering factors will obscure the real association between affective learning and the structure of the classroom network, thus reducing the validity and reliability of the research.

To address the above limitations, this study will use the innovative RP-GA algorithm to accurately identify the key elements affecting learning emotions in smart classroom scenarios, deeply analyze the dynamic changes between the elements and the network structure characteristics, and construct a multi-layer heterogeneous all-element classroom network model. This innovative approach not only improves the accuracy of feature selection and the convergence efficiency of the algorithm, but also more comprehensively grasps the complex impact of network structural characteristics on the propagation of learning emotions, providing new theoretical perspectives and technical support for the practice of smart classroom teaching.

III. Data Set and Experimental Parameters

A. Introduction to Data Sets

In the study of this paper, the historical academic dataset of students collected from Kaggle, an open machine learning database, and the constructed Student Sentiments for Learning in a Smart Classroom dataset (SCLE-Dataset) are used as the experimental datasets, and Table I demonstrates the important information about the use of public datasets in this paper.

The dataset of students' learning emotions under the smart classroom is composed of video data of 67 students studying a course in the smart classroom of a university as well as the results of a questionnaire survey, in which the learning emotions are acquired with reference to the method in the literature [22], and the learning emotions are classified into five grades: very positive, relatively positive, neutral, relatively negative, and very negative. These categories cover not only the intensity but also the positive and negative polarities of learning emotions, providing a multidimensional framework for analyzing the dynamics of learning emotions during the learning process. The SCLE-Dataset was divided into 63 sub-datasets based on the order of the length of time the 63 knowledge points were taught within the course, with each sub-dataset corresponding to an individual knowledge point and consistent factor dimensions. Table II shows the descriptions of the fields in the SCLE-Dataset.

TABLE I IMPORTANT CHARACTERISTICS OF THE PUBLIC DATASETS

Datasets	**Number of instances**	**Number of factors**	**Name of the target attribute in the dataset**
data[20]	4424	36	target
Student-mat[21]	395	32	G3
Student-pro[21]	649	32	G3
xAPI-Edu-Data[21]	480	16	Class
Students Performance[21]	1000	7	Writing score
Turkiye Student[21]	5820	32	Attendance

TABLE II DESCRIPTION OF FIELDS IN SCLE-DATASET

Name	Description	Name	Description	Name	Description
ID	Student number	Classroom test score	Classroom test scores	Interaction	Number of interactions
K point	Knowledge point number	Eng_grade	Grades in English	Use phone	Number of cell phone uses
Class	Date of class	Math_grade	math grade	Take notes	Number of notes taken
Gender	Gender of students	Position preferences	Seat Selection Preference Area	Lean on table	Number of times lying on the table
Character	Student Character	Head on rate	percentage of heads up	Prop up head	Number of headrests
Club	Participation in associations or not	Friend nomination	Number of friend nominations	Yawn	Number of yawns
Competition	Participation in competitions or not	Friend nominated	Number of times nominated by friends	Award rating	Scholarship level
Committee	Serve on a class council or not	Bad for learn	Number of adverse learning nominations	Fail	Number of subjects failed
Dorm	Dormitory number	Good for learn	Number of nominations for enabling learning	Learning emotions	Student Learning Emotions at a Knowledge Point

B. Data Standardization

Data standardization is a key step in the factor selection process, and standardization is the process of transforming each factor column Z_j into a new factor column X_{j} with a mean of 0 and a standard deviation of 1. It can eliminate scale differences between factors and improve the effectiveness of factor selection and model training. If the factors are not normalized, factors with a large range of scale values may dominate the calculation of factor importance, while factors with a small range of scale values may be ignored. This will lead to inaccurate results in the assessment of factor importance, ignoring the possibility that some small-valued factors are more important to the prediction results.

C. Classifier Parameters

The classifiers used in this paper are Support Vector Machine (SVM), Plain Bayes (NB), Multi-Layer Perceptron (MLP) and Random Forest (RF) classifiers, and the parameter settings of each classifier are shown in Table III.

TABLE III NAMES AND PARAMETERS OF THE CLASSIFIERS USED IN THE EXPERIMENT

Name	Parameters
Support Vector Machine (SVM)	C=1.0, kernel="rbf", gamma="scale", tol=1e3, cache_size=200
Plain Bayes (NB)	var_smoothing=1e-9
Multi-Layer Perceptron (MLP)	hidden_layer_ sizes=50, max_iter=1000, learning_rate_init=0.01, random_state=123
Random Forest (RF)	n_estimators=100, max_depth=None, min_samples_split=2, max_features="auto"

IV. RP-GA ALGORITHM

In this section, two improvements of the RP-GA algorithm over the GA algorithm will be presented: first, the use of factor importance versus random numbers to guide population initialization; and second, the dynamic selection of the number of components (i.e., the dimensionality of the PLSs) of the partial least squares PLS model in the fitness function, which allows the model to adaptively adjust its complexity to match the number and nature of the selected factors. The improved RP-GA algorithm optimizes not only the number of factors but also the prediction performance based on the selected set of factors. Fig. 1 illustrates the overall framework diagram of the RP-GA algorithm:

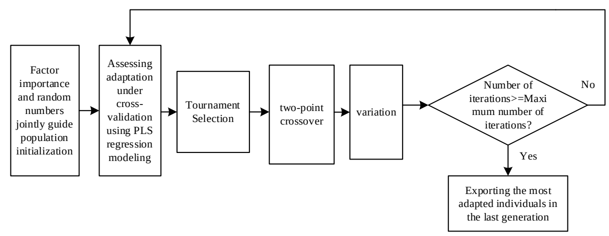


Fig. 1. Flowchart of RP-GA algorithm implementation.

