Study on Characteristic Raman Shift Screening Method Based on MPA for Raman Spectrum of Mine Water Inrush Source

Minghao Zhou and Rongying Dai

1 School of Science, Inner Mongolia University of Technology, Hohhot, Inner Mongolia 010051, China
2 School of Electrical and Information Engineering, Anhui University of Science and Technology, Huainan, Anhui 232001, China

Correspondence should be addressed to Minghao Zhou; 2971223348@qq.com

Received 16 April 2022; Accepted 14 July 2022; Published 23 August 2022

In the process of mine water inrush disaster prevention, accurate and rapid identification of water inrush source type is of great significance to coal mine safety production. However, traditional hydrochemical methods have shortcomings such as time-consuming and complex detection. Therefore, a new idea of identifying mine water inrush source by Raman spectroscopy is proposed. Goaf water, roof sandstone fissure water, Ordovician limestone water, Taiyuan limestone water, and surface water as well as their mixed water samples are selected as research objects, and Raman spectral data of different water samples are collected by the Raman spectroscopy system. To eliminate the influence of laser power fluctuation and spectrometer system noise in Raman spectrum acquisition, detrend correction (DC), multiplicative scatter correction (MSC), standard normal variate transformation (SNV), first derivative (FD), and mean centering (MC) were used to preprocess the raw Raman spectra. Due to the large dimension and long analysis time of Raman spectrum data, the marine predator algorithm (MPA) is used to screen the characteristic Raman shifts of the Raman spectrum of water samples, and the characteristic Raman shift information that can best characterize the mine water samples is obtained. Finally, to verify the feasibility of MPA screening the characteristic Raman shifts of Raman spectrum of mine water inrush source, the selected characteristic Raman displacement information is used as input to construct BP neural network (BP), k-nearest neighbor algorithm (KNN), support vector machine (SVM), and decision tree (DT) classification models, respectively. Experiments show that SNV has the best preprocessing effect on the raw Raman spectrum, which can effectively eliminate part of the noise in the Raman spectrum data and improve the accuracy of Raman spectrum identification. Using MPA, 226 characteristic Raman shifts can be screened from 2048 Raman data points, reducing the number of Raman shifts to 11.04%, and the modeling accuracy of characteristic Raman shift information screened by MPA is higher than that of full Raman data. In addition, the average analysis speed of BP, KNN, SVM, and DT water source identification models is 7.61 times faster than that of all Raman data. The results show that MPA is adopted to screen the characteristic Raman displacement of mine water source Raman spectrum, which can effectively reduce the redundancy of Raman spectral data and greatly improve the speed of Raman spectral analysis, which is of great significance to ensure the real-time detection of the mine water source.

1. Introduction

With the increasing depth of coal mining, the coal mine production process will face gas, water, fire, coal dust, roof disasters, and other coal mine disasters [1, 2], and mine water disaster has become the second largest disaster after the mine gas disaster [3, 4]. To effectively guarantee the safe production of coal mines and reduce the occurrence of mine water disasters, it is an important prerequisite for the prevention and control of water disasters to accurately identify the types of mine water inrush sources [5, 6]. At present, there are many methods used to distinguish mine water inrush sources, such as groundwater level dynamics [7], hydrochemical analysis [8], GIS theoretical analysis [9], and so on [10]. Among them, the hydrochemical analysis method is one of the most widely used method [11], but the hydrochemical method needs to obtain the hydrochemical parameters of the mine water source (such as pH value, ion
concentration, and conductivity) in the harsh experimental environment. In addition, the hydrochemical analysis takes a long time and is not suitable for the solution of online identification of mine water source.

