RESEARCH ARTICLE

A Prediction Model for Information Anxiety of CSs Based on CMA-ES and XGBoost

Bin Wang*, Li Shao

Harbin Institute of Technology, Harbin 150000, China

*Corresponding Author: 15164617312@163.com

Abstract: With the rapid advancement of information technology, college students are increasingly facing the problem of information overload. Information anxiety, an emerging mental health issue, significantly impacts their studies, daily lives, and mental well-being. Given the limitations of existing methods in handling complex nonlinear relationships and high-dimensional data, this study proposes an efficient and accurate prediction model for college students' information anxiety. The model integrates the covariance matrix adaptation evolution strategy (CMA-ES) with the extreme gradient boosting (XGBoost) algorithm to meet the high-precision prediction demands of college mental health early warning systems. CMA-ES, an advanced global optimization algorithm, effectively avoids local optima by dynamically adjusting the covariance matrix, thereby enhancing the model's global search capabilities and convergence speed. Building on this, the XGBoost algorithm further improves the model's prediction accuracy and generalization ability through ensemble learning, incorporating hyperparameters optimized by CMA-ES. Experimental results demonstrate that the proposed model outperforms other comparative algorithms in terms of accuracy, recall, F1 score, and area under the curve (AUC). Specifically, the model achieves an accuracy of 0.96, a recall of 0.97, Macro-F1 and Micro-F1 scores of 0.96 and 0.98, respectively, and an AUC close to 1. Additionally, the model maintains a prediction accuracy between 83.1% and 88.5% under varying experimental conditions, with reasonable execution times. The prediction model, which integrates the covariance matrix adaptation evolution strategy and the extreme gradient boosting algorithm, effectively addresses the challenge of accurately predicting information anxiety among college students. It provides a scientific basis and efficient tools for psychological health interventions in universities, holding significant application value and promotion potential.

Keywords: CMA-ES, XGBoost, College Students, Information Anxiety, Predictive Models

Received: 11-04-2025 | **Revised:** 09-07-2025 | **Accepted:** 24-10-2025 | **DOI:** 10.3844/ajbbsp.2026.22.01.013

Introduction

With the rapid advancement of information technology and the widespread use of the Internet, College Students (CSs) are exposed to an increasing number of information sources and content [1]. From academic resources to social media, and from career planning to daily trivialities, CSs encounter a rapidly growing volume of information in their daily lives [2]. This phenomenon of information overload has imposed significant psychological pressure on CSs during the processes of selecting, processing, and screening information. When faced with uncertain and ambiguous information, they are particularly prone to experiencing information anxiety [3]. Information anxiety refers to the stress and distress individuals feel when

DOI: 10.3844/ajbbsp.2026.22.01.013

processing and managing information, which can affect their emotional state, cognitive abilities, and daily decision-making skills, thereby impacting their academic, personal, and work performance [4]. In recent years, as mental health concerns among CSs have become increasingly prevalent, information anxiety has emerged as an undeniable mental health issue [5]. CSs' information anxiety is not only closely linked to their academic performance and psychological well-being but may also lead to more severe psychological problems, such as depression and anxiety disorders.

Li et al. proposed an elite-driven proxy-assisted Covariance Matrix Adaptation Evolution Strategy (CMA-ES) algorithm. This algorithm enhanced the computational efficiency and global optimization capability for predicting information anxiety among CSs through an improved lower confidence bound method. The algorithm leveraged the uncertainty term in the confidence bound formula under step size control to develop a model management approach that includes pre-screening strategies and competitive chaotic operators. This approach aimed to improve sampling quality, optimize convergence speed, and increase population diversity. Experimental results demonstrated that this algorithm outperformed other algorithms in terms of efficiency and optimization ability [6]. Van der Meersch et al. addressed the research problem that process-based models, compared to association models, are challenging to parameterize for a large number of species. They explored the feasibility of using CMA-ES to calibrate process-based models. Their method involved utilizing European tree species distribution data and adapting the CMA-ES algorithm to find optimal model parameter values. The results indicated that CMA-ES calibration was more effective than relying on expert knowledge, and the parameters were interdependent, with different combinations yielding high accuracy [7]. Li et al. proposed Knowledge-Extraction-based Variable-Fidelity Surrogate-Assisted (KE-VFS) CMA-ES to accelerate multi-objective optimization in high-cost engineering cases. The KE-VFS model was first established to extract knowledge from low-fidelity non-dominated solutions to determine high-fidelity sample points, which were then combined with other sample points to build the model. Additionally, they proposed a model management approach based on improved hypervolume improvement criteria and pre-screening strategies. The results showed that this method exhibited excellent efficiency, robustness, and applicability in testing [8]. Qiu et al. selected 150 sets of data containing 13 indicators as input and measured peak particle velocity as output to accurately predict ground vibration caused by blasting. They optimized the hyperparameters of the XGBoost model using the Grey Wolf Optimizer (GWO), Whale Optimization Algorithm (WOA), and Bayesian Optimization (BO), and compared the results with other models. The findings indicated that the WOA-XGBoost model was the most reliable, and the three improved hybrid models outperformed other compared models [9]. Asselman et al. aimed to improve the accuracy of student performance prediction from a technical perspective, as existing improvements in predictive failure analysis were mostly limited to the teaching level. The study focused on ensemble learning and proposed a predictive failure analysis method based on various models, such as random forest, AdaBoost, and XGBoost. The research results were evaluated on three datasets, and the scalable XGBoost model performed the best, significantly improving performance prediction compared to the original Predictive Failure Analysis (PFA) algorithm [10].

Parlak Sert et al. examined the relationship between COVID-19-related anxiety and the tendency toward social media overuse among College Students (CSs). The research methodology involved selecting 346 CSs in Türkiye as samples, collecting data online, and processing it using various statistical analysis methods. The results revealed that male students exhibited higher COVID-19 anxiety scores. During the epidemic, students used social media more frequently than before, with increased time consumption and addiction. Moreover, the level of addiction rose with increasing COVID-19 anxiety [11]. Mao B et al. investigated how information overload during the COVID-19 era affects individuals' information processing ability and motivation, studying 493 participants. The results indicated that information fatigue (a state of low motivation caused by overload) led to a reduction in the processing of COVID-19 prevention information. Information fatigue was associated with information overload (a state of low capacity caused by overload), and both motivation and capacity-related factors were interconnected [12]. Wang et al. explored the link between excessive information intake and employee anxiety in the workplace within the context of corporate social media. They constructed a theoretical model using 219 questionnaires collected online and validated it using Partial Least Squares Structural Equation Modeling (PLS-SEM). The results suggested that information overload on corporate social media had a positive impact on employee work anxiety. Instrumental connections weakened this relationship, while expressive connections strengthened it [13].

