


Article

Comprehensive Study on the Electrical Characteristics and Full-Spectrum Tracing of Water Sources in Water-Rich Coal Mines

Donglin Dong^{1,2,*}, Fangang Meng^{1,3}, Jialun Zhang¹, Enyu Zhang¹  and Xindong Lin¹

¹ College of Geoscience and Surveying Engineering, China University of Mining & Technology-Beijing, Beijing 100083, China; fangang@cumtb.edu.cn (F.M.); bqt2000203033@student.cumtb.edu.cn (J.Z.); bqt2200202001t@student.cumtb.edu.cn (E.Z.); bqt2200203039@student.cumtb.edu.cn (X.L.)

² Inner Mongolia Research Institute, China University of Mining and Technology (Beijing), Ordos 017004, China

³ Donghuantuo Mining Subsidiary, Kailuan (Group) Co., Ltd., Tangshan 063000, China

* Correspondence: ddl@cumtb.edu.cn; Tel.: +86-010-62339301

Abstract: This study addresses the complex hydrogeological conditions and frequent inrush water incidents in the Donghuantuo coal mine by proposing a novel spectral tracing technique aimed at rapidly and accurately identifying the sources of inrush water. Through the analysis of electrical data from the Donghuantuo mine, the electrical characteristics of the mine floor were examined. Systematic sampling of water from the primary aquifers within the mining area was conducted, followed by detailed spectral measurements, resulting in the establishment of a spectral database for inrush water sources in the Donghuantuo mine. The chaotic sparrow search optimization algorithm (CSSOA) was employed to optimize the key parameters of the random forest (RF) model, leading to the development of the CSSOA-RF spectral tracing identification model. This model demonstrated outstanding classification performance in the test set, achieving an accuracy of 100%. This research offers a novel, more accurate, and reliable method for identifying the sources of inrush water, facilitating the rapid identification of sources in coal-bearing regions of North China and reducing disaster losses. Although the geological structure of the study area is relatively simple, the research achieved significant results in identifying both single and mixed water sources. However, further validation and optimization are needed for its applicability in more complex geological conditions. The findings of this study provide crucial technical support for safe mining operations and hold significant reference value for water hazard prevention in similar regions.

Keywords: Donghuantuo mine; identification of inrush water sources; spectral tracing technology; chaotic sparrow search optimization algorithm (CSSOA)



Citation: Dong, D.; Meng, F.; Zhang, J.; Zhang, E.; Lin, X. Comprehensive Study on the Electrical Characteristics and Full-Spectrum Tracing of Water Sources in Water-Rich Coal Mines. *Water* **2024**, *16*, 2673. <https://doi.org/10.3390/w16182673>

Academic Editor: Andrea G. Capodaglio

Received: 24 July 2024

Revised: 5 September 2024

Accepted: 13 September 2024

Published: 19 September 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

As the mining activities in lower coal seams in China continue to advance, encounters with increasingly numerous and voluminous aquifers have become more frequent. This is particularly evident in the typical large-scale water-bearing coal mines of North China, such as the Donghuantuo mine, where the hydrogeological conditions are extremely complex and the maximum water influx has reached 62.84 m³/min. Since the commencement of mining operations, numerous inrush water incidents have occurred, including 32 instances with water flow exceeding 10 m³/min, with the highest single face water influx recorded at 6.01 m³/min. With the continuous deepening of coal seam mining, the threat of water hazards has further intensified.

Water inrush identification methods involve using data such as water level, water temperature, hydrogeochemical distribution, migration, and transformation characteristics in different aquifer water qualities. In the early 20th century, researchers primarily relied on experience or simple mathematical methods to determine the source of mine water inrush [1]. Subsequently, Italian researchers discovered that monitoring information such as changes in water level and gas concentration in mining areas could provide early warning of mine water inrush [2]. By the 1990s, scholars successfully identified the source

of water inrush by comparing the water quality characteristics of mine water with those of potential source waters, establishing a solid theoretical foundation for hydrochemical identification [3]. Building on this, some scholars have achieved significant practical effects in identifying water inrush sources by utilizing the unique hydrogeochemical properties of environmental isotopes or trace elements [4]. In addition, the theory of geothermal gradients has been employed to use water temperature as a factor in identifying sudden water inrush, proving effective in distinguishing water inrush sources [5]. With the development of basic theory and computer science, models for identifying water sources based on mathematical principles such as multivariate statistics and nonlinear analysis, combined with computer technology, have become widely used and have demonstrated high accuracy [6]. These developments have greatly enriched the theoretical framework for identifying sudden water inrush in mining operations.

Traditional methods for studying inrush water, which rely on inrush water data and physical parameters, have proven effective in distinguishing relatively simple water sources. However, in cases where the water quality is complex and physical parameters are similar, or where the water quality of different inrush sources is alike, traditional methods often fail. This necessitates the use of alternative approaches for identifying inrush water sources [7–13]. By analyzing the basic components of aquifers and the chemical composition of water samples from inrush points, suitable discriminant functions are selected, and discriminant formulas are established based on specific criteria. These formulas are then used to classify unknown samples, employing methods such as fuzzy mathematics, grey relational analysis, extension identification, geographic information systems (GIS), support vector machines, and artificial neural networks [14–23].

In recent years, with the rapid advancement of computer technology, neural networks have increasingly been used to identify inrush water sources in mines. Common artificial neural network models include BP neural networks, RBF neural networks, ELM, and Elman neural networks, which are valuable for accurately identifying multiple water sources in mines. Deep learning algorithms represent an evolution of artificial neural networks, with common approaches including deep neural network (DNN) analysis, convolutional neural network (CNN) analysis, and probabilistic neural network (PNN) analysis. Scholars often use these methods to determine the origin of water samples. Optimization algorithms such as genetic algorithms, ant colony algorithms, and particle swarm algorithms are frequently used for function optimization and combinatorial optimization of discriminant models [24–30]. For example, Zhang Di applied a genetic algorithm-optimized support vector machine to identify inrush water sources, which improved the accuracy of parameter selection for support vector machines in inrush water source identification. However, this method still has drawbacks, such as tracing detection times exceeding one hour, inability to trace mixed water and its proportions, and loss of the optimal rescue window post-inrush.

This study aims to explore a new method that abandons the traditional approach of concentration testing. Instead, it focuses on the rapid and accurate identification of inrush water sources, proposing a novel method for identifying mixed inrush water sources.

