Prediction of rock burst risk in deep coal mining based on time-frequency domain analysis

Dongping Sheng, Zhongyuan Ma, Chenqi Zhou, Haidong Feng, Hao Liu, Anbang Zhu, Hun Guo*

Changzhou Institute of Technology, Changzhou, China

ABSTRACT

Coal is an important energy and industrial raw material in China, and its exploitation plays a vital role in the stability and development of the national economy. This paper will use the data given in the title to explore more efficient data analysis methods and more advanced monitoring equipment to improve the accuracy and reliability of the early warning system. For the identification of interference signals, a detailed data preprocessing and analysis process is adopted. Through data cleaning, duplicate and invalid data are removed and data quality is ensured. Wavelet analysis is used to extract the features of interference signals in electromagnetic radiation and acoustic emission signals by using time-domain, frequency-domain and time-frequency domain methods, effectively identify and further eliminate interference data, and improve the accuracy of signal processing. In addition, support vector machine (SVM), neural network and random forest comparison model are constructed to effectively classify the interference signals. Precursor characteristic signal analysis, analyze precursor characteristic signal and identify the time interval of precursor characteristic signal. First of all, the same data processing method as work 1 is used to clean the data involved in work 2. Combining wavelet transform and statistical analysis method, the key statistical features of the signal are extracted, and a prediction model based on the sliding window of time series analysis is established to calculate the average value, standard deviation and energy of each window. In order to identify precursory feature signals as early as possible, a mathematical model based on historical and real-time data is developed. Smooth moving technology is used to capture the basic characteristics of signals in each window, and combined with the probability of having a small standard deviation in the prediction results, the method that the standard deviation is less than the average predicted value is calculated, so as to establish a prediction probability model and obtain the probability of precursor characteristics at the time of data collection. This model can realize timely early warning of disasters such as rock burst. It provides important guarantee for mine safety.

Keywords: Wavelet Analysis, SVM Model, Random Forest, Sliding Window, ARIMA Model, Feature Engineering

1 INTRODUCTION

With the deepening of China's coal mining work, the complexity and uncertainty of the underground coal mine environment are also increasing. In the process of deep mining, the significant increase of ground stress leads to a significant increase in the risk of underground coal and rock dynamic disasters, and rock burst, as one of the typical disasters, poses a serious threat to the safety of coal mine production.

Rock burst refers to the sudden release of energy and violent destructive dynamic phenomenon in the process of deep mining due to the influence of high ground stress and brittle failure of rock structure. This phenomenon may not only lead to huge economic losses, but also often accompanied by casualties.

In order to effectively prevent and control rock burst disasters, researchers and engineers have carried out a lot of research work in recent years, and developed a variety of monitoring and early warning technologies. Among them, acoustic emission (AE) and electromagnetic radiation (EMR) monitoring technologies are widely used in the dynamic disaster warning system of underground coal and rock. AE technology predicts the possibility of rockburst by monitoring the acoustic energy released during the propagation of rock microcracks, while EMR monitoring detects the electromagnetic wave changes caused by rock cracking [1]. In practice, these monitoring devices are installed in key parts of the mine, such as the face or roadway, and by continuously monitoring the relevant signal changes, researchers can analyze the area of stress concentration in the rock layer and the likely time of rock burst. These monitoring data are usually recorded at a fixed time interval and transmitted to the ground control center, through a special software system for data analysis and processing, to achieve real-time early warning of rock burst [2].

Although the current monitoring technology has made some progress, there are still many challenges in the prediction and control of rock burst. Due to the complexity of underground coal mine environment and the variability of geological conditions, the occurrence of rock burst is highly uncertain and random. In addition, the existing technology still has some limitations in data processing and interpretation. How to accurately interpret monitoring data and transform it into effective early warning information is the key to improve the level of coal mine safety production [3]. Through the implementation of these comprehensive measures, the risk of rock burst in deep mining can be effectively reduced, and the life safety of miners and the stable operation of mines can be guaranteed.

Work one: In mine operations, interference from other equipment and operations in the working environment is often encountered, affecting the processing and analysis of electromagnetic radiation (EMR) and acoustic emission (AE) signals, thereby reducing the accuracy of disaster warning systems. It is necessary to establish a mathematical model to analyze the difference between the interference signal and the normal signal, and capture the characteristic value of the interference signal. Using the characteristic value of the interference signal, the data in Annex I and II are analyzed, and the interval of the first five interference signals is obtained [4].

Work two: About 7 days before the earth burst, the electromagnetic radiation and acoustic emission signal will show a specific change trend, which is called precursor characteristic signal. Timely identification and analysis of these signals is essential so that timely measures can be taken to prevent potential disasters. By establishing a mathematical model, the trend characteristics of electromagnetic radiation and acoustic emission signal before the occurrence of danger are obtained respectively. According to the trend characteristics obtained, the data of specific time periods are analyzed, and the time interval of the earliest five precursor characteristic signals in these time periods is identified and recorded [5].

Work three: In order to issue a timely warning when the precursor feature signal first appears, it is necessary to conduct real-time analysis of the signal at each data collection to predict the probability of the precursor feature signal appearing. A mathematical model is established to analyze the electromagnetic radiation and acoustic emission signal data collected in each discontinuous time period given in the attachment. For the data at the last moment of each time period, the probability of the occurrence of the precursor characteristic signal is predicted and the result is recorded in the corresponding table.

2 RELATED WORK AND ASSUMPTION

2.1 Work analysis

Work one: First, the data set is cleaned to detect outliers and missing values. For electromagnetic radiation and acoustic emission signals, 13 characteristics such as mean value, standard deviation, skewness, kurtosis, maximum value, minimum value, peak value, root mean square, amplitude factor, waveform factor, shock factor, margin factor and energy are analyzed. Next, SVM, neural network and random forest models are built using these features. By comparing the performance of the models, it can be determined that the SVM model performs best in terms of accuracy [6]. Then, SVM interference recognition model is established to monitor the time period of electromagnetic radiation and acoustic emission signal, and finally determine the time period of interference signal.