A. Importance Calculation Based on Random Forest Model

Random Forest is an integrated learning method designed to improve the prediction accuracy of a model by constructing multiple decision trees. The main steps include Bootstrap sampling, decision tree construction, model training and prediction. In calculating factor importance, this paper focuses on quantifying the extent to which each factor contributes on average across all decision trees, rather than the relative proportion of each factor's contribution to the totality of all factors. The evaluation method uses mean square error (MSE) as an impurity metric by comparing the reduction in model prediction error (MSE) with and without node splitting using specific factors. Mean Square Error, a key indicator of predictive performance, calculates the average of the squares of the differences between the model's predicted values and the actual observed values, which can be accurately assessed to quantify the level of importance of each factor by comparing the change in MSE before and after factor use.

The steps for calculating the importance of factors are as follows:

1) Calculation of baseline MSE: The prediction of the dataset using the trained model and calculating the mean square error between the predicted and actual values can be expressed as:

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

Where y_i is the actual target value, $\hat{y}_i$ is the model predicted value, and N is the sample size.

2) Remove factor X_j from the model, then retrain the model and compute a new MSE

$$MSE_{exclude,j} = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_{i,j})^2 \tag{2}$$

Where $\hat{y}_{i,j}$ is the predicted value of the model with factor X_j excluded.

3) For each decision tree i, calculate the reduction in mean square error before and after excluding each factor X_j from that decision tree, denoted by $VIM_{ij}^{(MSE)}$

$$VIM_{ij}^{(MSE)} = MSE_{baseline} - MSE_{exclude,j} \tag{1}$$

4) Sum the mean squared error reductions of factor X_j across all decision trees and divide by the number of decision trees, n, to arrive at the average impurity reduction of the factor, i.e., the factor importance score

$$VIM_j^{(MSE)} = \frac{1}{n}\sum_{i=1}^{n} VIM_{ij}^{(MSE)} \tag{4}$$

If $VIM_j^{(MSE)}$ is larger, it indicates that factor has a greater impact on the predictive ability of the model and therefore the factor is more important. Table IV shows the results of the importance ratings of all the factors affecting learning emotions:

TABLE IV RESULTS OF FACTOR IMPORTANCE SCORES

Factor name	score	Factor name	score	Factor name	score
K point	0.183	Eng_grade	0.019	Use phone	0.010
Class	0.021	Math_grade	0.014	Take notes	0.038
Gender	0.022	Position preferences	0.048	Lean on table	0.022
Character	0.023	Head on rate	0.137	Prop up head	0.095
Club	0.010	Friend nomination	0.036	Yawn	0.010
Competition	0.008	Friend nominated	0.031	Award rating	0.005
Committee	0.009	Bad for learn	0.006	Fail	0.026
Dorm	0.065	Good for learn	0.030		
Classroom test score	0.101	Interaction	0.030		

B. Population Initialization Based on Importance and Random Numbers

In traditional genetic algorithms, the individuals of the initial population are usually generated by random generation, although the diversity of the initial population generated in this way is high, a large number of low-quality individuals will be introduced, which reduces the convergence speed and wastes the computational resources. To overcome this limitation, this paper customizes a method for generating a subset of factors that uses factor importance and random numbers to guide the initialization of individuals, with the initial value of each factor depending on its importance in the random forest model. For each factor, if the randomly generated number is less than the importance of the factor, its initial value is randomly determined by the importance of the factor and the specified minimum and maximum bounds. Otherwise, it is set to the minimum boundary value and the important factors are prioritized for model training.

The process in which factor importance and random numbers jointly guide population initialization is illustrated in Fig. 2. The random_threshold and random_value in the figure are random numbers generated between min_bound and max_bound, labeled differently in order to distinguish between the two generated random numbers. The number of factors in the dataset used in this paper is shown in Table 1. The number of individuals in the initialized population is set to 50, striking a balance between the speed of convergence and the maintenance of solution diversity.

When studying the complexity of factors affecting students' learning emotions in classroom networks, the use of factor importance and random numbers together to guide population initialization demonstrated the following significant advantages:

1) Avoidance of over-conditioning: If a fixed threshold is used to compare the importance of factors, the initial values of the factors can be influenced by the distribution of the data, potentially leading to over-adjustment issues, i.e., "biased initialization." By combining the importance of factors with dynamically generated random numbers for initialization, it becomes possible to consider the potential influence of factors on learning emotions in a more balanced manner, making the analysis more objective and scientific.

2) Enhanced exploration and adaptation: In the complex environment of the classroom network, learning emotions are influenced by a wide array of intertwined factors characterized by high degrees of dynamism and uncertainty. By incorporating random numbers into the initialization process, it not only boosts the diversity of individuals in the initial population but also mimics the randomness of learning emotional changes in diverse contexts. This enhances the algorithmic exploration flexibility and scope, enabling the algorithm to comprehensively investigate the interactions between factors and reducing the risk of converging to local optimal solutions.

