processes

MDPI

Article

Novel Method on Mixing Degree Quantification of Mine Water Sources: A Case Study

Qizhen Li [1,2], Gangwei Fan [1,2,*], Dongsheng Zhang [1,2], Wei Yu [1,2], Shizhong Zhang [1,2], Zhanglei Fan [1,2] and Yue Fu [3]

[1] School of Mines, China University of Mining and Technology, Xuzhou 221116, China; liqizhen@cumt.edu.cn (Q.L.); yuweitb2019@cumt.edu.cn (W.Y.); tbh252@cumt.edu.cn (S.Z.); fanzhanglei@cumt.edu.cn (Z.F.)
[2] State Key Laboratory of Coal Resources and Mine Safety, Xuzhou 221116, China
[3] School of Civil Engineering, Harbin Institute of Technology, Harbin 150006, China; a2177216755@163.com
* Correspondence: fangangwei@cumt.edu.cn

Abstract: After a mine water inrush occurs, it is crucial to quickly identify the source of the water inrush and the key control area, and to formulate accurately efficient water control measures. According to the differences in water chemical characteristics of four aquifers in the Fenyuan coal mine, the concentrations of K^+~Na^+, Ca^{2+}, Mg^{2+}, Cl^-, $SO_4{}^{2-}$, and $HCO_3{}^-$ were taken as water source identification indexes. A decision tree classification model based on the C4.5 algorithm was adopted to visualize the chemical characteristics of a single water source and extract rules, and intuitively obtained the discrimination conditions of a single water source with Mg^{2+}, Ca^{2+}, and Cl^- as important variables in the decision tree: $Mg^{2+} < 39.585$ mg/L, $Cl^- < 516.338$ mg/L and $Mg^{2+} \geq 39.585$ mg/L, $Ca^{2+} < 160.860$ mg/L. Factor analysis and Fisher discriminant theory were used to eliminate the redundant ion variables, and the discriminant function equations of the two, three, and four types of mixed water sources were obtained successively in turn. This paper puts forward MSE, RMSE, and MAE as the evaluation indexes of the water source mixing degree calculation models and obtains the ranking of the pros and cons of the mixed water source mixing degree calculation models. The results show that the minimum inscribed circle analytical method is the optimal model for the calculation of the mixing degree of two types of water sources, and the MSE, RMSE, and MAE are 0.17%, 4.13%, and 4.13%, respectively. The minimum inscribed circle clustering method is the optimal model for the calculation of the mixing degree of three types of water sources, and the minimum distance method is the optimal model for the calculation of the mixing degree of four types of water sources. The method of mine water source identification based on the decision tree C4.5 algorithm and mixing degree calculation has the characteristics of a simple calculation process, high efficiency, objective accuracy, and low cost, which can provide a scientific basis for the development of stope water control measures.

Keywords: mine water inrush source; decision tree; discriminant function equation; mixing degree of water sources

Citation: Li, Q.; Fan, G.; Zhang, D.; Yu, W.; Zhang, S.; Fan, Z.; Fu, Y. Novel Method on Mixing Degree Quantification of Mine Water Sources: A Case Study. *Processes* **2024**, *12*, 438. https://doi.org/10.3390/pr12030438

Academic Editor: Iqbal M. Mujtaba

Received: 20 January 2024
Revised: 8 February 2024
Accepted: 18 February 2024
Published: 21 February 2024

1. Introduction

In coal mining activities, mine water inrush disasters are an important factor restricting mine safety production [1–4]. Quickly and accurately identifying the causes of water damage and water source sources can effectively predict and determine the aquifer where water damage is expected to occur or the location of water damage accidents and can also provide guidance for formulating reasonable and effective treatment measures [5,6]. Therefore, the study of efficient and accurate water source identification methods and water source mixing degree calculation models has important guiding significance for safe mining.

Different aquifers in mines contain different chemical and isotopic compositions. Chemical characteristics analysis of aquifer water samples can clearly reflect the water quality characteristics of the aquifer [7–9]. The traditional water source identification method has limitations such as long time consumption and low efficiency. It also requires high human experience, knowledge reserve, and analytical judgment ability, and the general effect is not ideal [10,11]. With the advancement and development of science and technology, statistical methods, machine learning, and other technologies have made great progress in water source identification [12–15]. Among multivariate statistical methods, cluster analysis [16,17] and discriminant analysis are commonly used [18,19]. The introduction of hybrid calculations of multiple linear analyses, including fuzzy mathematics methods [20], gray correlation analysis [21,22], and extension identification methods provide new ideas for identifying water sources. Duan et al. used the maximum likelihood method to estimate the mixing ratios of water sources at each site and analyze the evolution rules of mine water [23]. Wei et al. tested six conventional hydrochemical indexes by using the multivariate mixed model and fuzzy comprehensive evaluation [24]. Huang et al. constructed a Piper–PCA–Bayes–LOOCV discrimination model of the water inrush source in mines that can increase the recognition accuracy effectively [25]. With the advent of the big data era, the development of artificial neural networks and the development of deep learning algorithms based on artificial neural networks, directly and independently learning and extracting features from a large amount of water sample information, fully and accurately reflecting the internal levels and corresponding rules of data [26–29]. Although these methods have significantly improved the accuracy of water source identification, the selection of relevant parameters and optimization algorithms of these network models is interfered with by human factors, and their internal recognition mechanism is always a "gray box" or "black box". The model structure is relatively complex [30–32]. At the same time, according to practical experience, the mine effluent is often mixed by different aquifers, that is, groundwater with a higher degree of mixing is relatively more serious to mine water inrush [33]. In the past, scholars rarely studied the mixing degree of groundwater in mines. Therefore, it is very necessary to study the source and mixing degree of water inrush, clarify the source of water inrush in the stope and key prevention areas, and formulate accurate and efficient water control measures.

Based on the hydrogeological information of the Fenyuan coal mine, this paper plans to build a decision tree classification model based on the C4.5 algorithm, visualizing and extracting the chemical characteristics of water sources, and obtaining corresponding intuitive discrimination conditions, thereby realizing a single water source identification. According to the sources of water inrush in different areas underground, various analysis models are used to calculate the mixing degree of different types of water sources successively and propose an objective and accurate optimal mathematical analysis. Research achievements are proposed to improve the scientific basis for clarifying the sources of water inrush, locating the key control areas, and developing accurate and efficient measures.

2. Original Data

2.1. Hydrogeological Conditions in Study Area

Fenyuan coal mine (China) is located in the northeast of Jingle County about 30 km, with geographical coordinates of 38°29′32″~38°32′58″ north latitude, 112°06′42″~112°08′30″ east longitude, the mineable coal seams are No.2 coal seam and No.5 coal seam. The coal seam thickness of the No. 2 coal seam is 0~9.25 m, with an average of 1.51 m, and the coal seam inclination angle is 30°. According to the location of the No.2 coal seam in the Shanxi formation, when mining the No.2 coal seam, the water inflow mainly comes from K4 sandstone fissure water (coal seam roof) and K3 sandstone fissure water (coal seam floor) of Shanxi formation, and some areas can be indirectly filled by L1 and L2 limestone karst water of Taiyuan formation. The No.5 coal seam is located in the Taiyuan formation, which thickness of the coal seam is 1.85~20.06 m, with an average distance of 10.29 m, and the spacing between two coal seams is 47.85–120.85 m (an average distance of 62.0 m). When the water-conducting fissures produced by mining No.5 coal in the minefield affect the Shanxi formation, the K4 and K3 sandstone fissure water of the Shanxi formation, and the L1 and L2 limestone karst water of the Taiyuan formation (coal seam roof), and the

K1 sandstone fissure water of the Taiyuan formation can be used as direct water-filling aquifers. Except for the mined area in the eastern part of the minefield, other sections are in the pressure mining area. The main threatening aquifer of the coal seam floor in the minefield is the Ordovician limestone aquifer, whose minimum water depth is less than 20 m from the floor of the No.5 coal seam. The specific stratigraphic information is shown in Figure 1.