As one of the branches of spectral analysis methods, Raman spectroscopy [12, 13] has many advantages, such as high precision, high sensitivity, and no consumption of samples to be tested. It has been widely used in the fields of agriculture, food industry [14], biomedical [15], and so on. Bock and Gierling [16] analyzed the Raman spectra of coniferyl alcohol, abietin, and coniferyl aldehyde, and found that the Raman spectra of coniferyl alcohol showed direction dependence. The Raman spectra of abietin and coniferyl alcohol were very similar, and the Raman spectra of coniferyl aldehyde were affected by the molecular crystal order. Gao et al. [17] compared the ability of near-infrared Fourier transform Raman (1064 nm) and visible Raman (532 nm) to distinguish animal fat. It was found that there were significant differences in the shoulder peak intensity ratio of animal fat in different species at 1653/1745 and 1653/1298 cm$^{-1}$, as well as 1298 and 1264 cm$^{-1}$. After collecting Raman spectra, Zhao et al. [18] integrated nine supervised machine learning algorithms into a system to realize the rapid analysis of edible oil and compared the analysis results with gas chromatography. It was found that these machine learning algorithms combined with Raman spectra had higher analysis accuracy and faster analysis speed. Han et al. [19] briefly described the latest development of machine learning in vibration spectrum prediction, and comprehensively discussed the relationship between machine learning methods and vibration spectrum, especially for Raman spectroscopy. In addition, the future development trend was analyzed. Considering that machine learning is shaping our lives in many ways, Lussier et al. [20] briefly summarized the most common machine learning techniques used in Raman, the guidelines for new users to implement machine learning in data analysis, and the modern applications of machine learning in Raman spectroscopy. In general, Raman spectroscopy has achieved many successful applications. However, Raman spectroscopy data usually have large dimensions and the analysis process is complex. To solve this problem, it is necessary to carry out the feature screening of Raman spectroscopy to reduce the redundancy of Raman spectroscopy data, to improve the real-time performance of spectral analysis, which is of great significance to ensure the real-time performance of Raman spectroscopy detection of mine water sources.

MPA is a new meta-heuristic optimization algorithm proposed by Faramarzi et al. [21] in 2020. Its inspiration comes from the survival of the sea fittest theory, that is, marine predators choose the best foraging strategy between the Levy walk or brown walk, which has the characteristics of strong optimization ability. MPA is widely used in model parameter optimization [22], optimal scheduling [23], feature selection [24], and many other fields. Ramezani et al. [25] improved the raw MPA based on the opposition learning method, chaotic map, population adaptation, and the switching between exploration and utilization stages, and evaluated the performance of the improved algorithm based on the actual optimization problem of proportion integral differential (PID) control applied to direct current (DC) motor. Aiming at the problem of parameter estimation of photovoltaic systems, Abdel-Basset et al. [26] proposed an improved algorithm based on the new meta-heuristic ocean predator algorithm to extract the optimal value of photovoltaic parameters. The specific idea is to improve the estimation accuracy of photovoltaic parameters by using two different ways to deal with the solutions within the population based on the average fitness of the population. Islam et al. [27] proposed a natural-inspired meta-heuristic MPA to solve the optimal power flow (OPF) problem. The proposed method was tested on IEEE 30 bus test system and proved to be feasible and effective. Considering that feature selection is a necessary process to reduce the high-dimensional data set, Abd Elaziz et al. [28] proposed a feature selection strategy based on MPA, which improved the search ability of the algorithm by introducing a sine cosine algorithm, and tested and verified on University of California Irvine (UCI) datasets.

After considering these factors comprehensively, this paper proposes a Raman spectrum feature Raman shift screening method based on MPA and applies it to the rapid identification of mine water source. After obtaining the Raman spectrum data of the mine water inrush source, the raw Raman spectra of mine water inrush sample are smoothed to reduce the noise interference. Then, MPA is used to screen Raman spectrum characteristic shifts of the mine water inrush source, removed the redundant information in the Raman spectrum, and reduced the dimension of Raman spectrum data. Taking the selected characteristic Raman shifts as the input of different classifiers, the recognition model of mine water inrush source is constructed, and the feasibility of MPA screening the characteristic Raman shifts of mine water inrush source Raman spectrum is discussed.

2. Experimental Part

The purpose of this paper is to screen the characteristic Raman shifts of Raman spectrum of mine water inrush source. Firstly, the Raman spectra of water samples are obtained by the Raman spectroscopy system, then the raw Raman spectra are preprocessed, and then the characteristic Raman shift information of the Raman spectrum of water samples is screened by MPA. Finally, the selected characteristic Raman displacement information is used as the input of different classifiers to verify the feasibility of MPA to screen the characteristic Raman displacement of the mine water inrush source. In particular, we use a laptop equipped with Intel Core i7-8565U as the data processing platform and use MATLAB R2020a to complete the data analysis process. The specific experimental process is shown in Figure 1.

