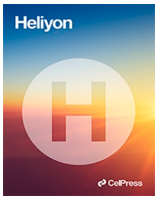
Heliyon 10 (2024) e26925

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon



Research article

Groundwater chemical characteristic analysis and water source identification model study in Gubei coal mine, Northern Anhui Province, China

Qilin Jiang [a], Qimeng Liu [a,*], Yu Liu [b], Huichan Chai [a], Jingzhong Zhu [c]

[a] School of Earth and Environment, Anhui University of Science and Technology, Huainan, 232001, China
[b] State Key Laboratory Mining Response and Disaster Prevention and Control in Deep Coal Mines, Anhui University of Science and Technology, Huainan, 232001, China
[c] School of Resources and Geosciences, China University of Mining and Technology, Xuzhou, 221000, China

ARTICLE INFO

Keywords:
Water chemical characteristics
Water source identification
Particle swarm optimization
Factor analysis
BP neural network

ABSTRACT

This study aims to accurately identify mine water sources and reduce the hazards caused by water inrush accidents in coal mines. Taking the Gubei coal mine as an example, the water quality results of the water samples from the Cenozoic unconsolidated aquifer, Permian sandstone fracture aquifer, and Carboniferous Taiyuan Formation limestone karst fracture aquifer in the mine area were tested, and K^++Na^+, Ca^{2+}, Mg^{2+}, Cl^-, SO_4^{2-}, HCO_3^-, TDS (Total Dissolved Solids), and pH were selected as the main indicators to study the water chemistry characteristics of the aquifer through water chemistry component analysis, major ion content analysis, Piper trilinear analysis, and correlation analysis. Thirty-five groups of water samples were randomly selected and imported into SPSS software for factor analysis (FA) and downsized to three main factors as the input variables of the artificial neural network model. The particle swarm optimization (PSO) code was written based on the MATLAB platform to improve the self-adjustment weights and acceleration factors for optimizing the initial weights and thresholds of the Back-Propagation (BP) neural network. The training and prediction samples were learned in the ratio of 8:2, and the recognition results were compared with the traditional BP neural network model. Results showed that the groundwater of the Gubei coal mine demonstrated a water quality vertical zoning pattern, and the chemical composition was dominated by cation K^++Na^+ and anion Cl^-. The FA-PSO-BP neural network model has a higher accuracy of water source discrimination compared with the cluster analysis and the FA-BP neural network model. The FA-PSO-BP neural network model is worthy of further application in the problem of water source identification in mine water inrush.

1. Introduction

China's coal reserves account for approximately 13% of the world's total and is the world's largest coal importer and producer [1]. Although the share of clean energy in China's energy mix has significantly increased, and the effective utilization rate and energy access rate of coal are more advantageous, coal will remain as the basic energy source of the country for some time to come [2,3]. Coal

* Corresponding author.
E-mail address: qmliu@aust.edu.cn (Q. Liu).

https://doi.org/10.1016/j.heliyon.2024.e26925
Received 7 August 2023; Received in revised form 21 February 2024; Accepted 21 February 2024
Available online 27 February 2024

mine resources, as an important non-renewable energy source, gradually depleted due to the increasing mining efforts. Water emergencies in coal mines frequently occur due to the complex mine water-filled hydrogeology and have become a major disaster second only to a gas explosion [4–6], which has a great effect on the society, economy, and personal safety and has become an important factor that restricts safe and efficient mining in coal mines. Therefore, accurate identification of the source of water inrush in mines has become an urgent problem to be solved to ensure the safety of coal mines [7,8].

The hydrogeological conditions of the mine site and the complexity of the water inrush mechanism have a certain influence on the difficulty of identifying the source of water inrush in the mine [9,10]. The study of water damage in mines has gradually evolved from simple water chemistry analysis to multivariate statistics. In recent years, a number of experts and scholars have flexibly applied a variety of mathematical modeling methods to conduct in-depth research on the identification of water sources of mine water emergencies. Isotope analysis can effectively determine the source of groundwater recharge, study the flow characteristics of groundwater, and determine the connection between aquifers [11,12]. Cluster analysis is to find the similarity and inherent structural differences between water quality data for classification. This method is not only economical and fast but also can understand the connectivity of aquifers and water rock action [13–15]. Simple water chemical analysis and multivariate statistical analysis provide favorable support for solving the problem of water source identification.

In addition, the nonlinear analysis method has been widely used in the problem of water source identification. Guan Zilong applied the hydrogeochemical analysis method and isotope analysis method to the Mindong I mine in Inner Mongolia. The calculation accuracy of fuzzy synthesis is influenced by the degree of affiliation, and the reliability of cluster and isotope analyses is higher, but they are subject to certain limitations [16]. The gray correlation method is used to analyze the water quality data of the water inrush point as a sequence, compare it with the parent sequence, and calculate the correlation to achieve the water source discrimination effect

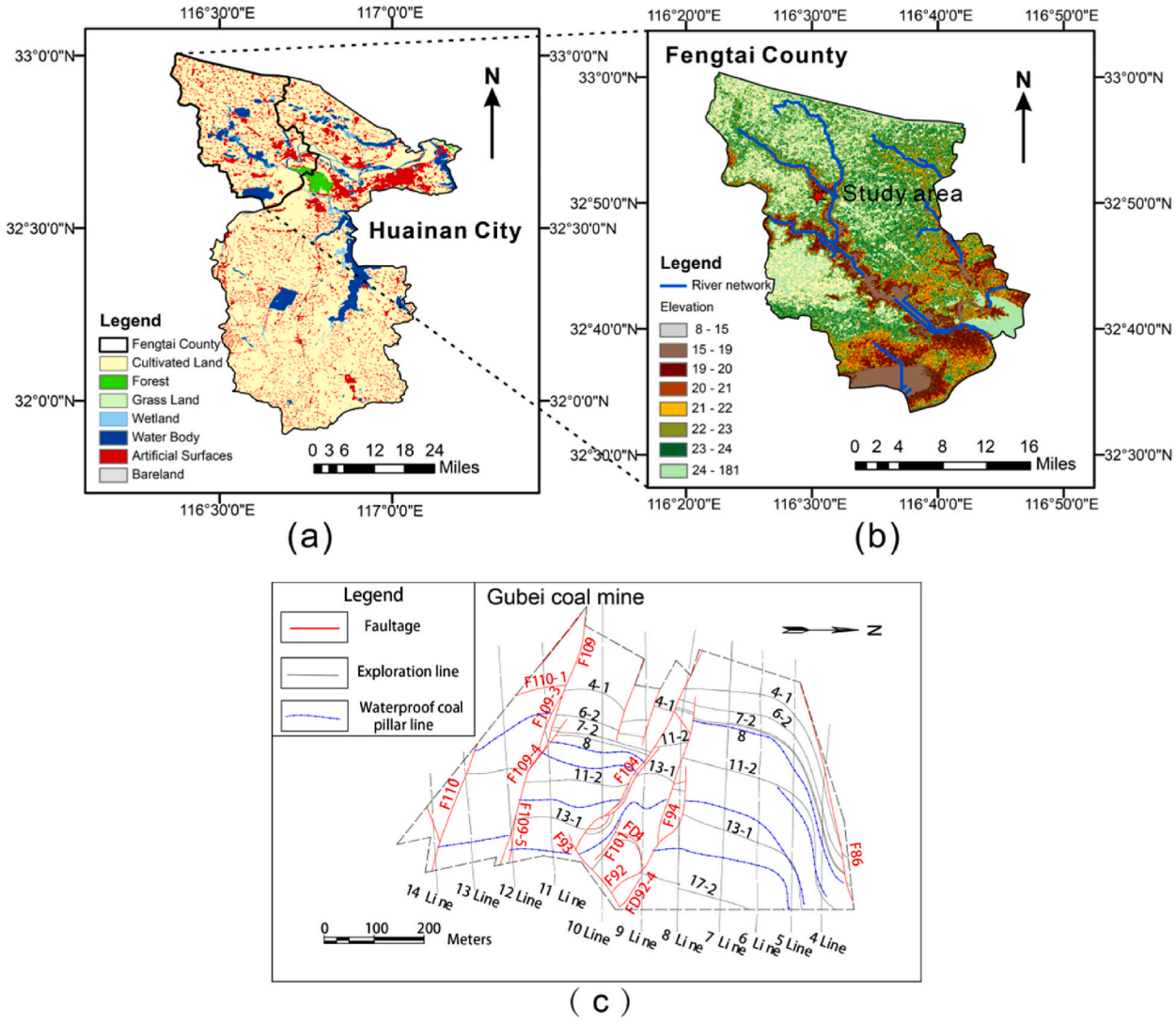


Fig. 1. (a):Land use cover map of Huainan City; (b):Geographical topographic map of Fengtai County; (c):Geological sketch of the study area.

[17]. The Geographic Information System (GIS) has a high practical value in the water source identification problem with its powerful data processing function. The visualization of the identification results is possible through the use of the GIS platform coupled with a variety of water source identification methods. GIS combines with hydrodynamics to visualize water chemistry information and analyze the spatial distribution and migration and diffusion patterns of ions [18]. The support vector machine method is used to classify the water quality data in the original space to find the optimal classification hyperplane of the predicted and real samples. The traditional support vector machine is only applicable to binary classification, which is not ideal for solving multi-classification problems. Fisher's discriminant method extracts feature variables and analyzes them using a support vector machine method, which has a higher accuracy for water source identification [19]. The artificial neural network method is a highly promising method for water source discrimination, based on the principles underlying neural networks in biology, which simulates the mathematical model of the neural system's processing mechanism of complex information [20]. The commonly used BP neural network is to back propagate the error between the predicted and the true values and continuously correct the weights and thresholds to improve the discriminative model [21]. The coupling of the laser-induced fluorescence technology and convolutional neural network model can effectively avoid the subjectivity of feature selection, eliminate the redundant information of raw data, and realize the dynamic online analysis of water source of mine water inrush [22]. The extensive study provides new ideas and methods for water source identification and useful guidance for future problem solving.