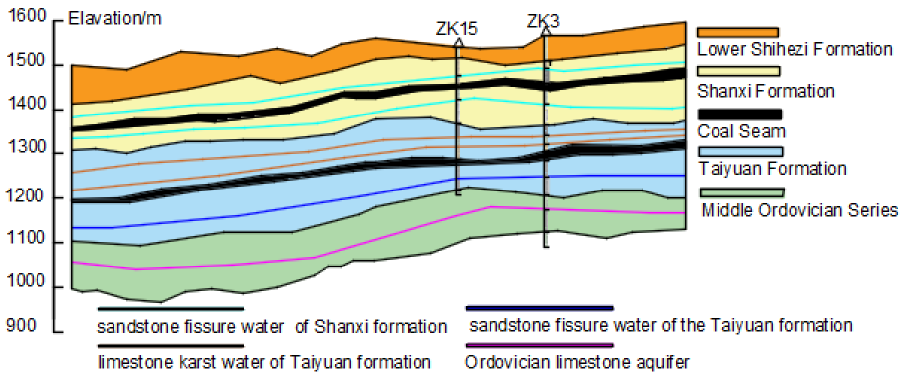


Figure 1. Comprehensive strata log diagram.

2.2. Water Source Data

When mining the No.2 or the No.5 coal seam, the mine water-filled aquifers are divided into four categories based on the mine water-filling conditions and aquifer types: I. sandstone fissure water of Shanxi formation (K3 and K4), II. limestone karst water of the Taiyuan formation (L1 and L2), III. sandstone fissure water of the Taiyuan formation (K1), IV. Ordovician limestone aquifer. Through quantitative monitoring of hydrochemical information at different locations in the Fenyuan coal mine during the same period. Using the concentrations of K^+~Na^+, Ca^{2+}, Mg^{2+}, Cl^-, $SO_4{}^{2-}$, and $HCO_3{}^-$ as evaluation variables, 100 water samples from four types of water sources were obtained. Among them, including 15 samples of Class I water sources (sandstone fissure water of Shanxi formation) and 40 samples of Class II water sources (karst water of Taiyuan formation). There are 24 samples of class III water sources (sandstone fissure water of the Taiyuan formation) and 21 samples of class IV water sources (Ordovician limestone water). The specific ion information of water source samples is shown in Table 1.

Table 1. Ion concentration of water source (mg/L).

Serial Number	Category	K^+~Na^+	Ca^{2+}	Mg^{2+}	Cl^-	$SO_4{}^{2-}$	$HCO_3{}^-$
1	I	199.29	128.84	89.15	420.21	202.23	510.72
2	I	180.55	95.66	63.53	64.79	231.63	419.27
3	I	410.37	99.02	70.98	78.65	910.68	349.76
4	I	91.04	99.75	59.33	66.15	228.90	408.79
5	I	299.78	95.97	68.36	224.91	422.21	510.09
…	…	…	…	…	…	…	…

"…" in Table 1 represents the remaining 95 sample data. Detailed sample data can be obtained by contacting the corresponding author.

3. Methods and Examples

3.1. Single Water Source Identification

According to the monitoring values of hydrochemical information of four types of water sources, in order to classify and predict these four types of water sources more intuitively and efficiently, the paper adopts the decision tree C4.5 algorithm model to automatically select the characteristics of hydrochemical information, and the information gain rate is used as the decision tree model to select the root node and internal node branch attribute [34]. The specific model principles: the data set is N, and the data set is divided into categories n. The number of sample data belonging to class i is c_i, and the total number of sample data in N is $|N|$, then the probability of a sample data belonging to class i is as follows [35,36]:

$$P(c_i) = \frac{c_i}{|N|} \tag{1}$$

The decision tree divides c information entropy into:

$$H(N,c) = H(N) = -\sum_{i=1}^{n} p(c_i)\log_2 p(c_i) \tag{2}$$

If attribute a is selected for testing, m is the number of subsets split by attribute a, and its conditional entropy is as follows:

$$H(N|a) = -\sum_{j=1}^{m} p(a=a_j)\sum_{i=1}^{n} p(c_i|a=a_j)\log_2 p(c_i|a=a_j) \tag{3}$$

Then the information content of attribute a is as follows:

$$I(N,a) = H(N) - H(N|a) \tag{4}$$

$$I(N,a) = -\sum_{i=1}^{m} p(c_i)\log_2 p(c_i) - \sum_{j=1}^{m} p(a=a_j)\sum_{i=1}^{n} p(c_i|a=a_j)\log_2 p(c_i|a=a_j) \tag{5}$$

The information gain rate of attribute a is as follows:

$$E(N,a) = \frac{I(N,a)}{H(N,a)} \tag{6}$$

According to the selection criteria, the $E(N,a)$ maximum attribute is selected as the test attribute. Due to the existence of noise and outliers in the data, many branches are abnormal when creating a decision tree, and the pruning method is often used to deal with this overfitting data problem. Firstly, reasonable parameters of minleaf are set to optimize the decision tree and analyze the impact of the minimum number of samples contained in leaf nodes on the performance of decision trees. Then, $leaf = \log space(1,2,10)$ is used to produce ten points between ten and one hundred, the number of cross-validation times is ten. As shown in Figure 2, when the minleaf is 13, the minimum cross-validation error is 0.040.

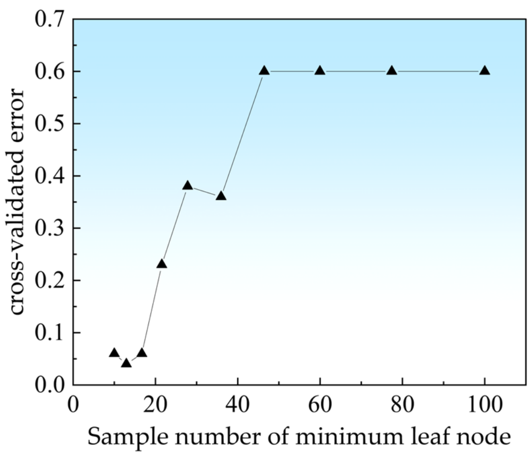


Figure 2. Decision tree parameter optimization diagram.

As shown in Figure 3, the CV loss pruning command is used to prune the optimized decision tree, and the important variables affecting the construction of the decision tree are

Mg^{2+}, Ca^{2+}, and Cl^-, whose parameter values are 0.68, 0.24, and 0.08, respectively. The clustering results, which are shown in Figure 4, obtained the discrimination conditions of a single water source with Mg^{2+}, Ca^{2+}, and Cl^- as important variables in the decision tree: $Mg^{2+} < 39.585$ mg/L, $Cl^- < 516.338$ mg/L, and $Mg^{2+} \geq 39.585$ mg/L, $Ca^{2+} < 160.860$ mg/L. With the help of discriminating conditions, the type of water source can be accurately identified.

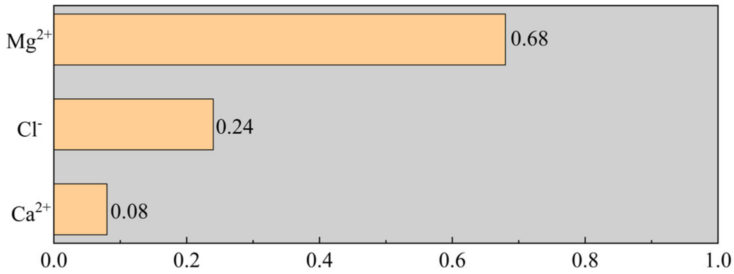


Figure 3. The importance of predictor variables.

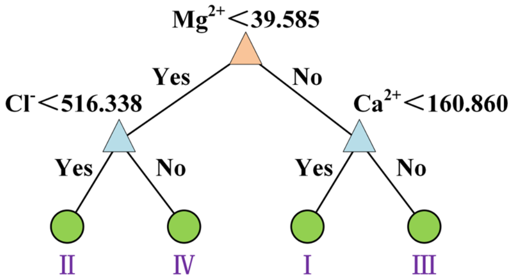


Figure 4. C4.5 decision tree construction.

The C4.5 algorithm decision tree is one hundred percent correct on 100 water samples from the Fenyuan coal mine. Therefore, the decision tree classification method based on the C4.5 algorithm has a good application effect in the identification of a single water source in the Fenyuan coal mine.