Work two: The electromagnetic radiation and acoustic emission signals in work two are analyzed, which may contain precursory characteristic signals indicating rock burst. A decision tree model is established to extract the statistical characteristics of the precursor signals, and the trend characteristics related to the occurrence of rock burst are determined by wavelet analysis. The sliding window method is used to calculate the mean value, standard deviation and energy of each window. Next, the classification model is trained using A random forest algorithm, aiming to learn patterns for categories A and B from these features. The electromagnetic radiation and acoustic emission data of work two in a specific time period are imported into the model, and the same pre-processing and feature extraction steps are performed on it as the training data. Finally, we will identify and merge the time Windows of continuous or near-continuous predictions as precursor features to form continuous time intervals.

Work three: For the development of real-time early warning system in work three, a real-time early warning system is established when the precursor characteristic signal first appears to predict the potential danger of rock burst. Using historical and real-time data, an ARIMA model is built and its stationarity is tested. The data is then differentially processed, the appropriate ARIMA parameters are selected, and predictions are made for the next 100 time points. The proportion of cases where the standard deviation is less than 10% of the average forecast is calculated, that is, the probability that the forecast result has lower uncertainty. If the proportion is found to be 78%, the prediction is considered reasonable and similar methods are applied to other data. Finally, the probability of precursor signal appearing in the sampling period is obtained.

2.2 Model assumption

In order to facilitate the construction of the model and ensure its feasibility, we propose the following assumptions to make the model more complete and the prediction results more reasonable:

(1) Assume that the data provided is true and reliable and has real validity;

(2) It is assumed that the occurrence of some abnormal data is reasonable and does not affect the accuracy of the overall model;

(3) Assume that our handling of missing values in the data will not cause significant errors in the predicted results.

3 MODEL BUILDING AND ANALYZING

3.1 Model establishing and solving for work one

3.1.1 Eigenvalue extraction

1. Data preprocessing

(1) Data reading: According to the data in Annex I and Annex II, the data of electromagnetic radiation and acoustic emission signals are read respectively, and the given time is converted into a standard date format for subsequent data analysis.

(2) Data screening: According to the classification of electromagnetic radiation and acoustic emission data in the title, the data is divided into normal data, interference data and other types of data.

(3) Data cleaning: First of all, the find function in Matlab is used to find the missing value of the given data set, and there is no missing value in the attachment given by the title. By observing the electromagnetic radiation data in Annex I, it is found that two data exist at the same time, as shown in Table 1.

Table 1 Outliers

EMR	Time	Class
29	2019-01-09 05:51:41	D/E
30.01	2019-01-09 05:51:41	D/E

The occurrence may be caused by the detector sending out multiple detection signals at the same time. Because the repeated data will interfere with the recognition of signal characteristics, 39277 duplicate values are found and deleted in the electromagnetic radiation data, 571,655 effective values are retained, and 39277 duplicate values are found and deleted in the acoustic emission data [7]. 571655 valid values are reserved.

The distribution diagram of AE and EMR on time series was obtained through data visualization, as shown in Fig.1.

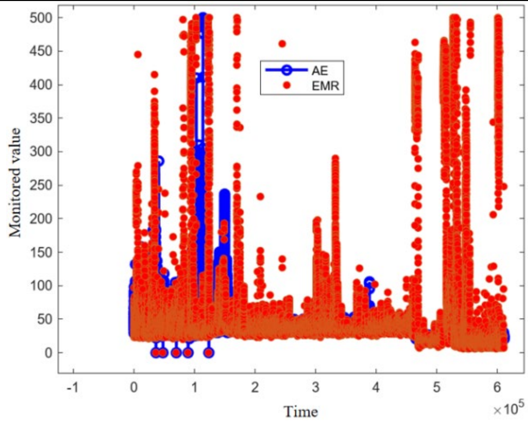


Fig.1 AE and EMR timing charts

As can be seen from Fig.1, AE and EMR fluctuate significantly with time transformation, which is assumed to be caused by interference signals. Further exploration of interference signals is carried out.

Class D and E are essentially invalid values because they are sensor wire break data and working face rest data. Therefore, class D and E data are deleted to obtain other values in order to obtain better visualization results.

2. Interference signal analysis and eigenvalue extraction

By analyzing the continuity of each interference signal, all signals are divided into 28 continuous interference signals and visualized.

Wavelet analysis: Wavelet analysis is a widely used technique in signal processing, especially for analyzing non-stationary signals and signals with time-frequency local characteristics [8]. In the analysis of electromagnetic radiation and acoustic emission signals, wavelet analysis can effectively extract the useful information and the characteristics of interference signals. The principle is as follows:

(1) Wavelet transform: the signal is decomposed into coefficients of different levels by wavelet basis function $\psi(t)$ and scale function $\phi(t)$.

$$s(t) = \sum_{j}\sum_{k} \alpha_{j,k}\, \phi_{j,k}(t) + \sum_{j}\sum_{k} d_{j,k}\, \psi_{j,k}(t) \tag{1}$$

Where, $a_{j,k}$ is the scale coefficient, $d_{j,k}$ is the wavelet coefficient, $\phi_{j,k}(t)$ and $\psi_{j,k}(t)$ is the scale function and the wavelet basis function respectively.