2. Materials and Methods

2.1. Study Area and Water Sampling

Donghuantuo mine is located in the city of Tangshan in North China's Hebei Province. The southeast wing of the mine extends 13.5 km in strike length and 3 km in dip width, while the northwest wing stretches 8 km in strike length and 0.5 km in dip width. The mining area covers 40.5 km². The terrain within the mining boundary is quite flat, with no rivers traversing the area. Additionally, there are no surface water systems within the Donghuantuo mining boundary, as illustrated in Figure 1.

The Donghuantuo mining field hosts multiple aquifers, including the Quaternary alluvial pore-confined aquifer, the Carboniferous–Permian sandstone fissure-confined aquifer, and the Middle Ordovician limestone karst fissure-confined aquifer. These aquifers are divided into seven aquifer groups, as shown in Figure 2.

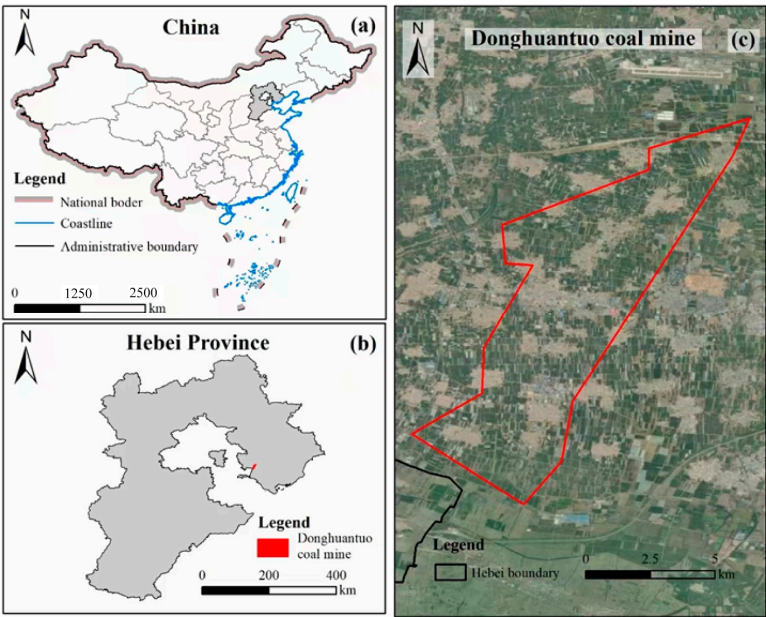


Figure 1. Geographical location of the study area.

VII		Quaternary alluvial pore-bearing aquifer	
VI	VI _b	Aquifer under A	
		Mafic rock	
	VI _a	Upper A aquifer	
		Light gray coarse sandstone dominates	
V	V _{a+b+c+d}	80m-5 Coal seam roof strong aquifer	
		5 Coal seams	
IV	IV _{b+c}	Aquifer	
		12 Coal seams	
	IV _a , IV' _a	Sandstone fissure aquifer	
		14 Coal seams	
III		K ₃ Aquifer	
		Dominated by siltstone and claystone	
II		K ₃ -G Unevenly water-rich aquifers	
		Dominated by siltstone and claystone	
I		Ordovician karst fissure pressurized aquifer	

Figure 2. Aquifer division map of Donghuantuo mining area.

In the preliminary phase of this study on spectral tracing of intrush water sources, four water samples were collected from the primary aquifers of the Donghuantuo mine. These samples were taken from the 12-2~14-1 aquifer, the 5 coal roof aquifer, the Quaternary aquifer, and the Ordovician limestone aquifer, with each sample measuring 1000 mL (Figure 3). Each of the four samples was labeled and numbered accordingly. The detailed information for each water sample is provided in Table 1.

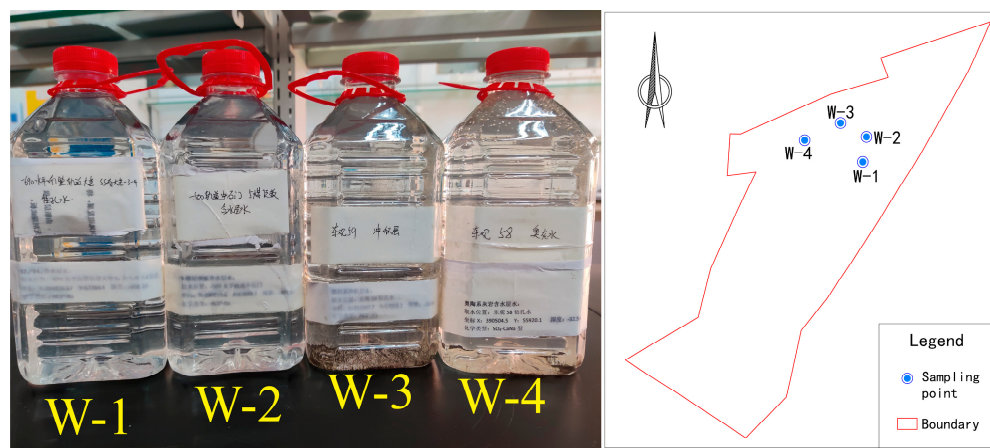


Figure 3. Water sample collection.

Table 1. Water sample collection information.

Sample ID	Aquifer	Sampling Coordinates (m)			Hydrochemical Type
		X	Y	Z	
W-1	12-2~14-1 Aquifer	389,826.67	57,764.4	−698.1	HCO ₃ -Na
W-2	5 Coal Roof Aquifer	390,610.8	57,880.7	−497.0	HCO ₃ -Na
W-3	Quaternary Aquifer	391,047.7	57,058.2	−22.4	HCO ₃ -Na
W-4	Ordovician Limestone Aquifer	390,504.5	55,920.1	−32.5	SO ₄ -CaNa

2.2. Transient Electromagnetic Method for Intrush Water Sources

In this study, the enhanced TEM67 transient electromagnetic method was employed. The transmitter used was the TEM57-Mk2, augmented with two TEM67 power modules, capable of reaching an emission voltage of up to 240 V and a maximum emission current of 28 A. This configuration allowed the use of large-size or large effective area transmitter coils, achieving greater exploration depths. The table below compares the transmission power of the TEM67 and the enhanced TEM67. It shows that with a transmission frame size of 2000 m × 2000 m (equivalent to an exploration depth of 2000 m), the magnetic moment $M(I \times S)$ of the enhanced TEM67 can reach up to 36,363,636 A·m².

2.3. Full-Spectrum Beer–Lambert Tracing Method Using Spectrophotometry

The spectrophotometer operates based on the Beer–Lambert law, which posits that molecular substances can selectively absorb certain wavelengths of light within the ultraviolet–visible (UV–Vis) spectrum under radiation, causing electronic transitions. Due to the unique electronic orbitals of different substances, the energy required for these transitions varies, leading to differences in the wavelengths of light they can absorb. Thus, the degree of light absorption at different wavelengths can be used to identify the substance.