3.2. Mixing Degree Quantification of Mine Water Sources

When mine water inrush occurs, the water inrush source often comes from the supply of multiple aquifers. In order to clarify the source of water inrush and the key control areas, formulating precise and efficient water control measures, the mixing degree of water sources is one of the important scientific bases. With reference to the water output characteristics of the Fenyuan coal mine, model calculations are conducted on the mixed water sources.

3.2.1. Mixing Degree Quantification of Two Types of Water Sources

Take mixed water sources containing II and III as an example, with a total of 65 sets of samples. The discriminant function equation is established based on factor analysis and Fisher discriminant theory [37,38]. The calculation results show that the determination coefficients of K^+~Na^+ and Ca^{2+} are less than 0.01, which is significant for the determination and classification of water sources. According to the calculation of Fisher's linear discriminant function coefficients (the original case classification accuracy rate is 100%), two types of discriminant functions need to be established to explain most of the information of water sources ion characteristics, and Fisher's linear discriminant function is established as follows:

$$F_a = 0.085K^+ \sim Na^+ + 0.259Ca^{2+} - 16.968 \tag{7}$$

$$F_b = 0.176K^+ \sim Na^+ + 1.306Ca^{2+} - 169.968 \tag{8}$$

As shown in Figure 5, the water sources ion data are brought in the discriminant function to solve the F_a and F_b of each sample, thereby obtaining the water sample discriminant clustering.

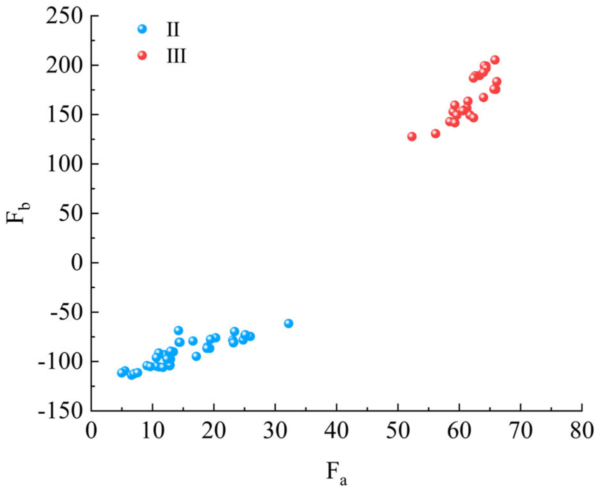


Figure 5. Clustering effect of water sources.

(1) Reference line intersection method

As shown in Figure 6, selecting one point from II water source samples (40) and III water source samples (24), respectively, and the line connecting the two points is used as the reference line, the mapping position of the other water source sample points on the reference line can be 100% included. Traversing 960 combinations by the loop algorithm to determine point A ($F_a = 4.96$, $F_b = -111.5$) and B ($F_a = 65.82$, $F_b = 205.35$), where point A represents 100% of II water source and decreases in the direction of point B successively and the position of point B indicates that the content of II water source is 0. The characterization direction of the content of the III water source is opposite to that of the II water source.

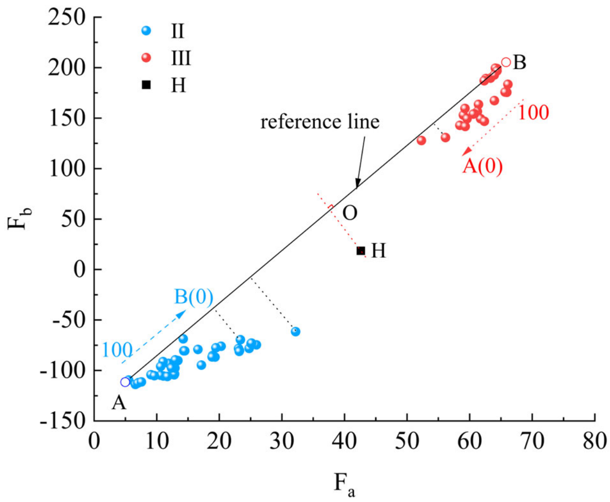


Figure 6. Reference line intersection method.

The new water source sample point is H, and its mapping point on the reference line (A–B) is point O. Therefore, the mixing degree of II water source and III water source of new sample point H calculated is as follows:

$$\omega_{\mathrm{II}} = \frac{OB}{AB} \times 100\% \tag{9}$$

$$\omega_{\mathrm{III}} = \frac{OA}{AB} \times 100\% \tag{10}$$

(2) Minimum inscribed circle analytical method

The algorithm is used to obtain the smallest inscribed circle that can cover the II water source sample point set and the III water source sample point set. The minimum inscribed circle center O_{II} (19.39, −87.66) and the minimum gathering radius R_{II} are obtained for the sample point set of II water sources, and the minimum inscribed circle center O_{III} (59.06, 166.57) and the minimum gathering radius R_{III} is 39.36, as shown in Figure 7. Taking the new water source sample point H, points A and B are the minimum distance positions from the two inscribed circles to point H, respectively. Then, the mixing degree of the II water source and the III water source of new sample point H is calculated as follows:

$$AH = \mathrm{O_{II}}H - R_{\mathrm{II}} \tag{11}$$

$$BH = \mathrm{O_{III}}H - R_{\mathrm{III}} \tag{12}$$

$$\omega_{\mathrm{II}} = \left(1 - \frac{AH}{AB}\right) \times 100\% \tag{13}$$

$$\omega_{\mathrm{III}} = \left(1 - \frac{BH}{AB}\right) \times 100\% \tag{14}$$

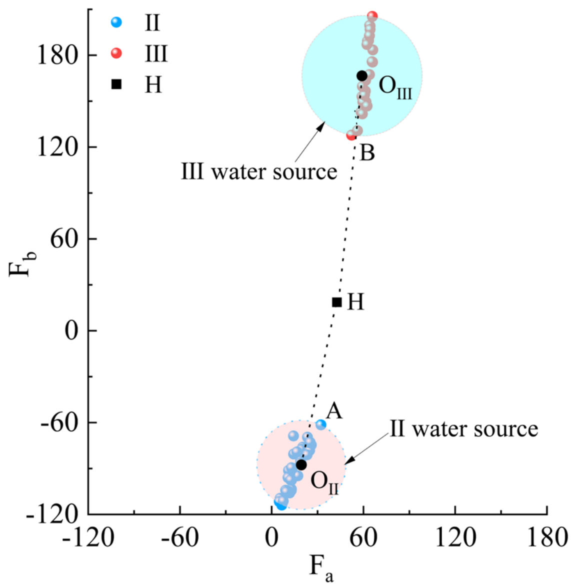


Figure 7. Minimum inscribed circle analytical method.

(3) Minimum inscribed circle comprehensive method

After obtaining the feature range of two types of water source samples II and III, the analysis method determined the center coordinates of the two minimum inscribed circles O_{II} and O_{III}, and took the new water source sample point H, whose mapping point on the reference line of $O_{\mathrm{II}}O_{\mathrm{III}}$ is point M, as shown in Figure 8. Therefore, the mixing degree of the II water source and the III water source of new sample point H is calculated as follows:

$$\omega_{\mathrm{II}} = \frac{\mathrm{O_{III}}M}{\mathrm{O_{II}}O_{\mathrm{III}}} \times 100\% \tag{15}$$

$$\omega_{\mathrm{III}} = \frac{\mathrm{O_{II}}M}{\mathrm{O_{II}}O_{\mathrm{III}}} \times 100\% \tag{16}$$

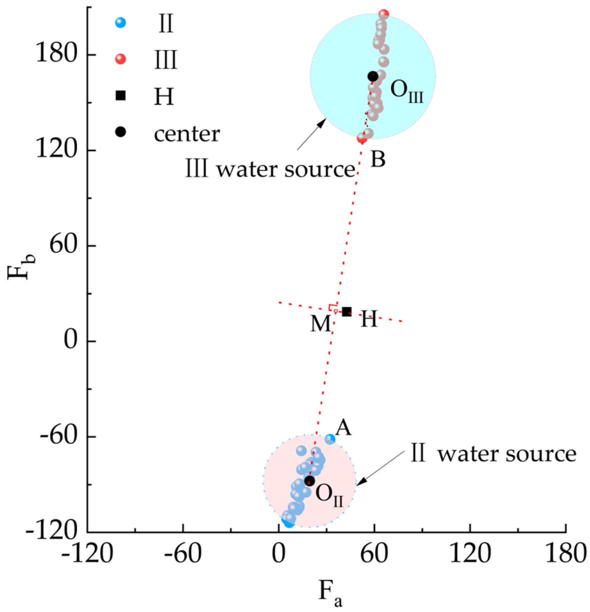


Figure 8. Minimum inscribed circle comprehensive method.