2.1. The Experimental Material. In this paper, the goaf water, roof sandstone fissure water, Ordovician limestone water, Taiyuan limestone water, and surface water collected in the...
Zhangji mine in Huainan in October 2021 were taken as experimental materials. The information of the mine water sample collection points is shown in Table 1. If the water inrush disaster occurs in the coal mines, the water source of different geological layers usually overlaps. Therefore, it is of great significance to study the Raman characteristics of mixed water sources of different geological layers. Based on a single mine water source, four kinds of mixed water samples were obtained by mixing roof sandstone fissure water, Ordovician limestone water, Taiyuan limestone water, and surface water with goaf water at the volume ratio of 1:1. Finally, we can get 4 types of mixed water and 5 types of single mine water source, a total of 9 watersources. All nine sources have 160 samples, so the total number of water samples is 1440. 112 samples from each water source type were selected as the training samples and 48 samples were selected as the test samples. The accurate collection of Raman spectrum data of different mine water samples is very important to ensure the reliability and authenticity of Raman spectrum analysis. Therefore, we stored all collected mine water inrush source samples in a dark and sealed way.

2.2. Raman Spectrum Acquisition. To collect the Raman spectrum of the mine water source, a Raman spectrum acquisition system was built, as shown in Figure 2. Laser 785-5HFUO laser (Shanghai Ruhai Photoelectric Technology Co. Ltd.) is used as the excitation light source of mine water sample Raman spectrum. The peak wavelength of the laser is 785 ± 0.5 nm and the excitation power is adjustable in the range of 0~500 mW. Considering the convenience of Raman spectrum measurement and the practicality of mine water disaster warning in the future, RBP-785-1.5-FS immersible Raman probe (Shanghai Ruhai Photoelectric Technology Co. Ltd.) is adopted to directly put into the samples to be measured to measure Raman spectra. The generated Raman spectra are collected by an XR3000 optical fiber spectrometer (Shanghai Ruhai Photoelectric Technology Co. Ltd.), which is equipped with a 2048 × 64 plane array near-infrared enhanced CCD, with a spectral detection range of 780~1070 nm. In the process of collecting the Raman spectra of mine water samples, the laser power is set to 25 mW and the integration time of the spectrometer is set to 500 ms. The Raman spectra are recorded by Uspectral-PLUS (version 5.2.0) software. At the same time, the whole spectrum acquisition process was carried out in the dark room.

2.3. Raman Spectrum Pretreatment. Because the spectral signals collected by the spectrometer contain not only useful information but also random errors, there is noise interference in the measured spectral curve, so it is necessary to preprocess the collected raw spectral data. Common spectral preprocessing methods include detrend correction (DC), multiplicative scatter correction (MSC), standard normal variable transformation (SNV), first derivative (FD), and mean centering (MC).

2.4. MPA Characteristic Raman Shift Screening. MPA is used to screen the characteristic Raman shifts of the mine water source. The basic idea is to determine the parameters to be optimized according to the identification problem of mine water inrush source, that is, the characteristic Raman shifts of mine water samples, and each spatial position of predator contains a group of characteristic Raman displacements. The spatial position of the predator is measured by the fitness function, and the predator-prey strategy is used to continuously update the predator position until the best predator spatial position is obtained, that is, the best set of characteristic Raman shifts of the problem to be optimized is obtained. The screening process of characteristic Raman shifts is as follows:

Step 1: define the fitness function. Since MPA is a process of solving the minimum value, the classification error (E) of the test set of the mine water source classification model is taken as the fitness function, that is, the objective function is

$$E = 1 - \frac{N_{\text{correct}}}{N_{\text{test}}}$$  \hspace{1cm} (1)$$

where $N_{\text{correct}}$ is the number of correctly predicted samples in the test set and $N_{\text{test}}$ is the total number of samples in the test set.

Step 2: set the algorithm parameters and initialize the population. Define the population number $n$, the maximum number of iterations $T$, effect coefficient of the fish coagulation device FAD and the constant $P$.

Step 3: build the prey matrix and elite matrix. The position of the randomly initialized population is shown as

$$X_{ij} = X_{ib} + \text{rand} \times (X_{ub} - X_{lb}), i = 1, 2, \ldots, n, j = 1, 2, \ldots, d,$$  \hspace{1cm} (2)$$

where $X_{ij}$ represents the $i$-th dimension coordinate of the $j$-th population, $X_{ub}$ and $X_{lb}$ is the upper and lower
Table 1: Location information of mine water sample collection.

| Water Type                           | Sampling Location                              | Coordinate information |
|-------------------------------------|------------------------------------------------|------------------------|
| Goaf water                          | 1222 (1) bottom pumping roadway                | 24805 53708 –680       |
| Roof sandstone fissure water        | 48.4m before Y18 of yunshun 1615A              | 28805.30 48188.84 –616.1 |
| Ordovician limestone water          | West no. 21 coal mining area –566m drainage roadway water bunker | 28718.5 51212.6 –566.5 |
| Taiyuan limestone water             | –536m drainage roadway closed wall             | 50746.6 29001.4 –536.0 |
| Surface water                       | Mining area surface                             | — — —                  |

![Image](image.png)

**Figure 2:** Diagram of the Raman spectrum acquisition system.

The boundaries of the search space, and rand is a random number in the [0, 1] interval.