The spectrophotometer measures the relative transmittance or absorbance (Abs, indicating the degree of light absorption) of the test sample, using a blank reference solution (commonly ultrapure water, air, or specific solutions) assumed to have 100% transmittance. By irradiating the measurement solution with incident light across a defined wavelength

range, the absorbance at each wavelength can be recorded. These data allow the creation of a wavelength–absorbance spectrum in a Cartesian coordinate system.

Different substances exhibit distinct absorption characteristics at various wavelengths, resulting in unique spectral line shapes and peak absorption wavelengths. The spectrophotometer consists of several key components: a light source, a monochromator, an absorption cell, a detector, and a processing and display system. The operational schematic of these components is illustrated in Figure 4.



Figure 4. Structure of the spectrophotometer.

The light source in the spectrophotometer provides a continuous spectrum. The monochromator’s function is to separate the continuous light emitted by the light source into various high-purity monochromatic lights of single wavelengths for irradiating the sample. The absorption cell, which holds the sample, uses a quartz cuvette for this experiment, as the detection range spans the visible and ultraviolet regions (320 nm to 1100 nm). The detector converts the light signal passing through the absorption cell into an electrical signal. The processing and display system is primarily used for instrument operation control and data processing, displaying the result data and the spectral line graph of the tested sample.

2.4. Chaos Sparrow Search Optimization Algorithm

The sparrow search algorithm (SSA) faces challenges such as a tendency to get trapped in local optima, a lack of randomness in the search process, and slow convergence. To address these issues, researchers including Xin introduced the tent chaotic sequence and Gaussian distribution into the SSA, forming the chaos sparrow search optimization algorithm (CSSOA). This method uses tent chaotic mapping during the population initialization stage to maintain a uniform distribution of initial individuals. When the population exhibits convergence or divergence, chaotic perturbations and Gaussian mutations are applied to alleviate local optima problems. The implementation steps are as follows:

1. **Population Initialization:** Randomly generate a group of sparrows (individuals), with each sparrow representing a potential solution. Each sparrow is randomly assigned an initial solution position.
2. **Objective Function Evaluation:** Define the optimization problem to be solved and calculate the objective function value at the current position of each sparrow, representing the quality of the solution.
3. **Chaotic Perturbation:** Introduce chaotic perturbation using chaotic mapping to disturb the positions of the sparrows, enhancing the diversity of solutions and reducing the likelihood of local optima.
4. **Neighborhood Search:** Each sparrow performs a dynamic search based on its current position to find better solutions. Sparrows communicate with each other to aggregate the best solutions found in their respective neighborhoods (optimal positions).
5. **Position Update:** Update the position of each sparrow based on the optimal positions found so far, generating new solutions.
6. **Global Best Solution Update:** Aggregate the optimization results from all sparrows to determine the best solution among them, representing the global optimal solution.
7. **Termination Condition Check:** Check whether the termination conditions are met, such as reaching the maximum number of iterations or the objective function value meeting a specified threshold. If the conditions are not met, return to step 3; otherwise, terminate the algorithm.
8. **Return Best Optimization Result:** Output the global optimal solution as the result of the algorithm.

This study focuses on optimizing four critical parameters that significantly impact the model generation of the random forest (RF) algorithm: ‘n_estimators’, ‘max_depth’, ‘min_samples_leaf’, and ‘min_samples_split’. The ‘n_estimators’ parameter is crucial as it controls the number of decision trees in the random forest. A too-small value can lead to underfitting, while a too-large value can slow down the model’s running speed, hence the need for an appropriate ‘n_estimators’ value. ‘Max_depth’ sets the maximum depth of the decision trees, controlling the number of features selected for each tree. Excessive depth can cause overfitting. ‘Min_samples_leaf’ specifies the minimum number of samples required to be at a leaf node, indicating the minimum number of samples each child node must have after a split. ‘Min_samples_split’ represents the minimum number of samples required to split an internal node, indicating that a node must have at least ‘min_samples_split’ samples to be split.

2.5. Random Forest (RF) Algorithm

Random forest (RF) is an ensemble learning algorithm commonly applied to classification tasks. It constructs a large number of decision trees from training samples to form a collective discriminative model, which is then used for classifying unknown samples (Equation (1)). In UV–Visible spectrum classification, RF can effectively handle non-linear relationships and, through its built-in feature importance metrics, select the most informative features from large datasets (Equation (2)), thus enabling efficient processing of high-dimensional spectral data while ensuring classification accuracy.

The classification result of RF is a synthesis of the individual classifications from all decision trees, with the mode of these results serving as the final RF classification output.

$$\hat{y} = \text{mode}(\hat{y}_1, \hat{y}_2, \dots, \hat{y}_T) \quad (1)$$

where T represents the number of decision trees, and \hat{y}_T is the classification result of the t -th decision tree. This method effectively reduces overfitting caused by individual trees, enhancing the model’s generalization capability.

RF also computes the contribution of each feature (dimension) to the classification decision.

$$\text{Importance}(f) = \frac{1}{T} \sum_{t=1}^T \Delta i_{f,t} \quad (2)$$

where f represents a specific feature, and $\Delta i_{f,t}$ denotes the consistency gain brought by feature f in the t -th decision tree. Higher consistency often leads to better classification performance. Analyzing feature contributions can also aid in dimensionality reduction for high-dimensional data.

3. Electrical Characteristics of Water-Rich Inrush Water Sources

3.1. Electrical Characteristics of Water-Rich Inrush Water Sources at the 20223 and 3015 Working Face

The current retreating working faces are 20223 and 3015. To ensure safe and efficient mining, it is necessary to conduct electrical resistivity surveys on the roof and floor of these working faces to accurately delineate their water-bearing properties. The enhanced TEM67 transient electromagnetic method was used to detect water-rich areas at the 20223 and 3015 working faces. These surveys were complemented by data from four drill holes and tunnel exposure data. The water-bearing characteristics of the roof and floor of these working faces are illustrated in the following Figures 5 and 6.

According to Figure 5, a relatively low-resistivity zone, designated as DF-1, was identified within the 0 to 80 m vertical detection range of the floor at the 20223 working face. Analysis of related hydrogeological data suggests that this zone corresponds to a sandstone fissure-confined aquifer. Comparisons with the fault lines in the profile and cross-section diagrams indicate that this zone is likely connected by faults, warranting

focused investigation. Verification through four drill holes confirmed that this is indeed a fault zone.

