Journal of
Water & Climate Change

 Journal of Water and Climate Change Vol 15 No 2, 593 doi: 10.2166/wcc.2023.604

Evaluation of mine water quality based on the PCA–PSO–BP model

Jiaqi Wang[a,*] and Yanli Huang[a,b]

[a] State Key Laboratory of Coal Resources and Safe Mining, School of Mines, China University of Mining & Technology, Xuzhou 221116, China
[b] College of Mining Engineering and Geology, Xinjiang Institute of Engineering, Urumqi 830000, China
*Corresponding author. E-mail: tb21020043b4@cumt.edu.cn

ABSTRACT

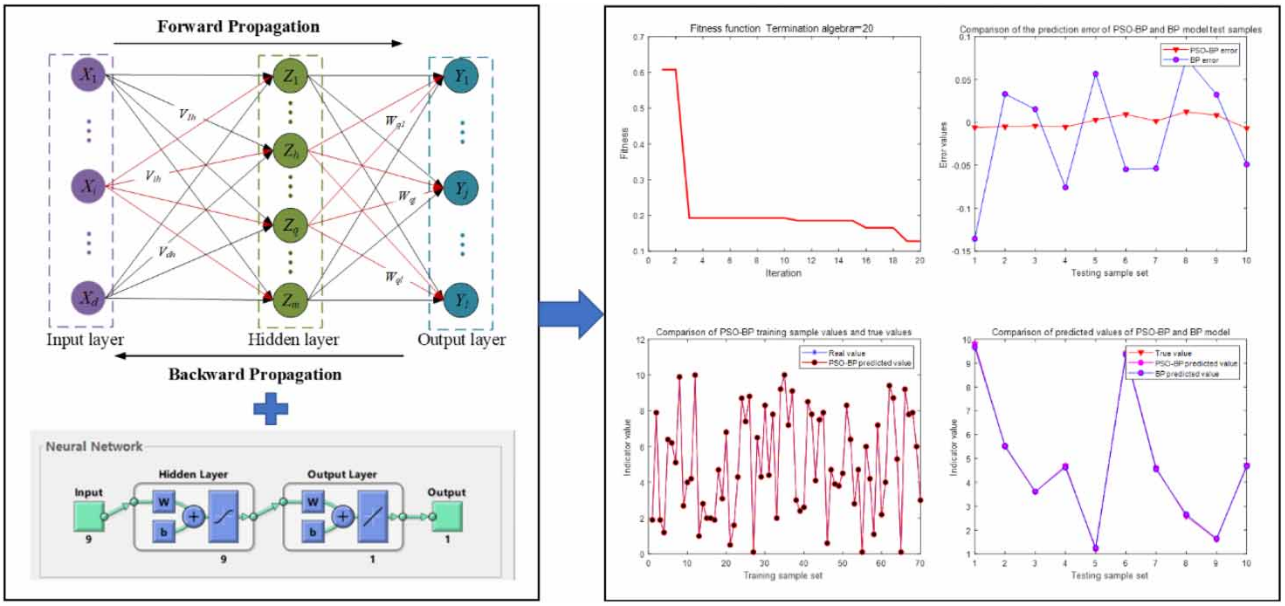
To enhance the mining area's overall use of mine water in the arid area of Western China and mitigate the current water scarcity problem, this paper introduces an intelligent optimization algorithm and neural network for mine water quality evaluation and proposes a principal component analysis (PCA)–particle swarm optimization (PSO)–back propagation (BP) mine water quality evaluation model. Firstly, the model uses PCA to identify the primary factors affecting mine water quality, then enhances the optimal weights and thresholds of the BP neural network based on the PSO algorithm, and the PCA–PSO–BP evaluation model with nine input layers, nine hidden layers, and one output layer is created. In addition, using the Shicaocun Mine as an example, the results demonstrate that the PCA–PSO–BP model has accurate mine water quality evaluation results, and the prediction accuracy reached 86.8255%. This exemplifies the PSO method's superiority to the BP neural network improvement. This study not only offers a novel theoretical framework for assessing and forecasting water quality in mining regions, but it also sets the stage for the possible broad use of state-of-the-art neural networks and optimization algorithms in the coal mining industry.

Key words: BP neural network, mine water quality evaluation, particle swarm optimization (PSO), principal component analysis (PCA), PSO–BP model

HIGHLIGHTS

- Intelligent algorithms and neural networks are introduced into mine water quality evaluation.
- Established a PCA–PSO–BP model for mine water quality evaluation.
- Realized the accurate evaluation and reasonable prediction against the background of big data.
- Provide reference for the in-depth research of optimization algorithms and neural networks in the field of water quality evaluation.

GRAPHICAL ABSTRACT



1. INTRODUCTION

The role of coal as the primary energy source in China will not change for a very long time (Wang *et al.* 2019), and the distribution of coal resources in China indicates a situation where more is in the West and less is in the East (Shang *et al.* 2016; Wang & Luo 2018), which is exactly the opposite of the distribution of water resources. Therefore, mining coal in arid mines in western China will deplete the water resources and exacerbate the current situation of water scarcity in the mines. (Wang *et al.* 2021). According to statistics, 2.1 t of mine water is created for every 1 t of raw coal mined, meaning that there is a huge waste of mine water resources. Mine water cannot be utilized directly due to the presence of suspended materials, salts, heavy metals, and other components. Instead, it must be graded based on its quality (Mitko *et al.* 2021; Zhang *et al.* 2022). In order to maximize the use of mine water resources and reduce adverse environmental consequences, it is crucial to accurately evaluate the different categories of mine water quality.

Several water quality evaluation methods have played a significant role in encouraging the development of water quality evaluation. At this point, the methods of mine water quality evaluation mostly include the single-factor evaluation method and the complete pollution index method. Although the single index evaluation technique is easy to use, it does not adequately reflect the full state of water quality, leading to significant differences. This is because it employs a single index as the reference standard (Zhao *et al.* 2021). The Nemerov pollution index is one of the most widely used comprehensive pollution index methods in evaluating mine water quality. The method is simple to calculate and the physical concept is understandable, but it emphasizes the impact of the maximum pollution index on the mine water quality, making it simple to exaggerate the effects of specific pollutants on water quality while at the same time indicating that the weight of each pollutant index is not objective when determining, which will eventually lead to inaccurate conclusions (Liu *et al.* 2022; Sharma & Krupadam 2022). In addition to the previously mentioned research methods, related theoretical techniques and models include principal component analysis (PCA) (Liu *et al.* 2020), Bayesian theory (Huang *et al.* 2019), fuzzy comprehensive evaluation method (Liu *et al.* 2021), gray system theory (Guo *et al.* 2020), cluster analysis method (Prasad *et al.* 2020), hierarchical analysis method (Yu *et al.* 2022), entropy power analysis method (Ju & Hu 2021), fuzzy variable set theory (Li *et al.* 2022), material element topological model (Shi *et al.* 2018), set pair analysis method (Qu *et al.* 2021), and multi-criteria group decision-making models (Baghapour *et al.* 2020). These models also support accurately classifying the water quality in mining areas and evaluating sudden water hazards. However, they have shortcomings such as instability in the calculation process, lack of universality, and a large gap between calculations and real outcomes in the definition of index thresholds.

