Journal of Engineering Science and Technology Review 16 (1) (2023) 161 - 169

Research Article

JOURNAL OF
**Engineering Science and Technology Review**

www.jestr.org

# A Prediction Method of Failure Depth of Coal Seam Floor Based on FA-GWO-SVM Model

**Zhongsi Dou[1,2,*], Ruili Han[3] and Yimeng Wang[4]**

[1]*School of Civil and Architectural Engineering, East China University of Technology, Nanchang 330013, China*
[2]*Engineering Research Center for Geological Environment and Underground Space of Jiangxi Province, Jiangxi Institute of Survey and Design, Nanchang,330015, China*
[3]*School of Earth Sciences, East China University of Technology, Nanchang 330013, China*
[4]*School of Science and Engineering, University of Dundee, Dundee DD1 4HN, UK*

Received 19 July 2022; Accepted 21 December 2022

***Abstract***

Accurately predicting the failure depth of the coal seam floor is an important premise to prevent water inrush from the coal seam floor and ensure safe and efficient mining operations. Based on the coal seam floor damage degree data collected from various mining areas in China, this study selected six indexes (coal seam mining thickness, coal seam dip angle, mining depth, working face slope length, floor damage resistance and presence of a cutting fault) to predict the value of the coal seam floor damage depth. Based on Support Vector Machine(SVM) model ,the factor analysis method was applied to reduce the dimension and extract the original data variables. The extracted variables were used as the input for the SVM model. Afterwards, the Grey Wolf Optimizer (GWO) algorithm was adopted to optimize the parameters C and g, and the Factor Analysis(FA)-GWO-SVM coal seam floor failure depth prediction model was established. The reliability of the model was verified before proceeding with the investigation. Results indicate that the prediction model of the coal seam floor failure depth based on the FA-GWO-SVM method has a good generalization ability and a strong prediction performance for the new sample data. Compared with the traditional SVM model and the GWO-SVM model, it has the minimum MAPE, RMSE and MAE values. Furthermore, the learning ability, stability and prediction accuracy of the model are significantly improved. The model does not only overcome the drawbacks of the traditional prediction methods that do not consider the interaction of various factors but also simplifies the input scale of the SVM model. The issue of affecting the prediction accuracy due to the difficulty of the parameter optimization in the SVM model is solved using the GWO optimization technique, and the model's prediction accuracy and operational performance are enhanced. This study provides an effective method for accurately predicting the failure depth of the coal seam floor.

*Keywords:* Bottom plate failure depth, Factor analysis, Support vector machine, Gray wolf optimization algorithm, Floor water inrush

## 1. Introduction

Coal accounts for 62% of China's total primary energy consumption [1]. In recent years, with the gradual depletion of shallow coal resources and the continuous increase of deep coal mining, the problem of water inrush from the coal mine floor has become increasingly problematic. This resulted in a large number of casualties and property losses [2]. The failure depth of the coal seam floor refers to the maximum depth achieved by the mutual penetration of cracks in the floor under the action of mine pressure [3]. During the coal mining operation, the waterproof rock strata of the coal seam floor deform, causing floor heaving and fissures, allowing groundwater to easily flow into the mine and causing water inrush accidents [4]. As a result, the prediction and research of floor failure depth have always been the research focus in the field of mine water disaster prevention for years. The failure depth of the stope floor is the critical data used to evaluate the water resistance of the coal seam floor and design a waterproof and safe coal pillar. Accurate prediction of coal seam floor failure depth is of great significance to effectively prevent floor water inrush and ensure safe and efficient mining [5].

Several scholars have performed extensive studies on the failure depth of coal seam floors. Mirjalili et al.[6] proposed a swarm intelligence optimization algorithm to simulate the hierarchy and hunting behavior of gray wolves in their natural habitat. This algorithm is characterized by simple operations, few adjustment parameters, easy programming and etc., but the parameter optimization operations need to be taken into consideration when applied to the destruction depth of the coal seam floor. At present, the most frequently used research techniques include the Theoretical analysis method, the empirical formula calculation method, the field measurement method [7], the numerical simulation test, the physical similarity simulation test [8] and the mathematical method. Many valuable insights and conclusions have been obtained by applying these methods. However, there are still some shortcomings. For example, on-site measurement provides more accurate results, however, it has some drawbacks, such as being time-consuming, labor-intensive, and requiring a large capital investment. Theoretical analysis and empirical formula calculations show large errors in comparison with the actual scenario due to a reduced number of influencing factors. On the other hand, due to the relatively ideal model configuration and complex parameter debugging, it is often difficult to obtain an exact estimate of the bottom plate failure depth when using numerical simulation and physical

*E-mail address: hnzyjsxydzs@163.com
ISSN: 1791-2377 
doi:10.25103/jestr.161.21

similarity simulation methods. Based on the research results of roadway convergence, deformation and failure, Piotr Malkowsk et al.[9] suggested that the failure depth of the floor was primarily affected by faults. Additionally, due to the complex mechanical and physical characteristics of the coal seam floor, predicting the depth of the floor water diversion failure zone becomes a complicated and nonlinear problem influenced by a variety of factors. Islam M R et al.[10] explored the stress characteristics and deformation surrounding faults in the conveyor band roadway of the Balapuria Mine (Bangladesh) through numerical simulations by using a boundary element method. They determined that the redistribution of mining-induced stress causes significant deformation in and around the two faults, and high stress concentrates near the ends of the two faults. In contrast, the mathematical prediction technique of predicting the failure depth of coal seam floor by studying the relationship between the elements impacting the failure depth of coal seam floor has certain advantages from the perspective of many factors [11]. With the advancement of computer technology in recent years, the method of establishing prediction models through machine learning has also become popular among researchers. At present, the common machine learning prediction models used in the prediction of floor failure depth mainly include grey theory, neural network, support vector machine, etc. On this premise, various new prediction models have been developed, yielding more accurate forecast results. Among them, the Support Vector Machine (SVM) follows the principle of structural risk minimization and offers great advantages in dealing with problems such as small samples, nonlinearity and large dimensions. However, SVM is highly dependent on parameter selection and the optimization method can be used to optimize the parameters of SVM. Xu et al. optimized the parameters of SVM through Particle Swarm Optimization (PSO) algorithm and established Least-Squares Support-Vector Machines (LS-SVM) model for predicting the failure depth of coal seam floor based on PSO optimization. Zhu et al. [12] and others used the artificial bee colony algorithm to optimize the parameters of SVM, and established a prediction model of floor failure depth based on the combination of the artificial bee colony algorithm and support vector machine. Jin et al.[13] developed a comprehensive application of the genetic algorithm and PSO algorithm to the SVM model, and established the evaluation model of coal seam floor damage degree based on Genetic Algorithm-Particle Swarm Optimization–SVM (GAPSO-SVM). The SVM model optimized by these algorithms achieved good prediction results in the prediction of coal seam floor failure depth [14], however, according to the existing research results, the convergence speed of the genetic algorithm is slow. The ant colony algorithm has strong adaptability to big data samples, but the convergence of the algorithm is slow and its implementation is complex. Gray Wolf Optimization (GWO) algorithm outperforms PSO algorithm, genetic algorithm and other intelligent optimization algorithms in global optimization. It has the advantages of easy implementation, fast convergence and high precision.