In this study, taking the Gubei mining area as an example, combined with the hydrogeological conditions of the study area, the hydrochemical characteristic composition analysis, water quality type analysis, and correlation analysis were carried out to comprehensively evaluate the hydrochemical data. Since hydrochemistry can reflect the essence of groundwater, six conventional ions (K^++Na^+, Ca^{2+}, Mg^{2+}, Cl^-, SO_4^{2-}, HCO_3^-), TDS (Total Dissolved Solids), and pH value were selected as the indicators for the analysis of groundwater chemical characteristics and water source discrimination in the study area. The simple hydrochemical analysis method is convenient, economical, and effective. However, this method also has certain limitations [23–25]. The multivariate statistical method and the nonlinear analysis method form complementary advantages. Therefore, the author performed factor analysis (FA) dimensionality reduction of the evaluation index and applied it to MATLAB software coupled particle swarm optimization (PSO) to determine the initial weight and threshold of backpropagation (BP) neural networks. The training and test samples randomly selected from the water samples in the study area were imported into the model for training. The accuracy, precision, and recall were evaluated. In summary, the purpose of this study is to: (1) explore the characteristics of groundwater chemical changes in the main aquifers of Gubei Coal Mine; (2) reveal groundwater recharge runoff and water storage conditions to a certain extent; (3) aiming at the problem of water source identification in Gubei Coal Mine, a FA-PSO-BP neural networks model for water inrush water source discrimination based on water chemical information was established.

2. Study area

The Gubei coal mine is located at the interface between the eastern flank of the Chenqiao backslope and the western part of the Panji backslope, and it is one of the main production mines in the Huainan coalfield (Fig. 1a). The study area is located in the Huaihe alluvial plain (Fig. 1b), with dense surface runoff, abundant groundwater resources, and flat topography, except for the low-lying area along the Xihe and Ganghe rivers, which is prone to flooding during the rainy season. The climate is monsoonal, warm, and semi-humid, with cold winters, hot summers, and abundant rainfall. The tectonic pattern is north–south, monoclinic tectonics dip eastward, and the stratigraphy is gently inclined. The structure is divided according to the development characteristics of the secondary folds and faults (Fig. 1c): simple monoclinic zone in the north, "X" conjugate shear zone in the middle, and monoclinic zone in the south.

The aquifers in the study area are mainly divided into Cenozoic unconsolidated aquifer, Permian sandstone fissure aquifer, and Carboniferous Taiyuan Formation limestone karst fracture aquifer from shallow to deep. The Cenozoic unconsolidated Layer consists of clay, sandy clay, and sand layer interlayer, and the water barrier mainly consists of four parts: upper, middle, lower, and bottom "red layer". The water barrier in the upper part is unstable, and the thickness of some sections becomes thin, resulting in the hydraulic connection between the upper and lower parts and the middle part. The distribution of the middle and lower parts of the water barrier is stable, and the lithology is mainly clay and sandy clay, with good water barrier. The lower part of the aquifer only develops a lenticular sand layer, and the bottom "red layer" demonstrates a lamellar distribution. The water barrier divides the unconsolidated aquifer in the new world into the upper, middle, and lower aquifers. At the bottom of the Permian coal system layer, the marine mudstone is in integrated contact with the underlying Taiyuan Formation. The sandstone at the top of this aquifer is mainly static storage, and the water-richness is weak. The thickness of the sandstone is unevenly distributed, and the fissure development varies. The local rock fissure development section and the vicinity of the fault influence zone may be relatively water-rich due to the water-richness is controlled by the thickness of the sandstone, the degree of fissure development, and the boundary conditions. The Carboniferous System and the underlying Ordovician System are mainly composed of sandstone, shale, limestone, and thin coal seam, and their contacts are parallel and unintegrated. Several hydraulic connections exist between Taiyuan Formation limestone water and Ordovician limestone water due to the existence of hidden faults, in which the limestone karst fissure aquifer of the Taiyuan Formation is pressurized water and has weak water-richness.

3. Materials

On the basis of the investigation of the geological background and hydrogeological conditions of the study area, groundwater sampling of coal mines in the study area was carried out to lay a scientific and reliable foundation for further research on the chemical

characteristics of groundwater and the establishment of a water source identification model [26]. The groundwater samples in the study area were sent to the key laboratory of mine water resource utilization within 10 min after collection, filtered using 0.45 μ cellulose filter, and stored in a refrigerator at 4 °C. The HCO_3^-, CO_3^{2-} content test was completed within 24 h, and the other parameters could be tested within 7 days. Ion chromatography, hydrochloric acid titration, and other methods were used to determine and obtain the water chemical analysis. The results meet the anion and cation balance check. A portable multi-parameter water quality meter was used to determine the pH of water in the field. Meanwhile, the weight method was utilized in the laboratory to determine the TDS. In this study, 35 groups of groundwater quality analysis data were randomly selected from the study area, including 5 groups of upper, middle, and lower water (UW, MW, and LW) samples in the unconsolidated aquifer; 10 groups of sandstone water (SW) samples; and 10 groups of limestone water (CW) samples. The data of different water samples were compiled and counted, and the main indexes, namely, K^++Na^+, Ca^{2+}, Mg^{2+}, Cl^-, SO_4^{2-}, HCO_3^-, TDS, and pH value, were selected to study and analyze the groundwater quality.

4. Methods

The study area location map and geological structure sketch were applied to ArcGIS (version 10.2) and AutoCAD 2021 software, respectively, and groundwater chemical analysis was performed using Excel 2018, Origin 2021 and AqQA (version 1.5), IBM SPSS Statistics 25 and Matlab 2020a for water source discrimination studies, and the images were retouched using CorelDRAW 2021 and Adobe illustrator 2021 for production.

4.1. Correlation analysis

Correlation analysis is a statistical method to describe the degree of correlation between random variables, reflecting the degree of similarity and dissimilarity between variables, and revealing the mechanism of water chemistry ion formation and the way of origin [27,28]. The groundwater chemical equilibrium system essentially determines that the changes between ions are mutually constrained and interconnected. The strength of correlation between ions can explain the groundwater chemical characteristics, and there is a certain influence on the effectiveness of water source discrimination.

4.2. Cluster analysis

Cluster analysis as a classical method in multivariate statistics, it is frequently cited in different problems in various fields [29,30]. The main principle of clustering is in finding the intrinsic structure of similarities and differences between data, dividing the evaluation factors of data into clusters based on intrinsic connections, with great similarity of data structure within the same cluster and large differences between different clusters. The clusters are merged again by the nature and distance between them, and so on until the final requirements are met. Cluster analysis is a kind of unsupervised learning of searching clusters, which does not rely on the pre-defined category information for classification, and the discriminative marker vertebrae depend only on the nature and distance of the data.

4.3. Factor analysis

Factor analysis is a multivariate statistical analysis method used to transform a large number of variables that may be correlated with each other into a relatively small number of composite variables that are uncorrelated with each other [31,32]. The correlation coefficient matrix of the original variables is transformed through dimensionality reduction to obtain a few unobservable composite variables that can control all variables without losing valid information, and the composite variables are used to describe the correlations of the original variables.

4.4. Particle swarm algorithm

Kennedy and Eberhart proposed a particle swarm optimization algorithm by simulating the foraging behavior of birds [33]. When a flock of birds gathers, the neighboring individuals transmit information to each other to maintain flight at the same speed, and the whole group flies together to the target. The flock foraging behavior is abstracted into a mathematical model, wherein each bird is treated as a particle. The target position is equivalent to the global optimal solution, and the trajectory of the birds to the target position is the objective function, thus solving the optimization complex problem.

4.5. BP neural network

BP neural network is a multi-layer feed-forward neural network based on the error backpropagation algorithm training in the field of machine learning. This network is one of the most widely used neural network models at present [34,35]. The operating principle is to initialize the relevant parameters and determine the topology, transfer the factor information from the output layer to the output layer through the implicit layer, back propagate the error between the output value and the expected value, and iterate through adjusting the weights and thresholds of each layer until the error reaches the minimum; thus, the training function converges.

5. Results

5.1. Determination of chemical characteristics of water

5.1.1. Water chemical component analysis

The statistical analysis of the main indicators of the selected groundwater samples in the study area (Table 1) shows that the proportion of cationic K^++Na^+ and anionic Cl^- contents in the UW medium is low, and its aquifer pore development and groundwater runoff conditions are relatively good, resulting in low UW mineralization and good water quality, and it is the main source of drinking water for domestic use [36]. Except for the UW, all other types of samples belong to the high mineralization range, and the salt content is much higher than 1000 mg/L. The average pH content of all types of water samples varies in the following order: CW > LW > SW > MW > UW. The water quality of each aquifer is alkaline–strongly alkaline, and the standard deviation of pH value is small. The overall difference is not evident. The standard deviation of the ion equivalent milligram concentration and TDS of SW is large, mainly due to the large thickness of the aquifer and the heterogeneity of the sandstone fracture development, resulting in the large variability of water quality space inside the sandstone aquifer and the poor mobility of the water body. The degree of SO_4^{2-} and HCO_3^- dispersion between SW and CW reflects the large difference of groundwater chemical environment and the degree of water–rock action in the corresponding aquifer. The standard deviations of the Ca^{2+} and Mg^{2+} ions indicate less volatility of the ion equivalent milligram concentrations in the groundwater samples at different depths.

5.1.2. Major ion content analysis

The order of the cation concentration of the groundwater samples in this area is $\rho(K^+{+}Na^+)>\rho(Ca^{2+})>\rho(Mg^{2+})$. The average content of Ca^{2+} and Mg^{2+} is low, and the K^++Na^+ ions are absolutely dominant (Fig. 2). In terms of the variation trend of the anions in the various types of water, the ion concentration distribution patterns of MW and LW are basically the same: $\rho(Cl^-)>\rho(SO_4^{2-})>\rho(HCO_3^-)$. The anion concentration trends of SW and CW are similar: $\rho(Cl^-)>\rho(HCO_3^-)>\rho(SO_4^{2-})$. The distribution pattern of the ion concentration in the UW is not the same as those of the other types of water samples: $\rho(HCO_3^-)>\rho(SO_4^{2-})>\rho(Cl^-)$. The upper aquifer of the Cenozoic unconsolidated layer is mainly recharged by atmospheric precipitation and infiltration of the surface water bodies and laterally recharged by regional paleochannels with strong evaporation [37], reducing the K^++Na^+ and Cl^- ion contents, and their ion concentration change characteristics are obscured. Consequently, the milligram equivalent concentration of ion HCO_3^- is higher than that ion K^++Na^+ in the first place, and the concentration of the remaining five ions are lower than 200 mg/L. A clay layer exists between the upper and the middle aquifer of the Cenozoic unconsolidated layer, and the extension is wide, blocking the vertical circulation of groundwater. The mineral composition of the coal-bearing sandstone makes the groundwater susceptible to leaching and decarbonation, resulting in a significant increase in the concentration of HCO_3^- ions and a decrease in the concentration of Ca^{2+} and Mg^{2+} ions. The lithology of the lower part of the Cenozoic unconsolidated layer is rich in calcium, which has a good water barrier effect, resulting in a significant difference in the trend of anion concentration in the SW compared with the LW.