Based on the location of the search agent, the prey matrix is constructed:

$$\text{Prey} = \begin{bmatrix} X_{1,1} & X_{1,2} & \cdots & X_{1,d} \\ X_{2,1} & X_{2,2} & \cdots & X_{2,d} \\ \vdots & \vdots & \ddots & \vdots \\ X_{n,1} & X_{n,2} & \cdots & X_{n,d} \end{bmatrix}^{n}.$$  

(3)

MPA, inspired by the natural selection law of survival of the fittest, believes that the top hunters are gifted in hunting. Therefore, the optimal predator is regarded as the ontology to copy n – 1 identical predators, and the matrix constructed by n top predators is called the elite matrix, as shown as follows:

$$\text{Elite} = \begin{bmatrix} X_{1,1}^i & X_{1,2}^i & \cdots & X_{1,d}^i \\ X_{2,1}^i & X_{2,2}^i & \cdots & X_{2,d}^i \\ \vdots & \vdots & \ddots & \vdots \\ X_{n,1}^i & X_{n,2}^i & \cdots & X_{n,d}^i \end{bmatrix}^{n}.$$  

(4)

where $X^i$ represents the optimal predator individual vector. Search agents include predators and prey. So, they are all looking for their own food. At the same time, after each iteration, the elite positions are updated according to the adaptation.

Step 4: according to the iteration stage, the predator selects the corresponding update method to update the predator position.

Stage 1: at the initial stage of iteration (the current generation is less than 1/3 of the maximum number of iterations, i.e., $t < (1/3)T$), conduct a global search and update Prey through Brownian random walk.

$$\text{stepsize}_i = R_B \odot (\text{Elite}_i - R_R \odot \text{Prey}_i), i = 1, 2, \ldots, n,$$

$$\text{Prey}_{i+1} = \text{Prey}_i + \text{stepsize}_i,$$

(5)

where $\text{stepsize}_i$ is the step size, and the vector $R_B$ is a random number vector caused by the normal distribution of Brownian motion, $\odot$ representing element by element multiplication. $P$ is a constant and $P = 0.5$ is a step factor, $R \in [0, 1]$ is a random number vector, $R_B \odot \text{Prey}_i$ and simulates the high-speed movement of prey.

Stage 2: the middle stage of the iteration (the current generation is greater than 1/3 of the maximum number and less than 2/3 of the maximum number of iterations, that is, $1/3T < t < (1/3)T$), and this stage is an intermittent stage between the exploration and development stages. The population is divided into two parts, in which the prey makes levy motion, which is responsible for the development of the algorithm in the search space, and the predator makes Brownian motion, which is responsible for the exploration of the algorithm in the search space. Development and exploration behaviors can be expressed as follows:

The location of the development team shall be updated as follows:

$$\text{stepsize}_i = R_L \odot (\text{Elite}_i - \text{Rey}_i), i = 1, 2, \ldots, \frac{n}{2},$$

$$\text{Prey}_{i+1} = \text{Prey}_i + P \cdot R \odot \text{stepsize}_i,$$

(6)

where $R_L$ is a random number vector based on Levy flight distribution, which is used to describe Levy flight motion.

The exploration group moves through equation (7) update mode:

$$\text{stepsize}_i = R_B \odot (R_B \odot \text{Elite}_i - \text{Prey}_i), i = \frac{n}{2}, \ldots, n,$$

$$\text{Prey}_{i+1} = \text{Elite}_i + P \cdot CF \odot \text{stepsize}_i,$$

(7)
Among them, $CF$ is an adaptive parameter used to control the moving step of the predator, and its calculation method is shown as follows:

$$CF = \left(1 - \frac{t}{T}\right)^{(2/3T)}, \quad (8)$$

where $t$ is the current number of iterations and $T$ is the maximum number of iterations.

Stage 3: in the later stage of iteration (the current generation is greater than $2/3$ of the maximum number of iterations, i.e., $t > 2/3T$), mainly improve the local development of the algorithm. At this time, the predator's strategy is levy motion:

$$\text{stepsizes}_i = R_q \otimes (R_q \otimes \text{Elite}_i - \text{Prey}_i), \quad i = 1, \ldots, n, \quad \text{Prey}_{i+1} = \text{Elite}_i + P \cdot CF \otimes \text{stepsizes}_i. \quad (9)$$

Step 5: calculate the fitness value and update the optimal position of the predator.