Furthermore, there are many factors influencing the failure depth of coal seam floor, the majority of which are accompanied by noise, which directly affects the accuracy of floor failure depth prediction When using numerous elements to construct a prediction model of floor failure depth in the past, most of them did not address the impact of the overlapping information between the influencing factors on the prediction results.

Considering the above-mentioned research gaps, this study uses the unique advantages of SVM in processing small samples and nonlinear data and employs the SVM model as the main body to predict the failure depth of coal seam floor. Firstly, factor analysis is used to reduce the dimension and process the influencing factors of coal seam floor failure depth, minimizing overlapping information between influencing factors. The processed data is used as the input vector of SVM. Simultaneously, GWO is used to optimize the penalty parameter C and kernel function parameter g in SVM in order to improve the fitting ability of the model. Finally, a prediction model of coal seam floor failure depth is developed based on FA-GWO-SVM.

## 2. Materials and Methods

### 2.1 Factor analysis

Factor analysis is a kind of multivariate statistical analysis technique. It is an advanced version of the principal component analysis method. It is a statistical method that uses a few factors to describe the relationship between numerous indicators or factors and uses a few factors to reflect the majority of the information of the original data [15]. Factor analysis is the best synthesis and simplification of multivariable plane data. Based on the principle of ensuring the least loss of data information, the dimensionality of high-dimensional variable space is reduced, which not only reasonably explains the correlation between factors but also simplifies the observation system [16]. According to the correlation matrix of variable x, the original P variables can be expressed as the linear combination of M (m < p) new variables, and its mathematical model is

$$\begin{cases} x_1 = a_{11}f_1 + a_{12}f_2 + \cdots + a_{1m}f_m + e_1 \\ x_2 = a_{21}f_1 + a_{22}f_2 + \cdots + a_{2m}f_m + e_2 \\ \vdots \\ x_p = a_{p1}f_1 + a_{p2}f_2 + \cdots + a_{pm}f_m + e_p \end{cases} \tag{1}$$

It can be expressed in the matrix form as,

$$X = AF + a\varepsilon \tag{2}$$

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \vdots & \vdots & \vdots & \vdots \\ a_{p1} & a_{p2} & \cdots & a_{pm} \end{bmatrix} \tag{3}$$

Where $f_j$ is the common factor of two orthogonal; $e_i$ is a special factor; $a_{ij}$ is a load of common factor and A is the load matrix of common factor.

The factor analysis has been conducted based on specific steps which are as follows [17-18]:

1) The data with Z-score are standardized.

2) The covariance matrix, i.e. correlation matrix R, is calculated according to matrix X.

3) According to the covariance matrix, the eigenvalues and their corresponding eigenvectors are calculated.

4) The cumulative percentage of variance of the previous q eigenvalues is found to be greater than 80% which is the

principle based on which the number of common factors is determined.

5) The factor is rotated and the factor load matrix A is calculated.

6) The factor score model is then established and solved.

### 2.2 SVM regression algorithm

Support vector machine (SVM) is a regression method based on statistical learning theory. The fundamental basis of this method is to map the input data (low dimensional space) into a high-dimensional feature space and then construct a kernel function to form the linear regression function [19]. It is capable of solving linear regression and nonlinear regression problems. It has unique advantages in solving small sample, nonlinear and high-dimensional problems.

Suppose the training sample is $\{(x_1, y_1), (x_2, y_2)...(x_n, y_n)\}$. For the prediction of floor failure depth, xi in the sample is the influencing factor and $y_i$ is the effect quantity. SVM will have to establish an optimal functional relationship $y = f(x)$ through a given sample to fit the relationship between the influence quantity and the effect quantity. The prediction of floor failure depth is a nonlinear problem, which can be mapped to high-dimensional space as a linear problem. The regression function is expressed as:

$$f(x_i) = \omega \bullet \varphi(x_i) + b \tag{4}$$

Where: $\omega \bullet \varphi(x_i)$ is the inner product of vectors $\omega$ and $\varphi(x_i)$; $\omega$ is the regression coefficient; $\varphi(x_i)$ is the mapping function from input space to feature space; b is the threshold.

Here, the relaxation variable $\xi, \xi \geq 0$ is introduced to solve w and b. according to the SRM criterion, equation (4) is transformed into a convex quadratic programming problem:

$$\min_{\omega,\xi} \frac{1}{2}\omega^T\omega + C\sum_{i=1}^{n}(\xi + \xi^*) \tag{5}$$

$$s.t. \begin{cases} \omega \bullet \phi(x) + b - y_i \leq \varepsilon + \xi_i \\ yi - \omega \bullet \phi(x) - b \leq \varepsilon + \xi^* \end{cases} (i = 1, 2, \cdots, n) \tag{6}$$

Where, the regularization parameter C is the penalty factor; ε is the insensitive loss function.

The Lagrange function solution formula (6) is introduced to convert the inner product operation of high-dimensional space into the original two-dimensional space calculation through kernel function $k(x_i, y_j)$, including:

$$\omega = \sum_{i=1}^{n}(\alpha_i - \alpha_i^*)\phi(x_i) \tag{7}$$

The regression function of the obtained SVM regression model is:

$$\begin{aligned} f(x) &= \omega \bullet \phi(x) + b \\ &= \sum_{i=1}^{n}(\alpha_i - \alpha_i^*)k(x_i, x_j) + b \end{aligned} \tag{8}$$

Where, $\alpha_i$ and $\alpha_i^*$ are Lagrange multipliers and weight vectors required for optimization respectively. If $\alpha_i$ is not zero or $\alpha_i^*$ is not zero, it means that this sample is a support vector.

In the case of nonlinearity, the general kernel function is Gaussian radial basis function (RBF), and its expression is as follows (9)

$$k(x_i, x_j) = \exp(-g\left\|x_i - x_j\right\|^2), g > 0 \tag{9}$$

Where, $g$ is the gamma parameter function setting (if $k$ is the number of attributes, g defaults to $1/k$).