The distribution of the six conventional ions in different water source groups can be visually reflected by radar diagrams (Fig. 3). The distribution patterns of cation K^++Na^+ and anion Cl^- are similar in the five water source types, and the two ions are absolutely dominant in MW, LW, SW, and CW. The distribution of Ca^{2+} and Mg^{2+} ions in the five aquifers is relatively stable, and their average content is relatively small and below 100 mg/L. The anion Cl^- alternately adsorbs with Ca^{2+} and Mg^{2+} ions to keep the two ions in a relatively balanced state. The ion equivalent milligram concentration of the UW contains a large number of anions HCO_3^-, the upper

Table 1
Statistics of the main chemical indicators of the groundwater samples.

Water source	Chemical composition	Ionic component content (mg/L)						TDS (mg/L)	pH
		K^++Na^+	Ca^{2+}	Mg^{2+}	Cl^-	SO_4^{2-}	$HCO3^-$		
UW (n = 5)	MAX	206.72	52.72	24.83	79.99	164.70	478.69	675	8.41
	MIN	122.22	24.35	11.53	35.00	25.10	425.00	469	7.90
	AVG	165.83	38.27	16.18	52.42	73.20	451.54	575	8.02
	RSD	39.23	10.54	5.27	18.19	56.92	20.68	92	0.22
MW (n = 5)	MAX	856.01	63.62	33.36	946.20	670.00	286.57	2571	8.40
	MIN	746.80	58.01	16.60	852.10	464.10	145.70	2311	7.86
	AVG	800.06	61.35	29.08	904.19	568.92	198.79	2456	8.23
	RSD	47.49	2.19	7.04	45.97	74.00	55.11	109	0.22
LW (n = 5)	MAX	843.44	62.46	24.96	1122.24	365.60	327.90	2486	8.59
	MIN	767.43	35.95	17.78	940.80	278.24	183.67	2332	8.18
	AVG	818.21	50.31	20.78	1016.49	336.07	274.43	2403	8.36
	RSD	31.13	11.53	3.05	68.94	35.36	53.85	60	0.17
SW (n = 10)	MAX	1029.03	41.68	31.36	1053.57	849.34	1494.99	2830	8.48
	MIN	623.02	4.61	0.85	510.48	4.73	350.86	2132	8.06
	AVG	865.85	21.87	14.60	714.49	497.16	598.42	2432	8.25
	RSD	129.16	16.28	11.31	217.59	322.37	354.05	271	0.15
CW (n = 10)	MAX	1132.00	33.67	16.53	1070.59	326.60	896.99	2714	9.19
	MIN	998.42	8.02	1.94	833.08	7.68	353.92	2439	7.95
	AVG	1055.21	15.23	6.22	1010.33	141.78	643.76	2597	8.65
	RSD	54.99	8.89	5.49	68.35	140.96	210.29	83	0.48

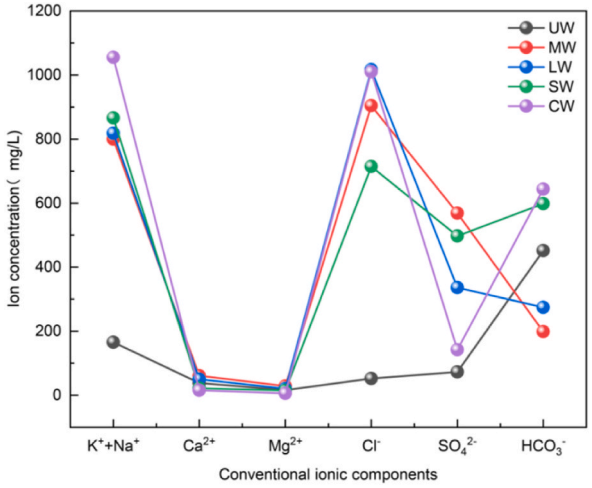


Fig. 2. Trend of the average content of conventional ions in groundwater.

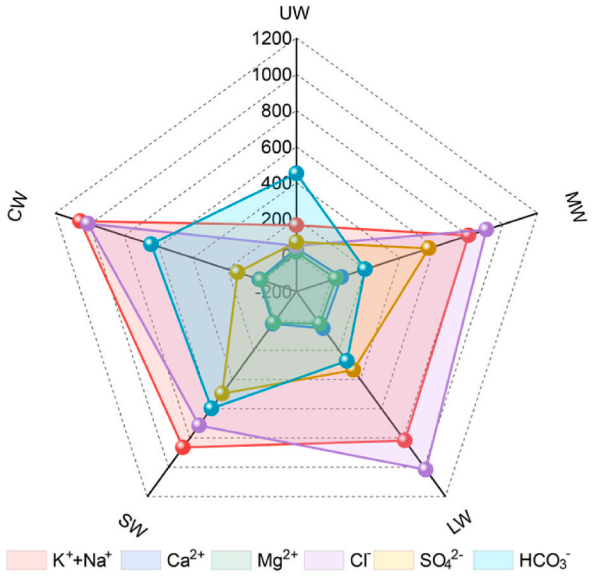


Fig. 3. Radar map of the average content of conventional ions in groundwater.

unconsolidated layer of the medium has considerable carbonic acid compounds, and the body of water has good runoff conditions, making it easy to produce bicarbonate ions in the water–rock action [38]. A large amount of gypsum and sulfate minerals exist in the rock of the lower and middle unconsolidated aquifers and sandstone aquifer, thereby generating sulfate ions that dissolved in the water through oxidation. A large amount of pyrite is frequently present in coal strata, which is more prone to oxidation reactions to produce anion SO_4^{2-}. The distribution pattern of the HCO_3^- and SO_4^{2-} ions is also particularly different because they have varying mechanisms of water–rock interaction.

5.1.3. Piper trilinear chart analysis

AqQA software was used to draw the Piper trilinear diagram shown in Fig. 4 with reference to the anion and cation data in Table 1 to visualize the water chemistry characteristics of the different aquifer waters. The sample data show that the UW is farther away from other aquifer water samples in the trilinear diagram. Some groundwater samples have more than 50% carbonate hardness, and the water quality type is mainly HCO_3–K+Na. The MW quality types are Cl–K+Na and SO_4–Cl–K+Na, and the LW quality types of the lower aquifer are Cl–K+Na and Cl–SO_4–K+Na. SW and CW are mainly along the HCO_3^- distribution, and the cation K^++Na^+ ion milligram equivalent percentage is the highest. Moreover, the water quality of carbonate and chloride content is higher and more alkaline. The SW quality types are Cl–SO_4–Na+K, Cl–HCO_3–Na+K, Cl–Na+K, and SO_4–Cl–HCO_3–Na+K. The CW quality types are Cl–Na+K and Cl–HCO_3–Na+K. Fourteen groups of Cl–K+Na type can be found in 35 groups of water samples, accounting for 40% of

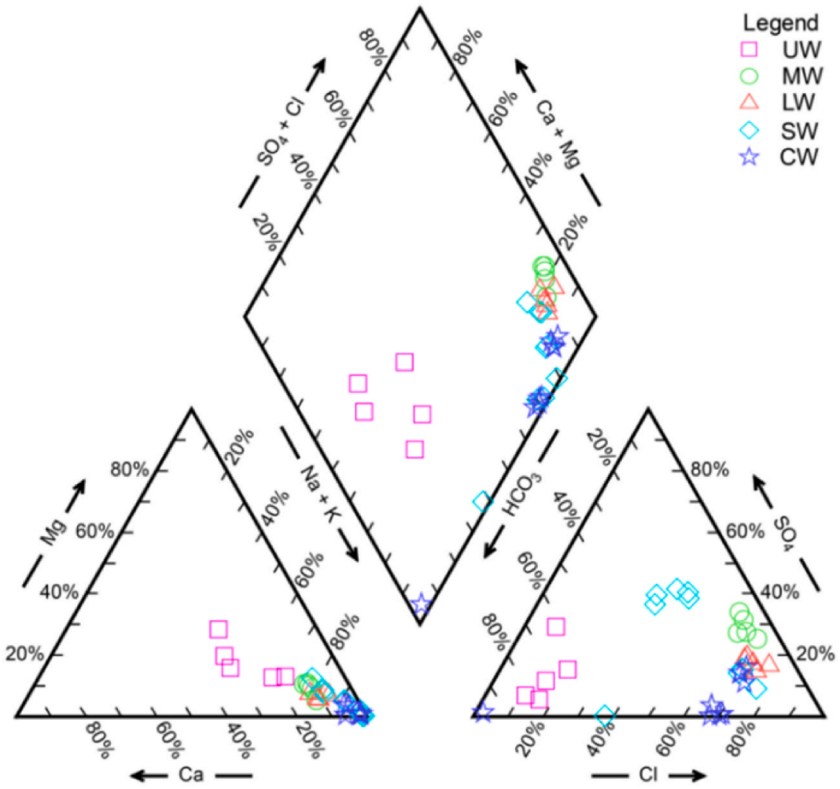


Fig. 4. Aquifer water chemistry Piper trilinear diagram.

the overall water quality type. The anion Cl^- content is high, and the main cation is K^++Na^+. The milligram equivalent concentrations of the ions in the upper water is significantly different from that in other aquifers.

5.1.4. *Correlation analysis*

A correlation analysis is performed for each indicator in the groundwater samples of the study area, and the degree of correlation between the indicators is described using a heat map (Fig. 5). The closer the correlation coefficient is to one, the greater the correlation will be. K^++Na^+ and Cl^- are highly positively correlated with TDS with correlation coefficients of 0.96 and 0.82, indicating that they are the main controlling ions for the variation of TDS in groundwater in the study area. The correlation coefficient of K^++Na^+ and Cl^- is 0.78, and a dissolved filtration dissolution effect exists. K^++Na^+ is negatively correlated, while Ca^{2+} and Mg^{2+} are negatively correlated, with correlation coefficients of −0.44 and −0.40, indicating the occurrence of direct alternate adsorption of cations. Ca^{2+} is significantly correlated with Mg^{2+} at the 0.01 level with a correlation coefficient of 0.83, indicating that they may have the same

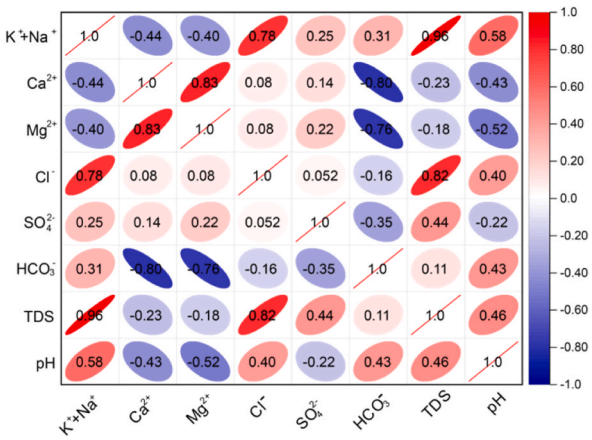


Fig. 5. Heat map of correlation coefficients among indicators.

source. The negative correlations of HCO_3^- with Ca^{2+} and Mg^{2+} are significant at the 0.01 level with correlation coefficients of −0.80 and −0.76, indicating that alternate cation adsorption and solubilization by filtration may co-exist.