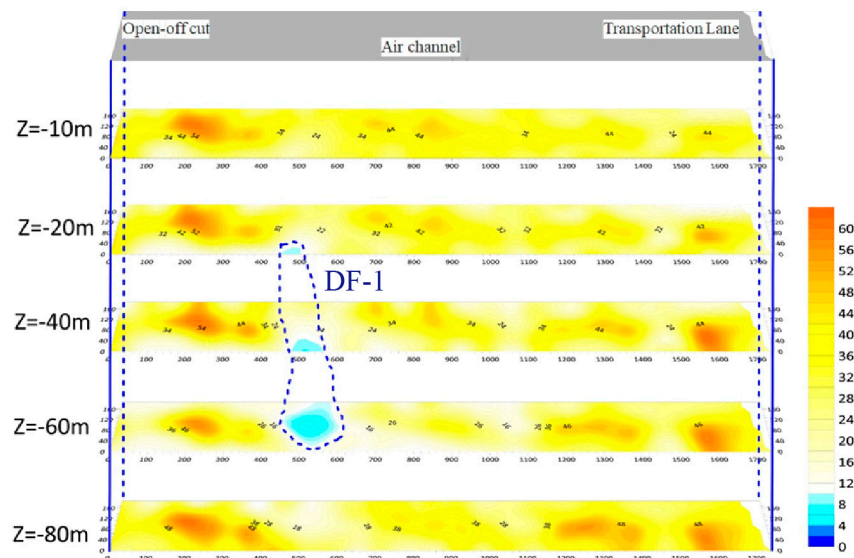


Figure 5. 3D Electrical resistivity tomography (ERT) image of the 20223 working face floor.

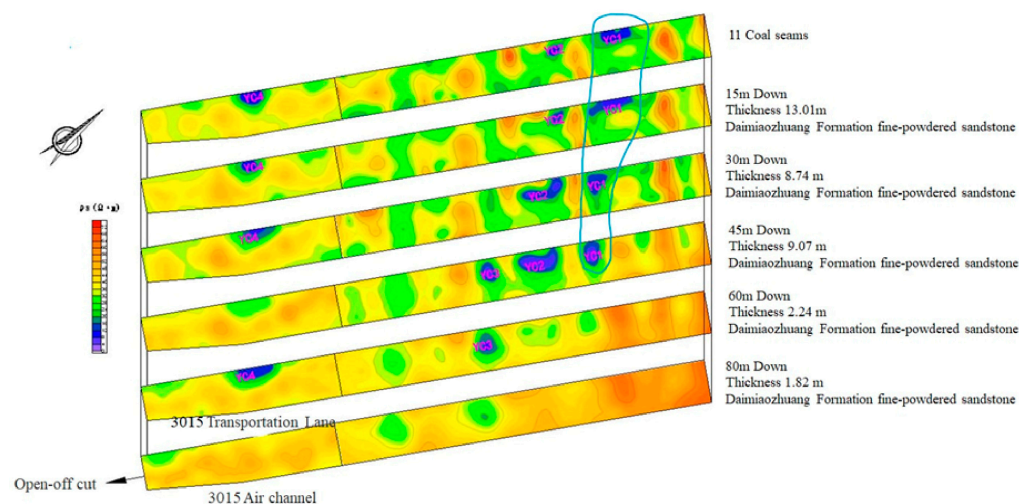


Figure 6. 3D resistivity imaging of the floor at the 3015 working face.

According to Figure 6, a fault f548 was exposed in the region where anomaly DF1 is located, suggesting that the anomaly may reflect the sandstone aquifer conducted by the fault. Anomaly DF2 corresponds to water from sandstone fissures in the roof and floor or water accumulated in the goaf of the overlying ninth coal seam, indicating weak to moderate water-bearing properties, which could impact the retreating operations at working face 3015.

In the area of anomaly DF3, fault f550 was identified, implying that this anomaly might also represent the sandstone aquifer conducted by the fault. Similarly, anomaly DF4 reflects water from sandstone fissures in the roof and floor or water accumulated in the goaf of the overlying ninth coal seam, with weak to moderate water-bearing properties, potentially affecting the retreating operations at the working face.

3.2. Spectral Characteristics of Various Inrush Water Sources

A vacuum filtration apparatus was used to filter each water sample solution through a 0.45 μm PTFE hydrophilic filter membrane to remove particulate impurities. The filtered water samples were then dispensed into sample bottles that had been washed with ultra-

pure water and air-dried. To increase the sample size for subsequent machine learning, each type of water sample was divided into eight sample bottles, resulting in eight sets of spectral data for each type of water sample.

Since the cuvette capacity of the DR-3900 spectrophotometer is 20 mL, each measurement was standardized to 15 mL to ensure measurement quality while conserving the sample. To ensure accuracy, each sample was measured three times, and the arithmetic mean of these measurements was used as the spectral curve data for that sample. For ease of data processing and result interpretation, samples were assigned hierarchical numbers, e.g., the first sample of water sample type 1 was labeled 1-1, and its two repeated measurements were labeled 1-1-1 and 1-1-2, respectively. The detailed preparation steps are illustrated in Figure 7. In total, 32 sets of water samples were prepared for subsequent classification model training in this tracing study.

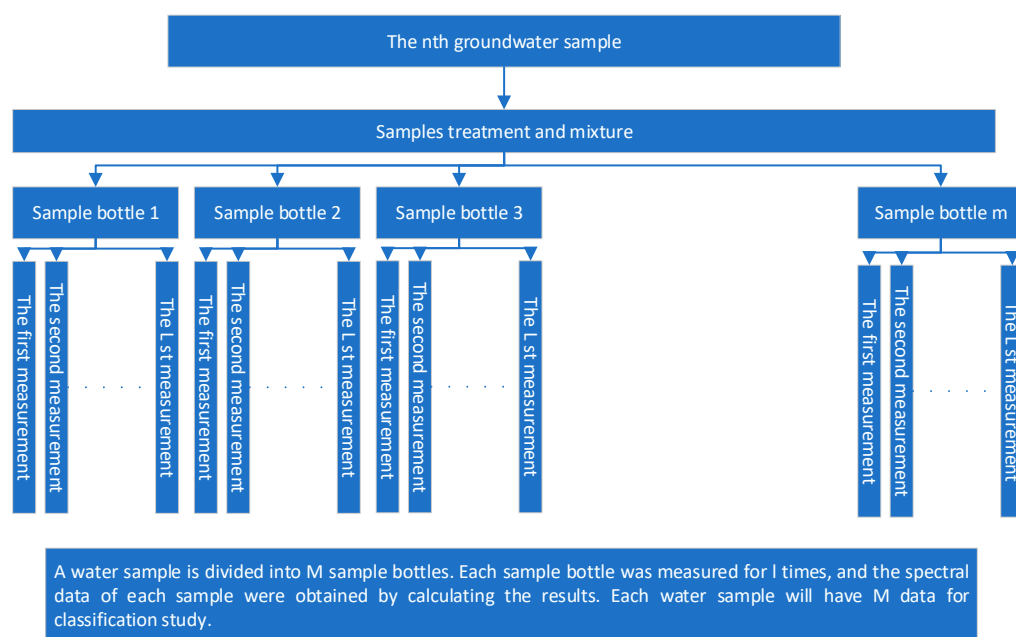


Figure 7. Sample preparation diagram.