The values of the penalty factor *C* and the kernel function parameter *g* in the support vector machine model have a significant impact on the accuracy of training and prediction data [20]. The tolerance of error is represented by penalty factor *C*. Too large or too small values of *C* can result in either over fitting or under fitting which is a bad impact. After mapping to the new feature space, parameter *g* determines the distribution of data implicitly. The size of g is proportional to the number of support vectors. The prediction ability of SVM is visibly influenced by penalty factor *C* and kernel function parameter *g*. This necessitates the optimization of the parameters, *C* and *g*, to obtain a support vector machine with high prediction accuracy.

### 2.3 GWO algorithm

#### 2.3.1 Overview of GWO algorithm

GWO algorithm is a swarm intelligence algorithm that constantly calculates the optimal value in an iterative way by replicating the hierarchy and predation strategy of wolves. The algorithm seeks the optimal position among multiple optimal solutions, has strong global search ability, effectively reduces the probability of falling into the local extremum, and requires fewer parameters to be adjusted. Further, it has the advantages of fast convergence and high precision. At present, it has been widely used in the engineering field. Existing research results show that the GWO algorithm performs better than the PSO algorithm, genetic algorithm, and other intelligent optimization algorithms in global optimization. The life habits of gray wolves are mostly gregarious, with obvious social attributes, and there is a strict hierarchy within them. According to the social level, individuals in the gray wolf group can be divided into four categories, thus forming a hierarchical pyramid structure, as shown in Figure 1. The wolf on the first floor of the pyramid is the leader wolf, which is defined as $\alpha$. They have the right to decide all major issues of the whole wolf pack. The wolf on the second layer is represented by $\beta$, who assists the leader to make decisions. The wolf on the third layer is represented by $\delta$. They are responsible for sentry, reconnaissance and other tasks. The wolf at the lowest level is represented by $\omega$. In terms of activity, the clan is always under the command of the first three levels of gray wolf. When hunting the prey, grey wolves divide their predation process into three stages: search, encirclement and attack. $\alpha, \beta$, and $\delta$ wolves can more accurately grasp the position of their prey. Whereas, ω, under the command of the first three wolves, jointly complete the task of herding up the prey.

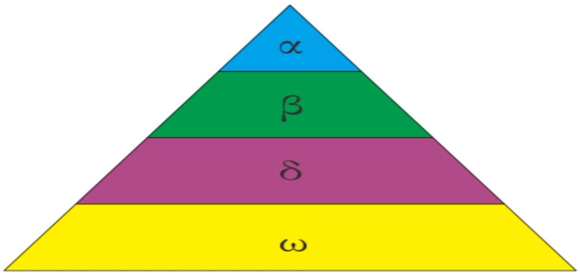


**Fig. 1.** Experimental wolf rank pyramid

### 2.3.2 Mathematical model of GWO algorithm
The gray wolf population size is defined as N and the search space dimension is D. The gray wolf population can be expressed as: $X = \{X_1, X_2, ..., X_n\}$ ,in which the position of the ith gray wolf can be expressed as $X_i = \{X_i^1, X_i^2, ..., X_i^D\}$, and the position of each gray wolf represents a solution of the problem. The gray wolf with the best current position is represented as α, the second best gray wolf is represented as $\beta$, the third best gray wolf is represented as $\beta$, and the remaining gray wolves are represented as $\omega$. The position corresponding to the globally optimal grey e wolf of the algorithm is the position of the prey. Let the distance between wolves and prey be D, then [24-25]:

$$D = \left| C * X_p(t) - X(t) \right| \tag{10}$$

$$X(t+1) = X_p(t) - A * D \tag{11}$$

$$A = 2 * q * r_2 - q \tag{12}$$

$$C = 2 * r_1 \tag{13}$$

Where, *t* is the number of iterations; $X_p(t)$ is the location of prey; $X(t)$ is the position of the t-generation gray wolf; $X(t+1)$ is the position of the $t+1$ gray wolf individual, so as to update the gray wolf position; $A$ and $C$ are coefficient vectors. By adjusting these two vectors, wolves can reach different positions around prey; $q$ is the number that decreases linearly from 2 to 0 in the iterative process; $r_1$ and y are random numbers between [0~1].

After each generation is updated, the first three solutions $(X_\alpha(t), X_\beta(t), X_\delta(t))$ with the lowest fitness value in history are selected as the position of current generation $(X_\alpha(t), X_\beta(t), X_\delta(t))$ through the calculation of fitness value. The next generation of gray wolf individuals takes $\alpha, \beta, \delta$ as the traction and update the position through equation (14). The predation process of gray wolf group is shown in Figure 2.

$$\begin{aligned} X_\alpha(t+1) &= X_\alpha(t) - A_1 \left| C_1 X_\alpha(t) - X(t) \right| \\ X_\beta(t+1) &= X_\beta(t) - A_2 \left| C_2 X_\beta(t) - X(t) \right| \\ X_\delta(t+1) &= X_\delta(t) - A_3 \left| C_3 X_\delta(t) - X(t) \right| \\ X(t+1) &= \frac{X_\alpha(t+1) + X_\beta(t+1) + X_\delta(t+1)}{3} \end{aligned} \tag{14}$$

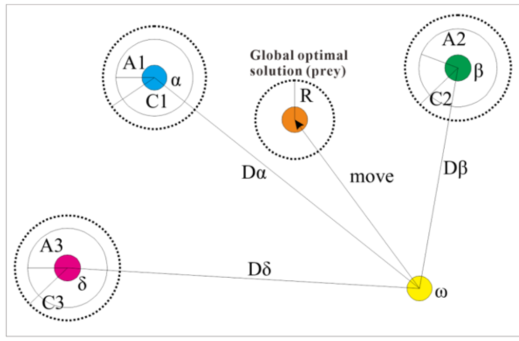


**Fig.2.** Location update during wolf hunting

### 2.4 FA-GWO-SVM model
FA-GWO-SVM model uses the SVM model as its main body. After determining the factors influencing the depth of water flowing through the fracture zone of the coal seam floor, factor analysis and processing are performed on the original sample data. A group of new variables is then extracted that can reflect the majority of the information of the original data. This will minimize the redundant information and noise, eliminate the interference of correlation between various influencing factors on the prediction results, and optimize the input variables of subsequent SVM models. The new variables generated from factor analysis are then fed into the SVM algorithm for training, and the GWO algorithm is used to optimize the penalty factor C and kernel function parameter g of SVM. The optimized parameters, C and g, are used in modeling, following which the test set is predicted, and the prediction results are analyzed. The flow chart for establishing the FA-GWO-SVM model is shown in Figure 3.