Machine learning and intelligent optimization algorithms have been used extensively in the area of artificial intelligence and big data. Zhang & Li (2019) designed a fuzzy genetic neural network model to realize the detection of atmospheric quality. Liu (2022) adopted particle swarm optimization (PSO) to improve the traditional BP neural network (BPNN) and finally constructed a PSO-powered BP neural network (PSO–BPNN) model for the intelligent emergency risk avoidance of sudden financial disasters. Nassif (2014) used evolutionary algorithms to optimize artificial neural network models for decreasing the energy consumption of air conditioning systems in the field of architectural design. A thorough IFPA-BP network model was built for the intelligent diagnosis of natural gas pipeline defects, realizing the intelligent diagnosis of natural gas pipeline defects. This was based on the IFPA algorithm, which is used to optimize the initial weights and thresholds of the BPNN (Liang *et al.* 2020). Yang (2022) introduced a BPNN optimization algorithm based on a multidirectional mutation genetic algorithm (MMGA-BP). The multidirectional mutation genetic BPNN method is used for the intelligent optimization of English-teaching courses. To effectively predict the stock index, Yang *et al.* (2019) proposed a hybrid intelligent algorithm based on brain storm optimization and PSO to optimize the parameters of the system model. Du *et al.* (2013) proposed an integrated learning algorithm, combining the RCDPSO_DM algorithm with a Kalman filtering algorithm, which was applied to optimize antecedent and consequent parameters of constructed T-S FNNs, for medical applications handling complex clinical pathway variances. In the study applying intelligent optimization algorithms to COVID-19 identification, Baghdadi *et al.* (2022) proposed a method for automatically and accurately classifying COVID-19 on CT lung images using convolutional neural networks (CNNs), pre-trained models, and the sparrow search algorithm (SSA).

As can be seen, intelligent optimization algorithms and neural networks have been used in a variety of fields. However, there have been relatively few studies on the evaluation of mine water quality using intelligent optimization algorithms and neural networks, even though these technologies have strong fault tolerance, high classification accuracy, and robustness. Therefore, this study proposes to incorporate an intelligent optimization algorithm and neural network into the evaluation of mine water quality. It also innovates and develops the mine water quality evaluation model. It can support the reasonable and scientific evaluation of mine water quality, as well as lay the theoretical scientific foundation for mine water treatment, utilization, and discharge.

2. THEORETICAL FOUNDATION

2.1. Principal component analysis

PCA is a mathematical transformation of multiple indicators with high correlation into several uncorrelated composite indicators, or principal components, by using the idea of dimensionality reduction (Marukatat 2022). Assume that the original variables are n-dimensional vectors $[f_1\ f_2\ \dots\ f_n]$, and the dimensionality reduction yields k-dimensional variables $[F_1\ F_2\ \dots\ F_k]$ $(k < n)$. PCA recombines the n-dimensional original variables to obtain a new k-dimensional linearly uncorrelated variable, as shown in Equation (1).

$$\begin{cases} F_1 = a_{11}f_1 + a_{12}f_2 + \cdots + a_{1n}f_n \\ F_2 = a_{21}f_1 + a_{22}f_2 + \cdots + a_{2n}f_n \\ \vdots \\ F_k = a_{k1}f_1 + a_{k2}f_2 + \cdots + a_{kn}f_n \end{cases} \tag{1}$$

The principal components obtained after PCA processing can reflect most of the information of the original variables with fewer indicators and are independent of each other, eliminating the information redundancy of the original numerous factors and reducing the complexity of the problem (Hsieh & Tung 2009). In addition, PCA can eliminate the influence of correlation among evaluation indicators and overcome the shortcomings of correlation among indicators and overlapping information reflected by indicators in multi-indicator evaluation, which makes the evaluation results more accurate.

2.2. The BPNN

BPNN is a neural network with reverse error transmission, which is usually composed of an input layer, hidden layer and output layer. The learning process is mainly divided into two parts: forward transmission of information and backward propagation of error. The training of the network is completed by continuously adjusting the weights and thresholds of the neural network through multiple cycles of training so that the output results tend to be close to the target value. A single hidden layer BPNN with d input units X_d, m hidden units Z_m, and l output units Y_l is given in Figure 1, where V_{ih} denotes the connection

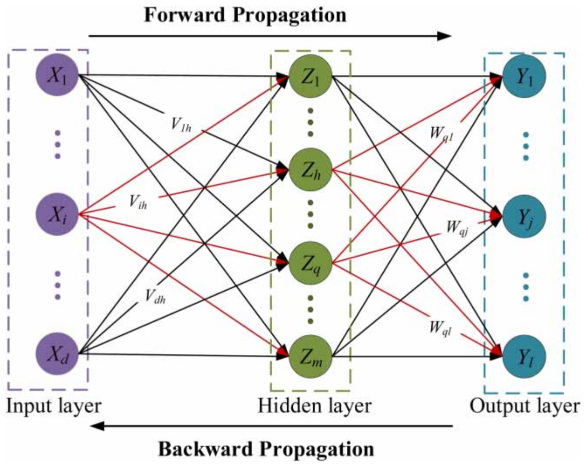


Figure 1 | Schematic diagram of BPNN structure.

weight between input unit X_i and hidden unit Z_h, and W_{qj} denotes the connection weight between hidden layer unit Z_q and output layer unit Y_j (Dong *et al.* 2013).

BPNN can handle non-smooth, non-time-series data related to water quality evaluation, but at the same time, it is easy to fall into the problem of local minima itself, and the difficulty in determining the weights and thresholds may lead to inaccurate results of mine water quality evaluation.

2.3. Particle swarm optimization

At the end of the 20th century, Kennedy and Eberhart proposed PSO capable of performing swarm intelligence searches (Mendes *et al.* 2004). This algorithm is a swarm-intelligent search method, mainly inspired by the predation process of birds, where the solution of each optimization problem is considered as a spatially flying bird, denoted as a particle, and each particle searches for an individual optimal solution in the local solution space, and all particles together form a swarm of particles, which obtain the swarm optimal solution by communicating with each other (Clarke *et al.* 2014).