Take a certain mixed water sample as a case study. The mixed water sample is composed of II water sources and III water sources at the ratio of 62.00% and 38.00% (II water source: K^+~Na^+: 555.35, Ca^{2+}: 7.56, Mg^{2+}: 3.15, Cl^-: 190.79, $SO_4{}^{2-}$: 427.88, $HCO_3{}^-$: 464.52; III water source: K^+~Na^+:267.23, Ca^{2+}: 207.90, Mg^{2+}: 82.01, Cl^-: 255.15, $SO_4{}^{2-}$: 644.70, $HCO_3{}^-$: 452.09; mixed water sample H: K^+~Na^+: 445.86, Ca^{2+}: 83.69, Mg^{2+}: 33.11, Cl^-: 215.24, $SO_4{}^{2-}$: 510.27, $HCO_3{}^-$: 459.80; the unit is mg/L). According to the method shown in Figure 6, $\omega_{II} = 46.33\%$ and $\omega_{III} = 53.67\%$ are obtained. The values of F_a and F_b calculated from the discriminant function are 42.61 and 18.67, respectively. According to Figure 7, $\omega_{II} = 57.87\%$ and $\omega_{III} = 42.13\%$ are calculated by the minimum inscribed circle analytical method. According to Figure 8, the coordinate position of point M and $\omega_{II} = 57.88\%$ and $\omega_{III} = 42.12\%$ are calculated by the minimum inscribed circle comprehensive method.

3.2.2. Mixing Degree Quantification of Three Types of Water Sources

Based on factor analysis and Fisher discriminant theory, the water samples are identified to show that the determination coefficients of K^+~Na^+, Ca^{2+} and Mg^{2+} are less than 0.01, which have significant significance for the determination and classification of water sources, while Cl^-, $SO_4{}^{2-}$, and $HCO_3{}^-$ are eliminated. As shown in Table 2, the variance contribution rate of the first type of discriminant function is 95.30%, indicating that this function can explain most of the information on the ion characteristics of water samples. The variance contribution rate of the second type of function is 4.70%, and it is tested by the Wilk Lambda function (the significance of function 1 to function 2 is $0.000 < 0.01$, and the significance of function 2 is $0.000 < 0.01$). The combination of two kinds of discriminant functions can obtain all the information of water samples.

Table 2. Eigenvalues of three types of water sources.

Function	Eigenvalues	Variance	Cumulative Percentage	Canonical Correlation
1	44.548 [a]	95.3	95.3	0.989
2	2.213 [a]	4.7	100.0	0.830

"a" in Table 2 represents the dependent variable.

Based on the statistics of typical discriminant function coefficients, the discriminant functions of the three types of water samples are obtained as follows:

$$F_a = 0.005K^+ \sim Na^+ + 0.079Ca^{2+} - 0.015Mg^{2+} - 8.370 \quad (17)$$

$$F_b = 0.004K^+ \sim Na^+ - 0.059Ca^{2+} + 0.147Mg^{2+} - 2.751 \quad (18)$$

As shown in Figure 9, according to the discriminant function, the water sample ion data are brought in to solve the a and b values of each sample, and to obtain the discriminant clustering of water samples.

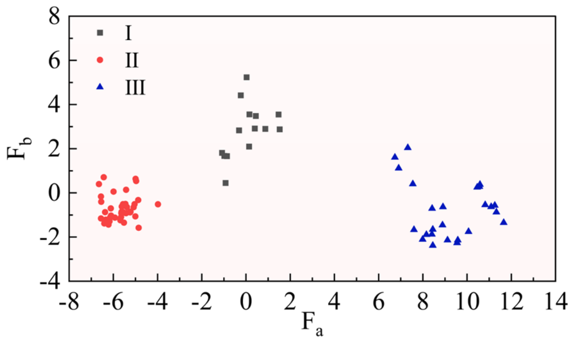


Figure 9. Cluster diagram of three types of mixed water sources.

(1) Maximum coverage traversal method

The principle of the mixing degree of water sources is quantified by the maximum coverage traversal method. As shown in Figure 10, a triangle is established with reference water sample points A, B, and C as endpoints, in which three sample points A, B, and C are successively selected from the I water source (15 samples), the II water source (40 samples) and the III water source (24 samples). By using the loop algorithm to traverse 14400 combinations, it is determined that the triangle established by the three sample points A, B, and C can maximize the coverage of the mixed water source sample points. When point A is the 8# water sample, point B is the 2# water sample, and point C is the 18# water sample, the triangle ABC covers a maximum of 68.35% (54) of the water sample points.

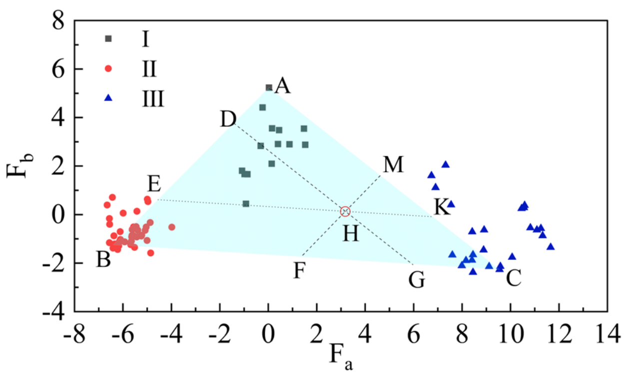


Figure 10. Maximum coverage traversal method.

Assuming that the mixed water sample point H to be determined exists inside the triangle, the mixing ratio of multiple water sources is determined according to the distance between the two points in the figure. The specific mixing degree is calculated as follows:

$$R_A = \frac{BE}{AB} \quad (19)$$

$$R_B = \frac{CG}{BC} \quad (20)$$

$$R_C = \frac{AM}{AC} \tag{21}$$

$$\omega_{\mathrm{I}} = \frac{R_A}{\sum\limits_{i=A}^{C} R_i} \times 100\% \tag{22}$$

$$\omega_{\mathrm{II}} = \frac{R_B}{\sum\limits_{i=A}^{C} R_i} \times 100\% \tag{23}$$

$$\omega_{\mathrm{III}} = \frac{R_C}{\sum\limits_{i=A}^{C} R_i} \times 100\% \tag{24}$$

(2) Minimum inscribed circle clustering method

The principle of the mixing degree of water sources is quantified by the maximum inscribed circle clustering method. As shown in Figure 11, using algorithm analysis to obtain the minimum inscribed circle O_{I} containing I water sources, the area within this circle can represent I water source type, with the center of the circle being O_{I} $(-0.451, 2.843)$, and the radius R_{I} is 2.441. Similarly, it can be obtained that the center of the minimum inscribed circle for II water source (40 samples) is O_{II} $(-5.420, -0.322)$ and the radius R_{II} is 1.449, the center of the minimum inscribed circle for III water source (25 samples) is O_{III} (9.196), 0.129), radius R_{III} is 2.870. Taking the mixed water sample point H to be determined, A, B, and C are the points on the three inscribed circles with the smallest distance to point H. The specific mixing degree calculation expression is as follows:

$$L_i = O_i H - R_i \quad (i = \mathrm{I}, \mathrm{II}, \mathrm{III}) \tag{25}$$

$$\omega_k = \frac{L_k}{\sum\limits_{i=\mathrm{I}}^{\mathrm{III}} L_i} \times 100\% \quad (k = \mathrm{I}, \mathrm{II}, \mathrm{III}) \tag{26}$$

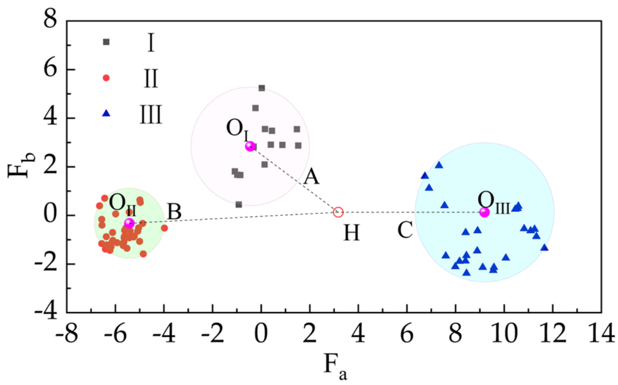


Figure 11. Minimum inscribed circle clustering method.