The prepared samples were subjected to spectrophotometric measurement. After the instrument completed its self-check and warm-up, the measurement parameters of the spectrophotometer were set: the starting wavelength was set to 320 nm, the ending wavelength to 650 nm, and the scanning interval to 1.0 nm. Zero-line calibration was then performed. Each sample was measured sequentially, with three repeated measurements. The data were then imported into a computer using a USB drive and labeled accordingly. After measuring each type of sample, the cuvette and other instruments were cleaned with ultrapure water.

Upon completing the measurements of all water samples, the repeated measurement data for each sample were organized and averaged arithmetically, resulting in 32 sets of single-source spectral data, as shown in Figure 8.

Analysis of Figure 8 reveals significant differences in the spectral data of the four aquifer types within the measured wavelength range of 320–650 nm. The absorbance fluctuations for water samples W-1 and W-2 are minimal, distributed within the range of 0 to 0.005, with most absorbance values being 0. Specifically, the absorbance for W-1 drops to 0 after 429 nm and remains unchanged, while for W-2, it remains at 0 beyond 335 nm. In contrast, water samples W-3 and W-4 exhibit significant absorbance variations, with both reaching maximum absorbance values around 320 nm and then gradually decreasing to their minimum values. The absorbance range for W-3 is between 0 and 0.015, while for W-4, it is between 0 and 0.0325. The absorbance variations for all four water samples are

concentrated in the ultraviolet region (<400 nm), aligning with the visual characteristics of colorless and transparent water samples.

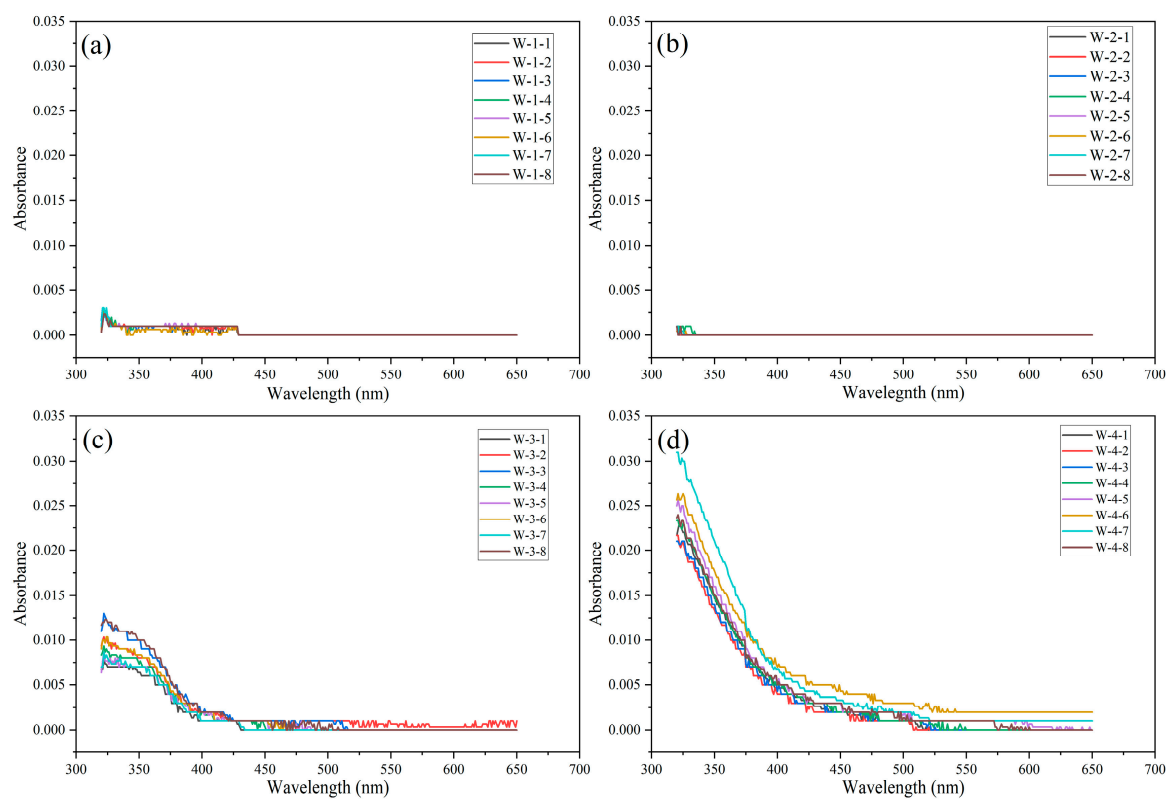


Figure 8. Original spectral data of each aquifer. (a) 12-2~14-1 aquifer water samples, (b) 5 coal roof aquifer water samples, (c) quaternary aquifer water samples, (d) Ordovician limestone aquifer water samples.

Water samples W-3 and W-4, which are from shallower locations, experience more frequent water–rock interactions compared to the deeper W-1 and W-2 samples. As a result, W-3 and W-4 have higher ion concentrations and more complex ionic compositions, leading to more pronounced spectral variations. The distinct spectral differences among the four water samples ensure the accuracy of the subsequent machine learning tracing model’s classification and identification.

Based on the hydrogeological conditions, hydraulic connections, and actual geological situations of the four main aquifers sampled from the Donghuantuo mine, the study investigated the mixing of the Quaternary water sample W-3 and the 12-2~14-1 aquifer water sample W-1. The preparation method for these mixed samples was consistent with that of the single-source samples. W-1 and W-3 samples were mixed in ratios of 1:9, 2:8, . . . , 8:2, and 9:1. To ensure sufficient training samples for the subsequent tracing model, 10 samples were prepared for each mixing ratio, resulting in a total of 90 mixed spectral data entries, as shown in Figure 9.

Figure 9 illustrates that as the concentration ratio of the two mixed water samples varies, the absorbance curve of the mixture increasingly resembles that of the sample with the higher concentration. The differences between the W-1 and W-3 mixtures become significant only in the wavelength range near the maximum absorbance (320–400 nm), while in the 500–650 nm range, the absorbance approaches zero. Similar to the spectral measurement results of the single-source samples, the measurements of the mixed samples exhibit considerable experimental error. Therefore, further data processing is necessary to eliminate noise interference generated during the experiments.