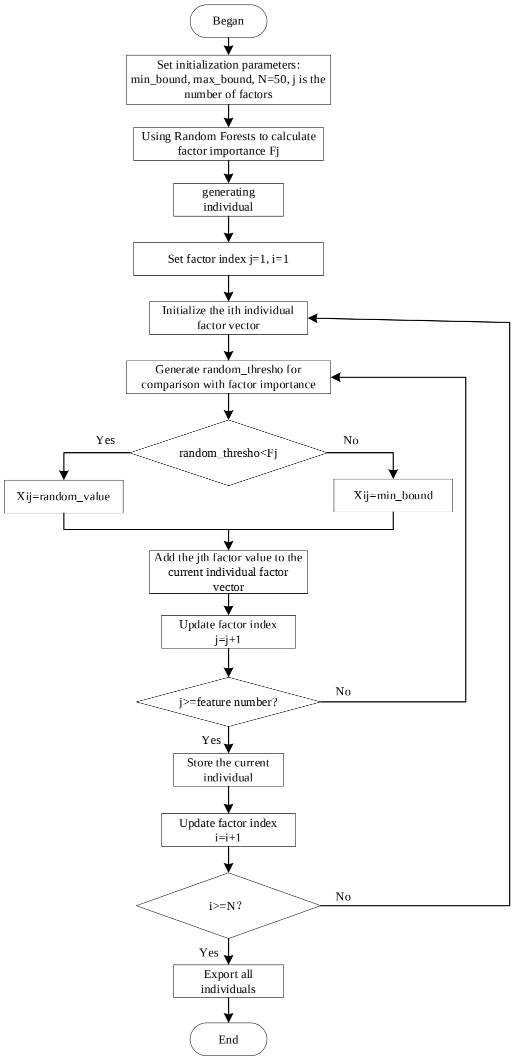


Fig. 2. Flowchart for initializing the population.

C. R²-Based Adaptation Assessment of PLS Regression

For each individual in the initialized population of individuals that have been guided by the importance of the factors, the factors whose factor weights exceed a set threshold are selected to form a subset of the factors. PLS regression models were applied to a subset of factors in each individual and performance was evaluated within a cross-validation framework. In the fitness function, the number of components (i.e., regression dimensions) of the PLS model is dynamically selected based on the rank of the selected subset of factors. The method improves the accuracy of the fitness assessment by finding the optimal number of PLS components on a selected subset of factors using a partial least squares regression model and evaluating the actual impact of the selected factors on the model performance in terms of maximizing the coefficient of determination (R^2).

In the 3-fold cross-validation framework, the following operations are performed for each fold: the model is trained using the training set; predictions are made on the validation set thus obtaining the predicted value $\hat{y}_i$; the R^2-value on the validation set for each fold is calculated; and the largest R^2-value is selected as the fitness value for that individual, which can be expressed as:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{5}$$

where y_i is the actual value of the ith sample in the validation set, $\hat{y}_i$ is the value predicted by the model for the ith sample, and $\bar{y}$ is the mean of the actual values of all samples in the validation set.

In the RP-GA algorithm, PLS regression modeling with cross-validation was chosen as the key innovation based on the following considerations:

1) Adaptive adjustment and optimization of model complexity: In the fitness function of the RP-GA algorithm, the dynamic selection of the number of components of the PLS model (i.e., the number of dimensions of the PLS) allows the model to adaptively adjust its complexity according to the number and nature of the selected factors. Through cross-validation, the optimal number of PLS model components is determined, thus ensuring model accuracy while avoiding over-complexifying the model and improving the optimization efficiency of the algorithm. This adaptive adjustment mechanism enables the RP-GA algorithm to balance model complexity and prediction performance more effectively.

2) PLS regression model: The PLS regression model reduces the impact of data noise on the model by extracting the most representative components, thus improving the robustness of the algorithm. Meanwhile, cross-validation further enhances the stability of the algorithm by dividing the dataset multiple times for training and validation, which reduces the bias caused by unreasonable dataset division. This combined use enables the RP-GA algorithm to show more stable performance when facing different datasets and problems.

V. EXPERIMENTAL RESULTS AND ANALYSIS

A. RP-GA Comparison Experiment

Among the current factor selection algorithms, Forward algorithm [23] and PSO algorithm [24] are more widely used. Forward algorithm can effectively and efficiently select key factors by virtue of its step-by-step forward strategy in the process of feature screening. PSO algorithm can quickly find the best in the search space by virtue of the principle of simulating the foraging behavior of bird flocks. The two algorithms have their own unique advantages, and have been widely used in a number of practical application scenarios. Each of them has its own unique advantages and has been widely used in many practical application scenarios. In addition, using the correlation Pearson coefficient as the evaluation criterion for the subset of factors is more stable. In this experiment, we use the Pearson correlation coefficient as the evaluation criterion for the CL-GA algorithm and as a comparative model, and the following is the formula for the correlation coefficient [21]:

$$R(f_i,c) = \frac{Cov(f_i,c)}{\sqrt{Var(f_i)Var(c)}} \tag{6}$$

where $R(f_i,c)$ denotes the correlation between factor f_i and target c, $Cov(f_i,c)$ denotes the covariance between factor f_i and target c, and $Var(f_i,c)$ denotes the variance between factor f_i and target c. In addition, the forward selection algorithm is also used as a comparison algorithm.