(1) Discrete wavelet transform: the signal is decomposed into scale coefficient and wavelet coefficient for the analysis of different frequency levels. The scale analysis represents the low frequency part of the signal, and the wavelet coefficient represents the high frequency part of the signal [9]. For the wavelet coefficient $d_{j,k}$ of the first stage, its calculation formula is as follows:

$$d_{j,k} = \int s(t)\psi_{j,k}(t)d \tag{2}$$

(1) Wavelet energy: the energy of each wavelet coefficient can be obtained by squaring the coefficient and summing:

$$E_j = \sum_{k} d_{j,k}^2 \tag{3}$$

This represents the energy of the signal at the J-order frequency level.

Comparison by wavelet analysis: The CWT graph of the normal signal and the signal that may contain interference is compared and analyzed. By comparing the differences between the two, it is easier to identify the characteristics of the interference signal for further analysis and processing [10]. Time domain feature extraction: Take electromagnetic radiation data as an example to extract the time domain feature of the above 28 interference signals, and extract 13 features including average value, standard deviation, skewness, kurtosis, maximum value, minimum value, peak-to-peak value, root mean square value, amplitude factor, waveform factor, shock factor, margin factor and energy. The results are shown in Table 2.

Table 2 Feature extraction results

Serial number	Mean value	Standard deviation	Skewness	Kurtosis	Maximum value	Minimum value
1	113.21	5.99	0.28	2.42	125.00	103.00
2	53.71	49.19	3.75	21.77	445.00	25.00
3	154.83	118.48	0.33	1.54	415.00	25.00
4	44.96	32.73	6.09	47.52	343.00	0.00
5	39.24	22.33	7.09	57.88	244.00	24.00

The analysis of Table 2 shows that the characteristic data signals show significant fluctuations at different time points. There are periods when the amplitude of the signal increases significantly, which may indicate some unusual activity or external interference.

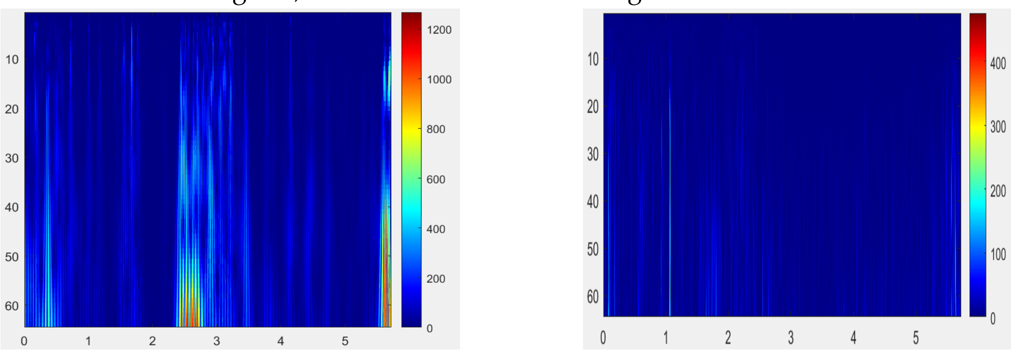
Frequency domain feature extraction: Fourier analysis was performed on the above 28 interference signals, and the spectrum diagram was obtained.

Taking electromagnetic radiation data as an example, its characteristics were extracted and quantified, and the results were shown in Table 3.

Table 3 Quantization of characteristic values of electromagnetic radiation

Item	1	2	3	4	5
MedianFreq	0.24	0.24	0.24	0.24	0.24
PeakFreq	0.00	0.00	0.00	0.00	0.00
Width	0.49	0.49	0.49	0.49	0.49

Feature extraction in time-frequency domain: Continuous wavelet transform (CWT) graph recognition is conducive to the identification of interference signals. Using continuous wavelet transform, the characteristic values of the above 28 interference signals are compared with those of normal signals, and the results shown in Fig.2 are obtained.



(a) Interference signal (b) Normal signal

Fig.2 Interference signal and normal signal

Comparative analysis, the normal signal and the signal that may contain interference are compared and analyzed. By comparing the differences between the two, the characteristics of the interference signal can be more easily identified. We choose the following four characteristic values.

(1) Frequency distribution: Observe the distribution of wavelet coefficients at different scales (frequencies). Interference signals usually produce prominent wavelet coefficients in a specific frequency range.

(2) Energy distribution: Pay attention to the energy distribution of the wavelet coefficient, especially in the high energy region. Interference signals tend to exhibit higher energy concentrations in certain frequency ranges.

(3) time-frequency concentration: Observe the time-frequency concentration of the wavelet coefficient. The interference signal usually has a strong time-frequency concentration, that is, it has a large wavelet coefficient value in some time periods and a specific frequency range.

(4) Change trend: The change trend of wavelet coefficient with time and frequency is analyzed. Interference signals may exhibit trends different from normal signals, such as frequency hopping, energy concentration in a specific frequency range, and so on.

3.1.2 Establishment and solution of SVM interference signal recognition model

1. Non-interference signal analysis

First of all, data processing and analysis of the non-interference signal are carried out, and its preliminary visualization is shown in Fig.3.

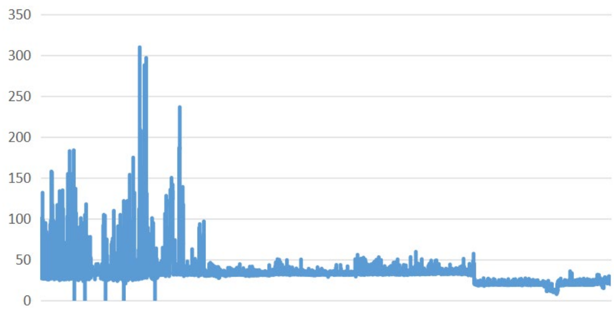


Fig.3 Visualization of interference signal

There are several sharp peaks in the diagram that may indicate some kind of emergency or sudden response of the device. Through the time domain feature extraction, as shown in Table 4.