(3) Minimum inscribed circle height ratio method

The principle of the mixing degree of water sources is quantified by the maximum inscribed circle height ratio method. As shown in Figure 12, after obtaining the center of the maximum inscribed circle representing the three types of water sources, the circle centers are connected successively to obtain $\Delta O_{\mathrm{I}} O_{\mathrm{II}} O_{\mathrm{III}}$, and the three heights are H_{I}, H_{II}, and H_{III}, respectively, and the distances from the mixed water sample point H to be determined to the three sides are h_{I}, h_{II} and h_{III}, respectively.

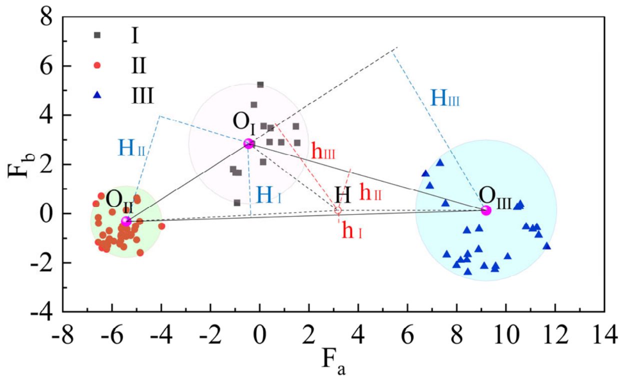


Figure 12. Minimum inscribed circle height ratio method.

The specific mixture degree calculation expression is as follows:

$$r_i = \frac{h_i}{H_i} \quad (i = \mathrm{I}, \mathrm{II}, \mathrm{III}) \tag{27}$$

$$\omega_k = \frac{r_k}{\sum_{i=\mathrm{I}}^{\mathrm{III}} r_i} \times 100\% \quad (k = \mathrm{I}, \mathrm{II}, \mathrm{III}) \tag{28}$$

Take a certain mixed water sample as a case study. The mixed water sample is composed of I, II, and III water sources at the ratio of 15.00%, 35.00%, and 38.00% (I water sample: K^+~Na^+:312.48, Ca^{2+}: 104.90, Mg^{2+}: 72.14, Cl^-: 257.99, $SO_4{}^{2-}$: 468.93, $HCO_3{}^-$: 471.56; II water sample: K^+~Na^+: 277.73, Ca^{2+}: 11.55, Mg^{2+}: 7.04, Cl^-: 84.74, $SO_4{}^{2-}$: 18.59, $HCO_3{}^-$: 610.16; III water sample: K^+~Na^+: 182.81, Ca^{2+}: 250.01, Mg^{2+}: 116.03, Cl^-: 269.01, $SO_4{}^{2-}$: 719.99, $HCO_3{}^-$: 455.07; mixed water sample H: K^+~Na^+: 235.48, Ca^{2+}: 144.78, Mg^{2+}: 71.29, Cl^-: 202.86, $SO_4{}^{2-}$: 436.84, $HCO_3{}^-$:511.82, the unit is mg/L), The values of and calculated from the discriminant function are 3.176 and 0.129, respectively. According to the method shown in Figure 10, the values of ω_{I}, ω_{II}, and ω_{III} are 17.39%, 41.27%, and 41.34%, respectively. According to Figure 11, the values of ω_{I},ω_{II}, and ω_{III} calculated by the maximum inscribed circle clustering method are 16.85%, 57.74%, and 25.41%. According to Figure 12, the values of ω_{I}, ω_{II}, and ω_{III} calculated by the maximum inscribed circle height ratio method are 8.41%, 39.19%, and 52.40%.

3.2.3. Mixing Degree Quantification of Four Types of Water Sources

Similarly, the discriminant function is established by performing the cluster analysis on water samples through factor analysis and Fisher's discriminant theory. The calculation results show that the determination coefficients of K^+~Na^+, Ca^{2+}, Cl^-, and Mg^{2+} are less than 0.01, which has important significance for the classification of water sources, while $SO_4{}^{2-}$ and $HCO_3{}^-$ are eliminated.

As shown in Table 3, the variance contribution rate of the first type of discriminant function is 64.6%, the variance contribution rate of the second type of function is 26.9%, and the variance contribution rate of the third type of function is 8.5%. The combination of the three types of functions can obtain all the information of the water sample. Based on the statistics of typical discriminant function coefficients, the discriminant functions of the four types of water samples are obtained as follows:

$$F_a = 0.032Ca^{2+} + 0.042Mg^{2+} - 0.002\mathrm{Cl}^- - 3.782 \tag{29}$$

$$F_b = 0.008K^+ \sim Na^+ + 0.012Ca^{2+} + 0.002Mg^{2+} + 0.003Cl^- - 5.323 \tag{30}$$

$$F_c = 0.008K^+ \sim Na^+ - 0.054Ca^{2+} + 0.138Mg^{2+} - 0.002Cl^- - 3.388 \tag{31}$$

Table 3. Eigenvalues of four types of water sources.

Function	Eigenvalues	Variance	Cumulative Percentage	Canonical Correlation
1	20.997 [a]	77.4	77.4	0.977
2	4.523 [a]	16.7	94.1	0.905
3	1.592 [a]	5.9	100.0	0.784

"a" in Table 3 represents the dependent variable.

As shown in Figure 13, the water sample ion data are brought into the discriminant function to solve the a and b values of each sample, and the water sample discriminant clustering is obtained.

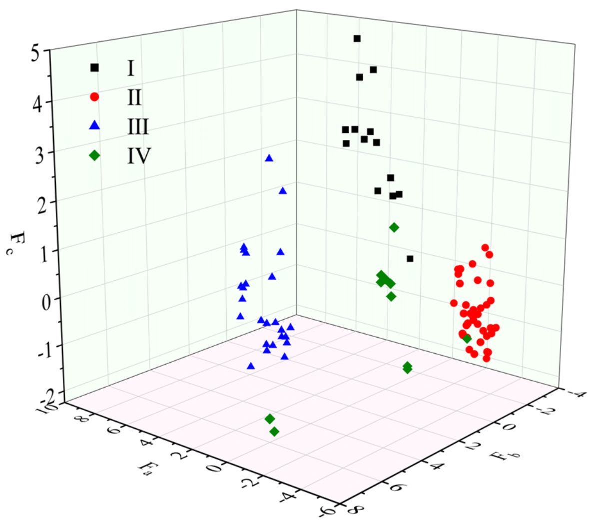


Figure 13. Cluster diagram of four types of mixed water sources.

(1) Maximum coverage tetrahedron method

As shown in Figure 14, a tetrahedron is established with reference water sample points A, B, C, and D as endpoints, and the four sample points A, B, C, and D are in the order of I water sources (15), II water sources (40), III water sources (24), and IV water sources (21). By using the loop algorithm to traverse 300,000 combinations, it is determined that tetrahedrons established at four water sample points A, B, C, and D could cover the sample points to the maximum extent. Tetrahedrons A-BCD covered the water sample points to a maximum of 72.