To summarize, in the current era of rapid information technology development, CSs are facing psychological pressure due to information overload, and the issue of information anxiety is becoming increasingly prominent. Although existing studies have made some progress in understanding the causes, impacts of information anxiety among CSs, and its relationship with mental health, there are still deficiencies in the construction of predictive models in current research. For instance, limitations in research methods lead to an insufficient consideration of the complex relationships among different variables, and optimization models for specific groups are not refined enough, making it difficult to meet the demand for precise predictions

DOI: 10.3844/ajbbsp.2026.22.01.013

in the early warning system for mental health in colleges and universities. Therefore, this study proposes a prediction model for CSs' information anxiety based on the combination of the CMA-ES and XGBoost algorithms. The aim is to construct an efficient and accurate prediction model for CSs' information anxiety to address the increasingly prominent information anxiety problem among CSs in the context of rapid information technology development. This model provides strong technical support for the early warning system for mental health in colleges and universities. The research innovation lies in the first-ever combination of the CMA-ES and XGBoost algorithms and their application in predicting information anxiety among CSs. The hyperparameters of XGBoost are optimized by dynamically adjusting the covariance matrix through CMA-ES, avoiding the problem of local optimal solutions, and improving the prediction accuracy and generalization ability of the model. This offers a new technical approach for mental health prediction. By introducing an online learning mechanism and regularly updating the model parameters, the model's adaptability to the dynamic changes in CSs' information behaviors has also been enhanced. The significance of this research lies in its efficient predictive performance, which enables it to be seamlessly integrated into the existing early warning system for mental health in colleges and universities. This model can serve as the core prediction module of the early warning system, receiving and processing multi-dimensional data such as students' basic information and information behavior characteristics in real-time, and quickly and accurately predicting students' anxiety tendencies. This, in turn, helps to effectively prevent and alleviate the information anxiety problem of CSs, ensure students' mental health, and improve the scientificity and effectiveness of mental health education in colleges and universities.

Methods

The study provides a detailed account of the construction process for a prediction model aimed at addressing CSs' information anxiety. This includes a comprehensive definition and feature analysis of CSs' information anxiety, an in-depth description of the mathematical model and optimization mechanism underlying the CMA-ES algorithm, and a step-by-step outline of the construction and implementation process for a prediction model that integrates the CMA-ES and XGBoost algorithms. These efforts collectively lay a solid foundation for the subsequent experimental validation of the model's performance.

Analysis of Information Anxiety Among CSs

Information anxiety, a prevalent mental health concern in contemporary society, is becoming increasingly prominent among CSs. With the rapid advancement of information technology and the widespread adoption of the Internet, CSs are inundated with a vast influx of learning, social, and lifestyle-related information. This information not only surpasses traditional information loads in quantity but also often poses challenges in terms of quality and credibility, significantly impacting the mental well-being of CSs [14, 15]. Consequently, the timely prediction of CSs' information anxiety is of paramount importance. The prediction process of CSs' information anxiety is shown in Figure 1.

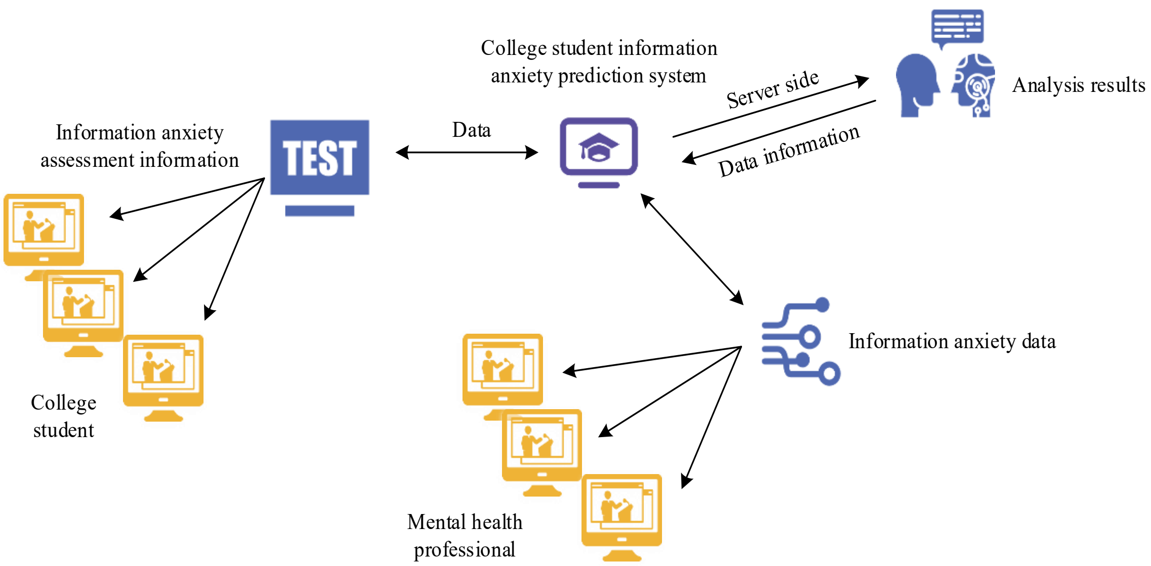

Fig. 1: Prediction Process of CSs Information Anxiety

Figure 1 depicts the data flow process of the information anxiety prediction system designed for CSs. CSs furnish personal information through information anxiety assessment tests, and the resulting assessment data is collected and fed into the system. Additionally, psychological health professionals contribute data, which is also gathered and processed. The information anxiety prediction system then receives and scrutinizes this data, with the server undertaking comprehensive data processing and analysis to ultimately yield prediction results. The fuzzy reasoning process of the prediction model for CSs' information anxiety is shown in Figure 2.

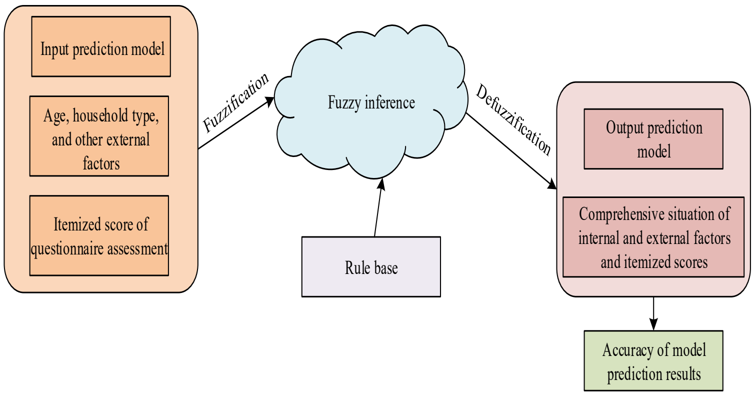

Fig. 2: Fuzzy Reasoning Process of Information Anxiety Prediction Model for CSs

Figure 2 illustrates the fuzzy inference process of the information anxiety prediction model tailored for CSs. This process comprises five sequential steps: input to the prediction model, fuzzification, fuzzy inference, defuzzification, and output from the prediction model. Initially, external factors such as the age and household registration type of CSs, along with the scores obtained for each item in the scale evaluation, are gathered as input data. Subsequently, this data undergoes fuzzification to align it with the requirements of fuzzy reasoning. Next, fuzzy rules and a rule library are employed for inference, with these rules being formulated based on the comprehensive analysis of both internal and external factors, as well as the scores of each item on the scale. Following this, the inference results are subjected to defuzzification to derive clear and interpretable prediction outcomes. Finally, the defuzzified prediction results, along with the accuracy of the model's predictions, are outputted.