The PSO algorithm runs initially to generate a random group of particles with vector dimension n. Then the position of a particle can be written as a point in the n-dimensional search space, that is, a solution in the n-dimensional optimization space. x_j denotes the vector of the current position of the jth particle.

$$x_j = x_{j1}, x_{j2}, \ldots, x_{jn}$$

v_j denotes the current velocity vector of the jth particle.

$$v_j = v_{j1}, v_{j2}, \ldots, v_{jn}$$

In the process of each iteration, the particle position vector needs to be substituted into the customized fitness function E_k first, and the fitness value of the particles is solved to find the best position p_j and the global best position g_j of the individuals of the particle population over generations by comparing the optimal fitness value, i.e.

$$p_j = p_{j1}, p_{j2}, \ldots, p_{jn}$$

$$g_j = g_{j1}, g_{j2}, \ldots, g_{jn}$$

The swarm of particles is updated and optimized by p_j, g_j, x_j and v_j to find the position vector and velocity vector after the iteration, and the evolution equation of this algorithm can be expressed as (Jiang *et al.* 2021).

$$v_j(k+1) = \omega \times v_j k + c_1 \times r_1 \times [p_j(k) - x_j(k)] + c_2 \times r_2 \times [g_j(k) - x_j(k)] \tag{2}$$

$$x_j(k+1) = x_j(k) + v_j(k+1) \tag{3}$$

where $j = 1,2,3, \ldots, N$, N denotes the total number of particle swarms; v_j denotes the velocity of the particle; x_i is the position of the particle; ω is the inertia factor; c_1 and c_2 are the learning factors; p_i is the optimal solution of the individual; g_i is the optimal solution of the whole; r_1 and r_2 are the random numbers between [0, 1]; k and $k+1$ denote the kth and $k+1$th generations, respectively.

The performance of the algorithm is strongly influenced by the ω inertia weight, the ω larger it is, the better it is for global optimization search, and the ω smaller it is, the better it is for local search. In this paper, we choose adaptively adjustable inertia weights, which are negatively correlated with the number of iterations, and the expressions are as follows.

$$\omega = \omega_{\max} - (\omega_{\max} - \omega_{\min}) \times \frac{k}{k_{\max}} \tag{4}$$

where k is the current iteration number; $k_{\max}$ is the maximum iteration number; $\omega_{\max}$ and $\omega_{\min}$ are the maximum and minimum values, respectively.

The PSO algorithm has the advantages of high efficiency and fast search speed, therefore, it is used to improve the BPNN to solve the defect that the BPNN is easy to fall into local optimum and achieves the purpose of improving model accuracy.

2.4. Construction of the BPNN model based on PCA and PSO

In the construction of the BPNN evaluation model of mine water quality, the accuracy of the evaluation results can be improved by appropriately increasing the network input variables. However, there are many factors affecting mine water quality, and there is often a high correlation between different factors, resulting in a large amount of redundancy of information, which not only makes the accuracy of the evaluation model deviated but also increases the difficulty of data processing. PCA is the main method to solve the correlation between factors. The PCA algorithm can get the main factors affecting the mine water quality by analyzing the contribution rate and cumulative contribution rate of each influencing factor to the mine water quality, and the accurate prediction of the mine water quality can be made by considering only the main factors.

The threshold and weights of the BPNN affect the accuracy of the neural network model; usually, the initial weights and thresholds are determined by random assignment. However, this method tends to make the BPNN fall into local optimal solutions. In order to overcome the defects and accelerate the convergence speed, this paper adopts the PSO algorithm to optimize the BPNN. The PSO algorithm can expand the space to search for the optimal solution and has a strong global search capability, and the search for the optimal weights and thresholds is completed through continuous iterative updating. The optimal weights and thresholds of the BPNN can be determined to improve the accuracy of the BPNN model and realize the scientific evaluation of the water environment in the mining area.

Thus, the BPNN is improved based on PCA and PSO methods, the mine water quality evaluation model of PCA–PSO–BP is proposed, and its implementation process is shown in Figure 2.

The steps for constructing a BPNN model based on PCA and PSO improvements are as follows.

(1) Determine the structure and relevant parameters of the BPNN.
(2) Set the cluster size, the initial flight speed of the particles and the corresponding point positions. The current best point position of the particle is selected as the initial point position, and the best point position of the swarm is the global best point position.
(3) Each particle contains a fitness value, which is used to reflect the superiority or inferiority of the particle. After training the BPNN, the training error is used as the fitness value of the current point position of the particle, and the result is

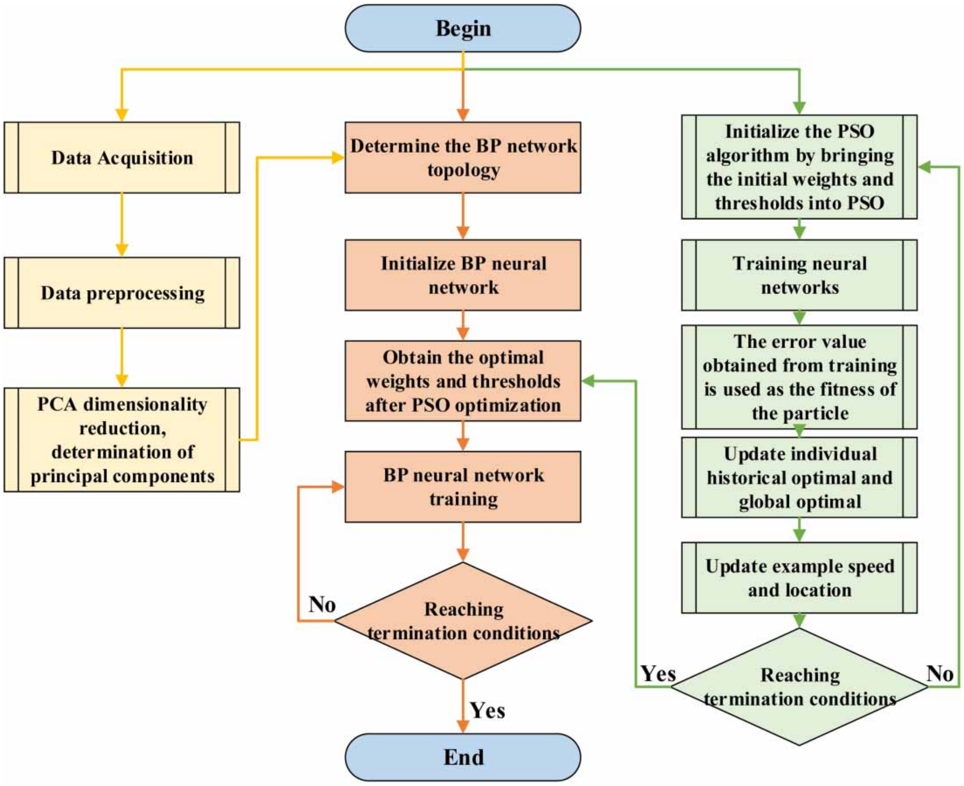


Figure 2 | Flow chart of PCA–PSO–BPNN.

compared with the value of the previous best point position, if it is better than the previous best point position, it is replaced, otherwise, it remains unchanged.