Step 6: fish aggregation device (FAD) and eddy current effect. Considering that the external environment may affect the movement of the population to a certain extent, Faramazi et al. [21] proposed a position update strategy based on vortex formation and fish aggregation device to avoid the result falling into the local optimal solution. By using this strategy, MPA effectively overcomes the problem of premature convergence and avoids local extremum in the optimization process. The strategy can be expressed as shown as follows:

$$\text{prey}_1 + CF[X_{lb} + R \otimes (X_{ub} - X_{lb})] \otimes U, r \leq FAD,$$

$$\text{prey}_1 + \lfloor FAD(1 - r) + r \rfloor(\text{prey}_{r1} - \text{prey}_{r2}), r > FAD. \quad (10)$$

Among them, FAD is the probability of affecting the search process, and the value is 0.2. $X_{lb}$ and $X_{ub}$ are the vectors composed of the minimum and maximum values of the search boundary, respectively, $r$ is a random value, $r \in [0, 1], r1$, and $r2$ represent the random index of the prey matrix, and $U$ is a binary vector.

Step 7: judge whether the stop conditions are met. If not, return to step 3 for repeated execution. Otherwise, output the optimal predator fitness value and its spatial position, which is the best set of characteristic Raman shifts.

### 2.5. Raman Spectrum Classification

To realize the classification and identification of Raman spectrum data, the basic classification algorithms such as back propagation neural network (BP), k-nearest neighbor (KNN), support vector machine (SVM), and decision tree (DT) are used to construct the Raman spectrum classification models, to realize the recognition of mine water source Raman spectrum. In particular, to ensure the reliability of the mine water source classification models, a 10-fold cross-validation method is adopted in the process of model training.

#### 3. Results and Analysis

##### 3.1. Raw Raman Spectrum

The Raman spectrum data of mine water samples are collected by XR3000 optical fiber spectrometer, and the Raman spectra of 1440 samples of nine water samples (160 samples of each water sample) are measured experimentally, as shown in Figure 3. It can be seen from the figure that unmixed goat water has a good identification degree, but it becomes difficult to distinguish when mixing occurs. The mixed water samples with similar mixing proportion overlap slightly, so it is necessary to classify and identify with the help of the chemometrics method.

To more intuitively understand the differences in the Raman spectrum of different water samples, basic classification algorithms such as BP neural network (BP), k-nearest neighbor (KNN), support vector machine (SVM), and decision tree (DT) are used as classifiers. Take the raw Raman spectrum data as the input, observe the classification effect of mine water source Raman spectrum under different classifiers, and count the classification accuracy and analysis time of different classifiers. The relevant results are shown in Table 2.

It can be seen from Table 2 that there are some differences in the classification results obtained by using different classifiers for the raw frontal Raman spectrum data. However, we note that BP, KNN, SVM, and DT have high classification accuracy, and the classification accuracy of all algorithms for the test set is higher than 94.00%. The difference in the analysis results of different classification algorithms for the raw Raman spectrum data is mainly reflected in the difference in the classifier model itself. In addition, the raw Raman spectrum data have a certain noise, and the anti-interference ability of different classification algorithms for noise is also different. In addition, the raw Raman spectrum data have a large dimension (2048 dimensions), and the analysis process takes a long time. It can be seen that there are significant differences in Raman spectrum analysis time between different classifiers. Among them, the analysis time of KNN is the shortest (0.31 s), the analysis time of BP is the longest (50.18 s), and the average analysis time of the four classifiers is 19.32 s. In general, although the four basic classification algorithms have good classification accuracy for the raw Raman spectra, the recognition accuracy of the Raman spectrum of mine water source still has room for further improvement. Therefore, it is very necessary to preprocess the raw Raman spectra.

##### 3.2. Smoothing Pretreatment

To eliminate the noise interference, reduce the error, and retain the useful information in the spectral curve, DC, SC, SNV, FD, and MC are used to preprocess the raw Raman spectral data of the mine water source, respectively, and the preprocessed spectral information is used as the input of classifier to obtain the classification accuracy of the preprocessed Raman spectra under the four algorithm models, as shown in Table 3.