Mathematical Model and Optimization Mechanism of CMA-ES Algorithm

Following an in-depth analysis of the characteristics and influencing factors of information anxiety among CSs, the study delved into strategies for constructing accurate prediction models through effective algorithm optimization. It introduced the mathematical model and optimization mechanism of the CMA-ES, thereby laying the groundwork for the subsequent development of prediction models that integrate CMA-ES with the XGBoost algorithm. In the study of predicting information anxiety among CSs, the CM of CMA-ES is dynamically adjusted using two methods: Rank-one-Update and Rank-μ-Update. Rank-one-Update uses the difference vector between the best individual in the population and the mean to adjust the CM and accelerate convergence speed. Rank-μ-Update utilizes the difference vector between the weighted average and mean of multiple best individuals in the population to adjust and enhance stability [16, 17]. The CM C is the key to describing the correlation between dimensions in a multidimensional space, and it is used in CMA-ES to guide the sampling of the current generation based on the distribution information of the earlier generation samples. Assuming there is an objective function $f(x)$ that needs to be optimized and $x=(x_1,x_2,...,x_N)^T$ is a vector of N dimensional parameters, it is necessary to find a global optimal solution that minimizes the value of the objective function. In each generation, CMA-ES estimates the mean and CM of the solution space based on the current population's solution, and then adjusts the generation method of the next generation population, as shown in Equation (1):

$$m_g=\frac{1}{\lambda}\sum_{i=1}^{\lambda}x_i \tag{1}$$

In Equation (1), x_i is the i^{th} solution, m_g is the mean of the current generation, and λ is the size of the population. Then, based on the current population sample, the CM is estimated, which represents the covariance between each pair of decision variables. Assuming there is a sample set $x_1,x_2,...,x_\lambda$, the CM is shown in Equation (2):

DOI: 10.3844/ajbbsp.2026.22.01.013

$$C_g = \frac{1}{\lambda}\sum_{i=1}^{\lambda}(x_i - m_g)(x_i - m_g)^T \quad (2)$$

In Equation (2), C_g is the CM of the generation *g*. Equation (2) calculates the deviation between each sample point and the current mean, and averages the product of these deviations to obtain the CM. One of the advantages of CMA-ES is the adaptive update mechanism of the CM, which enables the algorithm to continuously adjust the CM during the evolution process, thereby guiding the search process towards a better direction [18, 19]. CMA-ES ensures the correctness of the optimization direction by weighted updating of the CM, as shown in Equation (3):

$$C_{g+1} = (1-c_1)C_g + c_1\left(\frac{1}{\lambda}\sum_{i=1}^{\mu}(x_i - m_g)(x_i - m_g)^T\right) \quad (3)$$

In Equation (3), C_{g+1} is the updated CM. c_1 is the learning rate of CM update. μ is the number of selected individuals. In the research on predicting information anxiety among CSs, the update of CMA-ES algorithm is used to calculate the correlation of the population and strike a harmonious equilibrium between global search and local search. When the CM of the current solution approaches the identity matrix, it suggests a lack of a significant relationship between the various dimensions of the population, and the search process is relatively independent. At this point, CMA-ES gains more degrees of freedom to explore the entire search space, thereby achieving global search. On the contrary, if the disaggregation set of the current population is in a certain region, the CM will reflect the strong correlation between these solutions, and the search process will focus on the vicinity of the current optimal solution, thereby improving convergence speed and achieving local search [20, 21]. The step size update mechanism of CMA-ES utilizes evolutionary paths and evolutionary history to adjust the step size. In each generation, the update of step size is based on the current state of the evolutionary path, as shown in Equation (4):

$$p_\sigma(g+1) = (1-c_\sigma)p_\sigma(g) + \sqrt{c_\sigma(2-c_\sigma)}\cdot\frac{x_{best} - m(g)}{\sigma(g)} \quad (4)$$

In Equation (4), $p_\sigma(g)$ is the evolutionary path of the generation g . c_σ is the learning rate of evolutionary path updates. x_{best} is the best sample in the current population. $m(g)$ is the mean vector of the generation g $\sigma(g)$ is the step size of the generation *g*. During the step size adjustment process, CMA-ES dynamically adjusts the step size by calculating the length of the current evolutionary path, as shown in Equation (5):

$$\sigma(g+1) = \sigma(g)\cdot\exp\left(c_\sigma\cdot\frac{\| p_\sigma(g+1)\|}{\| N(0,1)\|} - 1\right) \quad (5)$$

In Equation (5), $\| p_\sigma(g+1)\|$ is the length of the evolutionary path. $\| N(0,1)\|$ is the norm of the standard normal distribution. In the context of predicting information anxiety among CSs, the step size update mechanism of the CMA-ES algorithm dynamically adjusts the step size to strike a balance between exploring local optima and conducting a global search. When the search process nears the optimal solution, CMA-ES accelerates convergence by incrementally increasing the step size, thereby enhancing the search speed while maintaining the accuracy of the final solution. Conversely, when the search space is expansive, CMA-ES retains a substantial step size to preserve its global search capabilities and prevent getting trapped in local optima during the search process. This adaptive mechanism empowers CMA-ES to dynamically tailor its search strategy at various stages of the search, effectively mitigating the risk of falling into local optima and enhancing the precision and reliability of the prediction model [22]. The CMA-ES algorithm flowchart is shown in Figure 3.

In Figure 3, key parameters such as population size, mean, CM, and learning rate are set during the initialization phase. Subsequently, an initial population is generated grounded on the current parameters, and fitness evaluation is performed to determine the performance of each individual. Individuals with high fitness are selected for recombination and mutation to increase population diversity and explore new solutions. Next, the algorithm updates the CM and stride to guide the population towards a more optimal solution. The algorithm continues to iterate until the maximum number of iterations or convergence condition is met. Ultimately, the optimal solution is output as the final answer to the problem.