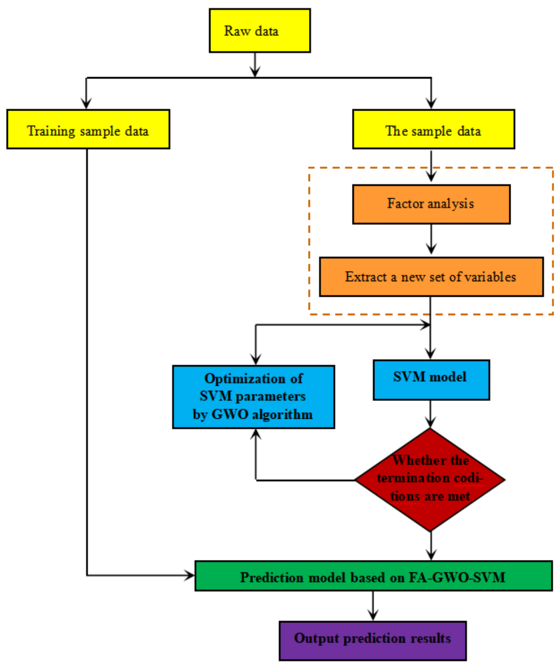


**Fig.3.** Flow chart of FA-GWO-SVM model

## 3. Result Analysis and Discussion

### 3.1 Determining influencing factors and data sources

### 3.1.1 Analysis of the influencing factors of failure depth of coal seam floor
On completion of coal seam mining, the floor rock mass gets damaged. The floor in front of the working face is affected by concentrated stress, which exceeds the ultimate strength of

floor rock mass and causes floor damage. The prediction of the depth of water flowing through the failure zone of the floor is a complex and nonlinear problem that is influenced by many factors [5]. In general, the failure depth of the floor is mainly affected by stress conditions, surrounding rock properties and geological structure [13]. The restrictive factors of stress conditions include mining depth, coal seam inclination, mining thickness, working face slope length, advancing speed, and pressure step. The natural conditions of surrounding rock are primarily the failure-resistance ability of floor rock which includes rock strength, rock combination, and original fracture rate. The geological structure conditions mainly consider whether there is a cutting through fault or fracture zone in the working face. The presence of cutting through fault or fracture zone in the floor has a significant impact on the failure of the floor. The factors affecting the failure depth of the bottom plate are as follows:

1) Mining depth: With the increase in mining depth, the self-weight of the overlying strata increases and the original rock stress of the coal seam increases. Further, the stress concentration of the floor becomes more obvious after mining, and the failure depth of the floor increases as well [21].

2) Coal seam dip angle: The dip angle of the coal seam affects the stress distribution and stress concentration of the floor, which in turn influences the failure depth of floor. Within a certain range, the greater the dip angle of the coal seam, the more obvious the mine pressure is, and the deeper the damage depth of the coal seam floor. This is because the greater the dip angle of the coal seam, the greater the tangential stress of the coal pillar on the side of the working face on the floor, and squeezing the floor easily causes it to bulge and break [5].

3) Thick mining: The greater the mining thickness of the coal seam, the bigger the deformation range of the roof, the greater the supporting stress that the coal wall and floor should bear, and the greater the damage depth of the floor [22].

4) Inclined length of working face: Within a certain range, the longer the inclined length of the working face, the larger the scope of the goaf. This in turn increases the range of rock strata in which the roof moves causing higher mine pressure and subsequent damage to the coal seam floor at higher depth. On the other hand, as the length of the working plane increases, so does the likelihood of involvement of the working plane [12].

5) The failure-resistant ability of coal seam floor: This index is a comprehensive reflection of the development of primary fractures in the floor, rock stratum combination, and rock strength. Under the same conditions, the stronger the failure-resistant ability of the floor, the less developed the primary cracks of the floor, the greater the rock strength, the less easy it is to destroy the floor and hence, the smaller the failure depth of the floor [5, 22].

6) Another influencing condition is Whether or not the working face has a cut-through fault or fracture zone. When there is a cut through fault or fracture zone in the coal seam floor, a floor fissure is developed and the overall strength is reduced. The maximum failure depth occurs near the fault zone or fracture zone. The failure depth of the floor near the fault or fracture zone increases due to the presence of a weak surface [23].

### 3.1.2 Research data sources

Six indexes are chosen as the key controlling factors affecting the failure depth of the coal seam floor based on prior study findings and the concepts of easy acquisition and unified quantification of influencing factor data [14,23]. These factors are assigned a nomenclature as follows:

Mining depth: $X_1$
Coal seam dip angle: $X_2$
Mining thickness: $X_3$
Inclined length of working face: $X_4$
Damage resistant ability of coal seam floor: $X_5$
Whether there is a cutting fault or fracture zone in the working face: $X_6$

Cut through fault or fracture zone in the working face is represented by 1, and no cut through fault or fracture zone in the working face is represented by 0. Data of coal seam floor failure depth from various mining areas in China are selected as sample data for the analysis and are presented in Table 1[23]. Among them, 1-27 groups of measured data are used as training samples and 28-30 groups are used as test samples to verify the prediction effect of the model on new samples.