By observing Table 3, it can be found that after smoothing and preprocessing the raw Raman spectrum of
the mine water source, the classification accuracy of the Raman spectrum under different classifiers has been improved to a certain extent. In particular, we find that the average recognition accuracy of the four classifiers is different from that of the processed Raman spectrum data. When using SNV processing, the average recognition accuracy of the five classifiers can reach the maximum, that is, 98.25%, which shows that SNV is more suitable for smoothing the raw Raman spectrum of mine water samples. When SVM is used as the classifier of Raman spectrum data, the four different classifiers can achieve 99.44% recognition accuracy, which shows that among the four classification algorithms, the SVM algorithm is more suitable for analyzing the preprocessed Raman spectra. Although the identification accuracy of the preprocessed Raman spectrum has been improved to a certain extent, its data dimension has not changed (still 2048 dimension), and the time of spectral analysis is still relatively long. Therefore, it is necessary to use a certain feature selection strategy to screen the Raman shifts of the Raman spectrum, effectively reduce the dimension of the spectral data, and improve the speed of data analysis.

### Table 2: Classification results of the raw Raman spectrum data.

|         | BP   | KNN  | SVM  | DT   | Mean value |
|---------|------|------|------|------|------------|
| Classification accuracy | 94.91% | 96.06% | 98.38% | 96.99% | 96.59% |
| Analysis time       | 50.18 s | 0.31 s | 24.87 s | 1.93 s | 19.32 s |

### Table 3: Classification accuracy (%) of the different pretreatment methods.

|         | BP   | KNN  | SVM  | DT   | Mean value |
|---------|------|------|------|------|------------|
| DC      | 93.29 | 94.91 | 99.44 | 97.40 | 96.26 |
| MSC     | 97.45 | 96.80 | 99.44 | 98.84 | 98.13 |
| SNV     | 97.92 | 96.80 | 99.44 | 98.84 | 98.25 |
| FD      | 97.22 | 96.80 | 99.44 | 98.61 | 98.02 |
| MC      | 97.45 | 97.45 | 99.44 | 98.61 | 98.24 |
| Mean value | 96.67 | 96.55 | 99.44 | 98.46 | — |

3.3. MPA Raman Shift Screening. There are 2048 spectral data points in the Raman spectrum data of mine water samples. Due to a large amount of data, long operation time, and certain collinearity, the prediction ability of the model may not be high enough when building the mine water source identification model. Therefore, MPA is used to screen the Raman shifts of water sample Raman spectrum data, and the characteristic Raman shift points with high importance in Raman spectrum data are extracted. After the raw Raman spectrum data is pre-processed by MA, MPA is used to screen the characteristic Raman shifts. Finally, the KNN classification model is established by using the screened characteristic Raman shift data (KNN is selected as the classifier to reduce the time of MPA iterative screening of characteristic Raman shifts). The number of Raman displacement variables corresponding to the smallest classification error value is the final screening result. In particular, in the process of characteristic Raman shift screening using MPA, the population number is set to 10, the maximum number of iterations is set to 100, the fish aggregating devices effect is set to 0.2, and the constant is set to 0.5. Figure 4 shows the variation trend of classification error when extracting different numbers of Raman displacement variables. From the figure, it can be seen that the classification error has reached the minimum value of 0.80% after 55 iterations. At this time, the number of characteristic Raman shifts is 226. In other words, when 226 characteristic Raman shifts are selected, the accurate identification of the Raman spectrum of mine water samples can be realized.

Compared with all 2048 Raman spectrum data points, 226 characteristic Raman shifts screened by MPA reduce the number of Raman shifts to the original 11.04%. The preferred 226 characteristic Raman shifts are shown in Figure 5. In the figure, they are represented by red squares. It can be seen that the screened characteristic Raman shifts effectively cover the positions of peaks and
3.4. Spectral Analysis after Screening. To further verify the effectiveness and reliability of the characteristic Raman displacement screened by MPA, the screened characteristic Raman displacement data is used as the input information for mine water source identification. The identification models of mine water source types are constructed by using BP, KNN, SVM, and DT classification algorithms, respectively, and the analysis results of the four classification models are recorded. The classification accuracy and analysis time are shown in Table 4.