Assuming that the mixed water sample point M to be determined exists inside the tetrahedron, the heights of the four vertices passing through *A*, *B*, *C*, and *D* in the tetrahedron are H_I, H_{II}, H_{III}, and H_{IV}, respectively. The distances from the mixed water sample point H to be determined to the four surfaces are h_I, h_{II}, h_{III}, and h_{IV}, respectively. The specific mixing degree is calculated as follows:

$$r_i = \frac{h_i}{H_i} \quad (i = A, B, C, D) \tag{32}$$

$$\omega_I = \frac{r_A}{\sum\limits_{i=A}^{D} r_i} \times 100\% \tag{33}$$

$$\omega_{II} = \frac{r_B}{\sum\limits_{i=A}^{D} r_i} \times 100\% \tag{34}$$

$$\omega_{\mathrm{III}} = \frac{r_C}{\sum\limits_{i=A}^{D} r_i} \times 100\% \tag{35}$$

$$\omega_{IV} = \frac{r_D}{\sum\limits_{i=A}^{D} r_i} \times 100\% \tag{36}$$

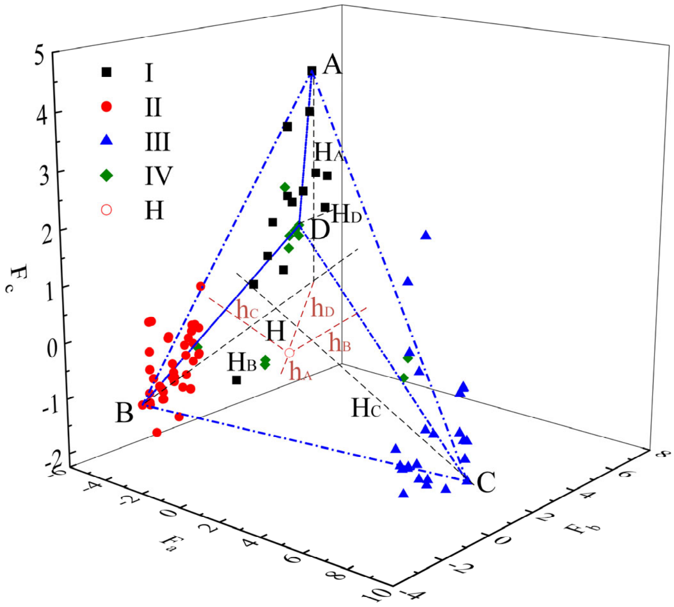


Figure 14. Maximum coverage tetrahedron method.

(2) Minimum inscribed sphere-centered tetrahedron method

The principle of the mixing degree of water sources is quantified by the minimum inscribed sphere-centered tetrahedron method. As shown in Figure 15, algorithm analysis is used to obtain the minimum inscribed sphere containing I water sources (15 samples), and the region within the sphere can represent the type of Class I water sources, with the sphere center being O_{I} (2.309,−1.998,2.461) and radius R_{I} being 2.735. Similarly, the minimum inscribed sphere center of the II water source (40 samples) is O_{II} (−3.030, −1.463, −0.064), and the radius R_{II} is 1.932. The minimum inscribed sphere center of the III water source (24 samples) is O_{III} (6.934, 0.642, −0.024), and the radius $\mathrm{R}_{\mathrm{III}}$ is 2.408. The minimum inscribed sphere center of the IV water source (21 samples) is O_{IV} (−2.885, 3.590, −1.026), and the radius R_{IV} is 3.803. A tetrahedron is constructed with O_{I}, O_{II}, O_{III}, and O_{IV} as vertices. The specific mixing degree calculation is consistent with the maximum coverage tetrahedron method.

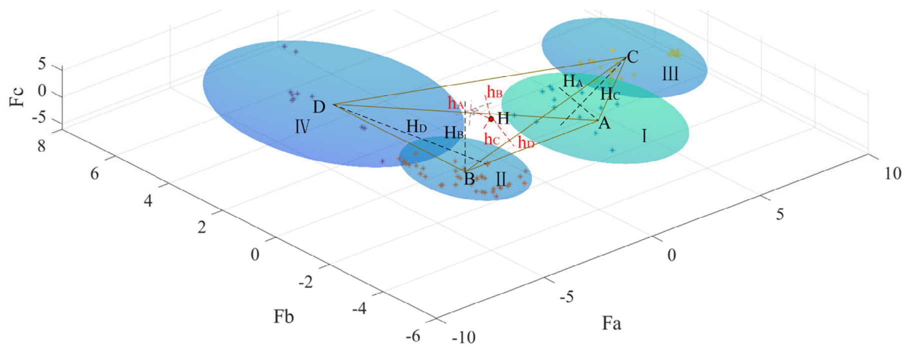


Figure 15. Minimum inscribed sphere-centered tetrahedron method.

(3) Minimum distance method

The principle of the mixing degree of water sources is quantified by the minimum distance method. As shown in Figure 16, the mixed water sample point H to be determined is connected to the sphere centers O_{I}, O_{II}, O_{III}, and O_{IV}, and the intersection points of each connecting line and the sphere are points A, B, C, and D, respectively. The distance between the mixed water sample point H and the four intersection points is the minimum distance from H to each spherical surface.

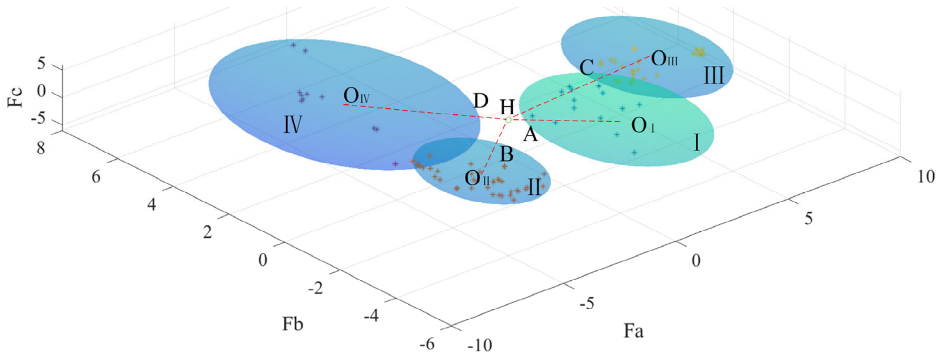


Figure 16. Minimum distance method.

The mixing degree is expressed by the proportion of each minimum distance. The specific expression is as follows:

$$L_i = O_i H - R_i \tag{37}$$

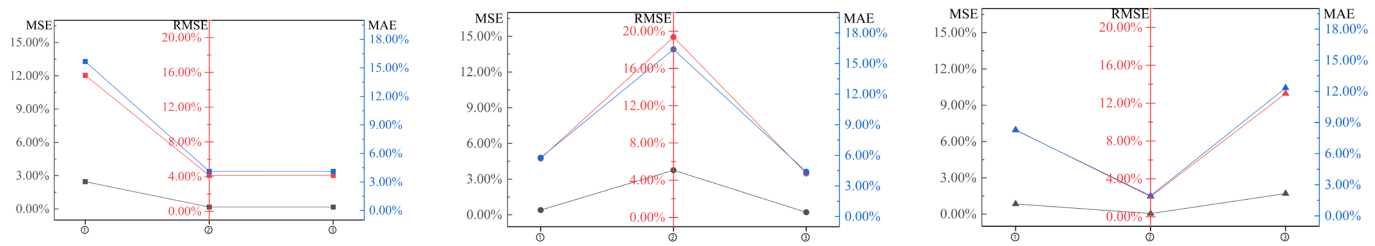
$$\omega_k = \frac{L_k}{\sum_{i=\mathrm{I}}^{\mathrm{IV}} L_i} \times 100\% \quad (k = \mathrm{I}, \mathrm{II}, \mathrm{III}, \mathrm{IV}) \tag{38}$$