Table 4 Time domain characteristics

Trait	Value	Trait	Value	Trait	Value
Mean number	41.1149	Maximum value	500	Crest factor	7.6392
Standard deviation	50.9265	Minimum value	0	Form factor	1.5919
skewness	6.5985	Peak value	500	Pulse factor	12.1610
Kurtosis coefficient	49.1517	root-mean-square	65.4518	Margin factor	0.2958

The frequency domain features are extracted, as shown in Table 5.

Table 5 Frequency domain characteristics

Trait	Maximum equal frequency	Maximum frequency	Gross power
Value	2.3293e+07	2.4270e+09	1.3749e+15

Feature extraction is fundamental to ensuring the effectiveness of data science and machine learning models. In this case, through amplitude, noise level, energy, and frequency analysis, we were able to construct a feature set that effectively reflects the characteristics of the signal, which are critical for training the classification model. Through these features, the model can learn how to distinguish between normal signals and precursor characteristic signals, thereby providing early warning in practical applications and reducing potential risks.

2. Establish interference recognition model

Through the analysis of the above eigenvalue extraction, it is decided to use SVM, neural network and random forest three different models for comprehensive comparison, to establish a more comprehensive interference recognition model.

(1) The training process of SVM model includes selecting the appropriate kernel function and adjusting the relevant parameters. By maximizing the interval to determine the hyperplane and classifying it according to the support vector, the SVM can be used to make predictions on new samples after the training is complete.

$$\max \sum_{i=1}^{n} \alpha_i - \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} \alpha_i \alpha_j y_i y_j \left(\Phi^T(x_i)\Phi(x_j)\right) \tag{4}$$

$$\sum_{i=1}^{n} \alpha_i, y_i = 0 \tag{5}$$

$$\alpha_i \geq 0, i = 1,2,\ldots,n \tag{6}$$

The first is to find the maximum value of $\min_{w,b} L(w,b,\alpha)$, which is:

$$\min_\alpha \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} \alpha_i \alpha_j y_i y_j \left(\Phi^T(x_i)\Phi(x_j)\right) - \sum_{i=1}^{n} \alpha_i \tag{7}$$

$$\sum_{i=1}^{n} \alpha_i, y_i = 0 \tag{8}$$

$$\alpha_i \geq 0, i = 1,2,\ldots,n \tag{9}$$

Calculate the extreme value of the above formula to find $\alpha^{*\prime}$, and bring α^* in equation, calculate w and b:

$$w^* = \sum_{i=1}^{n} \alpha_i^* y_i \Phi(x_i) \tag{10}$$

$$b^* = y_i - \sum_{i=1}^{n} a_i^* y_i \left(\Phi(x_i) \bullet \Phi(x_j)\right) \tag{11}$$

Find the hyperplane:

$$w^* \Phi(x) + b^* = 0 \tag{12}$$

To find the classification decision function:

$$f(x) = sign(w^* \Phi(x) + b^*) \tag{13}$$

(2) A neural network model is a machine learning model inspired by the biological nervous system, which consists of multiple artificial neurons (nodes) that are connected to each other by connecting edges (weights) and can be used to solve various tasks such as classification, regression, image recognition, speech processing, etc.

(3) Random Forest model is an ensemble learning method that performs prediction and

classification tasks by combining multiple decision tree models, each of which is trained on a randomly selected subset of features.

Anti-overfitting: By combining multiple decision trees to make integrated predictions, random forests reduce the risk of overfitting models. The comparison of BP neural network training set and test set is shown in Fig.4.

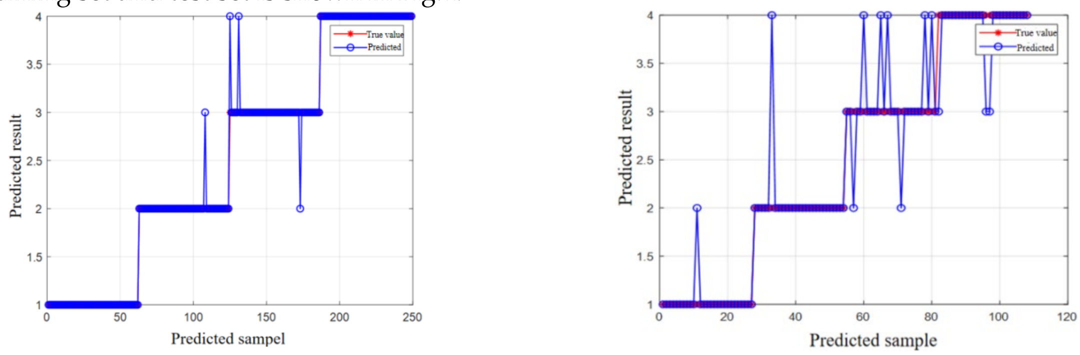


Fig.4 BP neural network analysis diagram

The accuracy results verified by SVM model training set and test set are shown in Fig.5.