(4) If the global best point position is less than the previous best point position of the current particle, the global best point position is replaced by the previous best point position, otherwise, it remains unchanged.
(5) Reprogram the flight speed of the particle and the corresponding point position according to Equations (2) and (3).
(6) Check whether the algorithm meets the termination condition (the number of iterations reaches the maximum number of iterations or the error accuracy reaches the initially set target error accuracy), if the termination condition holds, the best weight and threshold are output, and then the model is further simulated, otherwise skip to (3).

3. CASE STUDY

3.1. Background

The Shicaocun coal mine is a medium and large coal mine in the south of the Yuanyang Lake mining area of Ningxia Ningdong coalfield, and the administrative division is under the jurisdiction of Ningdong Town, Lingwu City, Ningxia. Ningxia is located in the western part of China, where water resources are scarce, and mining activities will aggravate the current situation of water scarcity in the mine area, so the efficient use of mine water is very important for the protection of mine water resources, and the comprehensive evaluation of mine water quality is a prerequisite for the efficient use of mine water resources.

Eighty sets of mine water observation points were selected in the Shicaocun coal mine, and the mine water was tested by laboratory testing at each testing point. The main testing indexes included sulfide ($f1$), oxygen consumption ($f2$), hexavalent chromium ($f3$), total hardness ($f4$), ammonia nitrogen ($f5$), volatile phenols ($f6$), sulfate ($f7$), chloride ($f8$), nitrate ($f9$), total dissolved solids ($f10$), fluoride ($f11$), turbidity ($f12$), mercury ($f13$), selenium ($f14$), total alpha radioactivity ($f15$), total beta radioactivity ($f16$), anionic surfactant ($f17$), sodium ($f18$), zinc ($f19$), iron ($f20$), septum ($f21$), lead ($f22$), aluminium ($f23$), total bacterial colony ($f24$), etc. Some data on specific test results are shown in Supplementary material, Appendix. According to the Groundwater Quality Standard (GB/T14848-2017) (Chen *et al.* 2022), the mine water quality was classified into the following five categories, and the mine water quality classification is shown in Table 1, detailed groundwater quality classification indicator ranges are provided in Supplementary material, Appendix.

We can determine the grade of each index by comparing the mine water test results with the Groundwater Quality Standard, but we are unable to determine the mine water quality overall. For instance, the monitoring points 01s lead and cadmium indicators meet the standard in class V, the indication of oxygen consumption meets class IV, and the total number of colonies indicator meets class I. 2160. However, it is not possible to determine the overall water quality of the monitoring site. It is evident that while the standard can assess the grade of a particular indicator, it is unable to assess the overall mining area's water quality. As a result, a set of models for evaluating the water quality in mining areas must be created.

Although there are relevant studies in the evaluation of mine water quality at this stage, mine water quality is affected by a variety of factors, and it is difficult to evaluate the water quality reasonably by a single indicator or multiple indicators. In addition, the mine water quality contains a large amount of data, and the traditional evaluation method can't make full use of all the data. Therefore, intelligent algorithms and neural networks can be introduced into the mine water evaluation to make full use of all the data to achieve accurate evaluation and prediction of mine water quality.

3.2. PCA reduction

Due to a large number of water quality testing indicators; direct water quality evaluation on the one hand, the fact that the amount of data is large, on the other hand, there may be a correlation between the indicators, which will also affect the evaluation results. Therefore, first of all, through the method of PCA, the water quality impact factors for dimensionality reduction processing. In the SPSS software for PCA, select the eigenvalue greater than 1 as the extraction conditions, the results of the PCA are shown in Table 2.

According to the results of principal component extraction, the final 24 indicators of water quality impact factors were downscaled into nine principal components, nine principal components are independent of each other, and cover more than 75% of the information of the original 24 indicators, to achieve the downscaling of indicators. According to the factor score matrix of the nine principal components in the PCA, the factor score of each principal component can be calculated, and the combined linear function of each principal component is shown in Equation (5).

$$\begin{cases} F_1 = 0.425f_1 + 0.334f_2 + 0.323f_3 + \cdots - 0.129f_{23} + 0.086f_{24} \\ F_2 = 0.175f_1 - 0.127f_2 + 0.265f_3 + \cdots + 0.254f_{23} + 0.195f_{24} \\ \vdots \\ F_9 = -0.035f_1 - 0.247f_2 + 0.087f_3 + \cdots + 0.205f_{23} + 0.404f_{24} \end{cases} \tag{5}$$

where f_1–f_{24} denote the 24 water quality impact indicators, F_1–F_9 denote the extracted nine principal component factor scores, respectively.

3.3. PSO–BPNN algorithm design

Based on the nine principal components obtained by dimensionality reduction of PCA, as the input parameters of the BPNN, and the mine water quality level as the output parameters of the neural network, the input unit d of the network is 9, the

Table 1 | Groundwater quality standard classification table

Classification	Meaning
Class I	Low groundwater chemical component content, suitable for various applications.
Class II	Lower groundwater chemical component content, suitable for various applications.
Class III	Medium content of groundwater chemical component, suitable for centralized domestic drinking water sources.
Class IV	The high content of groundwater chemical components, suitable for agriculture and some industrial water
Class V	Groundwater with high chemical component content is not suitable as a source of domestic drinking water.