It can be seen from the above table that the four different classification algorithms have excellent identification performance for the filtered characteristic Raman displacement data. Among them, the identification accuracy of SVM and DT can reach 100.00%, and that of BP can also reach 98.61%. When we observe the analysis time, it is obvious that the analysis time is greatly reduced, mainly because the dimension of Raman spectrum data is reduced, which makes the analysis speed faster. Compared with the average analysis time of 2048 spectral data in Table 2, it can be seen that the average analysis time of mine water source identification using the screened characteristic Raman displacement data is reduced from 19.32 s to 2.54 s, that is, the analysis speed of the screened characteristic Raman data is 7.61 times that of the full Raman displacement data. In conclusion, using MPA to screen the characteristic Raman shifts of the Raman spectrum of mine water source is effective, and can greatly improve the analysis speed, which is of great significance to speed up the research process of the Raman spectrum identification model of mine water source and realize the online identification of mine water inrush source.

4. Conclusion

Given the urgent need of mine water inrush disaster early warning, a new method of mine water source identification based on Raman spectrum and MPA is proposed in this paper. Taking the goaf water, roof sandstone fissure water, Ordovician limestone water, Taiyuan limestone water, surface water, and their mixed water samples as the research objects, the Raman spectrum data of different water samples are collected by Raman spectrum system. In conclusion, MPA is adopted to screen the characteristic Raman displacement of mine water source the Raman spectrum, which can effectively reduce the redundancy of Raman spectral data and greatly improve the real-time performance of Raman spectral analysis, which is of great significance to ensure the speed of mine water source Raman spectral detection.

At present, the research work of this paper still has some limitations, which are mainly reflected in that the current mine water source comes from the Huainan mining area, and the sample size of Raman spectrum data is relatively small. In the future, we will collect mine water samples from Shanxi and Inner Mongolia, collect their Raman spectra, and further enrich the Raman spectrum database of mine water sources.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.
References

[1] R. Gao, H. Yan, F. Ju, X. Mei, and X. Wang, "Influential factors and control of water inrush in a coal seam as the main aquifer," *International Journal of Mining Science and Technology*, vol. 28, no. 2, pp. 187–193, 2018.

[2] J. Lu, Y. Yin, H. Gao et al., "True triaxial experimental study of disturbed compound dynamic disaster in deep underground coal mine," *Rock Mechanics and Rock Engineering*, vol. 53, no. 5, pp. 2347–2364, 2020.

[3] W. Hu and C. Zhao, "Evolution of water hazard control technology in China’s coal mines," *Mine Water and the Environment*, vol. 40, no. 2, pp. 334–344, 2021.

[4] F. Hu, M. Zhou, P. Yan et al., "Selection of characteristic wavelengths using SPA for laser induced fluorescence spectroscopy of mine water inrush," *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, vol. 219, pp. 367–374, 2019.

[5] G. Lin, D. Jiang, D. Dong, J. Fu, and X. Li, "A multilevel recognition model of water inrush sources: a case study of the zhaozhehuang mining area," *Mine Water and the Environment*, vol. 40, no. 3, pp. 773–782, 2021.

[6] Y. Hu, P. Huang, H. Gao, and Q. Su, "State of the practice worldwide: HCA-PCA-EWM discrimination model of water inrush source in mines," *Groundwater Monitoring & Remediation*, vol. 42, no. 2, pp. 67–76, 2022.

[7] S. Shi, X. Xie, L. Bu, L. Li, and Z. Zhou, "Hazard-based evaluation model of water inrush disaster sources in karst tunnels and its engineering application," *Environmental Earth Sciences*, vol. 77, no. 4, p. 141, 2018.

[8] D. Wang, Q. Ju, Y. Wang et al., "Source identification of mine water inrush based on the exponential whitening function and the grey situation decision model," *Energy Exploration & Exploitation*, vol. 40, no. 4, pp. 1217–1235, 2022.

[9] H. Duan, S. Hao, J. Feng, Y. Wang, and D. Peng, "A prediction method for water enrichment in aquifer based on GIS and coupled AHP-entropy model," *Open Geosciences*, vol. 13, no. 1, pp. 1318–1327, 2021.

[10] Y. Wang, J. Yu, X. Feng et al., "A case study of direct current resistivity method for disaster water source detection in coal mining," *Elektronika ir Elektrotechnika*, vol. 27, no. 4, pp. 41–48, 2021.

[11] H. Zhang, G. Xu, X. Chen, J. Wei, S. Yu, and T. Yang, "Hydrogeochemical characteristics and groundwater inrush source identification for a multi-aquifer system in a coal mine," *Acta Geologica Sinica - English Edition*, vol. 93, no. 6, pp. 1922–1932, 2019.