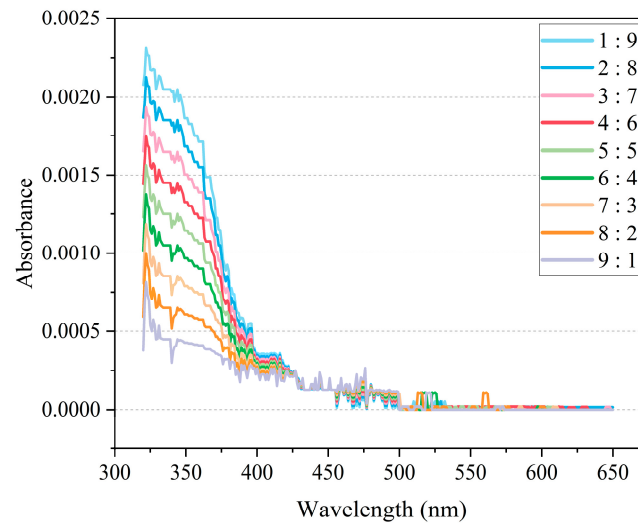


Figure 9. Original spectral data of different aquifers.

3.3. Construction and Evaluation of the Tracing Identification Model

Using the preprocessed sample data, we constructed the CSSOA-RF spectral tracing identification model based on the Python language. The CSSOA algorithm was employed to optimize the key parameters of the RF model, achieving adaptive parameter optimization for the tracing identification model. This phase involved steps such as dataset partitioning, CSSOA algorithm optimization, and the construction of the CSSOA-RF tracing identification model.

The parameters of the RF model, including 'n_estimators', 'max_depth', 'min_samples_leaf', and 'min_samples_split', were optimized based on the training set. The CSSOA algorithm's objective function was set to the average error obtained from five-fold cross-validation on the training set. The population size was set to 60, with a discoverer ratio of 0.7 and a warning reconnaissance ratio of 0.2. The optimization process is illustrated in Figure 10. After 67 iterations, the model achieved the minimum average error value of 0.1596, which remained stable. Therefore, the parameter combination at this point was chosen as the optimal parameter values for the RF model. The parameter search ranges and their optimal values are listed in Table 2, with other parameters retaining their default values.

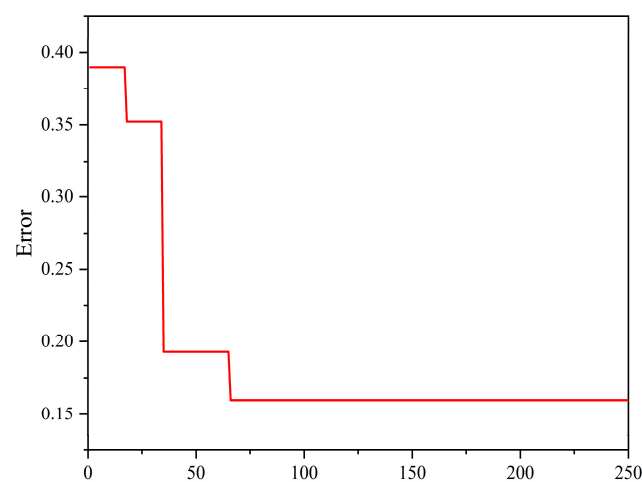


Figure 10. CSSOA parameter optimization curve.

Table 2. Optimal parameter values for random forest (RF).

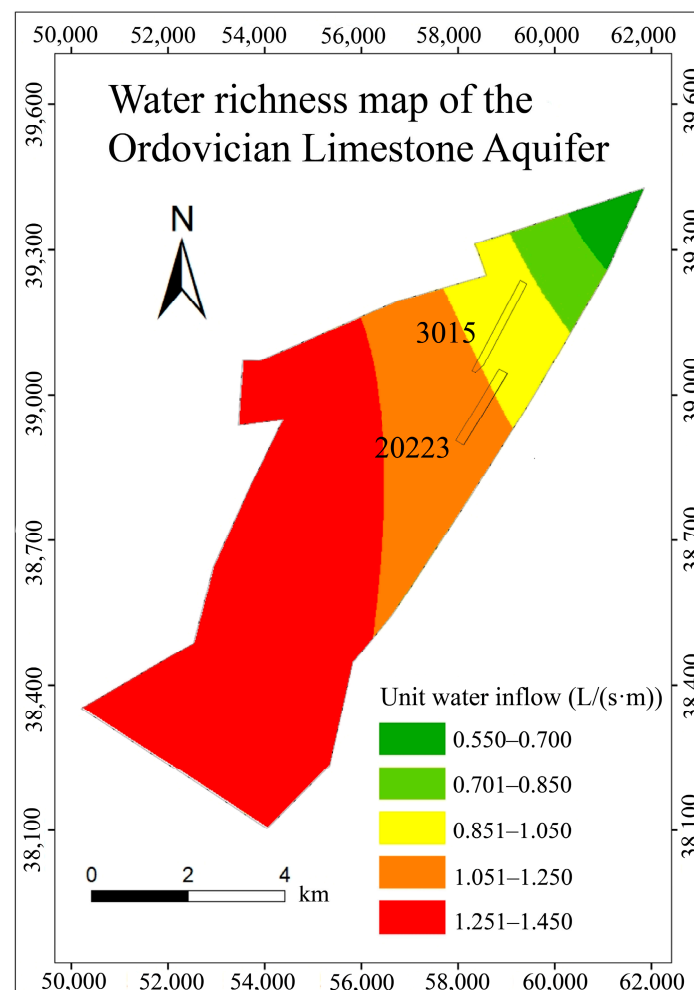
Parameter Name Search	Range	Optimal Value
n_estimators	[15, 200]	102
max_depth	[1, 12]	3
min_samples_leaf	[1, 15]	1
min_samples_split	[2, 15]	2

4. Discussion

4.1. Determination and Evaluation of Water-Richness of Inrush Water Sources

Based on actual observation data from the Donghuantuo mine and the revealed tunnel water sample detection data, the Ordovician limestone water has been identified as the most influential water source. Consequently, this study compared the water richness, characterized by the unit water inflow of the Ordovician limestone water, with the results obtained from the transient electromagnetic method and the direct current electrical method.

The verified water richness determined by electrical methods shows a generally similar trend, but the electrically determined water-rich areas are somewhat larger. For instance, at the current mining faces 20223 and 3015, the southern water-rich areas significantly exceed those to the north (Figure 11). Research indicates that this position coincides with the development of fault DF-1, and the water-conducting fissure zone induced by mining activity also contributes to this. During the excavation of the 20223 and 3015 working faces, it was found that the floor in this location was relatively fragmented, with well-developed fissures, thus significantly enlarging the range of geophysical water richness in this area.