Table 2 | Results of principal component dimensionality reduction analysis

	Initial eigenvalue			Extraction of the sum of squares of loads		
Components	Total	Percentage of variance	Accumulation, %	Total	Percentage of variance	Accumulation, %
1	3.225	13.437	13.437	3.225	13.437	13.437
2	2.955	12.313	25.750	2.955	12.313	25.750
3	2.309	9.619	35.369	2.309	9.619	35.369
4	2.161	9.004	44.373	2.161	9.004	44.373
5	1.732	7.215	51.588	1.732	7.215	51.588
6	1.720	7.167	58.754	1.720	7.167	58.754
7	1.478	6.157	64.912	1.478	6.157	64.912
8	1.372	5.718	70.630	1.372	5.718	70.630
9	1.191	4.963	75.593	1.191	4.963	75.593
10	0.896	3.734	79.327			
11	0.836	3.485	82.812			
12	0.804	3.352	86.164			
13	0.665	2.769	88.933			
14	0.574	2.393	91.326			
15	0.442	1.841	93.167			
16	0.433	1.805	94.972			
17	0.3030	1.377	96.348			
18	0.274	1.143	97.492			
19	0.229	0.955	98.446			
20	0.153	0.636	99.082			
21	0.105	0.437	99.519			
22	0.061	0.256	99.775			
23	0.030	0.127	99.901			
24	0.024	0.099	100.000			
Extraction method: Principal component analysis						

output unit l is 1, and the number of hidden layer units m can be determined according to Equation (6) (Li 2020).

$$m = \sqrt{d + l} + a \tag{6}$$

where m is the number of cells in the hidden layer, d is the number of cells in the input layer, l is the number of cells in the output layer, and a is an integer between [1,10].

Through several experiments, it is verified that the model accuracy of the BPNN is higher when $m = 9$, so the number of nine hidden layer units is set. The construction of the BPNN topology is carried out in MATLAB R2020b, as shown in Figure 3.

4. RESULTS

4.1. PSO–BP model training

The PSO–BP mine water quality evaluation prediction model is constructed based on MATLAB R2020b. The input nodes are the nine principal components that have been dimensioned down by PCA, the output nodes are the mine water quality levels, and the number of hidden layers is determined to be nine according to Equation (6).

The BPNN parameters were set with a training number of 1,000 times, a learning rate of 0.01, a minimum error of 1×10^{-6}, a Tansig function for the implicit layer, a Purelin function for the output layer, and a Trainlm for the training function.

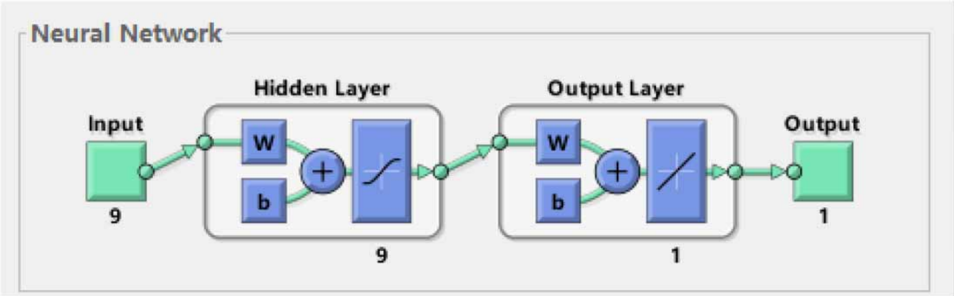


Figure 3 | BP neural network topology diagram.

The 80 ollected sets of data were randomly sorted, and the first 70 sets were selected as the training set for training and the last 10 sets as the testing set. Normalizing the data before training can eliminate the influence between the magnitude and order of magnitude of different indicators and help improve the learning speed of the network, and the formula for normalizing the data is as follows.

$$i = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \tag{7}$$

where i is the normalized input data, x is the actual data, and $x_{\max}$ and $x_{\min}$ are the maximum and minimum values of the actual data, respectively.

The model training results are shown in Figures 4 and 5. Figure 4 is a graph of the change in optimal individual fitness, the horizontal axis indicates the number of iterations and the vertical axis indicates its fitness value. From Figure 4, it can be seen that the PSO–BP model can converge to the optimal fitness value quickly, and during the iteration, the particles can jump out of the optimal value to search and avoid falling into the local optimum, and the accuracy is high, which shows the superiority of the PSO algorithm to improve the BPNN. Figure 5 shows the training results of the PSO–BP model, where the index value of the y-axis is a quantitative expression of the quality of mine water, according to the mine water quality classification standard, class I corresponds to [8, 10], class II corresponds to [6, 8], class III corresponds to [4, 6], class IV corresponds to [2, 4], and class V corresponds to [0, 2]. It can be seen that after 70 sets of sample data for training, the PSO–BP model has high accuracy and can be used to evaluate and predict the quality of mine water.

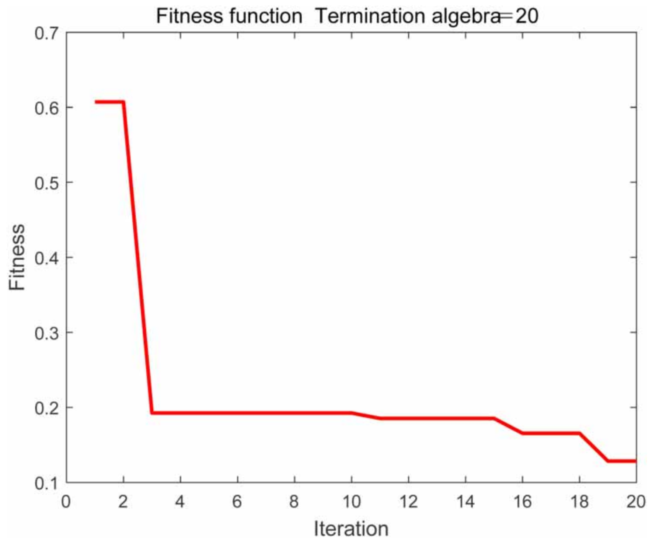


Figure 4 | Adaptation function.

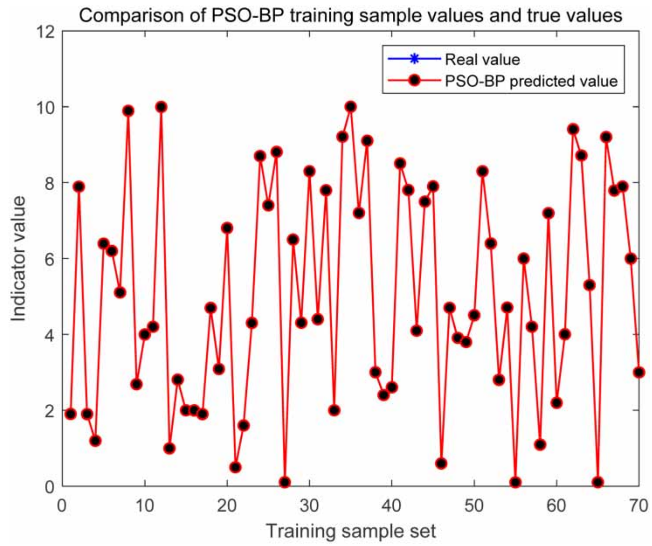


Figure 5 | PSO–BP model training results.