**Table 1.** Measured data of failure depth of coal seam floor in mining area

| Sample number | Mining depth $X_1$/m | Coal seam dip angle $X_2$/° | Mining thickness $X_3$/m | Inclined length of working face $X_4$/m | Damage resistant ability of coal seam floor $X_5$ | Whether there is a cutting fault or fracture zone in the working face $X_6$ | Bottom plate failure depth y/m |
|---|---|---|---|---|---|---|---|
| 1 | 123 | 15 | 1.1 | 70 | 0.2 | 0 | 7 |
| 2 | 123 | 15 | 1.1 | 100 | 0.2 | 0 | 13.4 |
| 3 | 145 | 16 | 1.5 | 120 | 0.4 | 0 | 14 |
| 4 | 130 | 15 | 1.4 | 135 | 0.4 | 0 | 12 |
| 5 | 110 | 12 | 1.4 | 100 | 0.4 | 0 | 10.7 |
| 6 | 148 | 18 | 1.8 | 95 | 0.8 | 0 | 9 |
| 7 | 225 | 14 | 1.9 | 130 | 0.8 | 0 | 9.75 |
| 8 | 308 | 10 | 1 | 160 | 0.6 | 0 | 10.5 |
| 9 | 287 | 10 | 1 | 130 | 0.6 | 0 | 9.5 |
| 10 | 300 | 8 | 1.8 | 100 | 0.4 | 0 | 10 |
| 11 | 230 | 10 | 2.3 | 120 | 0.6 | 0 | 13 |
| 12 | 230 | 26 | 3.5 | 180 | 0.4 | 0 | 20 |
| 13 | 310 | 26 | 1.8 | 128 | 0.2 | 0 | 16.8 |
| 14 | 310 | 26 | 1.8 | 128 | 0.2 | 1 | 29.6 |
| 15 | 259 | 4 | 3 | 160 | 0.6 | 0 | 16.4 |
| 16 | 320 | 4 | 5.4 | 60 | 0.6 | 0 | 9.7 |
| 17 | 520 | 30 | 0.94 | 120 | 0.6 | 0 | 13 |
| 18 | 400 | 9 | 7.5 | 34 | 0.4 | 0 | 8 |
| 19 | 400 | 9 | 4 | 34 | 0.4 | 0 | 6 |
| 20 | 227 | 12 | 3.5 | 30 | 0.4 | 0 | 3.5 |
| 21 | 227 | 12 | 3.5 | 30 | 0.4 | 1 | 7 |
| 22 | 900 | 26 | 2 | 200 | 0.6 | 0 | 27 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 23 | 1000 | 30 | 2 | 200 | 0.6 | 0 | 38 |
| 24 | 200 | 10 | 1.6 | 100 | 0.2 | 0 | 8.5 |
| 25 | 375 | 14 | 2.4 | 70 | 0.6 | 0 | 9.7 |
| 26 | 375 | 14 | 2.4 | 100 | 0.6 | 0 | 12.9 |
| 27 | 118 | 18 | 2.5 | 80 | 0.2 | 0 | 10 |
| 28 | 320 | 4 | 5.4 | 100 | 0.6 | 0 | 11.7 |
| 29 | 400 | 9 | 4 | 45 | 0.4 | 0 | 6.5 |
| 30 | 327 | 12 | 2.4 | 120 | 0.6 | 0 | 11.7 |

### 3.2 Establishment of prediction model

#### 3.2.1 Factor analysis on data preprocessing

In this study, factor analysis (FA) was performed on 27 groups of training sample data from Table 1 using SPSS26 software. Firstly, KMO (Kaiser Meyer Olkin) test statistics and the Bartlett's test are used to assess the correlation between indicators to determine whether the original variables are suitable for factor analysis [24-25]. The results are shown in Table 2. The measured value from KMO is 0.518 > 0.5. From the result of Bartlett's test, the sig value is found to be 0.002 which is less than 0.05, indicating that the selected indicators and measurement data are suitable for factor analysis [25-26].

**Table 2.** KMO and Bartlett test

| **Kmo sampling suitability quantity** | | **0.518** |
|---|---|---|
| Bartlett's test | Approximate chi square | 36.231 |
| | freedom | 15 |
| | Significance | 0.002 |

Table 3 displays the results of the correlation analysis conducted between the six main control factors which affect the failure depth of coal seam floor in 27 sets of training sample data. It can be seen from Table 3 that there is a definite correlation between various factors, among which the correlation coefficients between $X_1$, $X_2$ and $X_4$ are 0.473 and 0.438, respectively. The correlation coefficient between $X_2$ and $X_4$ is 0.507, whereas, the correlation coefficient between $X_3$ and $X_4$ is -0.481. This indicates that there is a strong correlation between these factors. Due to the existence of information redundancy, the prediction of the failure depth of the coal seam floor will become more complex, making prediction accuracy difficult to guarantee. Thus, it is necessary to use the factor analysis method to treat the relevant variables as new comprehensive variables that are low dimensional, uncorrelated and can retain most of the information of the original variables.

**Table 3.** Correlation matrix of main control factors

| | $X_1$ | $X_2$ | $X_3$ | $X_4$ | $X_5$ | $X_6$ |
|---|---|---|---|---|---|---|
| $X_1$ | 1 | 0.473 | 0.096 | 0.438 | 0.333 | -0.053 |
| $X_2$ | 0.473 | 1 | -0.336 | 0.507 | -0.093 | 0.143 |
| $X_3$ | 0.096 | -0.336 | 1 | -0.481 | 0.027 | 0.054 |
| $X_4$ | 0.438 | 0.507 | -0.481 | 1 | 0.264 | -0.174 |
| $X_5$ | 0.333 | -0.093 | 0.027 | 0.264 | 1 | -0.252 |
| $X_6$ | -0.053 | 0.143 | 0.054 | -0.174 | -0.252 | 1 |

Principal component analysis was used to extract factors for $\{X_1, X_2, X_3, X_4, X_5 \text{ and } X_6\}$ using the SPSS26 software, and four principal components $\{F_1, F_2, F_3, \text{and } F_4\}$ were extracted. The explanation of the total variance is given in Table 4. The interpretation degree of the extracted main components to the original variables reaches 90.02 percent > 80 percent, as shown in Table 4. After rotating by the maximum variance method, the score coefficient matrix of each component is obtained, as shown in Table 5. Based on this, the relationship expression between the extracted four new principal components and the original variables is obtained as follows:

$$\begin{cases} F_1=0.573x_1+0.491x_2+0.216x_3+0.204x_4-0.124x_5-0.028x_6 \\ F_2=0.311x_1-0.082x_2+0.733x_3-0.352x_4-0.071x_5-0.045x_6 \\ F_3=0.105x_1-0.32x_2-0.035x_3+0.141x_4+0.911x_5+0.190x_6 \\ F_4=-0.037x_1+0.055x_2-0.069x_3-0.081x_4+0.173x_5+1.011x_6 \end{cases} \tag{15}$$

After standardizing and substituting the original data of training and test samples into equation (15), four new variables are calculated. These four new variables are then used as the input variables for the SVM model.

**Table. 4.** Interpretation of total variance

| **Component** | **Initial eigenvalue** | | | **Extract the sum of squares of loads** | | |
|---|---|---|---|---|---|---|
| | **Total** | **Percentage variance** | **accumulate %** | **Total** | **Percentage variance** | **Accumulate %** |
| 1 | 2.188 | 36.475 | 36.475 | 2.188 | 36.475 | 36.475 |
| 2 | 1.394 | 23.24 | 59.715 | 1.394 | 23.24 | 59.715 |
| 3 | 1.134 | 18.896 | 78.611 | 1.134 | 18.896 | 78.611 |
| 4 | 0.686 | 11.439 | 90.05 | 0.686 | 11.439 | 90.05 |
| 5 | 0.329 | 5.491 | 95.541 | | | |
| 6 | 0.268 | 4.459 | 100 | | | |

**Table 5.** Component score coefficient matrix

| | **Component 1** | **Component 2** | **Component 3** | **Component 4** |
|---|---|---|---|---|
| $X_1$ | 0.573 | 0.311 | 0.105 | -0.037 |
| $X_2$ | 0.491 | -0.082 | -0.32 | 0.055 |
| $X_3$ | 0.216 | 0.733 | -0.035 | -0.069 |
| $X_4$ | 0.204 | -0.352 | 0.141 | -0.081 |
| $X_5$ | -0.124 | -0.071 | 0.911 | 0.173 |
| $X_6$ | -0.028 | -0.045 | 0.19 | 1.011 |

#### 3.2.2 Establish SVM prediction model optimized by GWO algorithm

The training sample and test sample data $\{F_1, F_2, F_3, F_4\}$ processed by factor analysis are used as the input data. The radial basis kernel function is selected and the GWO algorithm is used to optimize the parameters of the SVM model. The optimal penalty factor C and kernel function parameter g is then determined. The initial parameters of the GWO algorithm are set as follows: the number of wolves is 20, the maximum number of iterations is 20, and the optimization range of SVM parameters is set as [0,100].