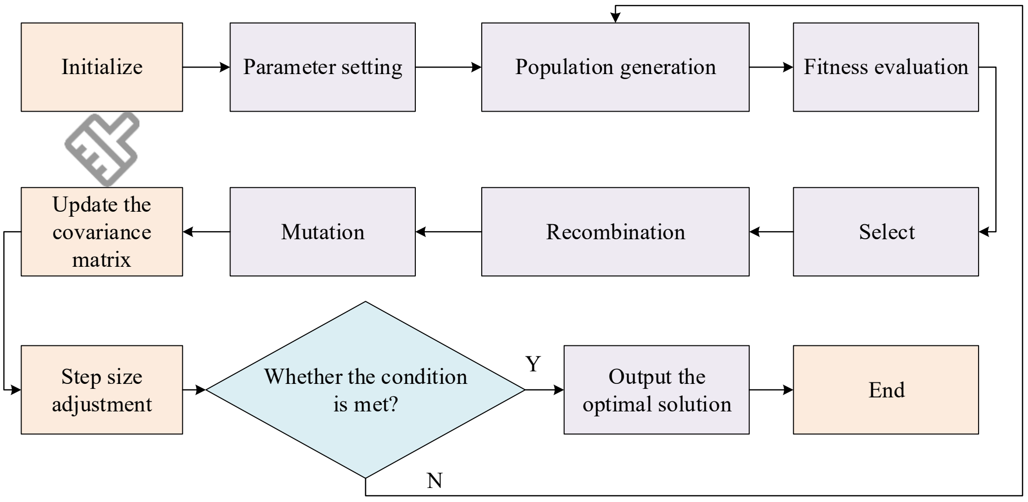

Fig. 3: Flowchart of the CMA-ES Algorithm

Construction of a CS Anxiety Prediction Model Integrating CMA-ES and XGBoost Algorithms

After a thorough analysis of the mathematical model and optimization mechanism of the CMA-ES algorithm, the study further explored how to combine this efficient optimization algorithm with the XGBoost algorithm to construct a fusion model that can accurately predict information anxiety among CSs. XGBoost is an ensemble learning method grounded on gradient boosting, which mainly uses iterative construction of decision trees for prediction. In each iteration, XGBoost learns complex patterns in the data by minimizing the Loss Function (LF). The advantage of XGBoost lies in its ability to handle missing values, support regularization, and have high training speed and prediction accuracy [23]. The training flowchart of XGBoost model is in Figure 4.

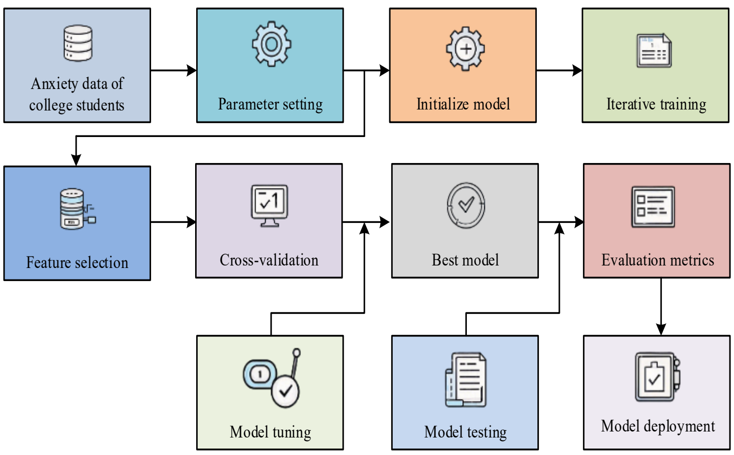

Fig. 4: XGBoost Model Training Flow Chart

The training process of XGBoost is considered as optimizing an objective function. The objective function is usually the accuracy or F1 value of XGBoost on the validation set, as shown in Equation (6):

$$f(\theta) = accuracy(X_{valid}, y_{valid}, \theta) \qquad (6)$$

In Equation (6), $\theta = (\theta_1, \theta_2, ..., \theta_k)$ represents the hyperparameter vector, X_{valid} and y_{valid} are the features and labels of the validation set, respectively. The goal of XGBoost is to minimize an LF, which includes two parts: the LF and the regularization term. The objective function of XGBoost is in Equation (7):

$$\mathrm{L}(\theta) = \sum_{i=1}^{n} \mathrm{L}(y_i, \hat{y}_i) + \sum_{k=1}^{K} \Omega(f_k) \qquad (7)$$

In Equation (7), $L(y_i, \hat{y}_i)$ is the LF, usually Mean Square Error (MSE) or cross entropy. $\Omega(f_k)$ is a regularization term used to control model complexity and avoid overfitting. The regularization term is defined as shown in Equation (8):

$$\Omega(f_k) = \gamma T + \frac{1}{2}\lambda \| \mathbf{w}_k \|^2 \quad (8)$$

In Equation (8), γ is the parameter that controls the complexity of the tree. T is the number of leaves on a tree. $\mathbf{w}_k$ is the weight of the tree. During the training process of XGBoost, the primary objective is to iteratively refine the LF to determine the optimal model parameters. In each iteration, XGBoost employs gradient boosting to compute the gradient of the LF with respect to the model parameters and subsequently adjusts the tree structure. To optimize the LF, XGBoost minimizes training errors by fitting residuals during the learning phase of each tree, as illustrated in Equation (9):

$$\theta_{t+1} = \theta_t - \eta \nabla_\theta L(\theta_t) \quad (9)$$

In Equation (9), θ_t is the hyperparameter of the generation t η is the learning rate. $\nabla_\theta L(\theta_t)$ is the gradient of the hyperparameters of the objective function. Then, the hyperparameters are optimized using CMA-ES, as shown in Equation (10):

$$\theta_{new} = \mu + \sigma \cdot z \quad (10)$$

In Equation (10), μ is the mean of the current population. $z \sim N(0, C)$ is a disturbance vector sampled from a multivariate normal distribution. Combining CMA-ES and XGBoost can improve the accuracy and robustness of mental health status prediction models. When dealing with high-dimensional and complex mental health data, this combination can provide more accurate predictions of anxiety among CSs. The flowchart of CMA-ES-XGBoost algorithm is shown in Figure 5.