In this paper, we conduct experiments on six publicly available datasets as well as on a specific sub-dataset of the SCLE dataset and compare the performance of the RP-GA algorithm with other algorithms when using SVM, NB, MLP and RF classifiers. We measure the performance of the model using the accuracy ratio (acc), which is defined as the ratio of the number of samples correctly predicted by the classifier to the total number of samples. It is calculated using the formula:

$$acc = \frac{TP+TN}{TP+TN+FP+FN} \tag{7}$$

From Table V, it can be seen that RP-GA algorithm outperforms CL-GA algorithm, Forward algorithm, and PSO algorithm on all experimental datasets. Differences in the performance of the algorithms not only depend on the characteristics of the algorithms themselves, but may also be closely related to the characteristics of the dataset. In the case of the Student-pro [21] dataset and the Turkiye Student [21] dataset, for example, although both have a total number of factors of 32, there is a significant difference in the number of samples, which has an impact on the algorithm's prediction accuracy. The Student-pro dataset with a sample size of 649 is derived from Portuguese language course grades in Portuguese secondary schools. The smaller sample size allows the algorithm to quickly capture key patterns and achieve higher prediction accuracy. The Turkiye Student dataset, with a sample size of 5820, is derived from Turkish university course evaluations, and the large number of samples brings rich information but also introduces more noise and uncertainty. In addition, the complexity of the university course evaluation data, including subjective student evaluations and diverse course content, puts the algorithm under computational pressure and the risk of overfitting, and it is easy to learn local non-universal patterns, which affects the prediction accuracy. In conclusion, RP-GA better adapts to various datasets, suits small-to-medium-sized samples with clear features, and shows superior experimental performance.

The four datasets data, Student-mat, Student-pro, and Turkiye Student, which have a higher number of factors, were selected to demonstrate the comparison of the number of factors after selection, and from Table VI it can be seen that compared to CL-GA algorithm, Forward algorithm, and PSO algorithm, the RP-GA algorithm reduces the number of factors after selection, improves the training speed and prediction accuracy of the classifier.

TABLE V COMPARING THE PREDICTION ACCURACY OF CL-GA, FORWARD, PSO AND RP-GA ON DIFFERENT CLASSIFIERS

Datasets	**Algorithms**	**Prediction accuracy of classifiers**			
		SVM	*NB*	*MLP*	*RF*
data	CL-GA	0.718	0.721	0.755	0.752
	Forward	0.751	0.736	0.715	0.735
	PSO	0.721	0.704	0.723	0.732
	RP-GA	**0.855**	**0.808**	**0.867**	**0.865**
Student-mat	CL-GA	0.807	0.798	0.815	0.849
	Forward	0.807	0.840	0.840	0.840
	PSO	0.816	0.823	0.832	0.832
	RP-GA	**0.832**	**0.874**	**0.874**	**0.899**
Student-pro	CL-GA	0.841	0.831	0.780	0.846
	Forward	0.846	0.862	0.856	0.862
	PSO	0.843	0.851	0.843	0.857
	RP-GA	**0.918**	**0.882**	**0.923**	**0.928**
xAPI-Edu-Data	CL-GA	0.701	0.688	0.694	0.806
	Forward	0.743	0.688	0.688	0.736
	PSO	0.717	0.688	0.679	0.742
	RP-GA	**0.764**	**0.701**	**0.771**	**0.833**
Students Performance	CL-GA	0.643	0.727	0.753	0.697
	Forward	0.697	0.717	0.710	0.717
	PSO	0.691	0.704	0.721	0.721
	RP-GA	**0.770**	**0.780**	**0.780**	**0.810**
Turkiye Student Evaluation	CL-GA	0.402	0.346	0.436	0.442
	Forward	0.410	0.418	0.416	0.416
	PSO	0.423	0.444	0.423	0.432
	RP-GA	**0.531**	**0.480**	**0.546**	**0.536**
SCLE-Dataset	CL-GA	0.667	0.619	0.714	0.571
	Forward	0.619	0.619	0.714	0.667
	PSO	0.623	0.619	0.714	0.651
	RP-GA	**0.700**	**0.700**	**0.750**	**0.700**

TABLE VI FACTOR RESULTS AFTER CL-GA, FORWARD, PSO AND RP-GA SEARCHES

Datasets	Results of factor selection			
	CL-GA	*Forward*	*PSO*	*RP-GA*
data	1,2,8,12,17,18,19,20,22,29,30,31,32,34,35	2,4,7,9,10,13,17,24,25,31,35	1,2,8,9,10,13,17,22,24,25,29,31,35	**3,12,17,20,21,23,24,25,26,28**
Student-mat	3,5,6,7,9,10,12,13,14,19,20,23,24,32	9,10,18,22,24,26,27	1,5,6,7,9,10,12,13,18,22,24,26	**3,8,15,16,17,31**
Student-pro	1,2,3,4,8,10,11,12,13,19,23,24,27,28,29,30,32	3,12,16,18,19,21,22,24,25	1,2,3,4,8,10,12,13,19,21,22,23,24,25	**2,15,25,31,32**
TurkiyeStudent Evaluation	3,4,7,9,11,14,16,17,19,20,29,30	1,2,3,14,17,22	1,2,3,4,7,9,11,14,16,17,29,30	**1,3,4,13,21,31**