The Mean Square Error (MSE) is selected as the fitness function of the model, The fitness function is an index that describes the parameter performance as well as an evaluation standard that determines whether the current target parameter value is the best [27]. MSE is defined as the expected value of the square of the difference between the predicted and measured. When MSE is the minimum value, it is considered that the target parameter value reaches its optimal standard. The SVM model is optimized by the GWO algorithm. The GWO-SVM model is trained using the training sample set, then the output is predicted with the help of the test sample set, and finally, the MSE is calculated. When the MSE reaches the minimum value, the model reaches the optimum. The whole process is implemented using MATLAB 2018b software. SVM is simulated using libsvm toolbox for design. The kernel function is defined using the radial basis function which uses the above-mentioned GWO algorithm to optimize the penalty parameter C and kernel function parameter g in SVM model until the iteration termination conditions are met. The fitness change curve during training is shown in Figure 4. It is observed that with the increase in the number of iterations, the optimal fitness value gradually decreases. When the FA-GWO-SVM model is iterated 10 times, the algorithm tends to be stable and the convergence accuracy is high. The parameters of SVM optimized using GWO algorithm are C = 52.8184 and g = 12.5487 respectively. Herein, the MSE value of the training sample is 9.47793e-05, and the goodness of fit R2 = 0.999772. The MSE value of the test sample is 0.00573678, and the goodness of fit R2 = 0.872324. The predicted values of the model for training samples and test samples are shown in Figure 5 and Table 6.

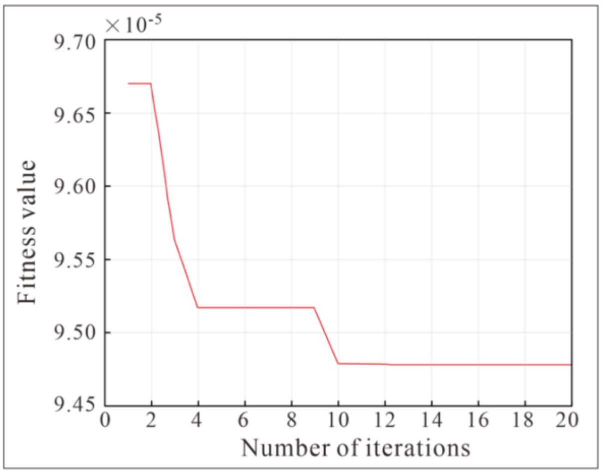


**Fig. 4**. Convergence curve of FA-GWO-SVM model

### 3.3 Model effect evaluation

To validate the optimization effect of the FA-GWO-SVM model, the traditional SVM model and GWO-SVM model are developed based on the data of training samples and test samples, and the prediction effects of each model are compared. Figure 5 shows the prediction outcomes of each model for training data, whereas Table 6 displays the prediction results for test samples. It can be seen from Figure 5 that the fitting effect of the FA-GWO-SVM model and GWO-SVM model on training samples is better than that of the traditional SVM model. Further, it can be seen from Table 6 that among the three test samples, FA-GWO-SVM model delivers two best prediction results, whereas, GWO-SVM model and traditional SVM model each provide one best prediction result.

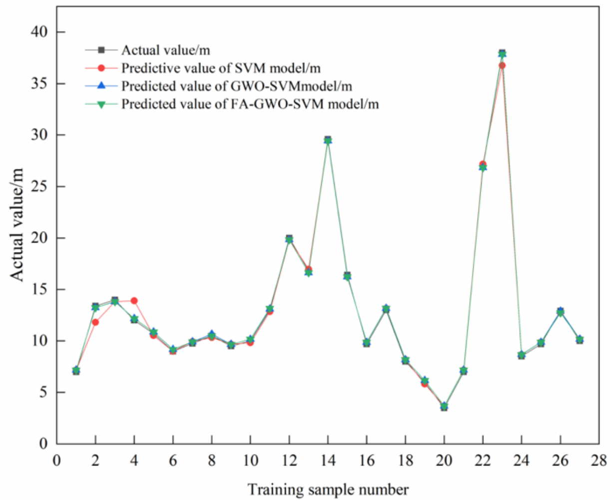


**Fig. 5**. Prediction results of each model using training samples

In order to quantitatively evaluate the prediction performance of the model, the Mean Absolute Percentage Error (MAPE), Root Mean Square Error (RMSE) and Mean Absolute Error (MAE) is introduced to compare the prediction accuracy of the model [28]. The smaller the RMSE index, the better the regression ability, learning ability and stability of the model. The smaller the MAE and MAPE index values, the higher the prediction accuracy of the model. The mathematical expression of each index is as follows [29]:

$$MAPE = \frac{100\%}{n}\sum_{i=1}^{n}\frac{\left|y_i - y_i^*\right|}{y_i^*} \tag{16}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - y_i^*)^2} \tag{17}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|y_i - y_i^*\right| \tag{18}$$

**Table 6**. Predicted values of three models using test samples

| Test samplenumber | Measuredvalue/m | SVME stimate/m | GWO-SVME stimate/m | FA-GWO-SVMEstimate/m |
|---|---|---|---|---|
| 28 | 11.7 | 14.0456 | 13.1543 | 11.1399 |
| 29 | 6.5 | 6.8435 | 8.1350 | 6.3010 |
| 30 | 11.7 | 13.7078 | 13.1568 | 13.8835 |

The comparison of the prediction performance results of the three models using the test samples is presented in Table 7. Table 7 shows that FA-GWO-SVM model has the smallest MAPE, RMSE and MAE values, indicating that it is capable of reducing the prediction error and improving the fitting degree with the actual data. Thus, its prediction effect is significantly better than the traditional SVM model and GWO-SVM model. Further, compared with the traditional SVM model, the GWO-SVM model has smaller RMSE and MAE index values, but larger MAPE values. In general, the prediction effect of the SVM model optimized with the GWO algorithm is better than that of the traditional SVM model. FA-GWO-SVM model performs factor analysis and processing on the original data. It extracts the comprehensive components that have a significant impact on the damage depth of the bottom plate eliminating the noise and information redundancy in the data. The information is then reflected by the data more objectively and effectively, reducing the input scale of SVM model through dimensionality reduction. This improves the operation efficiency of the model to a certain extent. The values of the three evaluation indices are reduced further when compared to the GWO-SVM model, and the model's learning ability, stability, and prediction accuracy are enhanced ultimately. The effectiveness of FA-GWO-SVM model is verified, which shows that the model has good prediction accuracy and operation efficiency and that it fully meets the actual needs of coal seam floor failure depth prediction.