**Figure 11.** Contour map of water richness for the Ordovician limestone aquifer.

4.2. Traceability Accuracy Analysis after Deep Learning

To further evaluate the constructed tracing identification model, we applied it to the designated test set for tracing identification. The results were visualized using a confusion matrix, as shown in Figure 12. The confusion matrix clearly illustrates the relationship between the model's predicted categories and the actual true categories within the test set. The CSSOA-RF spectral tracing identification model misclassified only one sample from category 2 as W-1, while accurately identifying the categories of the remaining seven samples, demonstrating the model's excellent classification performance.

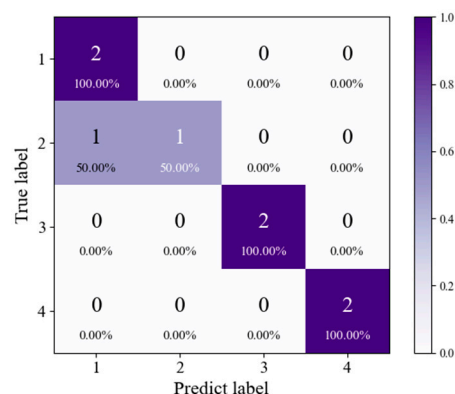


Figure 12. Process of determining the optimal conversion coefficient C value.

Additionally, the model's generalization performance needs further evaluation. Using the constructed database, a five-fold cross-validation method with 20 repetitions was employed to assess the generalization ability of the tracing identification model. The results showed that the accuracy of all repeated cross-validations reached 100%, further proving the excellent generalization performance of the CSSOA-RF tracing prediction model. This validation confirms the model's applicability to spectral tracing technology and indicates its superior identification performance when applied to the spectral water source database of the Donghuantuo mine.

In this study, spectral tracing technology was employed to measure the spectra of collected samples, leading to the construction of a spectral database for inrush water sources in the Donghuantuo mine. A machine learning adaptive parameter optimization tracing identification model was built using Python to complement this database. The performance of the constructed tracing identification model was evaluated based on existing data. The entire process, from data acquisition to result identification, took 34 min, significantly less than the time required by current traditional methods. The results indicate that this model could accurately identify the inrush water sources of the main aquifers in the Donghuantuo mine, demonstrating reliable performance.

The spectral tracing identification method was very successful in Donghuantuo Mine. In order to verify the application of this method in other mining areas, the method was applied to Huangyuchuan Mine, Daliuta Mine in Shaanxi Province, and Fengfeng Mining Area in Hebei Province. The accuracy rate of verification was always above 90%.

5. Conclusions

This study addressed the complex hydrogeological conditions and frequent inrush water incidents in the Donghuantuo coal mine by systematically collecting water samples from the primary aquifers in the mining area. A total of 40 samples from four main aquifers were collected and subjected to detailed spectral measurements. Based on these data, a spectral database for inrush water sources in the Donghuantuo mine was established. A novel spectral tracing technology was adopted, aiming to rapidly and accurately identify inrush water sources. Traditional methods often fail when dealing with complex water

quality types and similar physical parameters, prompting the need for a new, more accurate, and reliable method. The key research findings are as follows:

1. **CSSOA-RF Spectral Tracing Identification Model:** By using the chaos sparrow search optimization algorithm (CSSOA) to optimize the key parameters of the random forest (RF) model, the CSSOA-RF spectral tracing identification model was constructed. The CSSOA algorithm, through the introduction of chaotic perturbations and Gaussian mutations, overcame the limitations of traditional sparrow search algorithms that tend to get trapped in local optima. This improved optimization efficiency and the model's global search capability. The optimized RF model achieved optimal parameter selection, providing reliable technical support for tracing identification. In the test set, the CSSOA-RF model demonstrated excellent classification performance, achieving 100% accuracy, with only one sample misclassified, indicating the model's outstanding classification ability and generalization performance. The confusion matrix visualization further confirmed the model's accuracy and stability in practical applications. This result validates the effectiveness and practicality of spectral tracing technology in identifying mine inrush water sources.
2. **Innovative and Reliable Method for Inrush Water Source Identification:** This research provides a new, more accurate, and reliable method for identifying inrush water sources, addressing the shortcomings of traditional methods in handling complex water quality conditions. The model helps in quickly identifying inrush water sources in coal-bearing regions of North China, reducing disaster losses and enhancing mine safety. Additionally, the spectral database and CSSOA-RF model developed in this study offer valuable references for future related research.

Given the relatively simple geological structure of the study area, this research achieved high accuracy in identifying single-source water samples. For mixed water samples measured in the laboratory, accurate identification results were obtained for different mixing ratios, demonstrating the potential and reliability of this method in practical applications. However, the applicability and accuracy of this method need further validation and optimization under more complex geological conditions and real-world scenarios with multiple water sources. Future research should target more complex hydrogeological conditions, explore broader application scenarios, and further enhance the model's robustness and practicality.

Author Contributions: Methodology, E.Z.; Software, J.Z.; Validation, X.L.; Investigation, F.M.; Writing—original draft, D.D. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by National Key Research and Development Program of China, grant number 2023YFC3012100.

Data Availability Statement: If interested in the data used in the research work, contact ddl@cumtb.edu.cn for the original dataset.

Conflicts of Interest: Author Fangang Meng was employed by the company Donghuantuo Mining Subsidiary, Kailuan (Group) Co., Ltd. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

1. Sun, H.L. Characteristics of karst water inrush channels in coal seam fractures. *J. China Coal Soc.* **1965**, *2*, 51–57.
2. Hong, X.J. Judgment of Mine Water Inflow. *Saf. Coal Mines* **1978**, *6*, 18.
3. Sammarco, O. In rush prevention in an underground mine. *Int. J. Mine Water* **1988**, *7*, 43–52. [[CrossRef](#)]
4. Sun, F.Y.; Zheng, S.Y. Research on the characteristics of mine water quality and its application in identifying water sources. *Coal Sci. Technol.* **1996**, *10*, 27–30.
5. Li, Y. Application of Environmental Isotope Methods in Coal Mine Water Control. *Coal Sci. Technol.* **1991**, *4*, 8–10+63–64.
6. Yuan, W.H.; Gui, H.R. The characteristics of geothermal temperature and its application in distinguishing the source of water in renlou mine. *J. Anhui Univ. Sci. Technol. (Nat. Sci.)* **2005**, *4*, 9–11.