B. Analysis of Ablation Experiments

In order to verify the validity of population initialization through factor importance and random number co-guiding in RP-GA and the use of PLS regression model with cross-validation to assess the fitness. The GA algorithm is used as the baseline model for the ablation experiments, and the data preprocessed DATA dataset is used for the experiments, using the classifiers in Table III and keeping the parameters consistent, and ablating the improvement modules one by one to obtain four sets of experimental data, as shown in Table VII. In the table, A represents the improvement point in Chapter IV.B (using factor importance and random number to jointly guide population initialization), and B represents the improvement point in Chapter IV.C (using PLS regression model with cross-validation to assess fitness). From the ablation experimental data, it can be found that baseline improved by 0.091, 0.041, 0.05, and 0.061 after adding improvement point A with SVM, NB, MLP, and RF classifier prediction accuracies of 0.712, 0.720, 0.751, and 0.751, respectively, and the final effect improved by 0.143 after adding improvement point B. The final results were, 0.088, 0.116, 0.114. The above improvement points made significant contributions to the performance enhancement of the RP-GA algorithm in terms of the efficiency of the factor selection and the accuracy of the adaptation assessment, obtaining a large prediction accuracy enhancement, which fully proved the effectiveness of the two-point improvement strategy proposed in this paper. The visualization of the ablation experiment is shown in Fig. 3.

TABLE VII ABLATION EXPERIMENT

A	B	Prediction accuracy of classifiers			
		SVM	*NB*	*MLP*	*RF*
Baseline		0.712	0.720	0.751	0.751
✓	×	0.803	0.761	0.801	0.812
×	✓	0.793	0.758	0.796	0.784
✓	✓	**0.855**	**0.808**	**0.867**	**0.865**

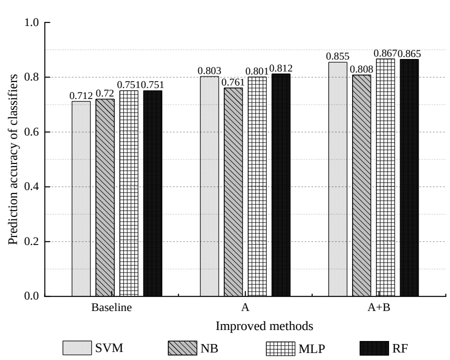


Fig. 3. Intuitive display of ablation experiments.

C. Convergence Analysis

The fitness function is used to measure the degree of superiority or inferiority of an individual (chromosome) in the problem environment. During the iteration of the algorithm, the fitness of individuals in the population changes in each generation. In order to further verify that the factor importance and random number co-guided population initialization proposed in this paper can improve the convergence speed of the RP-GA algorithm, we analyze the average fitness and standard deviation fitness change process of each generation of the population in the iterative process of $RP-GA_{lif}$ (lack of factor importance and random number co-guided population initialization) and RP-GA.

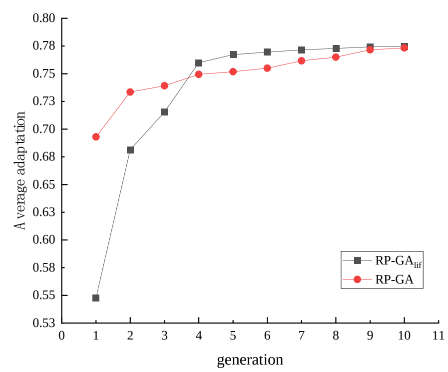


Fig. 4. The changing trend of the average fitness of RP-GA and $RP-GA_{lif}$.

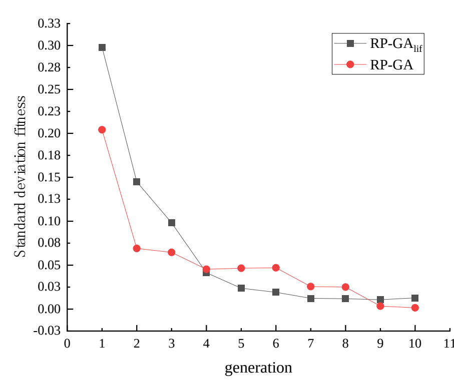


Fig. 5. The changing trend of the standard deviation fitness of RP-GA and $RP-GA_{lif}$.

From the average fitness curves in Fig. 4, RP-GA converges faster in the early stage, thanks to its method of factor importance and random number co-guiding population initialization, which enables the initial population to be more reasonably distributed in the solution space. As the iteration proceeds, although the average fitness of RP-GA and $RP-GA_{lif}$ both tend to stabilize, RP-GA converges to a region with higher fitness, indicating that it is better in convergence effect. And RP-GA finally reaches a higher average fitness, which indicates that RP-GA is more advantageous than in finding the optimal solution, and the special population initialization helps to avoid falling into the local optimal solution and easier to find the region of the global optimal solution.