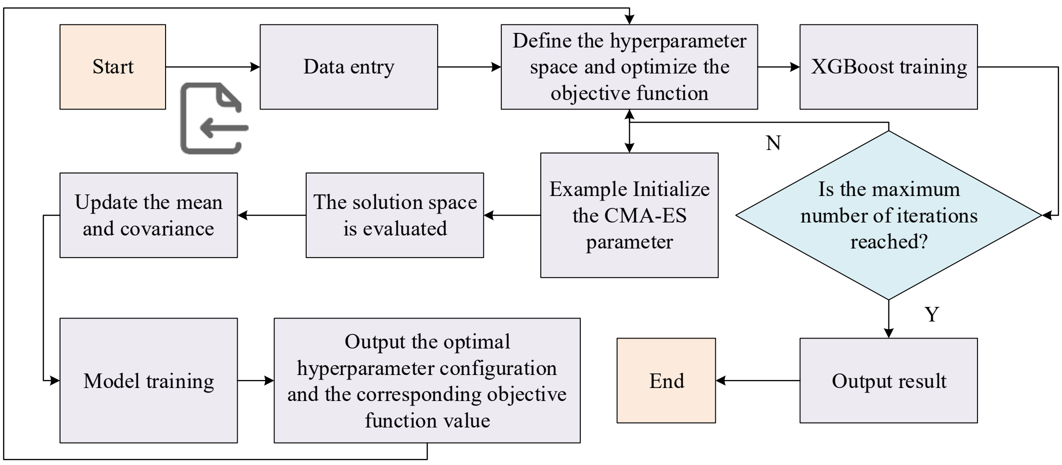

Fig. 5: Flowchart of the CMA-ES-XGBoost Algorithm

In Figure 5, starting from the starting node, data input is first performed, followed by defining a hyperparameter space and optimizing the objective function, followed by training the XGBoost model. Subsequently, the algorithm checks if the max number of iterations is reached. If not, it initializes the CMA-ES parameters, evaluates the solution space, updates the mean and covariance, and then returns to continue training the model. If the max number of iterations is reached, the optimal hyperparameter configuration and corresponding objective function values are output, and the prediction results of CS anxiety are finally ended and output.

Results

A comprehensive performance evaluation was conducted on a university student information anxiety prediction model constructed by integrating CMA-ES and XGBoost algorithms. Through comparative analysis with multiple algorithms, the performance of the model in key indicators such as accuracy, recall, F1 value, regression fit, and AUC value was demonstrated. Furthermore, the prediction effect and execution efficiency of the model under different experimental conditions were analyzed, providing data support for verifying the effectiveness and practicality of the model.

Performance Evaluation of Integrating CMA-ES and XGBoost Algorithms

The research used the fusion of CMA-ES and XGBoost algorithms to construct a prediction model for CSs' information anxiety, and improved the model's predictive performance by optimizing hyperparameters, to achieve accurate prediction of CSs' information anxiety status. The research was based on the data of 1,200 questionnaires on the information behavior of CSs in a certain university. The sample covers multiple grades and majors of the school. Among them, freshmen account for 25%, sophomores for 28%, juniors for 27%, and seniors for 20%. Students majoring in science and engineering account for 45%, those majoring in liberal arts account for 35%, and those majoring in other fields account for 20%. The questionnaire design underwent strict reliability and validity tests. The Cronbach's α value reached 0.85, indicating that the questionnaire has a high internal consistency reliability and can reliably measure the information anxiety level of CSs. Meanwhile, the structural validity of the questionnaire was verified through methods such as factor analysis, ensuring the scientificity and validity of the scale. In the data preprocessing stage, clear data cleaning standards were studied and formulated. For missing values, a similarity-based filling method was adopted, that is, reasonable filling was carried out based on other complete information of the same student and the average value of students in the same grade and major to ensure the integrity and consistency of the data. In addition, outliers were detected and processed, and data records that were obviously illogical were eliminated to ensure the accuracy of subsequent analysis. The experimental parameter configuration is in Table 1.

Table 1: Experimental Parameters

Parameter class	Parameter name	Parameter meaning	Initial value range	Optimized value
XGBoost	learning_rate	Learning rate	[0.01, 0.3]	/
	n_estimators	Number of base learners	[50, 500]	/
	max_depth	Maximum depth	[3, 10]	/
	min_child_weight	Minimum weight of a leaf node	[1, 10]	/
	gamma	Penalty parameter	[0, 1]	/
	subsample	Sample proportion	[0.5, 1]	/
	colsample_bytree	Feature sampling ratio	[0.5, 1]	/
	lambda	L2 regularization parameter	[0.1, 10]	/
	alpha	L1 regularization parameter	[0, 1]	/
	scale_pos_weight	Positive and negative sample weights	[0.5, 2]	/
CMA-ES	population_size	Population size	/	50
	max_iter	Maximum iterations	/	300
	tol	Convergence tolerance	/	1e-6
	sigma	Initial step size	/	0.5

In Table 1, the XGBoost algorithm was used as the core model for predicting information anxiety among CSs in the study, and the CMA-ES was introduced to optimize the hyperparameters of XGBoost. The setting of these parameters could balance the efficiency and accuracy of the optimization process, ensuring that the algorithm could find the global optimal solution within a limited number of iterations, while avoiding premature convergence or falling into local optima. During the experiment, the CMA-ES algorithm significantly improved the predictive performance of the model by iteratively optimizing the hyperparameters of XGBoost. The CMA-ES-XGBoost algorithm was compared and analyzed with Grid Search Gradient Boosting Tree (GS-GBT), Bayesian Optimization Gradient Boosting Tree (BO-GBT), Multilayer Perceptron (MLP), and Genetic Algorithm Gradient Boosting Tree (GA-GBT). The accuracy and recall of several algorithms are shown in Figure 6.

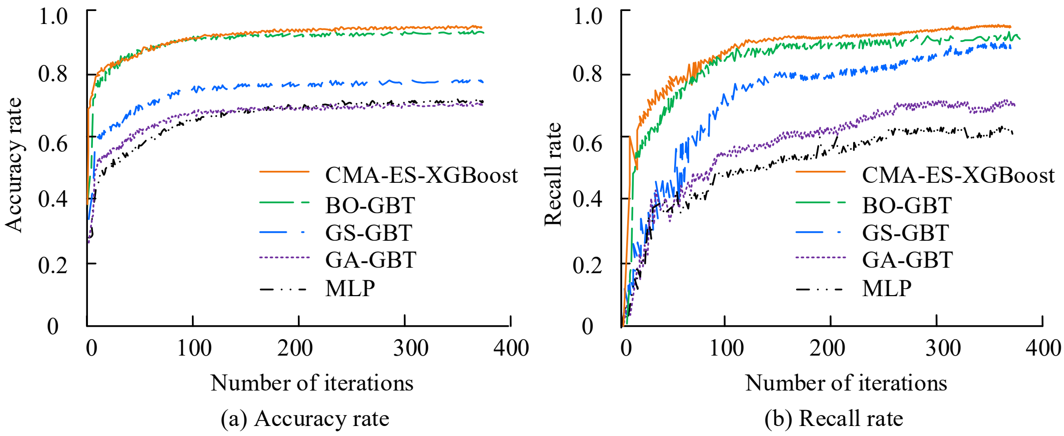

Fig. 6: Accuracy and Recall Rates of Several Algorithms

In Figure 6 (a), the CMA-ES-XGBoost algorithm had the fastest accuracy improvement speed, reaching a high accuracy even with fewer iterations, and stabilizing at a high level close to 0.96 after 400 iterations. In contrast, the accuracy improvement speed of BO-GBT, GS-GBT, and GA-GBT algorithms was slower and fluctuated greatly during the iteration process, and the final accuracy did not reach the level of CMA-ES-XGBoost. In Figure 6 (b), the recall rate of the CMA-ES-XGBoost algorithm also performed well, reaching a high level even with fewer iterations, and stabilizing at a high level close to 0.97 after 400 iterations. The CMA-ES-XGBoost algorithm had higher predictive performance and robustness when dealing with the problem of predicting information anxiety among CSs. The regression results of Macro-F1, Micro-F1, and CMA-ES-XGBoost algorithms for several algorithms are shown in Figure 7.