**Table 7.** Comparison results of prediction performance of each model on test samples

| Model | MAPE | RMSE | MAE |
|---|---|---|---|
| SVM | 14.16% | 1.7936 | 1.5657 |
| GWO-SVM | 16.68% | 1.5177 | 1.5154 |
| FA-GWO-SVM | 8.84% | 1.3065 | 0.9809 |

### 3.4 Discussion

The prediction of failure depth of coal seam floor is a complex and nonlinear problem affected by many factors. Based on the measured data of coal seam floor failure depth collected from many mining areas in China, this study uses the advantages of SVM in processing small samples and nonlinear data. Herein, the SVM model is made the main body and the factor analysis method is used to reduce the dimension and extract the original data variables. The factor analysis methods then use the extracted new variables as the input of SVM model, eliminating the noise and information redundancy between data and avoiding the influence of correlation between variables on the prediction results. Considering that the prediction ability of the traditional SVM model is significantly affected by the penalty factor $C$ and kernel function parameter $g$, GWO algorithm is used to optimize the parameters of SVM model, $C$ and $g$. This effectively avoids the problems of low search efficiency, poor convergence and the possibility of falling into local extremum. Among the SVM parameter optimization methods, GWO method has the advantages that traditional methods do not have in the global search of complex search space. It can accurately find the optimal SVM model parameters $C$ and $g$. Finally, the prediction model of coal seam floor failure depth based on FA-GWO-SVM is developed. The reliability of the model is verified by comparing the three model performance evaluation parameters the model with that of the traditional SVM model and GWO-SVM model. This study provides a novel method for accurately predicting the failure depth of a coal seam floor.

However, it should be noted that the depth of floor collapse is affected by a variety of conditions. This study solely takes into account six factors that have a direct impact on the depth of floor failure which are coal seam mining thickness, dip angle, mining depth, inclined length of working face, floor anti-failure ability, and the presence of cutting through fault or fracture zone. The evaluation indexes affecting the depth of floor failure are not comprehensive. For example, according to relevant research, the advancing speed of working face, the pressure step, coal mining method and roof management method also have a certain impact on the failure depth of the floor. Simultaneously, when considering the influence of geological structure factors on the failure depth of the coal seam floor, the deciding factor is whether there is a cutting through fault or fracture zone in the working face. The nature, fall and inclination of the cutting through fault also have an impact on the failure depth of the coal seam floor, however, the extent of the impact has not been investigated in detail. In the future, a more comprehensive and detailed evaluation index system should be established, and more measured data samples should be collected to further improve the prediction model of coal seam floor failure depth. The results of the present study show that the prediction performance of the GWO-SVM model optimized by the GWO algorithm is improved to a certain extent compared to the traditional SVM model. Further, the prediction performance of the FA-GWO-SVM model after factor analysis is improved compared with the GWO-SVM model. In the future, the prediction performance of the FA-GWO-SVM model should be compared to that of other optimization models (such as genetic algorithms, particle swarm optimization algorithms, etc.). Further discussion and verification of the prediction effect of the FA-GWO-SVM model need to be done.

## 4. Conclusions

Based on the measured data of coal seam floor failure depth collected from various mining areas in China, factor analysis was performed to reduce the dimension of the original data and eliminate the influence of the correlation between factors on the prediction results. GWO algorithm is used to optimize the parameters, $C$ and $g$, of the SVM model. Finally, a prediction model of coal seam floor failure depth based on FA-GWO-SVM is established. The main conclusions are as follows:

1) The dimension of the measured data of the floor failure depth is reduced using factor analysis. The six factors affecting the floor failure depth are transformed into four new variables that can reflect most of the original data information. These four new variables are used as the input of the SVM model, which eliminates the noise and information redundancy in the data. It reduces the input scale of the SVM model and improves the prediction accuracy and operation efficiency of the model.

2) GWO algorithm is used to optimize the parameters, $C$ and g, in the SVM model. The optimized parameters of the support vector machine are found to be $C = 52.8184$ and $g = 12.5487$, to obtain an SVM with high prediction accuracy.

3) The prediction performance of the FA-GWO-SVM model, traditional SVM model and GWO-SVM model using test samples are quantitatively compared. Three indexes are used in the comparison, namely, Mean Absolute Percentage Error (MAPE), Root Mean Square Error (RMSE) and Mean Absolute Error (MAE). The results show that the FA-GWO-

SVM model has the smallest MAPE, RMSE and MAE values, indicating that the prediction performance of the FA-GWO-SVM model is significantly better than that of the traditional SVM model and GWO-SVM model.

4) The prediction model of coal seam floor failure depth based on FA-GWO-SVM has good generalization ability and strong prediction performance. It can thus meet the actual needs of coal seam floor failure depth prediction in modern mine production.

The obtained conclusions can effectively improve the prediction accuracy of the destruction depth of the coal seam floor. In the future, a more comprehensive prediction index system and a prediction model with excellent forecasting performance will be developed to further improve the prediction accuracy of the destruction depth of the coal seam floor. The research conclusions can accurately infer the destruction depth of the coal seam floor and provide guidance for mining safety procedures.