From the standard deviation fitness curve in Fig. 5, the standard deviation fitness of RP-GA decreases faster in the early stage, which is due to the fact that it adopts the factor importance and random number to guide the population initialization, so that the population quickly concentrates in the more optimal direction. With the increase of iteration, the standard deviation of both RP-GA and $RP-GA_{lif}$ tends to stabilize and have smaller values, and the standard deviation of RP-GA is slightly smaller. This indicates that the population convergence of RP-GA is better, the individual adaptations are more consistent, and it can effectively guide the population to converge in the process of evolution, reduce the differences between individuals, and help to find a more optimal solution.

D. Analysis of SCLE-Dataset Experimental Results

SCLE-Dataset is composed of data from different knowledge points of freshman students at a university while studying a course. Considering the existence of different elements in each knowledge point that affect students' learning emotions, this paper evaluated each knowledge point using the RP-GA algorithm. Table VIII shows all the knowledge points elements results and the prediction accuracy of the NB classifier. The elements of all knowledge points were counted statistically, from which the top five elements were filtered as the key elements affecting students' learning emotions, which were gender, head-up rate, preferred area for seat selection, number of friend nominations, and number of favorable study nominations.

TABLE VIII RESULTS OF ELEMENTS FILTERED WITHIN ALL KNOWLEDGE POINTS AND ACCURACY

Knowledge point number	Element name	Accuracy of NB
Point 1	Classroom test score， Award rating	0.85
Point 2	Classroom test score， Award rating	0.85
Point 3	Gender, Friend nominated	0.60
⋮	⋮	⋮
Point 61	Position preferences, Head on rate, Lean on table	0.70
Point 62	Math grade, Head on rate, Use phone, Lean on table	0.95
Point 63	Gender, Committee, Math_grade, Head on rate, Good for learn, Take notes	0.75

E. Characterization of the Structure of the Total Element Classroom Network

In order to further reveal the synergistic effect of key elements and enhance the classroom learning experience, the key elements affecting learning emotions are integrated into the framework of analyzing the structural properties of classroom networks, and the potential influence of key elements on the propagation of students' learning emotions is explored by quantifying and analyzing the structural properties of classroom networks. The structural characteristics of the multi-layer heterogeneous all-element classroom network G_{MHLA} based on SCLE-Dataset influencing the elements of students' learning emotions are analyzed experimentally, and its multi-layer heterogeneous network G_{MHLA} construction method is utilized in the literature [19].

Table IX shows that the all-factor classroom network G_{MHLA} constructed on the basis of the key elements (students' gender, head-up rate, seating information, friends' nomination, and favorable learning nomination) has the properties of low network diameter, low average path length, high network density, high average degree, high average weighting, and low clustering coefficient. These structural properties reflect more connections of nodes in the multilayer heterogeneous all-factor classroom network, shorter paths for learning emotions, and easier cross-cluster dissemination of learning emotions among small groups. Fig. 6 shows a visualization of the structural changes in the classroom network caused by the change of student seating information in the first, fourth, and seventh classes, and it can be found that the denseness of the connections between nodes in the network increases significantly with the increase in the number of classes. Fig. 7 shows a slight increase in the density of connections between the nodes in the first class as the number of points taught increases and the students' head-up rate changes, due to the fact that the students are unfamiliar with each other and have established fewer connections with each other in the first class.

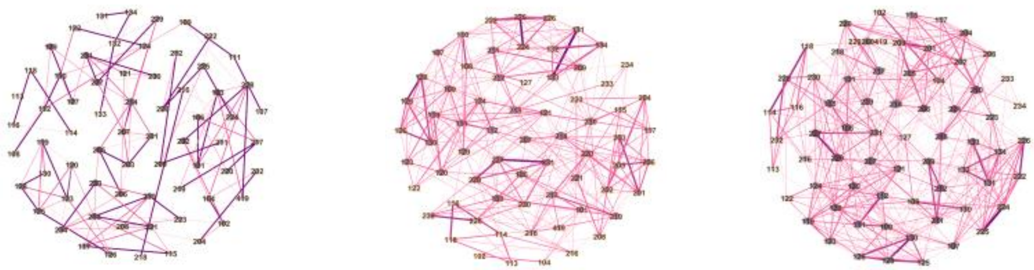
Table X shows the analysis of the structural characteristics of the total classroom network with different classroom seating variations. It can be found that as the number of classes increases, the structural characteristics of the all-factor classroom network change significantly, which is manifested in the reduction of network diameter and average path length, the enhancement of network density, the growth of average degree and average weighting degree, and the decrease of clustering coefficient, and the structural all-factor classroom network is more compact and efficient in the seating relationship of the last classroom, and all these changes together constitute a network environment more favorable to emotion dissemination and network environment for emotion transmission.

Table XI shows the characterization of the whole-factor classroom network structure in terms of changes in head-up rates for different knowledge points in the first class. It can be found that as the number of knowledge points increases, the structural characteristics of the whole-factor classroom network change, which is manifested in the reduction of the average path length, the enhancement of the network density, the growth of the average degree and the average weighting degree, and the decrease of the clustering coefficient.