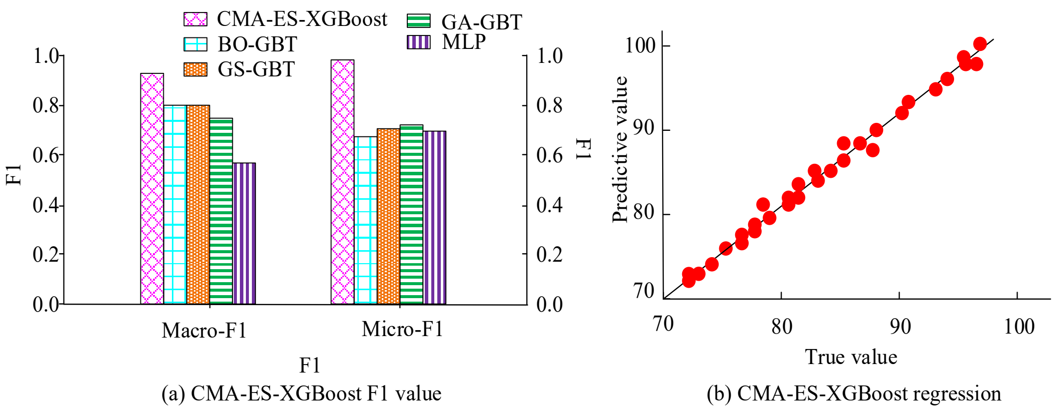

Fig. 7: Macro-F1 and Micro-F1 of Several Algorithms

In Figure 7 (a), on the Macro-F1 metric, the CMA-ES-XGBoost algorithm led other algorithms with a score of 0.96, demonstrating its superior performance in handling imbalanced datasets. On the Micro-F1 metric, the CMA-ES-XGBoost algorithm also led with a high score of 0.98, indicating its outstanding performance on the overall dataset. The CMA-ES-XGBoost algorithm performed well in both category balance and overall performance when dealing with the problem of predicting information anxiety among CSs, demonstrating its potential for application in this field. In Figure 7 (b), the CMA-ES-XGBoost algorithm had high prediction accuracy. The data points were tightly clustered around the regression line, indicating that the deviation between the predicted values and the true values was very small, thus reflecting the superior performance of the CMA-ES-XGBoost algorithm in fitting data. This tight fitting indicated that the algorithm could effectively capture patterns and trends in the data while maintaining good generalization ability to new data. The error of CMA-ES-XGBoost algorithm is shown in Figure 8.

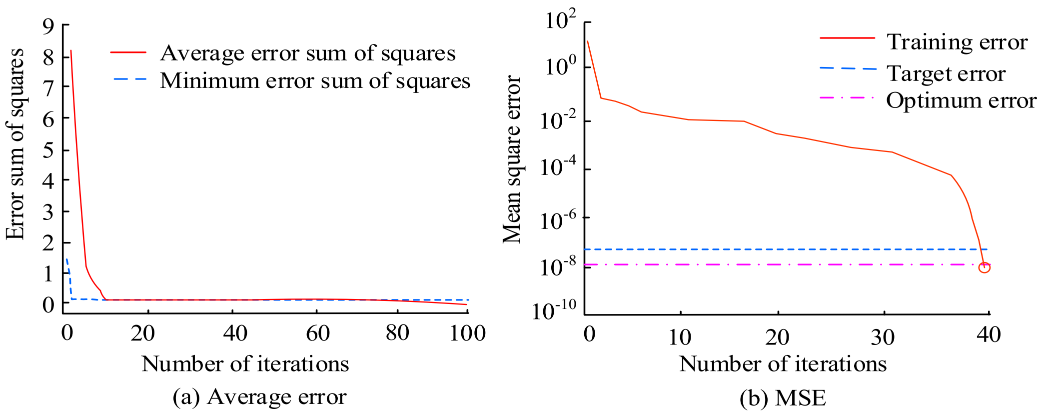

Fig. 8: Error Diagram of CMA-ES-XGBoost Algorithm

In Figure 8 (a), as the number of iterations rose, the average error sum of squares rapidly decreased and stabilized after approximately 20 iterations, approaching the minimum error sum of squares. This indicated that the CMA-ES-XGBoost algorithm could converge quickly and achieve lower error levels in fewer iterations. In Figure 8 (b), the training error gradually decreased with the rise of iteration times and tended to stabilize after about 40 iterations, approaching the target error and optimum error. This indicated that the CMA-ES-XGBoost algorithm could not only fit training data, but also maintained good generalization ability to new data, avoiding overfitting. The AUC value of several algorithms are in Figure 9.

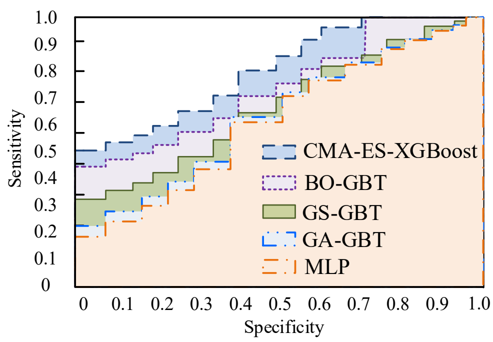

Fig. 9: AUC Value of Several Kinds of Algorithm

In Figure 9, the AUC value of the CMA-ES-XGBoost algorithm was close to 1, confirming its superior classification performance. The CMA-ES-XGBoost algorithm performed the best among several algorithms, with the highest AUC value, indicating its best classification ability in predicting information anxiety among CSs. To demonstrate more comprehensively the performance advantages of the CSs' information anxiety prediction model integrating the CMA-ES and XGBoost algorithms, the study further compared it with logistic regression and random forest. The convergence of the models of several algorithms is shown in Figure 10.