**Acknowledgements**

The authors are grateful for the support provided by the Doctoral research startup fund project of East China University of Technology (No. DHBK2019011) and Open Fund from Engineering Research Center for Geological Environment and Underground Space of Jiangxi Province (No.JXDHJJ2022-014) and Academic and Technical Leader Training Program of Jiangxi Province (No.20212BCJ23003)



## References

1. Yu, G. F., Yuan, L., Ren, B., Li, L. C., Cheng. G. W., Han, Y. C., Wang, S. X., Wei, T. S., Zheng, Q. & Ma, J. G., "Big data prediction and early warning platform for floor water inrush disaster". *Journal of China Coal Society*, 46(11), 2021, pp.3502-3514.
2. Liu, S. Q., Wu, Q., Li, Z., Zeng, Y. F., Yuan, Q. D. & Yu, Y. L., "Vulnerability evaluation and application of floor water inrush in mining area with multiple coal seams and single aquifer based on variable weight". *Journal of China University of Mining & Technology*, 50(03), 2021, pp.587-597.
3. Bascmpta, M., Sanmtquel, L. & Zhang, H., "Airflow stability and diagonal mine ventilation system optimization a case study", *Journal of Mining Science*, 54(5), 2018,pp.813-820.
4. Shi, L. Q., Xu, D. J., Qiu, M., Jing, X. & Sun, H. H., "Improved on the formula about the depth of damaged floor in working area". *Journal of China Coal Society*, 38(S2), 2013, pp.299-303.
5. Liu, C. L., Tan, Z. X., Li, P. X., Bai, L. G. & Deng, K. Z., "Calculation Methods for Depth of Floor Water-conductive Fissure Zone Induced by Mining". *Journal of Mining And Strata Control*, 15(05), 2010, pp. 38-41 .
6. Mirjalili, S., Mirjalili, S. M. & Lewis, A., "Grey wolf optimizer". *Advances in Engineering Software,* 2014(3), 2014, pp.46-61.
7. Zhang, P. S., Wu, J. W. & Liu, S. D., "Study on dynamic observation of coal seam floor's failure law". *Chinese journal of rock mechanics and engineering*, 25(S1), 2006, pp.3009-3013.
8. Jiang, Y. D., LV., Y. K., Zhao, Y. X. & Zhang, D. Y., "Similar simulation test for breakage law of working face floor in coal mining above aquifer". *Chinese Journal of Rock Mechanics and Engineering*, 30(08), 2011, pp.1571-1578.
9. Piotr, M., Lukasz, Q. & Piotr B., "The Impact of the Low Throw Fault on the Stability of Roadways in a Hard Coal Mine". *Studia Geotechnica et Mechanica*. 39(1), 2017, pp.63-72.
10. Islam, M. R., "Mining-induced fault reactivation associated with the main conveyor belt roadway and safety of the Barapukuria Coal Mine in Bangladesh: Constraints from BEM simulations". *International Journal of Coal Geology*, 79(4), 2009, pp.115-130.
11. Mee, S. H., Sulaiman, M. H. & Mohamed, M. R., "An application of grey wolf optimizer for solving combined economic emission dispatch problems". *International Review on Modelling and Simulations*, 7(5), 2014, pp. 838-843.
12. Zhu, Z. J., Zhang, H. W. & Wang, C. M., "Prediction of floor damaged depth in working area based on support vector machine and artificial bee colony algorithm". *Journal of Chongqing University*, 38(06), 2015, pp.37-43.
13. Jin, C. C., Feng, X. W., Ruan, M. & Li, J.Y., "Prediction of Damage Degree of Coal Seam Floor Based on GAPSO-SVM". *Safety in Coal Mines,* 50 (03), 2019, pp.208-211.
14. Zhang, Y. H., Li, W.Y. & Dong, F. G., "Medium and Long-Term Power Demand Forecasting Based on DE-GWO-SVR". *Electric Power,* 54 (09), 2021, pp.83-88.
15. Antczak, T., "A lower bound for the penalty parameter in the exact minimax penalty function method for solving nondifferentiable extremum problems". *Journal of Optimization Theory and Applications,* 159(2), 2013, pp.437-453.
16. Ma, S.X. & Li, X.J., "Study on Prediction Model of Coal Mine Gas Emission by Improved BP Neural Network". *Mining Research and Development,* 39(10), 2019, pp. 138-142.
17. Gao, F., Bao, Y. P., Wang, M., Liu, Y., Huang, Y.S. & Sun, G.T., "Prediction model of end-point phosphorus content of converter based on FA-ELM". *Iron & Steel,* 55 (12), 2020, pp. 24-30.
18. Mirjalili, S., Mirjalili, S. M. & Lewis. A., "Grey wolf optimizer". *Advances in Engineering Software*, 69, 2014, pp.46-61.
19. Hu, J., Qiu, J. B., Luan, C. Q. & Zhang, H. D., "Deformation Prediction Based on IGWO-SVM for Open-Pit Mine Slopes". *Mining and Metallurgical Engineering*, 42(01), 2022, pp.15-18.
20. Chen, Y. Y., Suo, Y. F. & Yang, S. H., "Ship trajectory prediction based on grey wolf optimization support vector regression". *Journal of Shanghai Maritime University*, 42(04), 2021, pp.203-209.
21. Xue, X. C. & Tian, F. F., "Coal Floor Failure Depth Prediction Based on Data Optimized BP Neural Network". *Coal Geology of China,* 33(09), 2021, pp. 8-12.
22. Yu, X. G., Han, J., Shi, L. Q., Wei, J. C., Zhu, L. & Li, S. C., "Forecast of destroyed floor depth based on BP neural networks". *Journal of China Coal Society*, 34(06), 2009, pp. 731-736.
23. Cheng, A. P., Gao, Y. T., Liang, X. W., Ji ,M. W., Wang, C. W. & Gao, Y. H., "Dynamic forecasting of mining-induced failure depth of floor based on unascertained clustering method". *Journal of Mining and Safety Engineering,* 31(05), 2014, pp.739-744.
24. Xu ,Y. M. & Zhang, R. Q., "Evaluation of comprehensive benefits of edible fungi cultivation in Hebei Province based on factor analysis combined with entropy method". *Northern Horticulture,* 2022, pp.138-144 .
25. Zhou, J., Guan, W. S. & Fu, L. T., "Water quality assessment and pollution source analysis of Xi'an river based on multivariate statistics". *Water Resources Protection,* 36(02), 2020, pp.79-84.
26. Chen, Y., Zhang, C., Xiao, C.Y., Zhao, X. L., Shi, Y. X., Yang, H., Liu, Z. Y. & Li ,S. H., "Study on prediction model of soil cadmium content moisture content correction based on GWO-SVR". *Acta Optica Sinica*. 40(10), 2020, pp.180-187.
27. Mirjali, S., Mirjaici, S M. & Lewis, A., "Greg wolf optimizer". *Advances in Engineering software,* 69(3), 2014, pp.46-61.
28. Turson, M. M. T., Zhao, M.J ., Ning, C. B. & Kong, Q. H., "Prediction of Diesel Engine Exhaust Emissions Based on Deep Extreme Learning Machine". *Science Technology and Engineering,* 21(36), 2021, pp.15646-15654.
29. Chen, H. Y., Liu, J. H., Sheng, W. M., Huang, H. & Zhao, Y.J., "Application of SVM Based on Grey Wolf Optimizer in Measurement Error Analysis of Infrared Methane Sensor". *Acta Metrologica Sinica,* 42 (09), 2021, pp.1244-1249.