5.2. Analysis of the mine water source identification model

5.2.1. Cluster analysis

The systematic cluster analysis used in this study is a natural upward aggregation clustering, which classifies the data at different levels until the requirements are met, thus forming a tree-like clustering structure for analysis. The sample data were imported into SPSS software, opened in the order of analysis, classification, and systematic clustering. Eight main factors, namely K^++Na^+, Ca^{2+}, Mg^{2+}, Cl^-, SO_4^{2-}, HCO_3^-, TDS, and pH, were used as variables in the systematic cluster analysis dialog box, and the intergroup linkage and squared Euclidean distance calculation methods were selected to obtain the systematic cluster analysis of the water source of the water inrush. The resulting spectrum diagram can be seen in Fig. 6. Samples were divided into 5 groups using a relative distance of 3 as the criterion for division. The correlation of the upper water, partly sandstone water and partly limestone water samples is strong, and they are concentrated in one group respectively in the spectrogram. However, the second group mixed multiple water samples to the extent that it was difficult to distinguish the origin of test water sample 2 from test water sample 3. Test water sample 4 was grouped separately, and its water source type could not be accurately determined from the results of systematic cluster analysis alone.

5.2.2. Factor analysis

The above correlation analysis results show that the correlation coefficients among some variables are too large. The existence of redundant information in the data will affect the discrimination result [39,40]. Therefore, factor analysis is used to reduce the dimension of variables. K^++Na^+, Ca^{2+}, Mg^{2+}, Cl^-, SO_4^{2-}, HCO_3^-, TDS and pH in this factor analysis were numbered X_1, X_2, X_3, X_4, X_5, X_6, $X_7$and X_8, respectively. The main control variables (X_1–X_8) of the 35 sets of groundwater sample data collected in the laboratory were imported into the SPSS software for factor analysis. First, the raw data were standardized, and the KMO (Kaiser-Meyer-Olkin) sampling aptness measure was 0.655 > 0.5, which showed a medium degree of suitability for factor analysis. The Bartlett sphericity test significance was 0.000 < 0.005, and the data demonstrated an overall normal distribution. Each variable was independent of each other to a certain extent, which was suitable for further analysis. Variables X_1, X_2, X_3, X_4, X_5 and X_7 have a variable commonality of 82% or more, and the explanatory power of this variable is strong. Meanwhile, variables X_6 and X_8 have variable commonalities of 75.5% and 47.7%, respectively.

In Table 2, the eigenvalues of the first three variables of the groundwater samples were 3.393, 2.151, and 1.114, which were all

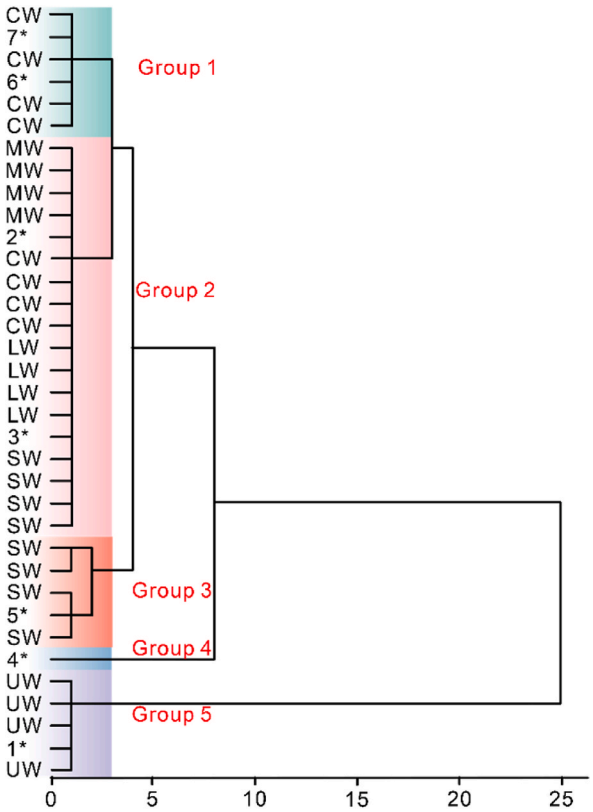


Fig. 6. Cluster analysis tree graph.

greater than one, and the cumulative contribution rate reached 83.223%. The first three variables could explain most of the information of the original variables. Accordingly, the three principal factors were extracted. The three variables were extracted using principal component analysis. The factor variables were rotated and transformed by using Caesar's normalized maximum variance method to make them more interpretable, and the principal factor score coefficient matrix was calculated, as shown in Table 3.

The evaluation factors were reduced from eight to three dimensions, eliminating the correlation between the indicators to a certain extent. The integrated decision model for the three main factors was obtained from the main factor score matrix (Table 3). Specifically, the expression for the integrated score of each sample was derived as follows (Equations (1)–(3):

$$Y_1 = -0.082X_1 + 0.338X_2 + 0.322X_3 + 0.224X_4 - 0.011X_5 - 0.333X_6 - 0.020X_7 - 0.058X_8, \tag{1}$$

$$Y_2 = 0.291X_1 + 0.049X_2 + 0.025X_3 + 0.462X_4 - 0.077X_5 - 0.096X_6 + 0.287X_7 + 0.226X_8, \tag{2}$$

$$Y_3 = 0.127X_1 - 0.051X_2 + 0.039X_3 - 0.282X_4 + 0.740X_5 + 0.008X_6 + 0.257X_7 - 0.323X_8. \tag{3}$$

5.2.3. Particle swarm optimization BP neural network model

The optimization variables in the optimization problem are determined by the positions of the particles. Meanwhile, the flight direction and rate magnitude are determined by the velocities of the particles, and the fitness function constructed by the optimization objective determines the degree of superiority or inferiority of the performance of each particle. The author writes the PSO algorithm iterative code based on computer knowledge (Table 4). The PSO algorithm flow is as follows:

First, the position and velocity of the randomly initialized particle in the solution space and the position x_{im} and velocity v_{1m} of the initialized particle in the mth dimension ($1 \leqq m_d$) are:

$$x_{im}(1) = x_{min} + p_1(x_{max} - x_{min}), \tag{4}$$

$$v_{1m}(1) = \frac{p_2\ (\ x_{max} - x_{min}\ )}{\Delta t}, \tag{5}$$

In equations (4) and (5): x_{max} and x_{min} are the maximum and minimum values of the optimized variable X, respectively; Δt is the time interval; and p_1 and p_2 are the random numbers in the interval [0,1].

Second, the particle population will generate individual extrema p_{best} and global extrema G_{best} at the tth iteration generation, whose positions are denoted as follows (Equations (6) and (7)):

$$p_i(t) = (p_{i1}(t)\ p_{i2}(t)p_{i3}(t).....p_{iN}(t)), \tag{6}$$

$$p_g(t) = \left(p_{g1}(t)\ p_{g2}(t)p_{g3}(t).....p_{gN}(t)\right). \tag{7}$$

Finally, after the above-mentioned two extremes are obtained, the mth dimensional ($1 \leqq md$) velocity v_{im} (t+1) and position x_{im} (t+1) of the particle at iteration (t+1) are:

$$v_{im}(t+1) = \omega\ v_{im}(t) + \varphi_1 q_1(p_{im}(t) - x_{im}(t)), \tag{8}$$

$$x_{im}(t+1) = x_{im}(t) + v_{im}(t) * \Delta t, \tag{9}$$

In equations (8) and (9): ω is the inertia weight, and φ_1 is the acceleration coefficient.

The BP neural network is trained using a gradient descent algorithm. Meanwhile, the backpropagation error signal is utilized to continuously correct the weights and thresholds, resulting in a slow learning speed of the aforementioned algorithm and easy to fall into local minimal values. The global optimization capability of the particle swarm optimization algorithm can effectively solve this drawback. The initialization of the particle velocity, position, inertia factor, and other parameters is carried out through several iterations to continuously update the particle velocity and position and calculate the individual and group extreme values of the particles to minimize the error, the output to obtain the global optimal particle position, and the initial weights and thresholds of the BP neural network to initialize the process according to the reverse error to continuously update the weights and thresholds until all sample errors reach the termination condition and complete the particle swarm of the BP neural network. Finally, the optimization of the BP neural network by particle swarm is completed. The particle swarm algorithm is used to optimize the traditional BP neural

Table 2
Variance interpretation of the initial eigenvalues.

Variables	Eigenvalues	Percentage of variance	Accumulation rate (%)
X_1	3.393	42.410	42.410
X_2	2.151	26.882	69.292
X_3	1.114	13.931	83.223
X_4	0.767	9.585	92.808
X_5	0.290	3.622	96.429
X_6	0.215	2.682	99.111
X_7	0.067	0.841	99.953
X_8	0.004	0.047	100.000

Table 3
Main factor score coefficient matrix.

Variable category	Y_1	Y_2	Y_3
K^++Na^+	−0.082	0.291	0.127
Ca^{2+}	0.338	0.049	−0.051
Mg^{2+}	0.322	0.025	0.039
Cl^-	0.224	0.462	−0.282
SO_4^{2-}	−0.011	−0.077	0.740
HCO_3^-	−0.333	−0.096	0.008
TDS	−0.020	0.287	0.257
pH	−0.058	0.226	−0.323

Table 4
PSO algorithm iteration code summary.