Take a certain mixed water sample as a case study. The mixed water sample is composed of I, II, III, and IV water sources at the ratio of 9.00%, 36.00%, 34.00%, and 21.00% (I: K^+~Na^+: 312.48, Ca^{2+}: 104.90, Mg^{2+}: 72.14, Cl^-: 257.99, $SO_4{}^{2-}$: 468.93, $HCO_3{}^-$: 471.56; II: K^+~Na^+: 261.35, Ca^{2+}: 20.79, Mg^{2+}: 14.07, Cl^-: 72.24, $SO_4{}^{2-}$: 60.06, $HCO_3{}^-$: 576.77; III: K^+~Na^+: 271.95, Ca^{2+}: 212.73, Mg^{2+}: 91.88, Cl^-: 578.34, $SO_4{}^{2-}$: 267.02, $HCO_3{}^-$: 739.94; IV: K^+~Na^+: 766.65, Ca^{2+}: 28.62, Mg^{2+}: 10.91, Cl^-: 943.78, $SO_4{}^{2-}$: 107.87, $HCO_3{}^-$: 359.70; mixed water sample H: K^+~Na^+: 375.67, Ca^{2+}: 95.26, Mg^{2+}: 45.08, Cl^-: 444.05, $SO_4{}^{2-}$: 177.26, $HCO_3{}^-$: 577.19, the unit is mg/L). According to the calculation of the discriminant function, F_a, F_b, and F_c are 0.272, 0.248, and -0.193, respectively, According to the method shown in Figure 14, the values of ω_{I},ω_{II},ω_{III}, and ω_{IV} calculated by the maximum coverage tetrahedron method are 12.08%, 49.45%, 23.53%, and 14.94%. According to the method shown in Figure 15, the values of ω_{I}, ω_{II}, ω_{III}, and ω_{IV} calculated by the minimum inscribed sphere-centered tetrahedron method are 10.45%, 38.43%, 30.91, and 20.21%. According to the method shown in Figure 16, the values of ω_{I}, ω_{II}, ω_{III}, and ω_{IV} calculated by the minimum distance method are 15.76%, 21.76%, 51.91%, and 10.57%.

4. Discussion

In this paper, a decision tree classification model based on the C4.5 algorithm is used to identify a single water source. This method is easy to use and explain and can visualize the chemical characteristics of water sources and extract rules. It has a good application effect in the identification and classification of 100 water sources in the Fenyuan coal mine. This model is susceptible to sample data disturbance resulting in large changes in the performance of the decision tree classification model. Therefore, with the increase in the water source sample data, neural networks can be used to optimize the decision tree in later research, building a single water source identification neural decision tree classification model.

Using MSE, RMSE, and MAE as the evaluation indexes of the water source mixing degree calculation model, the smaller the value, the more accurate the calculation results and the stronger the applicability of the model. It can be seen from Figure 17a–c that when calculating the mixing degree of two types of water sources, the relationship between the MSE, RMSE, and MAE index values is as follows: minimum inscribed circle analytical method > minimum inscribed circle comprehensive method > reference line intersection method, that is, minimum inscribed circle analytical method can better meet the accuracy requirements of two types of water sources mixing degree, and the model has good performance and strong applicability. When calculating the mixing degree of three types of water sources, the relationship between MSE, RMSE, and MAE index values is as follows: minimum inscribed circle clustering method > maximum coverage traversal method > minimum inscribed circle height ratio method, that is, minimum inscribed circle clustering method can be better applied to calculate the mixing degree of three types of water sources. When calculating the mixing degree of four types of water sources, the relationship between MSE, RMSE, and MAE index values is as follows: minimum distance method > maximum coverage tetrahedron method > minimum inscribed sphere-centered tetrahedron method, that is, the minimum distance method has the highest calculation accuracy and the best performance.



(**a**) (**b**) (**c**)

Figure 17. Mixing degree quantization model error: (**a**) mixing degree quantization model error of two types of water sources; (**b**) mixing degree quantization model error of three types of water sources; (**c**) mixing degree quantization model error of four types of water sources.

The mixing degree calculation model of water sources obtained from the comparison has a stable structure and simple calculation process. However, the 100-sample data used by the model are only obtained in different regions at a certain time point. In order to better test the accuracy and applicability of the calculation model of water source mixing degree, it is necessary to further increase the collection of hydrogeological data, grasp the corresponding data at different time points in the region, and conduct long-term observation of the dynamic hydrochemical information at the same outlet point. In addition, the model construction is only based on conventional water chemistry indicators, and elements (trace elements, isotopes, and temperature) in the water source are planned to be introduced into the model in the later stage, improving the accuracy of the model.

5. Conclusions

This study built a decision tree classification model based on the C4.5 algorithm, and obtained the intuitive discrimination conditions, realizing a single water source identification. Various analysis models are used to calculate the mixing degree of different types of water sources successively and propose an objective and accurate optimal mathematical analysis. The specific conclusions are as follows:

(1) A decision tree classification model was constructed based on the C4.5 algorithm, and the chemical characteristics of water sources were visualized and rules extracted. After pruning optimization of the model, it was intuitively found that the important variables affecting the construction of decision trees were Mg^{2+}, Ca^{2+}, and Cl^{-}, whose parameter values were 0.68, 0.24, and 0.08, respectively. According to the classifica-

tion and discrimination conditions, $Mg^{2+} < 39.585$ mg/L, $Cl^{-} < 516.338$ mg/L and $Mg^{2+} \geq 39.585$ mg/L, $Ca^{2+} < 160.860$ mg/L.

(2) The study focuses on the quantitative calculation of conventional water chemistry indicators of water sources, using factor analysis and Fisher's discriminant theory to eliminate redundant ion variables, and the discriminant function equations of two, three, and four types of mixed water sources are obtained successively. Taking MSE, RMSE, and MAE as the evaluation indexes, the optimal calculation methods of the two, three, and four types of water sources obtained are the minimum inscribed circle analytical method, minimum inscribed circle clustering method, and the minimum distance method, respectively.

(3) The mathematical analysis method used for the identification and classification of a single water source and the calculation of the mixing degree of two, three, and four types of mixed water sources has the characteristics of a simple calculation process, high efficiency, objective accuracy, and low cost. However, this study is based on only a certain number of water samples, and more water samples should be collected to improve the accuracy of the model. In addition, the model should consider the effects of temperature, hydrogeological conditions, and human activities on aquifers to further improve its applicability.

Author Contributions: Conceptualization, methodology, and writing, Q.L. and G.F.: writing—original draft, investigation, and data curation, W.Y. and S.Z.; writing—reviewing and editing, and funding, D.Z.; formal analysis and validation, Z.F. and Y.F. All authors have read and agreed to the published version of the manuscript.

Funding: This research is funded by the Shanxi Province Unveils Bidding Project (grant No. 20201101009); the National Natural Science Foundation of China (No. 52204161), the Jiangsu Funding Program for Excellent Postdoctoral Talent (No. 2022ZB511), and the Fundamental Research Funds for the Central Universities (No. 2022QN1008; 2021ZDPY0226).

Data Availability Statement: Data are contained within the article.

Conflicts of Interest: We declare that we do not have any commercial or associative interests that represent conflicts of interest in connection with this submitted manuscript.

References

1. Sun, W.; Zhou, W.; Jiao, J. Hydrogeological Classification and Water Inrush Accidents in China' s Coal Mines. *Mine Water Environ.* **2016**, *35*, 214–220. [CrossRef]
2. Cui, F.; Wu, Q.; Lin, Y.; Zeng, Y.; Zhang, K. Damage Features and Formation Mechanism of the Strong Water Inrush Disaster at the Daxing Co Mine, Guangdong Province, China. *Mine Water Environ.* **2018**, *37*, 346–350. [CrossRef]
3. Hu, Y.; Li, W.; Liu, S.; Wang, Q.; Wang, Z. Risk assessment of water inrush from aquifers underlying the Qiuji coal mine in China. *Arab. J. Geosci.* **2019**, *12*, 98. [CrossRef]
4. Hu, W.; Zhao, C. Evolution of Water Hazard Control Technology in China's Coal Mines. *Mine Water Environ.* **2021**, *40*, 334–344. [CrossRef]
5. Li, G.; Meng, Z.; Wang, X.; Yang, J. Hydrochemical Prediction of Mine Water Inrush at the Xinli Mine, China. *Mine Water Environ.* **2017**, *36*, 78–86. [CrossRef]
6. Shi, L.; Wang, Y.; Qiu, M.; Wang, M. Assessment of water inrush risk based on the groundwater modeling system—A case study in the Jiaojia Gold Mine Area, China. *Arab. J. Geosci.* **2019**, *12*, 807. [CrossRef]
7. Gu, H.; Ma, F.; Guo, J.; Li, K.; Lu, R. Assessment of Water Sources and Mixing of Groundwater in a Coastal Mine: The Sanshandao Gold Mine, China. *Mine Water Environ.* **2018**, *37*, 351–365. [CrossRef]
8. Jiang, Q.; Zhao, W.; Zheng, Y.; Wei, J.; Wei, C. A Source Discrimination Method of Mine Water-Inrush Based on 3D Spatial Interpolation of Rare Classes. *Arch. Min. Sci.* **2019**, *64*, 321–333.
9. Adhikari, K.; Mal, U. Application of multivariate statistics in the analysis of groundwater geochemistry in and around the open cast coal mines of Barjora Block, Bankura District, West Bengal, India. *Environ. Earth Sci.* **2019**, *78*, 72. [CrossRef]
10. Wang, X.; Ji, H.; Wang, Q.; Liu, X.; Huang, D.; Yao, X.; Chen, G. Divisions based on groundwater chemical characteristics and discrimination of water inrush sources in the Pingdingshan coalfield. *Environ. Earth Sci.* **2016**, *75*, 1. [CrossRef]
11. Jin, W.; Xibing, L.; Tao, C.; Jinlin, Y. Application of distance discriminant analysis method to headstream recognition of water-bursting source. *Procedia Eng.* **2011**, *26*, 374–381. [CrossRef]