4.2. Analysis of evaluation results

To illustrate the advantages of the PSO–BPNN model, the prediction results of the traditional BPNN are compared with those of the PSO–BP model to discuss the advantages of the constructed PSO–BP model. The prediction values of the traditional BPNN and the PSO-optimized BPNN for 10 sets of test samples are given in Figure 6. It is obvious that the prediction effect of the BPNN optimized by the PSO algorithm is more accurate than that of the traditional BPNN, and the prediction values of the PSO–BP model for 10 sets of samples almost overlap each other, while the conventional BPNN has a poor fit between predicted and actual values compared to PSO–BP.

To analyze the prediction results of the PSO–BP model separately, 10 of these samples were selected as the test set and validated in the PSO–BP model, and the results of the predicted and real values of PSO–BP for each group of samples are shown in Figure 7. It can be seen that the predicted values in the test set are very close to each other and the prediction results are more accurate.

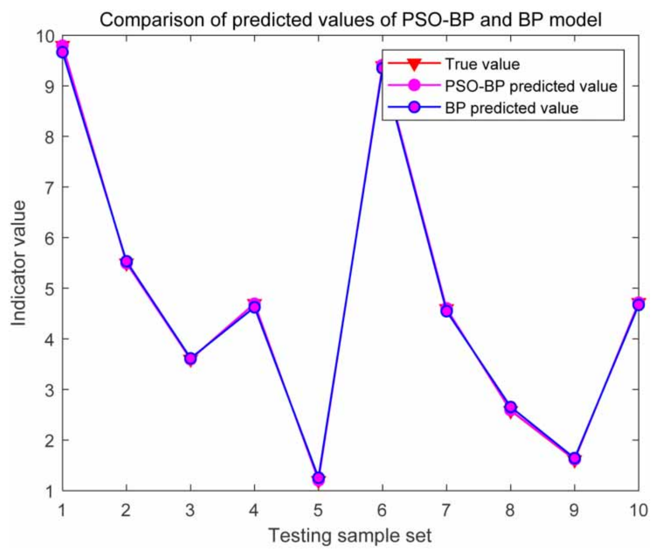


Figure 6 | PSO–BP and BP predicted values.

4.3. Discussion

Figure 8 gives a comparison of the prediction error values between the traditional BPNN and the PSO–BP model, and it can be found that the error value of the BPNN fluctuates more, and the error values of all test sample points are higher than those of PSO–BP model, while the prediction error of PSO–BP is between ± 0.02, and the stability and error value is better than that of BPNN. In order to more clearly analyze the accuracy of the constructed PSO–BP prediction model, the prediction errors of the test set samples are shown in Figure 9, from which it can be seen that the errors of all test samples are between ± 0.01, among which the error of test sample 8 is the largest, reaching 0.015, and most of the samples have errors between ± 0.01, indicating that the prediction results of the PSO–BP model are more accurate and can achieve accurate prediction and evaluation of mine water quality.

In order to evaluate and compare the model prediction accuracy objectively, this study uses the evaluation indexes of mean square error (MSE), mean absolute error (MAE), root mean square error (RMSE), and mean percentage error (MAPE) to analyze the prediction accuracy of the two models, and the calculation results are shown in Table 3. It is obvious from

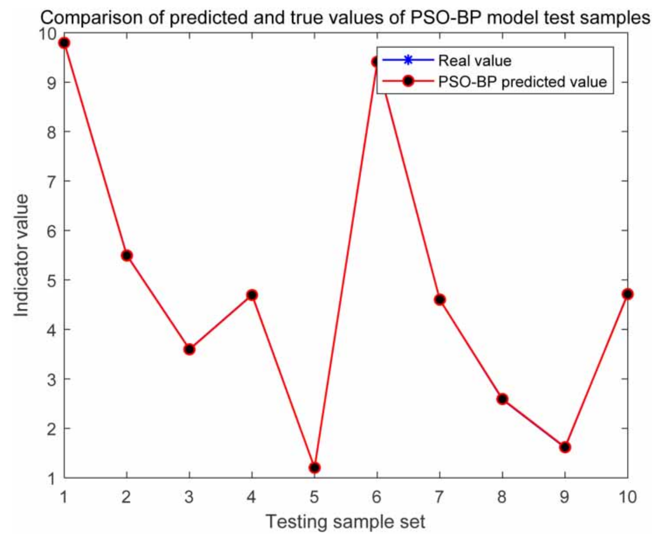


Figure 7 | PSO–BP predicted and actual values.

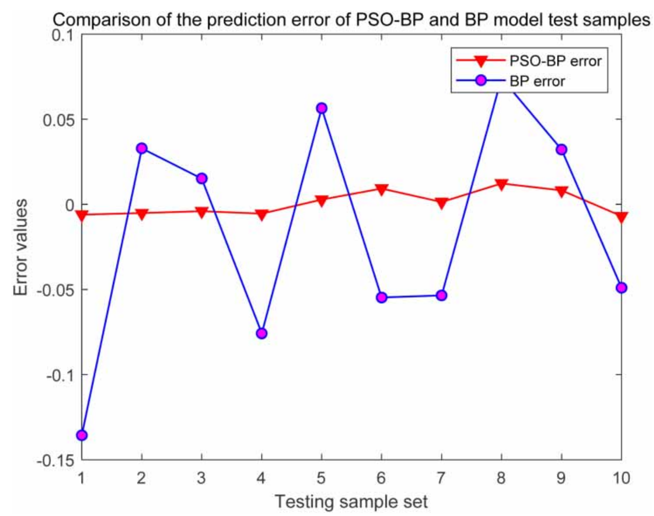


Figure 8 | Comparison of PSO–BP and BP errors.

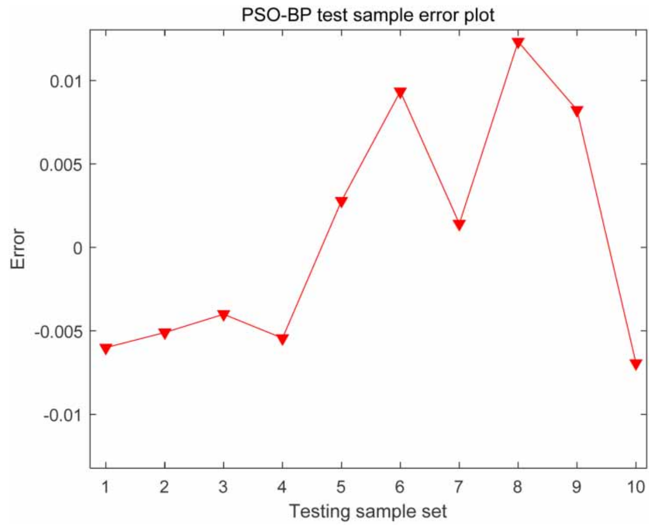


Figure 9 | Error plot of the test sample set.