PSO algorithm iterative process

k = 1
while k ≤ MaxNum
ObjV=Objfun (x,P,T,hiddennum, P_test,T_test);
for i = 1:particlesize
f(i) = fitness (x (i,1));
 	
if ObjV(i) < personalbest_faval(i)
personalbest_faval(i) = ObjV(i);
personalbest_x (i,) = x (i,);
end
end
[globalbest_favalN,i] = min (personalbest_faval); globalbest_xn = personalbest_x (i,);
trace (1:N,k) = globalbest_xn;
trace (end,k) = globalbest_favalN;
for i = 1:particlesize
v (i,) = w*v (i,) + c1*rand*(personalbest_x (i,) - x (i,)) + c2*rand*(globalbest_x -x (i,));
for j = 1:narvs
if v (i,j) > vmax
v (i,j) = vmax;
elseif v (i,j) < -vmax
v (i,j) = -vmax;
end
end
x (i,) = x (i,) + v (i,);
end
globalbest_faval = globalbest_favalN;
globalbest_x = globalbest_xn;
k = k + 1;
end

network, thus greatly accelerating the convergence speed and improving the learning efficiency [41–43]. The model flow of coupling the particle swarm optimization algorithm with BP neural network is shown in Fig. 7.

5.2.4. FA-PSO-BP neural network model

The code is written in the editor dialog box using MATLAB software, and the parameters, such as the number of particle iterations, particle swarm size, and inertia factor, are initialized with an allowable error of 0.001 (Table 5). The three main factors extracted after factor analysis are used as input variables for water source identification, the number of implied layers is set to seven, and the output values of UW, MW, LW, SW, and CW are set to one, two, three, four, and five (Fig. 8). Meanwhile, and the training and test sets are randomly divided in the ratio of 8:2.

The performance of the neural network model is evaluated through several iterations (Fig. 9), and the FA-BP neural network model achieves the minimum mean square error of 0.072065 for the validation and test sets at epochs = 7 (Fig. 9a). The minimum mean square error of the FA-PSO-BP neural network model is 0.057592 at epoch = 5 (Fig. 9b) , MSE is optimized by 0.014473. The FA-PSO-BP neural network model achieves the minimum mean square error of 4.002×10^{-5} at epoch = 10. When the epoch = 10, the lowest point of the gradient image is 4.002×10^{-5}, which is the optimal position, and the error accuracy parameter Mu = 1.0×10^{-10}. The error curve of the sample no longer decreases for the five consecutive iterations, and the training error is not improved. Accordingly, the training is stopped (Fig. 9d). By contrast, the FA-BP neural network model was cycled 13 times, and the lowest gradient point was 4.4602×10^{-3}. The error accuracy parameter reached 1.0×10^{-9}, and the error was minimized by six consecutive iterations (Fig. 9c). The FA-PSO-BP neural network model has a faster merit seeking speed and a smaller mean square error. The R-values of the regression coefficients for the training, validation, test, and overall data of the FA-PSO-BP neural network model were 0.98850, 0.84175,

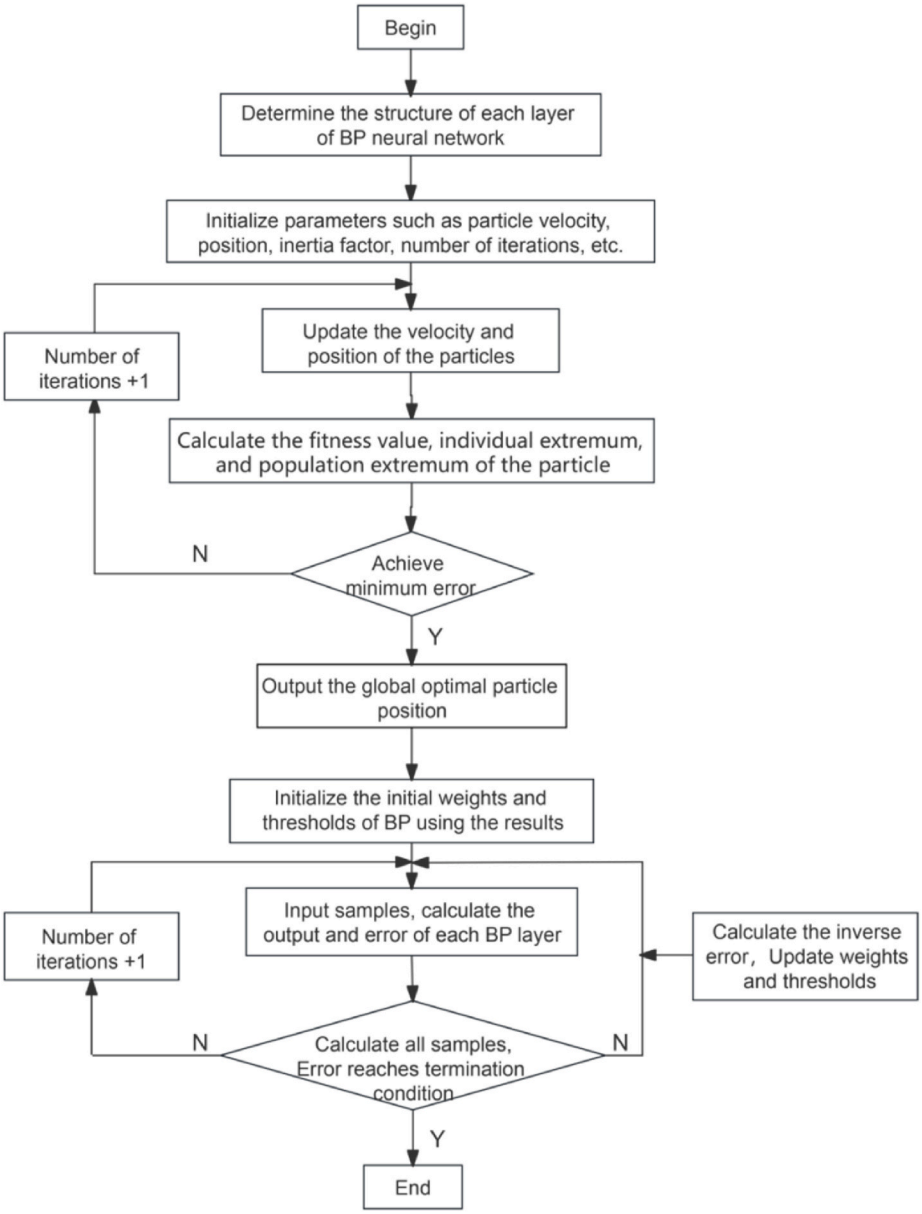


Fig. 7. PSO-BP neural network model flow.

Table 5
PSO algorithm parameter setting summary.

Parameters	Related settings
E0	0.001
MaxNum	50
narvs	N
particlesize	15
c1	1.5
c2	1.5
w	0.8
vmax	1
x	$-5 + 10$ * rand (particlesize, narvs)
v	2 * rand (particlesize, narvs)

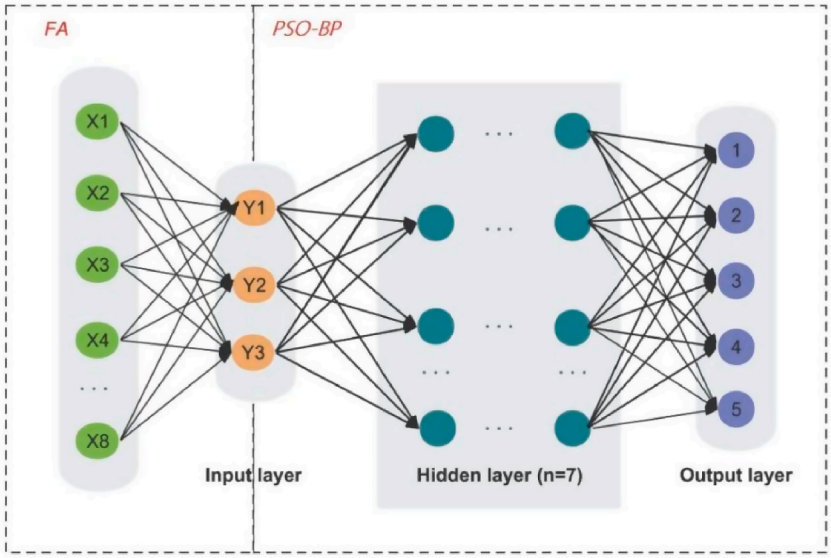


Fig. 8. FA-PSO-BP neural networks model structure diagram.

0.62223, and 0.89758 (Fig. 9f), respectively. Meanwhile, the R-values of the regression coefficients for each data of the FA-BP neural network model respectively were 0.98703, 0.74495, 0.43175, and 0.88885 (Fig. 9e), indicating that the water source discrimination model constructed by the FA-PSO-BP neural network was significantly more accurate.

The confusion matrix is a common evaluation metric for the classification models [44,45]. Majority of the evaluation metrics are derived from the confusion matrix. In this work, we choose three secondary metrics of the confusion matrix, namely, accuracy, precision, and recall, to evaluate the prediction results of the neural network model (Figures (10) and (11)). The numbers of samples of categories 1–5 were selected from the training set of the FA-BP neural network model as 4, 2, 5, 8, and 9 (Fig. 10a), among which two samples of category 3 were discriminated as category 5. One sample of category four was discriminated as category three, and the accuracy rates were 75% and 81.8%. The recall rates of category 3 and 4 were 60% and 87.5%, respectively, and the overall accuracy rate reached 89% (Fig. 10c). The number of samples from categories 1 to 5 was randomly selected in the test set as 1, 3, 0, 1, and 2, respectively (Fig. 10b). One sample of category 4 was discriminated as category 5, the accuracy rate of category 5 was 50%, the recall rate of category 4 was 50%, and the overall accuracy rate reached 85.7% (Fig. 10d). The training set of the FA-PSO-BP neural network model was randomly selected from class 1 to 5 sample numbers 4, 3, 5, 10, and 6, respectively (Fig. 11a). One sample of class 4 was discriminated as class 3, and the precision rate of class 3 was 83.3%. The recall rate of class 4 was 90%, and the overall accuracy rate reached 96.4% (Fig. 11c). The test set was randomly selected from class 1 to 5 sample numbers 1, 2, 0, 2, and 4, respectively (Fig. 11b). The precision, recall, and accuracy rates were 100% (Fig. 11d). The ratio of the training set to the test set is fixed at 8:2, and the samples are randomly generated. The FA-PSO-BP neural network model regression fits better by analyzing the precision, recall, and accuracy of the confusion matrix of the prediction results.