[12] V. S. Gorelik, S. D. Abdurakhmonov, N. V. Sidorov, and M. N. Palatnikov, "Raman scattering in lithium niobate and lithium tantalate single crystals and ceramics," *Inorganic Materials*, vol. 55, no. 5, pp. 524–532, 2019.

[13] H. Werheit, "Systematic error in conventionally measured Raman spectra of boron carbide-A general issue in solid state Raman spectroscopy," *Review of Scientific Instruments*, vol. 90, no. 4, Article ID 043114, 2019.

[14] L. Yu, S. Sen Meng, K. Lin, S. Q. Duan, Z. Q. Wang, and R. T. Zhang, "Ethanol concentration in Chinese liquors by absolute Raman difference spectra," *Spectroscopy and Spectral Analysis*, vol. 40, pp. 2143–2147, 2020.

[15] C. Chen, F. Chen, B. Yang, K. Zhang, X. Lv, and C. Chen, "A novel diagnostic method: FT-IR, Raman and derivative spectroscopy fusion technology for the rapid diagnosis of renal cell carcinoma serum," *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, vol. 269, Article ID 120684, 2022.

[16] P. Bock and N. Gierlinger, "Infrared and Raman spectra of lignin substructures: coniferyl alcohol, abietin, and coniferyl aldehyde," *Journal of Raman Spectroscopy: JRS*, vol. 50, no. 6, pp. 778–792, 2019.

[17] F. Gao, D. Ben-Amotz, S. Zhou, Z. Yang, L. Han, and X. Liu, "Comparison and chemical structure-related basis of species discrimination of animal fats by Raman spectroscopy using near-infrared and visible excitation lasers," *Lebensmittel-Wissenschaft und -Technologie*, vol. 134, Article ID 110105, 2020.

[18] H. Zhao, Y. Zhan, Z. Xu et al., "The application of machine-learning and Raman spectroscopy for the rapid detection of edible oils type and adulteration," *Food Chemistry*, vol. 373, Article ID 131471, 2022.

[19] R. Han, R. Kjetaew, and S. Lubier, "A concise review on recent developments of machine learning for the prediction of vibrational spectra," *The Journal of Physical Chemistry A*, vol. 126, no. 6, pp. 801–812, 2022.

[20] F. Lussier, V. Thibault, B. Charonn, G. Q. Wallace, and J.-F. Masson, "Deep learning and artificial intelligence methods for Raman and surface-enhanced Raman scattering," *Trends in Analytical Chemistry*, vol. 124, Article ID 115796, 2020.

[21] A. Faramarzi, M. Heidarinejad, S. Mirjalili, and A. H. Gandomi, "Marine predators algorithm: a nature-inspired metaheuristic," *Expert Systems with Applications*, vol. 152, Article ID 113377, 2020.

[22] N. Panagant, M. Yildiz, N. Pholdee, A. R. Yildiz, S. Burcenter, and S. M. Sait, "A novel hybrid marine predators-Nelder-Mead optimization algorithm for the optimal design of engineering problems," *Materials Testing*, vol. 63, no. 5, pp. 453–457, 2021.

[23] J. Pan, J. Shan, S. Chu, S. Jiang, S. Zheng, and L. Liao, "A multigroup marine predator algorithm and its application for the power system economic load dispatch," *Energy Science & Engineering*, vol. 10, no. 6, pp. 1840–1854, 2021.

[24] D. S. A. Elminaam, A. Nabil, S. A. Ibraheem, and S. H. Houssein, "An efficient marine predators algorithm for feature selection," *IEEE Access*, vol. 9, pp. 60136–60153, 2021.

[25] M. Ramezani, D. Bahmanyar, and N. Razmjooey, "A new improved model of marine predator algorithm for optimization problems," *Arabian Journal for Science and Engineering*, vol. 46, no. 9, pp. 8803–8826, 2021.

[26] M. Abdel-Basset, D. El-Shahat, R. K. Chakrabortty, and M. Ryan, "Parameter estimation of photovoltaic models using an improved marine predators algorithm," *Energy Conversion and Management*, vol. 227, Article ID 113491, 2021.

[27] M. Z. Islam, M. L. Othman, N. I. Abdul Wahab et al., "Marine predators algorithm for solving single-objective optimal power flow," *PLoS One*, vol. 16, no. 8, Article ID e0256050, 2021.

[28] M. Abd Elaziz, A. A. Elwasif, D. Youssri, L. Abualigah, and M. A. Al-qaness, "Modified marine predators algorithm for feature selection: case study metabolomics," *Knowledge and Information Systems*, vol. 64, no. 1, pp. 261–287, 2022.