Table 3 | Comparison of prediction accuracy indexes of two models

	MSE	MAE	RMSE	MAPE
BP	4.3987	1.6615	2.0973	49.943%
PSO–BP	0.25972	0.46655	0.50963	13.1745%

the table that the PSO–BP model has a prediction error of 13.1745% and a prediction accuracy of 86.8255%, while the BPNN model has a MAPE of 49.943% and a prediction accuracy of only 50.057%, which also proves that the PSO–BP model has a higher prediction accuracy than that of the BPNN.

In addition, we analyze the training time of the model by changing the number of training data sets. When the training set data is 80, the training time of the model is 32.9816 s, when the training data of the model is reduced to 50, the training time shrinks to 27.7433 s, and when the training data is 30, the training time is 26.3589 s. It can be seen that the amount of training data is proportional to the training time of the model, and when the number of the training set is reduced, the training time is reduced as well, but as the number of training continues to decrease, the rate of training time reduction becomes slower and slower.

By comparing the traditional BPNN with the PSO–BP model, it is found that the BPNN has the problems of slow convergence speed and long training time. While this study makes full use of the global search capability of the PSO algorithm to optimize the BPNN, it can improve network prediction performance and computational efficiency. Moreover, for the problem of multi-indicator evaluation, it may be that the number of indicators is too large, leading to too much computation. This paper proposes a pre-processing method of indicator dimensionality reduction by using PCA, which can achieve the reduction of the number of evaluation indicators, but also retain the main information to ensure the accuracy of the evaluation results. In conclusion, the PCA–PSO–BPNN model proposed in this paper can solve the problem of multi-indicator evaluation and has the advantages of high computational efficiency and accurate evaluation results. It has accurate prediction results for mine water quality, which can propose a new assessment system for mine water quality evaluation.

5. CONCLUSIONS

An intelligent optimization algorithm and neural network are introduced into mine water quality evaluation, and a mine water quality evaluation model of PCA–PSO–BPNN is proposed, which can provide an intelligent evaluation prediction model for the research related to mine water quality, and the main conclusions are as follows.

(1) The quality of mine water is influenced by several things. The PCA method is used in this study to minimize the dimensionality of the factors influencing the quality of mine water. It achieves dimensionality reduction of the data and minimizes information loss by reducing the original 24 evaluation indexes to nine primary components.

(2) The PSO algorithm enhances the BPNN, and to make up for the difficulty of figuring out the thresholds and weights of the conventional BPNN, the optimal weights and thresholds of the BPNN are determined by using PSO search.
(3) Using the water testing data from the Shicaocun Coal Mine, the developed PCA–PSO–BP mine water quality evaluation prediction model was validated and contrasted with the conventional BPNN prediction outcomes. The outcomes show that the PCA–PSO–BP water quality evaluation model has an 86.8255% prediction accuracy compared to the conventional BPNN.

ACKNOWLEDGEMENTS

We appreciate the comments and suggestions from anonymous reviewers.

AUTHOR CONTRIBUTIONS

J.W. was involved in conceptualization, writing – original draft, data curation, analyses of results. Y.H. was involved in conceptualization, methodology, supervision.

FUNDING

This work was supported by the National Natural Science Foundation of China (Nos 52104103, 52022107, and 52174128) and the Natural Science Foundation of Jiangsu Province (Nos BK20210499 andBK20190031).

CONSENT FOR PUBLICATION

Written informed consent for publication was obtained from all participants.

DATA AVAILABILITY STATEMENT

All relevant data are included in the paper or its Supplementary Information.

CONFLICT OF INTEREST

The authors declare there is no conflict.

REFERENCES

Baghapour, M., Shooshtarian, M. & Zarghami, M. 2020 Process mining approach of a new water quality index for long-term assessment under uncertainty using consensus-based fuzzy decision support system. *Water Resour. Manage.* **34** (3), 1155–1172. doi:10.1007/s11269-020-02489-5.

Baghdadi, N., Malki, A., Abdelaliem, S., Balaha, H., Badawy, M. & Elhosseini, M. 2022 An automated diagnosis and classification of COVID-19 from chest CT images using a transfer learning-based convolutional neural network. *Comput. Biol. Med.* **144**, 105383. doi:10.1016/j.compbiomed.2022.105383.

Chen, K., Zhang, Q., Tao, Y., Luo, K. & Chen, Q. 2022 The slope safety, heavy metal leaching, and pollutant diffusion prediction properties under the influence of unclassified cemented paste backfill in an open pit. *Int. J. Environ. Res. Public Health* **19** (19), 12772. doi:10.3390/ijerph191912772.

Clarke, J., McLay, L. & McLeskey, J. 2014 Comparison of genetic algorithm to particle swarm for constrained simulation-based optimization of a geothermal power plant. *Adv. Eng. Inf.* **28** (1), 81–90. doi:10.1016/j.aei.2013.12.003.

Dong, S., Zhou, D., Zhou, W., Ding, W. & Gong, J. 2013 Research on network traffic identification based on improved BP neural network. *Appl. Math. Inf. Sci.* **7** (1), 389–398. doi:10.12785/amis/070148.

Du, G., Jiang, Z., Diao, X. & Yao, Y. 2013 Intelligent ensemble T-S fuzzy neural networks with RCDPSO_DM optimization for effective handling of complex clinical pathway variances. *Comput. Biol. Med.* **43** (6), 613–634. doi:10.1016/j.compbiomed.2013.02.007.

Guo, Y., Dong, S., Hao, Y., Liu, Z., Yeh, T., Wang, W., Gao, Y., Li, P. & Zhang, M. 2020 Risk assessments of water inrush from coal seam floor during deep mining using a data fusion approach based on grey system theory. *Complexity* **2020**, 8205370. doi:10.1155/2020/8205370.

Hsieh, P. & Tung, P. 2009 A novel hybrid approach based on sub-pattern technique and whitened PCA for face recognition. *Pattern Recognit.* **42** (5), 978–984. doi:10.1016/j.patcog.2008.09.024.

Huang, P., Yang, Z., Wang, X. & Ding, F. 2019 Research on Piper-PCA-Bayes-LOOCV discrimination model of water inrush source in mines. *Arabian J. Geosci.* **12** (11), 334. doi:10.1007/s12517-019-4500-3.

Jiang, J., Wei, W., Shao, W., Liang, Y. & Qu, Y. 2021 Research on large-scale bi-level particle swarm optimization algorithm. *IEEE Access.* **9**, 56364–56375. doi:10.1109/ACCESS.2021.3072199.