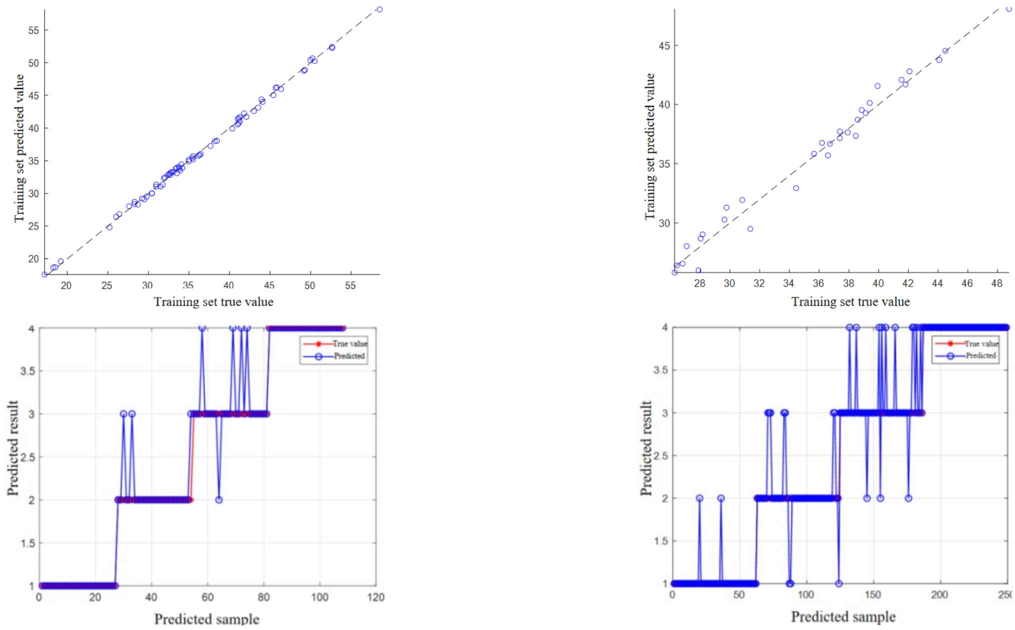


Fig.5 SVM model analysis diagram

The comparison results of eigenvalue training set and test set obtained by random forest model are shown in Fig.6.

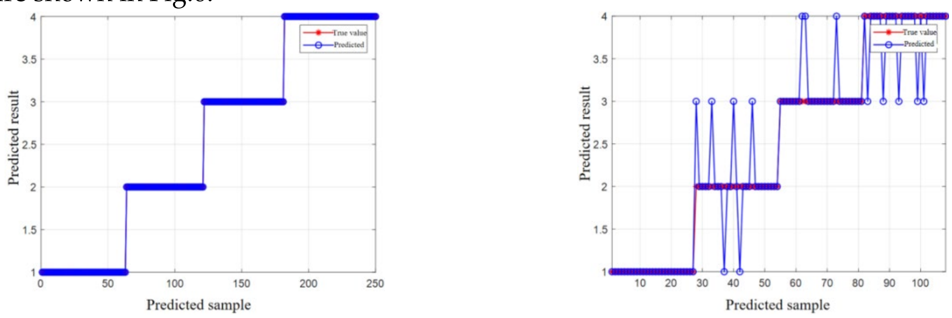


Fig.6 Random forest analysis diagram

According to the analysis in Fig.6, 7 and 8, the fit degree of training data using BP neural network is 90.39%, the fit degree of SVM model training data is 92.59%, and the fit degree of random forest training data is 90.5%. Compared with the SVM model, the highest fit degree of data is the most perfect prediction result, so the SVM model is selected for fault identification.

3. Solve the interference recognition model

The SVM model was used to identify the electromagnetic radiation from May 1, 2022 to May 30, 2022 and the time intervalwhere the interference signals in acoustic emission signals from April 1, 2022 to May 30, 2022 and from October 10, 2022 to November 10, 2022. The results are shown in Fig.7.

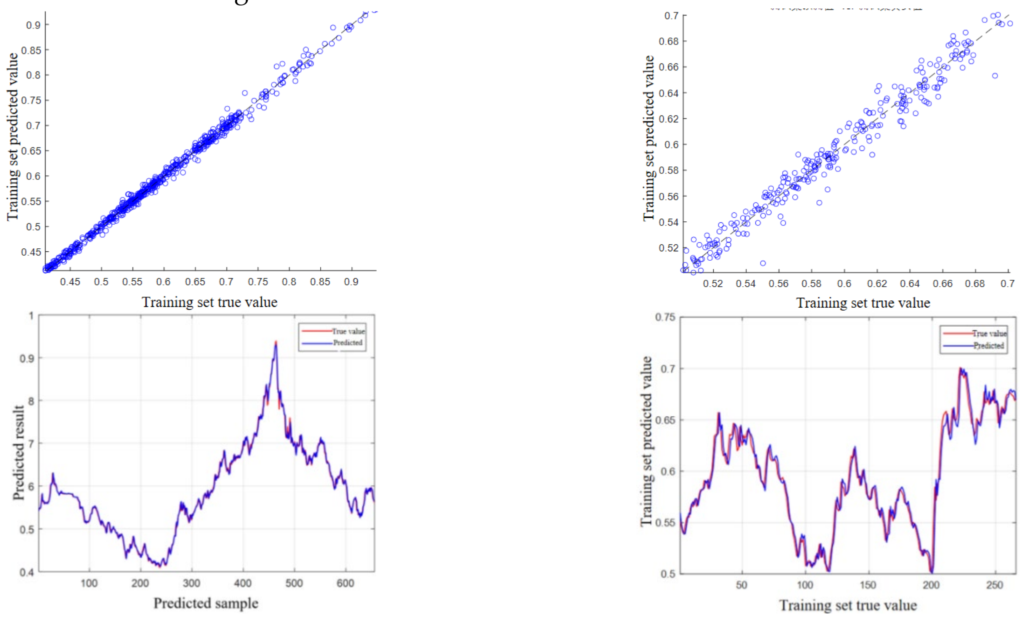


Fig.7 SVM interference recognition model solution diagram

Table 6 Time interval of electromagnetic radiation interference signal

Number	1	2	3	4	5
Start of time interval	2022/5/11 0:57:50	2022/5/11 0:57:50	2022/5/11 0:57:50	2022/5/11 0:57:50	2022/5/11 0:57:50
End of time interval	2022/5/1 12:18:19	2022/5/1 12:18:19	2022/5/1 12:18:19	2022/5/1 12:18:19	2022/5/1 12:18:19

Table 7 Time interval of acoustic emission interference signal

Number	1	2	3	4	5
Start of time interval	2022/4/11 10:33:36	2022/4/11 10:33:36	2022/4/11 10:33:36	2022/4/11 10:33:36	2022/4/11 10:33:36
End of time interval	2022/4/11 15:18:16	2022/4/11 15:18:16	2022/4/11 15:18:16	2022/4/11 15:18:16	2022/4/11 15:18:16

3.2 Model establishing and solving for work two

3.2.1 Eigenvalue extraction

1. Data preprocessing

In order to analyze precursor characteristic signals in electromagnetic radiation and acoustic emission signals, data visualization of precursor signal information is performed, as shown in Fig.8. It can be seen from Fig.10 that the precursory signals have the characteristics of aggregation, and at the same time, invalid values are deleted. Visualization results of the remaining data are shown in Fig.9.