On the basis of analyzing the structural characteristics of the network, this paper constructs a multidimensional and comprehensive classroom network model using the key elements screened by the RP-GA algorithm, and analyzes the mechanism of the key elements and the structural characteristics of the classroom network in depth. The results of the analyses show that the dynamics of the key elements in the classroom network significantly influence and change the structural characteristics in the classroom network. This strongly verifies the effectiveness and scientificity of the RP-GA algorithm in identifying the elements affecting learning emotions, and further proves that the elements selected by the RP-GA algorithm provide an important basis and support for the calculation and dissemination of learning emotions in the future.

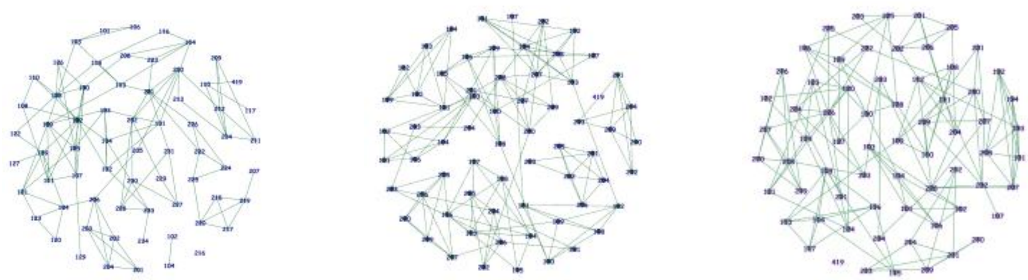
TABLE IX STRUCTURAL CHARACTERIZATION OF THE MULTILAYERED HETEROGENEOUS ALL-FACTOR CLASSROOM NETWORK

Network	Network diameter	Average path length	Network density	Average degree	Average weighted degree	Minimum degree	Clustering factor
G_{MLHA}	3	2.842	0.271	8.657	20.49	7	0.454



N=1 N=4 N=7

Fig. 6. Network structures under different numbers of classes N.



K=1 K=3 K=6

Fig. 7. Network structures under different numbers of knowledge points K.

TABLE X CHARACTERIZATION OF THE NETWORK STRUCTURE AT DIFFERENT NUMBER OF N HOURS OF CLASSES

Network	Network diameter	Average path length	Network density	Average degree	Average weighted degree	Minimum degree	Clustering factor
$G_{MHLA}^{N=1}$	5	3.324	0.153	6.857	15.01	7	0.624
$G_{MHLA}^{N=2}$	5	3.186	0.173	7.089	15.75	7	0.612
$G_{MHLA}^{N=3}$	5	3.169	0.186	7.282	16.45	7	0.610
$G_{MHLA}^{N=4}$	4	3.143	0.195	7.749	17.65	7	0.594
$G_{MHLA}^{N=5}$	4	3.015	0.203	8.102	19.65	7	0.540
$G_{MHLA}^{N=6}$	3	2.943	0.224	8.371	20.23	7	0.471
$G_{MHLA}^{N=7}$	3	2.842	0.271	8.657	20.49	7	0.454

TABLE XI CHARACTERIZATION OF THE NETWORK STRUCTURE AT DIFFERENT NUMBERS OF KNOWLEDGE POINTS K

Network	Network diameter	Average path length	Network density	Average degree	Average weighted degree	Minimum degree	Clustering factor
$G_{MHLA}^{k=1}$	5	3.224	0.122	6.695	14.32	7	0.625
$G_{MHLA}^{k=2}$	5	3.221	0.135	6.701	14.66	7	0.625
$G_{MHLA}^{k=3}$	5	3.221	0.135	6.701	14.66	7	0.625
$G_{MHLA}^{k=4}$	5	3.214	0.140	6.705	14.86	7	0.623
$G_{MHLA}^{k=5}$	5	3.214	0.140	6.743	15.01	7	0.623
$G_{MHLA}^{k=6}$	5	3.121	0.142	6.743	15.01	7	0.616
$G_{MHLA}^{k=1}$	5	3.224	0.122	6.695	14.32	7	0.625

VI. CONCLUSION

In this study, the RP-GA algorithm is innovatively employed to screen key elements influencing learning emotions in smart classroom teaching, and subsequently, a multi - layer heterogeneous all-element classroom network model is built, which, through in-depth exploration of the relationship between elemental variations and network structural features, lays a scientific groundwork for future calculation and dissemination of learning emotions. As the research process was limited by the scope of data acquisition, the collected data could not completely cover all types of smart classroom teaching scenarios, which might have an impact on the comprehensiveness of the RP-GA algorithm in screening the key elements. In factor selection and model optimization, the RP - GA algorithm, with its unique initialization and dynamic adjustment, avoids local optima and boosts the model's prediction accuracy. In this study, we mainly focus on small and medium-sized datasets, for large datasets there may be problems of lower prediction accuracy and higher computational pressure. Even so, this study substantiates the high validity and reliability of the RP-GA algorithm in optimizing classroom network structural characteristics. It not only broadens the theoretical research on classroom networks but also offers a scientific basis and practical guidelines for enhancing teaching quality and evolving learning emotions.

ACKNOWLEDGMENT

The National Natural Science Foundation of China (62177012, 62267003).

Guangxi Natural Science Foundation (2024GXNSFDA010048).

The Project of Guangxi Wireless Broadband Communication and Signal Processing Key Laboratory (GXKL06240107).