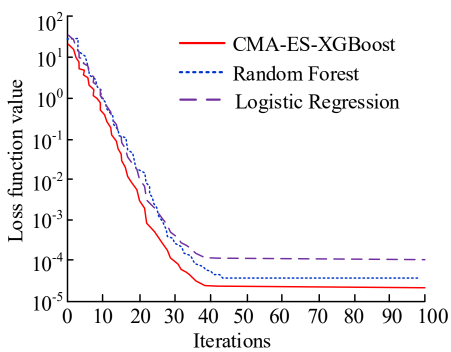

Fig. 10: Convergence curve of the model

In Figure 10, the CMA-ES-XGBoost model converged to 10-5 after 40 iterations, significantly lower than other algorithms, indicating that this model had extremely high efficiency and rapid convergence ability during the optimization process. This characteristic of rapid convergence was mainly attributed to the dynamic adjustment mechanism of the CMA-ES algorithm and the efficient optimization ability of the XGBoost algorithm. The performance comparison of several algorithms is shown in Table 2.

Table 2: Performance Comparison Table

Model type	Logistic Regression	Random Forest	CMA-ES-XGBoost
Accuracy	0.80	0.85	0.96
Recall	0.78	0.83	0.97
Macro-F1	0.79	0.84	0.96
Micro-F1	0.81	0.86	0.98
AUC value	0.85	0.90	0.99
Model convergence time (seconds)	100-120	120-140	130-155

In Table 2, the accuracy rate of the CMA-ES-XGBoost model reached 0.96 and the recall rate was 0.97, both of which were higher than 0.80 and 0.78 of logistic regression and 0.85 and 0.83 of random forest. In terms of F1 values, the Macro-F1 and Micro-F1 of CMA-ES-XGBoost were 0.96 and 0.98 respectively, which were much higher than 0.79 and 0.81 of logistic regression and 0.84 and 0.86 of random forest. This indicated that it had stronger performance when dealing with imbalanced datasets. Furthermore, the AUC value of the CMA-ES-XGBoost model was close to 0.99, which was significantly better than 0.85 of logistic regression and 0.90 of random forest, further verifying its excellent classification ability. Although the convergence time of the CMA-ES-XGBoost model was slightly longer, ranging from 130 to 155 seconds, its improvement in prediction accuracy and generalization ability made it more valuable in the task of predicting information anxiety among CSs. These results indicated that the CMA-ES-XGBoost model could provide more efficient and accurate technical support for the early warning system of mental health in colleges and universities, and was significantly superior to the traditional logistic regression and random forest models.

Analysis of the Performance of a Prediction Model Integrating CMA-ES and XGBoost Algorithms

Finally, the study analyzed the performance of the prediction model that integrates CMA-ES and XGBoost algorithms. The experimental results of predicting information anxiety among CSs based on CMA-ES-XGBoost are in Table 3.

Table 3: Experimental Outcomes of Information Anxiety Prediction of CSs Grounded on CMA-ES-XGBoost

Experiment ID	Age (years)	Study time (hours/week)	Social media usage (hours/day)	Anxiety level (scale 0-10)	Sleep duration (hours/night)	Focus score (0-100)	Predicted anxiety level (CMA-ES-XGBoost)	Prediction accuracy (%)	Model convergence time (seconds)
1	20.45	15.50	3.5	7.4	6.5	60.4	7.6	87.3	145.2
2	21.23	18.75	4.2	6.8	7.0	70.1	6.9	85.5	139.8
3	19.67	12.30	2.9	8.1	6.2	58.3	8.0	88.2	151.5
4	22.11	14.80	3.7	6.5	6.8	64.7	6.4	83.1	132.4
5	21.89	16.60	3.0	7.2	7.1	67.2	7.3	84.7	142.7

As shown in Table 3, age and learning time were used as independent variables during model training, while taking into account social media usage time and sleep patterns. Anxiety level (0-10 points) is the self-reported level of anxiety among participants, used to evaluate the accuracy of model predictions. The anxiety level predicted by CMA-ES-XGBoost was the predicted anxiety value generated after model training. The prediction accuracy reflected the degree to which the model's predicted value was close to the actual reported anxiety level, with an accuracy range of 83.1% to 88.2%. The convergence time of a model referred to the time it took for the model to converge during the training process, ranging from 132.4 seconds to 151.5 seconds.

Table 4: Summary of Experimental Results of Stress Prediction of CSs

Experiment ID	GPA	Workload (hours/week)	Social media engagement (time/week)	Mental health support (0-10)	Hobby time (hours/week)	Sleep efficiency (%)	Predicted stress (CMA-ES-XGBoost)	Prediction accuracy (%)	Model execution time (seconds)
1	3.7	10.2	15.3	7.3	8.5	83.2	6.8	86.7	125.8
2	3.4	15.4	18.2	6.9	6.3	78.9	7.2	85.2	139.4
3	3.9	12.5	12.4	8.2	7.8	87.4	6.4	88.5	118.7
4	3.6	14.8	14.5	7.1	5.9	79.6	6.9	84.1	130.2
5	3.8	13.1	17.0	7.4	9.1	84.1	6.6	86.9	121.6

Table 4 summarizes the experimental results of using the CMA-ES-XGBoost model to predict stress among CSs. The table lists the students' GPA, weekly workload, social media participation time, mental health support score, weekly hobby time, sleep efficiency, model predicted stress level, prediction accuracy, and model execution time for five experiments. The outcomes indicated that the prediction accuracy ranged from 84.1% to 88.5%, and the model execution time varied from 118.7 seconds to 139.4 seconds, indicating that the model had high prediction accuracy and acceptable execution efficiency in different situations.

Conclusion

The research aimed to construct an efficient and accurate prediction model for information anxiety among CSs, providing technical support for early warning systems for mental health. The study adopted a combination of CMA-ES and XGBoost algorithms to dynamically adjust the CM to optimize hyperparameters, thereby improving the prediction accuracy and generalization ability of the model. The experiment was conducted based on 1200 questionnaires on information behavior of CSs from a certain university, covering multidimensional data such as students' basic information, information behavior characteristics, and self-assessment scores of information anxiety scales. The research results indicated that the CMA-ES-XGBoost model performed well in multiple key indicators. During the iteration process, the model accuracy quickly improved to 0.958 and the recall rate reached 0.967, significantly better than other compared algorithms. On the Macro-F1 and Micro-F1 metrics, the model achieved 0.955 and 0.978 respectively, demonstrating superior performance in handling imbalanced datasets. In addition, the AUC value of the model was close to 0.99, further verifying its excellent classification ability. In actual prediction, the prediction accuracy of the model remained between 84.2% and 89.3% under different experimental conditions, with an average prediction accuracy of 86.9%. The convergence time of the model was between 130 seconds and 155 seconds, demonstrating good execution efficiency. These results fully demonstrated the efficiency and accuracy of the CMA-ES-XGBoost model in predicting information anxiety among CSs. Given that the information behavior and psychological state of CSs may change significantly over time and in the environment, the predictive ability of the model may also be affected accordingly. Therefore, it is recommended that the model be updated and retrained regularly in practical applications to ensure that it can adapt to the dynamic changes in students' behaviors. Although significant progress has been made in the field of predicting information anxiety among CSs, there are still some limitations. Firstly, the research data were only derived from 1,200 questionnaires of a certain university. The regions and types of institutions of the samples were relatively single, which might have limited the universality of the model. Future research can consider introducing more data from universities, covering institutions in different regions and at different levels, to further verify the generalization ability of the model. Although the dynamic adaptability of the model has been improved to a certain extent through the online learning mechanism, when facing the rapidly changing information environment and student behavior patterns, the real-time update mechanism of the model may need to be further optimized to ensure its long-term effectiveness. In the future, combined with the actual needs of mental health education, further research can be conducted on how to better transform the prediction results into effective intervention strategies.