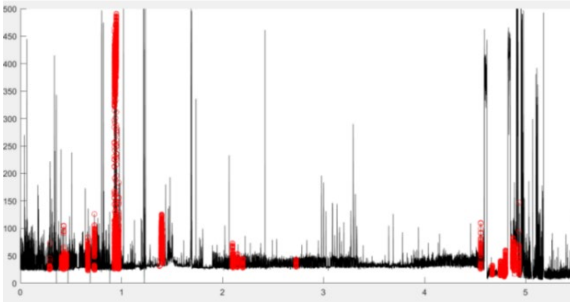


Fig.8 Visualization of precursor signals

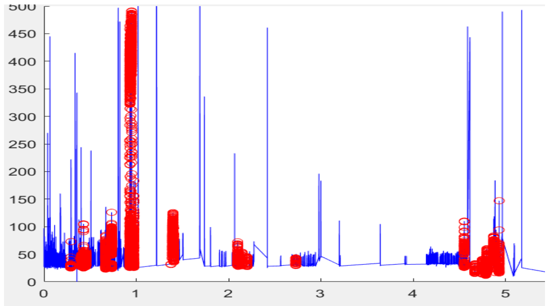


Fig.9 Invalid value culling visualization

Through the observation of Fig.11, the fluctuation and aggregation characteristics of the precursor signal after the rise or fall of the value can be obtained.

2. Statistical feature extraction

The random forest model was used to extract its statistical characteristics. A random forest is an ensemble learning method that builds multiple decision trees for classification or regression prediction. Its algorithm works as follows:

(1) Random sampling: A certain number of samples are extracted from the original training set with a place to form a new training subset. This process is called bootstrapping. At the same time, for each decision tree, some features are randomly selected for training.

(2) Decision tree construction: Based on the training subset obtained from the above sampling and the randomly selected features, multiple decision trees are constructed.

Random sampling process:

$$D_i = \{(x_1, y_1), (x_2, y_2), \ldots, (x_N, y_N)\} \tag{14}$$

Where, D*i* represents the training subset of the i^th^ decision tree, and (x_j, y_j) represents the features and labels of the j^th^ sample.

Feature random selection:

$$F_k = \{f_{k1}, f_{k2}, \ldots, f_{km}\} \tag{15}$$

Where, F_k represents the k^th^ feature selected by the decision tree.

Construction of decision tree:

$$T_i = (D_i, F_k) \tag{16}$$

Where, T_i represents the i^th^ decision tree model.

Prediction results for classification works:

$$\hat{y} = \left(\sum_{i=1}^{M} (T_i(x) = c) \right) \tag{17}$$

Prediction results for regression work:

$$\hat{y} = \frac{1}{M} \sum_{i=1}^{M} T_i\,(x) \tag{18}$$

Where, $\hat{y}$ represents the final prediction result, M is the number of decision trees in the random forest, T_t(x) represents the classification result of the i^th^ decision tree for sample x, and represents the category.

The importance of each feature in the decision tree and the difference between the predicted value and the real value of the training set and the test set can be obtained by the feature extraction of the random forest model. As shown in Fig.10.

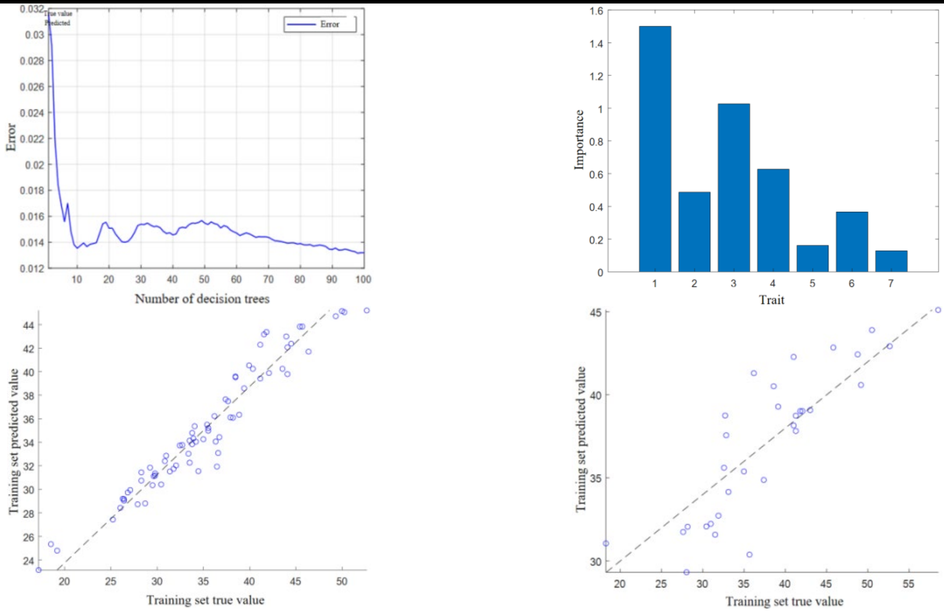


Fig.10 Feature value extraction diagram of random forest model

Six time-domain statistical indicators were extracted by random forest model, as shown in Table 8.