Ju, Q. & Hu, Y. 2021 Source identification of mine water inrush based on principal component analysis and grey situation decision. *Environ. Earth Sci.* **80** (4), 157. doi:10.1007/s12665-021-09459-z.

Li, X. 2020 Artificial intelligence neural network based on intelligent diagnosis. *J. Ambient Intell. Hum. Comput.* **12** (1), 923–931. doi:10.1007/s12652-020-02108-6.

Li, X., Zhang, W., Wang, X., Wang, Z. & Pang, C. 2022 Evaluation on the risk of water inrush due to roof bed separation based on improved set pair analysis-variable fuzzy sets. *Acs Omega* **7** (11), 9430–9442. doi:10.1021/acsomega.1c06700.

Liang, X., Liang, W. & Xiong, J. 2020 Intelligent diagnosis of natural gas pipeline defects using improved flower pollination algorithm and artificial neural network. *J. Clean. Prod.* **264**, 121655. doi:10.1016/j.jclepro.2020.121655.

Liu, L. 2022 Research on digital economy of intelligent emergency risk avoidance in sudden financial disasters based on PSO-BPNN algorithm. *Comput. Intell. Neurosci.* **2021**, 7708422. doi:10.1155/2021/7708422.

Liu, W., Liu, S., Tang, C., Qin, W., Pan, H. & Zhang, J. 2020 Evaluation of surface water quality after mine closure in the coal-mining region of Guizhou, China. *Environ. Earth Sci.* **79** (18), 427. doi:10.1007/s12665-020-09167-0.

Liu, X., Liu, H., Wan, Z., Wang, L. & Chen, Q. 2021 Study on evaluation index system of sustainable development of mine water resources based on PSO-AHP model and fuzzy comprehensive evaluation. *J. Intell. Fuzzy Syst.* **41** (3), 4253–4264. doi:10.3233/JIFS-189686.

Liu, T., Ji, X. & Gong, Y. 2022 Wetland functional area division method: A correlation analysis of water quality and landscape structure. *Sustainability* **14** (21), 14015. doi:10.3390/su142114015.

Marukatat, S. 2022 Tutorial on PCA and approximate PCA and approximate kernel PCA. *Artif. Intell. Rev.* doi:10.1007/s10462-022-10297-z.

Mendes, R., Kennedy, J. & Neves, J. 2004 The fully informed particle swarm: Simpler, maybe better. *IEEE Trans. Evol. Comput.* **8** (3), 204–210. doi:10.1109/tevc.2004.826074.

Mitko, K., Turek, M., Jaroszek, H., Bernacka, E., Sambor, M., Skóra, P. & Dydo, P. 2021 Pilot studies on circular economy solution for the coal mining sector. *Water Resour. Ind.* **26**, 100161. doi:10.1016/j.wri.2021.100161.

Nassif, N. 2014 Modeling and optimization of HVAC systems using artificial neural network and genetic algorithm. *Build. Simul.* **7** (3), 237–245. doi:10.1007/s12273-013-0138-3.

Prasad, B., Soni, A., Vishwakarma, A., Trivedi, R. & Singh, K. 2020 Evaluation of water quality near the Malanjhkhand copper mines, India, by use of multivariate analysis and a metal pollution index. *Environ. Earth Sci.* **79** (11), 259. doi:10.1007/s12665-020-09002-6.

Qu, X., Shi, L., Qu, X., Bilal, A., Qiu, M. & Gao, W. 2021 Multi-model fusion for assessing risk of inrush of limestone karst water through the mine floor. *Energy Rep.* **7**, 1473–1487. doi:10.1016/j.egyr.2021.02.052.

Shang, Y., Wang, J., Liu, J., Jiang, D., Zhai, J. & Jiang, S. 2016 Suitability analysis of China's energy development strategy in the context of water resource management. *Energy* **96**, 286–293. doi:10.1016/j.energy.2015.12.079.

Sharma, M. & Krupadam, R. 2022 Adsorption-desorption dynamics of synthetic and naturally weathered microfibers with toxic heavy metals and their ecological risk in an estuarine ecosystem. *Environ. Res.* **207**, 112198. doi:10.1016/j.envres.2021.112198.

Shi, S., Xie, X., Wen, Z., Zhou, Z., Li, L., Song, S. & Wu, Z. 2018 Intelligent evaluation system of water inrush in roadway (tunnel) and its application. *Water* **10** (8), 997. doi:10.3390/w10080997.

Wang, S. & Luo, K. 2018 Life expectancy impacts due to heating energy utilization in China: Distribution, relations, and policy implications. *Sci. Total Environ.* **610**, 1047–1056. doi:10.1016/j.scitotenv.2017.08.195.

Wang, W., Li, Z., Lyu, J. & Ni, W. 2019 Eliminating outdated capacity to promote energy conservation in China's coal-fired power industry. *Engineering* **5** (2), 194–196. doi:10.1016/j.eng.2018.11.028.

Wang, X., Gao, Y., Jiang, X., Zhang, Q. & Liu, W. 2021 Analysis on the characteristics of water pollution caused by underground mining and research progress of treatment technology. *Adv. Civ. Eng.* **2021**, 9984147. doi:10.1155/2021/9984147.

Yang, Z. 2022 Application of multidirectional mutation genetic algorithm and its optimization neural network in intelligent optimization of English teaching courses. *Comput. Intell. Neurosci.* **2021**, 4297600. doi:10.1155/2021/4297600.

Yang, B., Zhang, W. & Wang, H. 2019 Stock market forecasting using restricted gene expression programming. *Comput. Intell. Neurosci.* **2019**, 7198962. doi:10.1155/2019/7198962.

Yu, S., Ding, H. & Zeng, Y. 2022 Evaluating water-yield property of karst aquifer based on the AHP and CV. *Sci. Rep.* **12** (1), 3308. doi:10.1038/s41598-022-07244-x.

Zhang, B. & Li, W. 2019 Intelligent air quality detection based on genetic algorithm and neural network: An urban China case study. *Concurrency Comput. Pract. Exper.* **31** (10), e4673. doi:10.1002/cpe.4673.

Zhang, S., Wu, Q. & Ji, H. 2022 Research on zero discharge treatment technology of mine wastewater. *Energy Rep.* **8** (2), 275–280. doi:10.1016/j.egyr.2022.01.014.

Zhao, W., Xiao, C., Chai, Y., Feng, X., Liang, X. & Fang, Z. 2021 Application of a new improved weighting method, ESO method combined with fuzzy synthetic method, in water quality evaluation of Chagan Lake. *Water* **13** (10), 1424. doi:10.3390/w13101424.

First received 4 September 2023; accepted in revised form 18 December 2023. Available online 30 December 2023