12. Yang, Z.; Lv, H.; Wang, X.; Yan, H.; Xu, Z. Classification of Water Source in Coal Mine Based on PCA-GA-ET. *Water* **2023**, *15*, 1945. [CrossRef]
13. Wang, J.; Zhao, W.; Wang, X.; Hersi, N.A.M.; Zhang, P.; Sang, X. Determining the Quality of Mine Gushing and Mixed Water Using Coupled AHP and Fuzzy Comprehensive Evaluation Methods. *J. Water Resour. Prot.* **2018**, *10*, 1185–1197. [CrossRef]
14. Jafarbeigi, E.; Ayatollahi, S.; Ahmadi, Y.; Mansouri, M.; Dehghani, F. Identification of novel applications of chemical compounds to change the wettability of reservoir rock: A critical review. *J. Mol. Liq.* **2023**, *371*, 121059. [CrossRef]
15. Ahmadi, Y.; Mansouri, M.; Jafarbeigi, E. Improving Simultaneous Water Alternative Associate Gas Tests in the Presence of Newly Synthesized γ-Al_2O_3 /ZnO/Urea Nano-Composites: An Experimental Core Flooding Tests. *Acs Omega* **2023**, *8*, 1443–1452. [CrossRef] [PubMed]
16. Ragno, G.; Luca, M.D.; Ioele, G. An application of cluster analysis and multivariate classification methods to spring water mon itoring data. *Microchem. J.* **2007**, *87*, 119–127. [CrossRef]
17. Li, C.; Sun, L.; Jia, J.; Cai, Y.; Wang, X. Risk assessment of water pollution sources based on an integrated k-means clustering and set pair analysis method in the region of Shiyan, China. *Sci. Total Environ.* **2016**, *557–558*, 307–316. [CrossRef]
18. Singh, K.P.; Malik, A.; Sinha, S. Water quality assessment and apportionment of pollution sources of Gomti river (India) using multivariate statistical techniques—A case study. *Anal. Chim. Acta* **2005**, *538*, 355–374. [CrossRef]
19. Li, B.; Wu, Q.; Liu, Z. Identification of Mine Water Inrush Source Based on PCA-FDA: Xiandewang Coal Mine Case. *Geofluids* **2020**, *2020*, 2584094. [CrossRef]
20. Guan, Z.; Jia, Z.; Zhao, Z.; You, Q. 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.* **2019**, *128*, 200. [CrossRef]
21. Song, C.; Yao, L.; Gao, J.; Hua, C.; Ni, Q. Identification model of water inrush source based on statistical analysis in Fengyu minefield, Northwest China. *Arab. J. Geosci.* **2021**, *14*, 518. [CrossRef]
22. Chen, M.; Wu, Y.; Gao, D.; Chang, M. Identification of coal mine water-bursting source using multivariate statistical analysis and tracing test. *Arab. J. Geosci.* **2017**, *10*, 28. [CrossRef]
23. Duan, X.; Ma, F.; Zhao, H.; Guo, J.; Gu, H.; Lu, R.; Liu, G. Determining mine water sources and mixing ratios affected by mining in a coastal gold mine, in China. *Environ. Earth Sci.* **2019**, *78*, 299. [CrossRef]
24. Wei, J.; Li, G.; Xie, D.; Yu, G.; Man, X.; Wang, J. Discrimination of mine water-inflow sources based on the multivariate mixed model and fuzzy comprehensive evaluation. *Arab. J. Geosci.* **2020**, *13*, 873. [CrossRef]
25. Huang, P.; Yang, Z.; Wang, X.; Ding, F. Research on Piper-PCA-Bayes-LOOCV discrimination model of water inrush source in mines. *Arab. J. Geosci.* **2019**, *12*, 334. [CrossRef]
26. Li, J.; Zhang, H. Application of BP neural network to determine of mine water inrush sources based on Matlab. In Proceedings of the 2011 2nd International Conference on Artificial Intelligence, Management Science and Electronic Commerce (AIMSEC), Dengleng, Deng Feng, China, 8–10 August 2011.
27. Huang, G.; Zhou, H.; Ding, X.; Zhang, R. Extreme Learning Machine for Regression and Multiclass Classification. *IEEE Trans. Syst. Man Cybern. Part B Cybern.* **2011**, *42*, 513–529. [CrossRef] [PubMed]
28. Ji, Y.; Dong, D.; Gao, J.; Wei, Z.; Ding, J.; Hu, Z. Source Discrimination of Mine Water Inrush Based on Spectral Data and EGA–PNN Model: A Case Study of Huangyuchuan mine. *Mine Water Environ.* **2022**, *41*, 583–593. [CrossRef]
29. Alizamir, M.; Sobhanardakani, S. A Comparison of Performance of Artificial Neural Networks for Prediction of Heavy Metals Concentration in Groundwater Resources of Toyserkan Plain. *Avicenna J. Environ. Health Eng.* **2017**, *4*, 11792. [CrossRef]
30. Qiu, M.; Shi, L.; Teng, C.; Zhou, Y. Assessment of Water Inrush Risk Using the Fuzzy Delphi Analytic Hierarchy Process and Grey Relational Analysis in the Liangzhuang Coal Mine, China. *Mine Water Environ.* **2017**, *36*, 39–50. [CrossRef]
31. Chang, Q.; Sun, X.; Zhou, H.; Shi, X. A multivariate matrix model of analysing mine water bursting and its application. *Desalin. Water Treat.* **2018**, *123*, 20–26. [CrossRef]
32. Fang, B. Method for Quickly Identifying Mine Water Inrush Using Convolutional Neural Network in Coal Mine Safety Mining. *Wirel. Pers. Commun.* **2022**, *127*, 945–962. [CrossRef]
33. Song, Y.; Guo, J.; Ma, F.; Liu, J.; Li, G. Interpreting Mine Water Sources and Determining Mixing Ratios Based on the Spatial and Chemical Characteristics of Bedrock Brines in a Coastal Mine. *Water* **2023**, *15*, 2105. [CrossRef]
34. Kaur, P.; Singh, S.; Garg, S. Harmanpreet. In Proceedings of the International Conference on Methods and Models in Science and Technology, Chandrigarh, India, 25–26 December 2010; Volume 1324.
35. Cherfi, A.; Nouira, K.; Ferchichi, A. Very Fast C4.5 Decision Tree Algorithm. *Appl. Artif. Intell.* **2018**, *32*, 119–137. [CrossRef]
36. Grzymala-Busse, J.; Mroczek, T. A Comparison of Four Approaches to Discretization Based on Entropy. *Entropy* **2016**, *18*, 69. [CrossRef]
37. Hajigholizadeh, M.; Melesse, A.M. Assortment and spatiotemporal analysis of surface water quality using cluster and discrimi nant analyses. *Catena* **2017**, *151*, 247–258. [CrossRef]
38. Bi, Y.; Wu, J.; Zhai, X.; Wang, G.; Shen, S.; Qing, X. Discriminant analysis of mine water inrush sources with multi-aquifer based on multivariate statistical analysis. *Environ. Earth Sci.* **2021**, *80*, 144. [CrossRef]