Table 8 Time domain statistical indicators

Item	Mean	Std	Skewness	Kurtosis	Max	Min
1	28.31	1.6497	0.074326	2.5132	32	25
2	47	18.421	0.75812	2.0955	73	30
3	28	1.5323	-0.074074	1.5309	30	26
4	37.069	3.9254	0.0020856	3.1496	50	27
5	39.703	9.5034	2.5988	17.533	105	25

3. Trend feature extraction

Wavelet decomposition is applied to the sample data. Usually, wavelet decomposition can be used to analyze the different frequency components of the signal. Discrete wavelet transform (DWT) will be used.

Fig.11 shows the results of wavelet decomposition. On the left is the approximation coefficient, which reflects the main trend or smoothing part of the signal; On the right is the detail coefficient, which shows the rapidly changing or noisy part of the signal. Such a view can help to understand the characteristics of signals at different frequency levels.

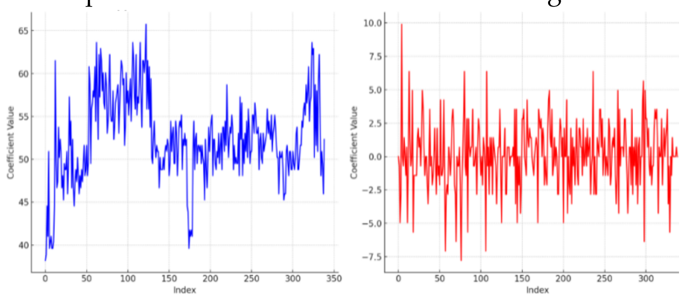


Fig.11 Wavelet transform

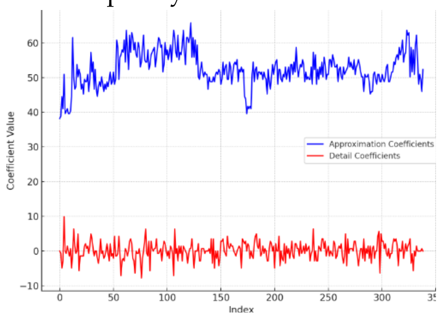


Fig.12. Results of wavelet transform

Wavelet transform result graph showing approximation coefficients (blue) and detail coefficients (red). These coefficients can show the low and high frequency characteristics of the signal. Through the analysis of the approximate coefficient, the main trend of the signal can be obtained. As can be seen from the figure, the data has certain volatility, but the overall trend is relatively stable, with no obvious long-term upward or downward trend. Fluctuation features show periodic changes in data, which may indicate the influence of external cyclical factors or internal data rules. This analytical method can be used for long-term trend analysis to help predict future movements or identify turning points in trends.

4, the establishment of smooth movement

Based on the analysis of the feature results in the time-frequency domain of the precursory features, the precursory signals are predicted by means of smooth movement. Trends can be observed by smoothing time series data. Common moving average methods include simple moving average (SMA) and weighted moving average (WMA). When performing a moving average analysis, adjustments need to be made based on the interval of data points. According to the information provided, the interval of each data point is 30 seconds, and the window size of the moving average is set to 10 in order to analyze the trend over a longer time period (since 5 minutes equals 300 seconds and each data point is spaced by 30 seconds, 300 seconds /30 seconds =10 data points). By adjusting the above moving average window, you can smoothly perform moving average analysis and observe the trend of the data. According to the average analysis, the data trend as shown in Fig..13 and Fig.14 can be obtained.

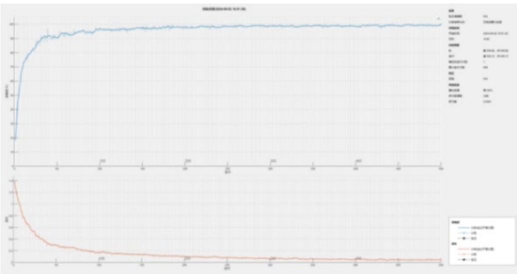


Fig.13 Smooth movement

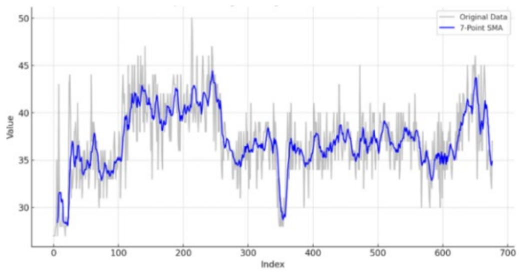


Fig.14 Data trends

3.2.2 Establishment and solution of smoothing window

The standard deviation is calculated within the selected window by smoothing the move, and a window with a smaller standard deviation indicates that the data is relatively smooth within this paragraph. The mean, standard deviation, skewness, kurtosis, maximum and minimum values calculated from the smoothing window. For example, from April 8, 2020 to June 8, 2020, the results of the smoothing window are shown in Fig.15.

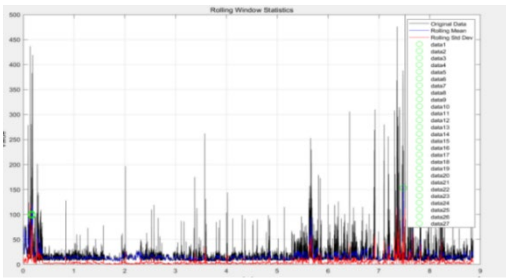


Fig.15 Smoothing window result diagram

By comparing the calculated value in the smoothing window with the time-frequency domain trend characteristics and statistical characteristics of the time where the electromagnetic radiation and acoustic emission precursory features are located, the occurrence time segment of the electromagnetic radiation and acoustic emission precursory features is obtained, and the results are shown in Table 9 and Table 10.