Among the 35 groups of water samples, seven groups of samples to be tested were randomly selected and substituted into the cluster analysis model, FA-BP water source discrimination model and FA-PSO-BP water source discrimination model for validation (Table 6). In three of the seven groups of samples to be tested, cluster analysis model could not accurately discriminate, and the accuracy rate was only 57.1%. The FA-BP water source discrimination model produced a misjudgment between the LW and the CW with a correct rate of 85.7%, and the FA-PSO-BP water source discrimination model had a correct rate of 100%, which improved by 14.3%. The correct rate of discrimination is significantly improved, and it is more instructive for the identification of water sources of water inrush in the Gubei mining area.

6. Discussion

6.1. Water chemistry characterization

There are many factors that affect the chemical characteristics of groundwater, such as hydrogeological conditions, human activities, water-rock interaction, etc. The author can only simply reveal the hydrochemical characteristics of groundwater through conventional hydrochemical analysis methods, and conduct appropriate analysis and reasoning on hydrogeological conditions. A comprehensive grasp of the main controlling factors of groundwater chemical characteristics in the study area still needs further exploration and research. The surface water in the study area is relatively scarce, while the atmospheric precipitation recharge is very abundant. The pore development of the upper aquifer of the unconsolidated formation provides a good runoff condition for the recharge of atmospheric precipitation, so that the ion concentration and salinity of the upper water are low. The chemical composition

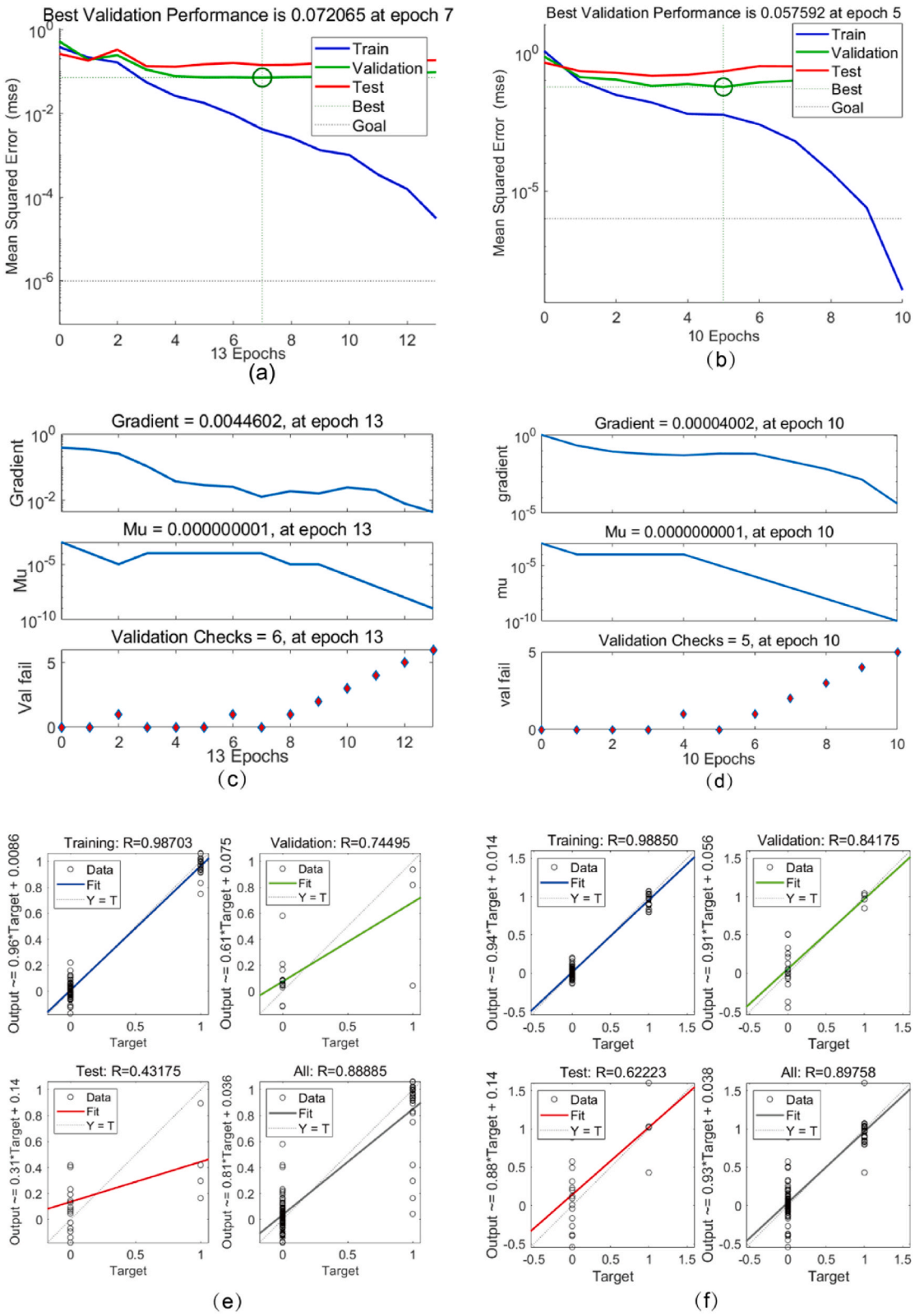


Fig. 9. Neural network model performance evaluation graph (a):FA-BP neural network model performance graph; (b):FA-PSO-BP neural network model performance graph; (c):FA-BP neural network model training state graph; (d):FA-PSO-BP neural network model training state graph; (e):FA-BP neural network model regression graph; (f):FA-PSO-BP neural network model regression graph.

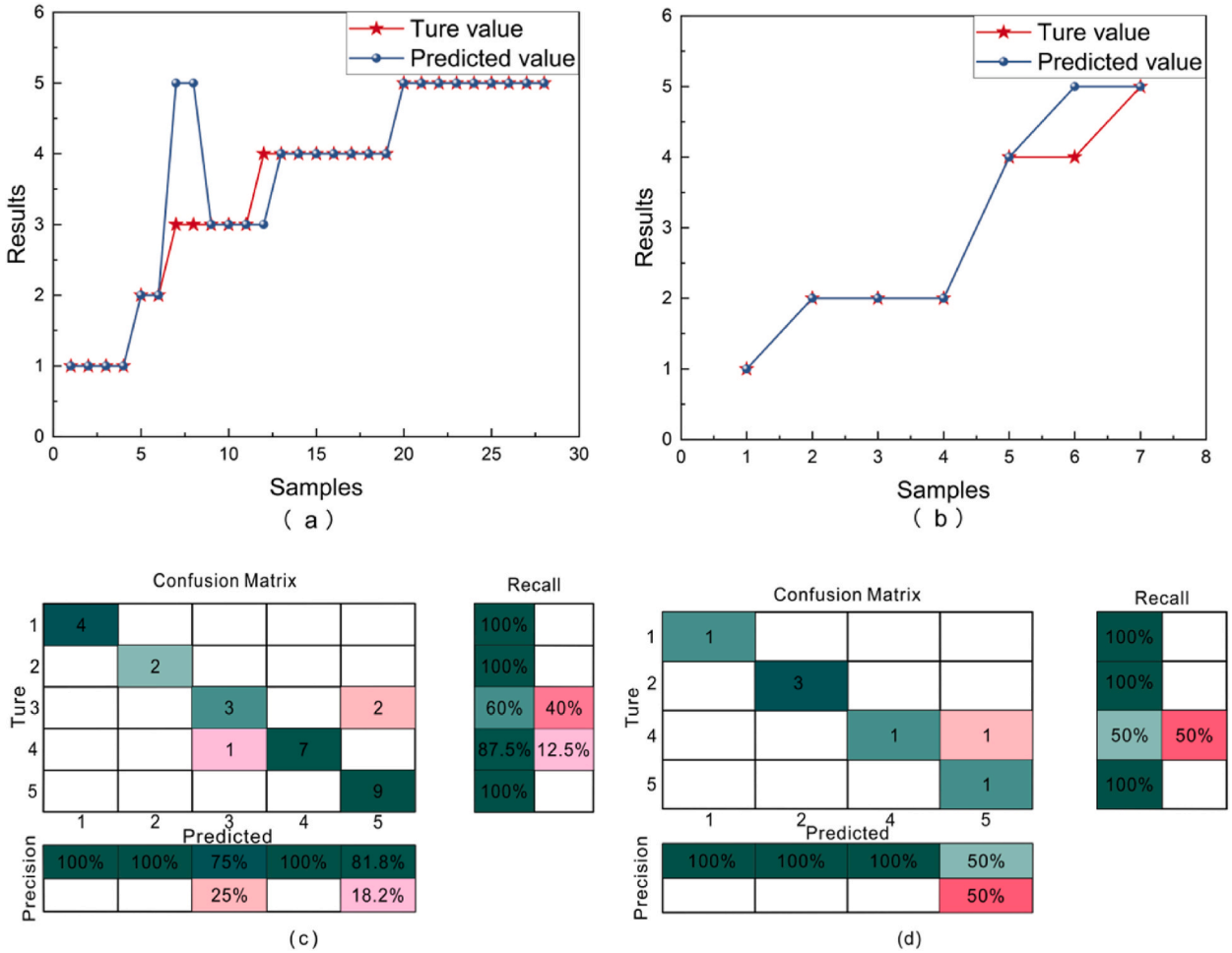


Fig. 10. The prediction results of the FA-BP neural network model.(a):FA-BP neural network model train results; (b):FA-BP neural network model test results; (c):FA-BP neural network model train results confusion matrix; (d):FA-BP neural network model test results confusion matrix.

of upper water and other kinds of water samples is quite different, it is inferred that the water barrier between upper water and middle water prevents the downward flow of upper water, making it difficult to produce hydraulic connection between upper water and middle water, and eventually leading to the sudden increase in the TDS of middle water, lower water, sandstone water and limestone water. Sandstone fissure aquifer is thick, and the development of sandstone fissure is very uneven, and there are large differences in water quality variation between sandstone aquifer and limestone aquifer, resulting in huge differences in some ions in the same aquifer.