REFERENCES

[1] X. Jia. "Research on the role of big data technology in the reform of English teaching in universities," Wireless Communications and Mobile Computing, 2021..

[2] G. Putnik, E. Costa, C. Alves, et al. "Analysing the correlation between social network analysis measures and performance of students in social network-based engineering education," International Journal of Technology and Design Education, vol. 26(3), 2016.

[3] A. Charitopoulos, M. Rangoussi and D. Koulouriotis, "On the use of soft computing methods in educational data mining and learning analytics research: A review of years 2010–2018," International Journal of Artificial Intelligence in Education, vol. 30(3), 2020.

[4] N. Romanov, L. M. Culci, A. I. Daniel, et al. "Artificial intelligence applications and tools IN higher education: an overview," Proceedings of the SESYR Sustainable Education through European Studies for Young Researchers Jean Monnet Module, 2020.

[5] L. You, "Research on learning sentiment based on group interaction behavior," Wuhan: Central China Normal University, 2022.

[6] Z. J. Qing, "Intelligent education visualization system based on social network analysis,"2020 International Conference on Robots & Intelligent System (ICRIS). 2020: IEEE, pp. 291-294.

[7] K. Vignery, "From networked students centrality to student networks density: What really matters for student performance?," Social Networks, vol. 70, pp. 166-186, 2022.

[8] J. K. Yang, "Incorporating network and propagation properties for source identification on social networks," Hangzhou: Huazhong Dianai Univerity, 2023.

[9] G. X. Dong, Z. Xia, G. Y. Mei, "A study of affective factors affecting college students' autonomous english learning," Advances in Educational Technology and Psychology, vol. 7(18), pp. 6-17, 2023.

[10] L. X. Deng, H. Y. Chen, H. Y. Liu, H. Zhang, Y. Zhao, "Overview of UAV path planning algorithms," 2021 IEEE International Conference on Electronic Technology, Communication and Information (ICETCI). 2021: IEEE, pp. 520-523.

[11] R. Izabela, L. Krzysztof, "GAAMmf: genetic algorithm with aggressive mutation and decreasing feature set for feature selection," Genetic Programming and Evolvable Machines, vol. 24(2), 2023.

[12] K. Y. Aram, S. S. Lam and M. Khasawne, "Cost-sensitive max-margin feature selection for SVM using alternated sorting method genetic algorithm," Knowledge-Based Systems, vol. 267, 2023.

[13] S. Deng, Y. Li, J. Wang, R. Cao, M. Li. "A feature-thresholds guided genetic algorithm based on a multi-objective feature scoring method for high-dimensional feature selection," Applied Soft Computing, vol. 148, 2023.

[14] A. Razmjoo, P. Xanthopoulos, Q. P. Zheng, "Feature importance ranking for classification in mixed online environments," Annals of Operations Research, vol. 276, pp. 315-330, 2019.

[15] H. Kaneko, "Cross-validated permutation feature importance considering correlation between features," Analytical Science Advances, vol. 3(9-10), pp. 278-287, 2022.

[16] S. K. Du, J. Zhang, Z. J. Han, M. Y. Gong, . "Armed conflict risk prediction and influencing factors analysis based on the random forest model at the grid-month scale: a case study of indochina peninsula," Journal of Geo-information science, vol. 25(10), pp. 2026-203, 2023.

[17] Z. Kong, Q. Sun, X. Y. Kou, L. F. Wang. "Research on the importance of network nodes based on attribute information and structural characteristics," Journal of Northeastern University (NatralScience), vol. 43(05), pp. 625-631, 2022.

[18] Z. Y. Shou, M. Tang, H. Wen, et al. "Key student nodes mining in the in-class social network based on combined weighted GRA-TOPSIS method," International Journal of Information and Communication Technology Education (IJICTE), vol. 19(1), pp.1-19, 2023.

[19] Z. Y. Shou, H. Wang, H. B. Zhang, J. L. Xie, J. H. Tang. "The NEDC-GTOPSIS node influence evaluation algorithm based on multi-Layer heterogeneous classroom networks," International Journal of Information and Communication Technology Education (IJICTE), vol. 20(1), pp. 1-24, 2024.

[20] R. Valentim, M. Jorge, B. Luis, et al. "Predict students' dropout and academic success," UCI Machine Learning Repository, vol. 10, 2021, C5MC89.

[21] W. Xiao , P. Ji, J. Hu, "RnkHEU: A hybrid feature selection method for predicting students' performance," Scientific Programming, vol. 2021(1), 2021, 1670593.

[22] Z. Y. Shou, N. Zhu, W. H. Wang, et al. "A method for analyzing learning sentiment based on classroom time-series images," Mathematical Problems in Engineering, vol. 2023(1), 2023, 6955772.

[23] T. Nakanishi, P. Chophuk, K. Chinnasarn, "Evolving Feature Selection: Synergistic Backward and Forward Deletion Method Utilizing Global Feature Importance," IEEE Access,12[2025-01-22].DOI:10.1109/ACCESS.2024.3418499.

[24] X. F. Song, Y. Zhang, D. W. Gong, X. Z. Gao. "A fast hybrid feature selection based on correlation-guided clustering and particle swarm optimization for high-dimensional data," IEEE Transactions on Cybernetics, vol. 52(9), pp. 9573-9586, 2021.