Funding Information

The authors received no external funding for this research.

DOI: 10.3844/ajbbsp.2026.22.01.013

Ethics

This article is original and contains unpublished material. The corresponding author confirms that all of the other authors have read and approved the manuscript and no ethical issues involved.

Data Availability Statement

All data generated or analysed during this study are included in this article.

Conflict of Interest

The authors declare that they are no competing interests.

References

1. Lyngdoh T, El-Manstrly D, Jeesha K. Social isolation and social anxiety as drivers of generation Zs willingness to share personal information on social media. Psychology & Marketing, 2023, 40(1): 5-26.
2. Brailovskaia J, Miragall M, Margraf J, Herrero R, Bañ os R M. The relationship between social media use, anxiety and burden caused by coronavirus (COVID-19) in Spain. Current Psychology, 2022, 41(10): 7441-7447.
3. Nayan M I H, Uddin M S G, Hossain M I, Alam M M, Zinnia M A, Haq I, Methun M I H. Comparison of the performance of machine learning-based algorithms for predicting depression and anxiety among University Students in Bangladesh: A result of the first wave of the COVID-19 pandemic. Asian Journal of Social Health and Behavior, 2022, 5(2): 75-84.
4. Hong J C, Cao W, Liu X, Tai K H, Zhao L. Personality traits predict the effects of Internet and academic self-efficacy on practical performance anxiety in online learning under the COVID-19 lockdown. Journal of Research on Technology in Education, 2023, 55(3): 426-440.
5. Kaya F, Aydin F, Schepman A, Rodway P, Yetişensoy O, Demir Kaya M. The roles of personality traits, AI anxiety, and demographic factors in attitudes toward artificial intelligence. International Journal of Human–Computer Interaction, 2024, 40(2): 497-514.
6. Li Z, Gao T, Tian K, Wang B. Elite-driven surrogate-assisted CMA-ES algorithm by improved lower confidence bound method. Engineering with Computers, 2023, 39(4): 2543-2563.
7. Van der Meersch V, Chuine I. Estimating process-based model parameters from species distribution data using the evolutionary algorithm CMA-ES. Methods in Ecology and Evolution, 2023, 14(7): 1808-1820.
8. Zengcong L I, Kuo T I A N, Zhang S, Bo W A N G. Efficient multi-objective CMA-ES algorithm assisted by knowledge-extraction-based variable-fidelity surrogate model. Chinese Journal of Aeronautics, 2023, 36(6): 213-232.
9. Qiu Y, Zhou J, Khandelwal M, Yang H, Yang P, Li C. Performance evaluation of hybrid WOA-XGBoost, GWO-XGBoost and BO-XGBoost models to predict blast-induced ground vibration. Engineering with Computers, 2022, 38(5): 4145-4162.
10. Asselman A, Khaldi M, Aammou S. Enhancing the prediction of student performance based on the machine learning XGBoost algorithm. Interactive Learning Environments, 2023, 31(6): 3360-3379.
11. Parlak Sert H, Başkale H. Students increased time spent on social media, and their level of coronavirus anxiety during the pandemic, predict increased social media addiction. Health Information & Libraries Journal, 2023, 40(3): 262-274.
12. Mao B, Jia X, Huang Q. How do information overload and message fatigue reduce information processing in the era of COVID-19? An ability–motivation approach. Journal of Information Science, 2024, 50(5): 1242-1254.
13. Wang C, Yuan T, Feng J, Peng X. How can leaders alleviate employees workplace anxiety caused by information overload on enterprise social media? Evidence from Chinese employees. Information Technology & People, 2023, 36(1): 224-244.
14. Wu Y, Peng X, Wang H, Jin Y, Xu D. Cooperative coevolutionary CMA-ES with landscape-aware grouping in noisy environments. IEEE Transactions on Evolutionary Computation, 2022, 27(3): 686-700.
15. Miyauchi Y, Maki A, Umeda N, Rachman D M, Akimoto Y. System parameter exploration of ship maneuvering model for automatic docking/berthing using CMA-ES. Journal of Marine Science and Technology, 2022, 27(2): 1065-1083.
16. Zhao F, Bao H, Wang L, He X, Jonrinaldi. A hybrid cooperative differential evolution assisted by CMA-ES with local search mechanism. Neural Computing and Applications, 2022, 34(9): 7173-7197.
17. Liao H P, Wang J L. The impact of epidemic information on the public's worries and attitude toward epidemic prevention measures during the COVID-19 outbreak. Current Psychology, 2023, 42(1): 145-153.
18. Guo G, Zhao S. 3D multi-object tracking with adaptive cubature Kalman filter for autonomous driving. IEEE Transactions on Intelligent Vehicles, 2022, 8(1): 512-519.
19. Feng X, Kong X, Du B, Luo J. Adaptive LII-RMPLS based data-driven process monitoring scheme for quality-relevant fault detection. Journal of Control and Decision, 2022, 9(4): 477-488.
20. Mao Y, Xu C, Chen J, Pu Y. An efficient conjugate gradient based Cholesky CMA-ES estimation algorithm for nonlinear systems. International Journal of Robust and Nonlinear Control, 2024, 34(3): 1610-1628.
21. Gheisari M, Hamidpour H, Liu Y, Saedi P, Raza A, Jalili A, Rokhsati H, Amin R. Data Mining Techniques for Web Mining: A Survey. Artificial Intelligence and Applications, 2023, 1(1): 3-10.
22. Popov A A, Zanetti R. An adaptive covariance parameterization technique for the ensemble Gaussian mixture filter. SIAM Journal on Scientific Computing, 2024, 46(3): A1949-A1971.

DOI: 10.3844/ajbbsp.2026.22.01.013

23. Wang Y Y, Wang Y S. Development and validation of an artificial intelligence anxiety scale: An initial application in predicting motivated learning behavior. Interactive Learning Environments, 2022, 30(4): 619-634.