6.2. *Mine water source identification model*

Correlation analysis not only reflects the chemical characteristics of groundwater, but also shows that there are correlations and multicollinearity among the variables of groundwater samples. The redundant information generated by the sample data will affect the accuracy of the water source discrimination results. Therefore, it is necessary to use factor analysis to reduce the dimension of groundwater sample data.The traditional BP artificial neural network model can achieve self-optimization through backpropagation. The model has strong self-learning function and strong robustness, but it is easy to fall into the local optimal solution and slow convergence speed.Particle swarm optimization has the ability of global optimization, which can effectively improve the traditional BP artificial neural network model.The research shows that the BP artificial neural network model after particle swarm optimization has better performance than the traditional BP artificial neural network model in terms of performance and fitting effect. FA-PSO-BP artificial neural network model is more accurate than FA-BP artificial neural network model and cluster analysis. Therefore, FA-PSO-BP artificial neural network model is worthy of further promotion in groundwater source identification.However, the model needs to learn and train a large amount of data to establish a database in order to accurately identify different water sample sources, which increases the difficulty of further promotion of the model. In this paper, the identification model of mine water inrush source in the study area is only based on groundwater chemical information, without considering the influence of groundwater temperature,

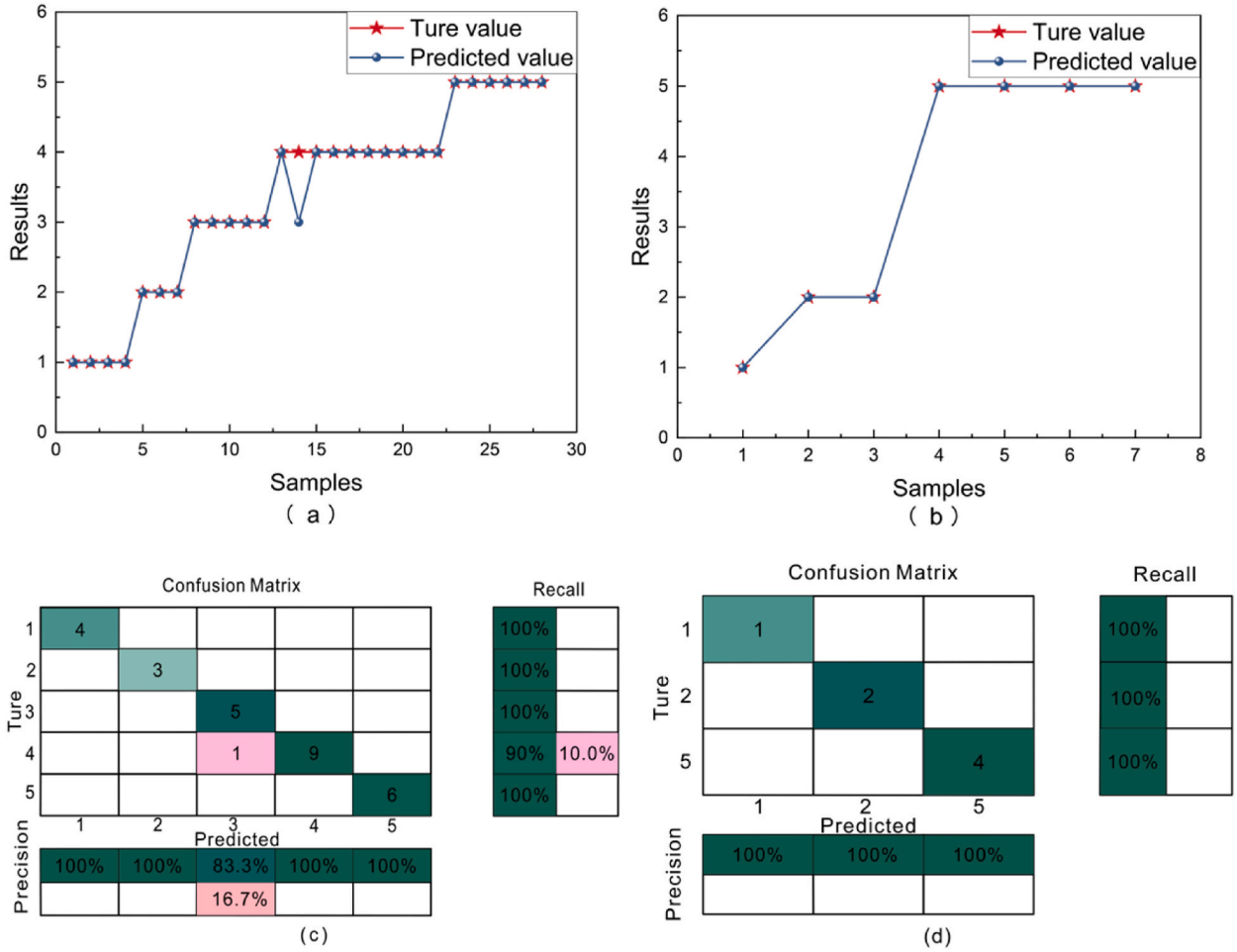


Fig. 11. The prediction results of the FA-PSO-BP neural network model.(a):FA-PSO-BP neural network model train results; (b):FA-PSO-BP neural network model test results; (c):FA-PSO-BP neural network model train results confusion matrix; (d):FA-PSO-BP neural network model test results confusion matrix.

Table 6
Comparison of the regeneration results.

Sample ID	FA-BP neural network model results	FA-PSO-BP neural network model results	Cluster analysis model results	Real results
1*	UW	UW	UW	UW
2*	MW	MW	/	MW
3*	LW	LW	/	LW
4*	LW	SW	/	SW
5*	SW	SW	SW	SW
6*	CW	CW	SW	CW
7*	CW	CW	CW	CW

human activities and other factors. In addition, limited by the quantity of groundwater samples, the applicability of this model needs to be further improved.

7. Conclusion

The statistical analysis, major ion analysis, and correlation analysis of $K^{+}+Na^{+}$, Ca^{2+}, Mg^{2+}, Cl^{-}, SO_4^{2-}, HCO_3^{-}, TDS, and pH of the water quality data of the different aquifers in the Gubei mine area were carried out to obtain the water chemistry characteristics of the different aquifers in the study area. The change of hydrochemical characteristics of groundwater shows obvious vertical division of water quality. The surface water source is scarce, but the atmospheric precipitation is abundant, so the shallow layer has abundant recharge. The groundwater chemical characteristics of the upper aquifer significantly differ from the water quality patterns of the other

aquifers. The proportion of cation K^++Na^+ and anion Cl^- occupy a clear advantage and are the key ions affecting the change of TDS. The water quality type is dominated by Cl–K+Na type, the overall water quality is weakly alkaline, and the alternate cation adsorption and dissolution filtration dissolution easily occur in the aquifer.

The BP neural network exhibits self-learning ability, generalization ability, and fault tolerance ability, and it learns and stores the information characteristics according to the input data reflection. The BP neural network in the learning process of the randomly selected weights and thresholds easily falls into local minima. Accordingly, a particle swarm optimization BP neural network model is proposed, and the particle swarm algorithm continuously optimizes the initial weights and thresholds of the BP neural network through the global optimal parameters of the particles to obtain the self-fitness value and the optimal extreme value. The performance and training effect of FA-PSO-BP neural network model are better than the FA-BP neural network model. The training learning accuracy reaches 96.4%, and the back generation results are all correct. We need to extensively collect the actual measurement data and establish a training sample library to further enhance the applicability of the model. Applying the model to the problem of mine water source identification can greatly improve the efficiency and accuracy, which is of great practical significance to ensure safe production in coal mines.

Data availability

Data included in article/supp.material/referenced in article.

Ethical approva

The authors confirm that accepted principles of ethical and professional conduct have been followed.

Consent to participate

The authors declare that the manuscript does not involve human subjects, does not describe human transplantation studies, and does not report on studies involving vulnerable groups.

Consent to publish

All authors agree to the publication of their work in the Heliyon.

Additional information

No additional information is available for this paper.

CRediT authorship contribution statement

Qilin Jiang: Writing – original draft, Software, Methodology, Formal analysis. **Qimeng Liu:** Project administration, Funding acquisition, Data curation, Conceptualization. **Yu Liu:** Resources, Investigation, Data curation. **Huichan Chai:** Visualization, Validation, Formal analysis. **Jingzhong Zhu:** Writing – review & editing, Supervision, Funding acquisition.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Qimeng Liu reports financial support was provided by Natural Science Foundation of Anhui. Jingzhong Zhu reports financial support was provided by the Postgraduate Research & Practice Innovation Program of Jiangsu Province. Jingzhong Zhu reports financial support was provided by the Fundamental Research Funds for the Central Universities. Jingzhong Zhu reports financial support was provided by Graduate Innovation Program of China University of Mining and Technology.

Acknowledgments

This work was financially supported by the Natural Science Foundation of Anhui Province under grant number 1908085ME145, the Postgraduate Research & Practice Innovation Program of Jiangsu Province (KYCX23_2760), the Fundamental Research Funds for the Central Universities (2023XSCX003) and the Graduate Innovation Program of China University of Mining and Technology (2023WLKXJ003). We sincerely thank the editors and reviewers for their valuable comments that greatly improved this paper.

References

[1] H. Ping-hua, W. Xin-yi, H. Su-min, Recognition model of groundwater inrush source of coal mine: a case study on Jiaozuo coal mine in China, Arabian J. Geosci. 10 (2017) 323, https://doi.org/10.1007/s12517-017-3099-5.