7. Jin, D.; Chen, J.; Wang, Y.; Zeng, Y.; Wang, X.; Zhou, C.; Zhang, G. Study on Artificial Neural Network System of Water Inrush Forecast. *J. Xi'an Univ. Sci. Technol.* **2000**, *3*, 214–217.
8. Zhang, X.; Zhang, Z.; Peng, S. Application of Quantitative Theory in Identifying Inrush water Sources in Mines. *J. China Univ. Min. Technol.* **2003**, *3*, 42–45. (In Chinese)
9. O'Shea, B.; Jankowski, J. Detecting Subtle Hydrochemical Anomalies with Multivariate Statistics: An Example from Homogeneous Groundwaters in the Great Artesian Basin, Australia. *Hydrol. Process.* **2010**, *20*, 4317–4333. [[CrossRef](#)]
10. Wang, Y.; Shi, L.; Wan, G.M. A New Evaluation Model for Discrimination of Mine Water Quality Based on EWM, Fuzzy, Fisher, and DS Evidence Theory: A Case Study in the Jiaojia Gold Mine Area, China. *Arab. J. Geosci.* **2021**, *14*, 1954. [[CrossRef](#)]
11. Zhang, H.; Xing, H.; Yao, D.; Liu, L.; Xue, D.; Guo, F. The Multiple Logistic Regression Recognition Model for Mine Water Inrush Source Based on Cluster Analysis. *Environ. Earth Sci.* **2019**, *78*, 612. [[CrossRef](#)]
12. Panagopoulos, G.; Angelopoulou, D.; Tzirtzilakis, E.; Giannouloupoulos, P. The Contribution of Cluster and Discriminant Analysis to the Classification of Complex Aquifer Systems. *Environ. Monit. Assess.* **2016**, *188*, 591. [[CrossRef](#)] [[PubMed](#)]
13. Wang, X.; Xu, T.; Huang, D. Application of Distance Discrimination Method in Identifying Water Inrush Sources in Similar Mining Areas. *J. China Coal Soc.* **2011**, *36*, 1354–1358. (In Chinese)
14. 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](#)]
15. Tiantian, W.; Dewu, J.; Jian, Y.; Liu, J.; Wang, Q. Assessing Mine Water Quality Using a Hierarchy Fuzzy Variable Sets Method: A Case Study in the Guojiawan Mining Area, Shaanxi Province, China. *Environ. Earth Sci.* **2019**, *78*, 264. [[CrossRef](#)]
16. Ju, Q.; Hu, Y. Source Identification of Mine Water Inrush Based on Principal Component Analysis and Grey Situation Decision. *Environ. Earth Sci.* **2021**, *80*, 157. [[CrossRef](#)]
17. Wu, Q.; Fan, S.; Zhou, W.; Liu, S. Application of the Analytic Hierarchy Process to Assessment of Water Inrush: A Case Study for the No.17 Coal Seam in the Sanhejian Coal Mine, China. *Mine Water Environ.* **2013**, *32*, 229–238. [[CrossRef](#)]
18. Donglin, D.; Wenjie, S.; Sha, X. Water Inrush Assessment Using a GIS-Based Bayesian Network for the 12-2 Coal Seam of the Kailuan Donghuanuo Coal Mine in China. *Mine Water Environ.* **2012**, *31*, 138–146. [[CrossRef](#)]
19. Yue, M. Application of Grey System Correlation Analysis in Identifying Water Inrush Sources in Mines. *Coal Sci. Technol.* **2002**, *4*, 37–39. (In Chinese)
20. Zhang, R.; Qian, J.; Ma, L.; Qin, H. Application of Extension Recognition Method in Identifying Water Inrush Sources in Mines. *J. China Coal Soc.* **2009**, *34*, 33–38. (In Chinese)
21. Wu, Q.; Hua, X.; Wei, P. GIS and ANN Coupling Model: An Innovative Approach to Evaluate Vulnerability of Karst Water Inrush in Coalmines of North China. *Environ. Geol.* **2008**, *54*, 937–943. [[CrossRef](#)]
22. Yan, Z.; Du, P.; Guo, D. Support Vector Machine Model for Analyzing Water Inrush Sources in Mines. *J. China Coal Soc.* **2007**, *8*, 842–847. (In Chinese)
23. Jiang, C.; Zhu, S.; Hu, H.; An, S.; Su, W.; Chen, X.; Li, C.; Zheng, L. Deep Learning Model Based on Big Data for Water Source Discrimination in an Underground Multiaquifer Coal Mine. *Bull. Eng. Geol. Environ.* **2022**, *81*, 26. [[CrossRef](#)]
24. Yan, P.; Shang, S.; Zhang, C.; Yin, N.; Zhang, X.; Yang, G.; Zhnag, Z.; Sun, Q. Research on the Processing of Coal Mine Water Source Data by Optimizing BP Neural Network Algorithm with Sparrow Search Algorithm. *IEEE Access* **2021**, *9*, 108718–108730. [[CrossRef](#)]
25. Chen, Y.; Tang, L.; Zhu, S. Comprehensive Study on Identification of Water Inrush Sources from Deep Mining Roadway. *Environ. Sci. Pollut. Res.* **2021**, *29*, 19608–19623. [[CrossRef](#)]
26. Zhang, J.; Chen, L.; Chen, Y.; Ge, R.; Ma, L.; Zhou, K.; Shi, X. Discrimination of Water Inrush Source and Evolution Analysis of Hydrochemical Environment under Mining in Renlou Coal Mine, Anhui Province, China. *Environ. Earth Sci.* **2020**, *79*, 61. [[CrossRef](#)]
27. Wang, Y.; Shi, L.; Wang, M.; Liu, T. Hydrochemical Analysis and Discrimination of Mine Water Source of the Jiaojia Gold Mine Area, China. *Environ. Earth Sci.* **2020**, *79*, 123. [[CrossRef](#)]
28. Ruan, Z.; Li, C.; Wu, A.; Wang, Y. A New Risk Assessment Model for Underground Mine Water Inrush Based on AHP and DS Evidence Theory. *Mine Water Environ.* **2019**, *38*, 488–496. [[CrossRef](#)]
29. Yu, X.; Han, J.; Shi, L.; Wang, Y.; Zhao, Y. Application of a BP Neural Network in Predicting Destroyed Floor Depth Caused by Underground Pressure. *Environ. Earth Sci.* **2017**, *76*, 535. [[CrossRef](#)]
30. Zhang, D.; Meng, L.; Dong, F.; Liu, X.; Shao, Q. Research on GA-SVM Method for Identifying Water Inrush Sources in Mines. *Coal Technol.* **2018**, *37*, 144–147. (In Chinese)

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.