[2] X. Bai, H. Ding, J. Lian, D. Ma, X. Yang, N. Sun, W. Xue, Y. Chang, Coal production in China: past, present, and future projections, Int. Geol. Rev. 60 (2018) 535–547, https://doi.org/10.1080/00206814.2017.1301226.
[3] Y. Zhang, G. Feng, M. Zhang, H. Ren, J. Bai, Y. Guo, H. Jiang, L. Kang, Residual coal exploitation and its impact on sustainable development of the coal industry in China, Energy Pol. 96 (2016) 534–541, https://doi.org/10.1016/j.enpol.2016.06.033.
[4] H. Gui, X. Song, M. Lin, Water-inrush mechanism research mining above karst confined aquifer and applications in North China coalmines, Arabian J. Geosci. 10 (2017) 180, https://doi.org/10.1007/s12517-017-2965-5.
[5] H. Gui, M. Lin, Types of water hazards in China coalmines and regional characteristics, Nat. Hazards 84 (2016) 1501–1512, https://doi.org/10.1007/s11069-016-2488-5.
[6] S. Zhang, W. Guo, Y. Li, Experimental simulation of water-inrush disaster from the floor of mine and its mechanism investigation, Arabian J. Geosci. 10 (2017) 503, https://doi.org/10.1007/s12517-017-3287-3.
[7] K. Chen, L. Sun, J. Xu, Statistical analyses of groundwater chemistry in the Qingdong coalmine, northern Anhui province, China: implications for water–rock interaction and water source identification, Appl. Water Sci. 11 (2021) 50, https://doi.org/10.1007/s13201-021-01378-5.
[8] S.Y. Ganyaglo, B. Banoeng-Yakubo, S. Osae, S.B. Dampare, J.R. Fianko, Water quality assessment of groundwater in some rock types in parts of the eastern region of Ghana, Environ. Earth Sci. 62 (2011) 1055–1069, https://doi.org/10.1007/s12665-010-0594-3.
[9] S.J. Li, Q. Wu, F.P. Cui, Y.F. Zeng, G.R. Wang, Major characteristics of China's coal mine water disaster occurred in recent years, AMM 501 (2014) 336–340, https://doi.org/10.4028/www.scientific.net/AMM.501-504.336, 504.
[10] J. Zhang, Investigations of water inrushes from aquifers under coal seams, Int. J. Rock Mech. Min. Sci. 42 (2005) 350–360, https://doi.org/10.1016/j.ijrmms.2004.11.010.
[11] Q. Guo, Y. Yang, Y. Han, J. Li, X. Wang, Assessment of surface–groundwater interactions using hydrochemical and isotopic techniques in a coalmine watershed, NW China, Environ. Earth Sci. 78 (2019) 91, https://doi.org/10.1007/s12665-019-8053-2.
[12] J. Yang, S. Dong, H. Wang, G. Li, T. Wang, Q. Wang, Mine water source discrimination based on hydrogeochemical characteristics in the northern ordos basin, China, Mine Water Environ. 40 (2021) 433–441, https://doi.org/10.1007/s10230-020-00723-5.
[13] H. Bu, X. Tan, S. Li, Q. Zhang, Water quality assessment of the Jinshui River (China) using multivariate statistical techniques, Environ. Earth Sci. 60 (2010) 1631–1639, https://doi.org/10.1007/s12665-009-0297-9.
[14] Q. Liu, Y. Sun, Z. Xu, G. Xu, Application of the comprehensive identification model in analyzing the source of water inrush, Arabian J. Geosci. 11 (2018) 189, https://doi.org/10.1007/s12517-018-3550-2.
[15] H. Zhang, G. Xu, X. Chen, J. Wei, S. Yu, T. Yang, Hydrogeochemical characteristics and groundwater inrush source identification for a multi-aquifer system in a coal mine, Acta Geol. Sin. 93 (2019) 1922–1932, https://doi.org/10.1111/1755-6724.14299.
[16] Z. Guan, Z. Jia, Z. Zhao, Q. You, Identification of inrush water recharge sources using hydrochemistry and stable isotopes: a case study of Mindong No. 1 coal mine in north-east Inner Mongolia, China, J. Earth Syst. Sci. 128 (2019) 200, https://doi.org/10.1007/s12040-019-1232-4.
[17] N. Xie, Explanations about Grey Information and Framework of Grey System Modeling, GS 7 (2017) 179–193, https://doi.org/10.1108/GS-05-2017-0012.
[18] Q. Lu, Z. Bian, N. Tsuchiya, Hydrotransport-oriented Zn, Cu, and Pb behavior assessment and source identification in the river network of a historically mined area in the hokuroku basin, northeast Japan, IJERPH 16 (2019) 3907, https://doi.org/10.3390/ijerph16203907.
[19] D. Dong, Z. Chen, G. Lin, X. Li, R. Zhang, Y. Ji, Combining the Fisher feature extraction and support vector machine methods to identify the water inrush source: a case study of the wuhai mining area, Mine Water Environ. 38 (2019) 855–862, https://doi.org/10.1007/s10230-019-00637-x.
[20] P. Singh, R.M. Singh, Identification of pollution sources using artificial neural network (ANN) and multilevel breakthrough curve (BTC) characterization, Environ. Forensics 20 (2019) 219–227, https://doi.org/10.1080/15275922.2019.1629548.
[21] Y. Chen, S. Zhu, C. Yang, S. Xiao, Analysis of hydrochemical evolution in main discharge aquifers under mining disturbance and water source identification, Environ. Sci. Pollut. Res. 28 (2021) 26784–26793, https://doi.org/10.1007/s11356-021-12639-w.
[22] Y. Yang, J. Yue, J. Li, Z. Yang, Mine water inrush sources online discrimination model using fluorescence spectrum and CNN, IEEE Access 6 (2018) 47828–47835, https://doi.org/10.1109/ACCESS.2018.2866506.
[23] Q. Jiang, Q. Liu, Y. Liu, J. Zhu, H. Chai, K. Chen, Chemical composition of groundwater and its controlling factors in the Liuzhuang coal mine, northern Anhui Province, China, Water Supply 23 (2023) 4937–4956, https://doi.org/10.2166/ws.2023.290.
[24] L. Sun, H. Gui, Hydro-chemical evolution of groundwater and mixing between aquifers: a statistical approach based on major ions, Appl. Water Sci. 5 (2015) 97–104, https://doi.org/10.1007/s13201-014-0169-7.
[25] J. Zhang, L. Chen, Y. Chen, R. Ge, L. Ma, K. Zhou, X. Shi, Discrimination of water-inrush source and evolution analysis of hydrochemical environment under mining in Renlou coal mine, Anhui Province, China, Environ. Earth Sci. 79 (2020) 61, https://doi.org/10.1007/s12665-019-8803-1.
[26] Y. Bi, J. Wu, X. Zhai, G. Wang, S. Shen, X. Qing, Discriminant analysis of mine water inrush sources with multi-aquifer based on multivariate statistical analysis, Environ. Earth Sci. 80 (2021) 144, https://doi.org/10.1007/s12665-021-09450-8.
[27] P. Mohana, P.M. Velmurugan, Evaluation and characterization of groundwater using chemometric and spatial analysis, Environ. Dev. Sustain. 23 (2021) 309–330, https://doi.org/10.1007/s10668-019-00581-4.
[28] H. Su, W. Kang, Y. Xu, J. Wang, Assessment of groundwater quality and health risk in the oil and gas field of dingbian county, northwest China, Expo Health 9 (2017) 227–242, https://doi.org/10.1007/s12403-016-0234-6.
[29] S. Bencer, A. Boudoukha, L. Mouni, Multivariate statistical analysis of the groundwater of Ain Djacer area (Eastern of Algeria), Arabian J. Geosci. 9 (2016) 248, https://doi.org/10.1007/s12517-015-2277-6.
[30] A. Tiri, N. Lahbari, A. Boudoukha, Multivariate statistical analysis and geochemical modeling to characterize the surface water of oued chemora basin, Algeria, Nat. Resour. Res. 23 (2014) 379–391, https://doi.org/10.1007/s11053-014-9239-7.
[31] N. Hao, J. Yang, H. Liao, W. Dai, A unified factors analysis framework for discriminative feature extraction and object recognition, Math. Probl Eng. (2016) 1–12, https://doi.org/10.1155/2016/9347838, 2016.
[32] J.L. Horn, A rationale and test for the number of factors in factor analysis, Psychometrika 30 (1965) 179–185, https://doi.org/10.1007/BF02289447.
[33] D. Wang, D. Tan, L. Liu, Particle swarm optimization algorithm: an overview, Soft Comput. 22 (2018) 387–408, https://doi.org/10.1007/s00500-016-2474-6.
[34] X. Li, S. Xiang, P. Zhu, M. Wu, Establishing a dynamic self-adaptation learning algorithm of the BP neural network and its applications, Int. J. Bifurcation Chaos 25 (2015) 1540030, https://doi.org/10.1142/S0218127415400301.
[35] Yu Xiao-Hu, Guo-An Chen, Shi-Xin Cheng, Dynamic learning rate optimization of the backpropagation algorithm, IEEE Trans. Neural Network. 6 (1995) 669–677, https://doi.org/10.1109/72.377972.
[36] B. Zhang, X. Song, Y. Zhang, Y. Ma, C. Tang, L. Yang, Z.-L. Wang, The interaction between surface water and groundwater and its effect on water quality in the Second Songhua River basin, northeast China, J. Earth Syst. Sci. 125 (2016) 1495–1507, https://doi.org/10.1007/s12040-016-0742-6.
[37] C.M. Kazezyelmaz-Alhan, M.A. Medina, The effect of surface/ground water interactions on wetland sites with different characteristics, Desalination 226 (2008) 298–305, https://doi.org/10.1016/j.desal.2007.01.246.
[38] K. Chen, Q. Liu, W. Peng, X. Liu, Source apportionment and natural background levels of major ions in shallow groundwater using multivariate statistical method: a case study in Huaibei Plain, China, J. Environ. Manag. 301 (2022) 113806, https://doi.org/10.1016/j.jenvman.2021.113806.
[39] S. Dragović, N. Mihailović, Analysis of mosses and topsoils for detecting sources of heavy metal pollution: multivariate and enrichment factor analysis, Environ. Monit. Assess. 157 (2009) 383–390, https://doi.org/10.1007/s10661-008-0543-8.
[40] Y. Wang, L. Shi, M. Wang, T. Liu, Hydrochemical analysis and discrimination of mine water source of the Jiaojia gold mine area, China, Environ. Earth Sci. 79 (2020) 123, https://doi.org/10.1007/s12665-020-8856-1.
[41] E.T. Mohamad, D.J. Armaghani, E. Momeni, A.H. Yazdavar, M. Ebrahimi, Rock strength estimation: a PSO-based BP approach, Neural Comput. Appl. 30 (2018) 1635–1646, https://doi.org/10.1007/s00521-016-2728-3.
[42] Y. Song, D. Yang, M. Li, F. Cao, Investigations on optimal discharge pressure in CO2 heat pumps using the GMDH and PSO-BP type neural network—part B: experimental study, Int. J. Refrig. 106 (2019) 248–257, https://doi.org/10.1016/j.ijrefrig.2019.06.008.

[43] P. Wan, H. Zou, K. Wang, Z. Zhao, Research on hot deformation behavior of Zr-4 alloy based on PSO-BP artificial neural network, J. Alloys Compd. 826 (2020) 154047, https://doi.org/10.1016/j.jallcom.2020.154047.
[44] S. Ameen, S. Vadera, A convolutional neural network to classify American Sign Language fingerspelling from depth and colour images, Expet Syst. 34 (2017) e12197, https://doi.org/10.1111/exsy.12197.
[45] A.E. Maxwell, T.A. Warner, L.A. Guillén, Accuracy assessment in convolutional neural network-based deep learning remote sensing studies—Part 1: literature review, Rem. Sens. 13 (2021) 2450, https://doi.org/10.3390/rs13132450.