Table 9 Time interval of electromagnetic radiation precursor characteristics

Number	1	2	3	4	5
Start of time interval	2020-4-8 22:10:37	2020-4-8 22:10:37	2020-4-8 22:10:37	2020-4-8 22:10:37	2020-4-8 22:10:37
Start of time interval	2020-4-9 19:12:38	2020-4-9 19:12:38	2020-4-9 19:12:38	2020-4-9 19:12:38	2020-4-9 19:12:38

Table 10 Time interval of acoustic emission precursor characteristics

Number	1	2	3	4	5
Start of time interval	2021-11-1 0:01:01	2021-11-1 0:01:01	2021-11-1 0:01:01	2021-11-1 0:01:01	2021-11-1 0:01:01
End of time interval	2021-11-11 3:58:42	2021-11-11 3:58:42	2021-11-11 3:58:42	2021-11-11 3:58:42	2021-11-11 3:58:42

In order to explore the accuracy of the classification results, the accuracy and accuracy of the random forest classification model were calculated. Among them, the accuracy and accuracy of the electromagnetic radiation precursor feature classification model were 0.82 and 0.85. The accuracy of the classification model of acoustic emission precursor features is 0.94, and the accuracy is 0.95. The model has a good effect.

3.3 Model establishing and solving for work three

3.3.1 The establishment of ARIMA model

ARIMA model is the combination of difference operation and ARMA model, denoted as ARIMA(p,d,q), in ARIMA(p,d,q), AR is autoregressive, p is the number of autoregressive terms; MA is the moving average, q is the number of terms of the moving average, and d is the number of differences (orders) made to make it a stationary sequence. It can simulate the autoregressive part, differential stability and moving average part of time series. The ARIMA model can be expressed as:

$$\psi(B) \quad (1-B)^{\mathrm{d}} y_t = \theta(B)\varepsilon_t \tag{19}$$

Where, y_t is the time series of the historical observed values, d is the order of the difference, and ε_t is an independent uniformly distributed white noise sequence with zero mean and constant variance, and B is the lag factor. Satisfy the following expression:

$$\begin{aligned} \psi(B) &= 1 - \psi_1 B - \cdots - \psi_p B^p \\ \theta(\mathrm{B}) &= 1 - \theta_1 B - \cdots - \theta_{\mathrm{q}} B^{\mathrm{q}} \end{aligned} \tag{20}$$

$$BIC = Accuracy\ (m) - \frac{p}{2}\log N \tag{21}$$

The key point of establishing ARIMA(p, d, q) model is the selection of three parameters (p, d, q), where (BIC) is selected to select p and q. Bayesian information criteria can provide a simple approximation of the logarithmic model evidence, as shown below, where, P is the number of parameters and N is the number of points. Model parameter tuning, using ACF and PACF graphs to estimate parameters of ARIMA models. Use automated model selection methods such as PMDARMIa. auto_arima to automatically find the optimal parameters. Forecast and analysis: Execute the forecast for a certain time period in the future, analyze the trend and standard deviation of the forecast results, and determine the time period with small trend growth and standard deviation in the forecast results. Calculate the probability, determine the threshold of the standard deviation, and calculate the proportion of predicted outcomes within this standard deviation range. If you need to calculate the probability of a growth trend, you can compare the predicted values at successive time points to determine whether there is a growth trend.

3.3.2 Model solving

Through the ADF unit root detection method for the electromagnetic radiation time series, the ADF null hypothesis is that the series has a unit root, and the p-value we get is 0.1859, which is higher than the usual significance level (usually 5% or 1%). Therefore, the null hypothesis cannot be rejected, that is, the series has a unit root, then the series is considered non-stationary. In order to make the sequence stationary, we perform a difference processing and run the stationarity test again. The results show that the p-value of the differential sequence is very small (close to 0), well below any commonly used significance level, indicating that the differential sequence is stationary. Now, let's start fitting an ARIMA model. Now that we know that the difference order is 1, we need to determine the order of the autoregressive term (p) and the order of the moving average term (q). We will use automated methods to select the appropriate p and q, as shown in Fig.16.

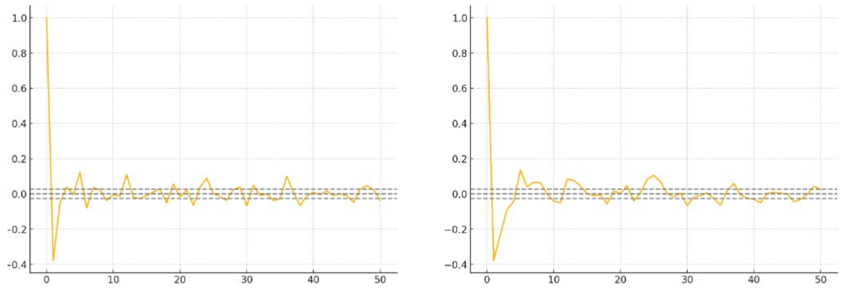


Fig.16. Autocorrelation function

Analyzing the autocorrelation (ACF) and partial autocorrelation (PACF) graphs can help us estimate the parameters of the ARIMA model: The PACF graphs show truncation after a delay of 1, suggesting that the autoregressive term p might be appropriate to set to 1. The ACF plot decays gradually, indicating that the moving average term q may require more lag. In this case, we can consider setting q to 1. Next, the ARIMA(1,1,1) model is used to make predictions on the data. Then, the standard deviation of the predicted result is analyzed and its probability is estimated within some smaller range.

When analyzing the electromagnetic radiation data, the stationarity of the data is ensured first, and the differential processing is carried out to adapt to the requirements of ARIMA model. The data were then fitted using the ARIMA(1,1,1) model and predictions were made for the next 100 points in time. The proportion of the standard deviation less than 10% of the average predicted value, that is, the probability that the uncertainty of the forecast result is low, is calculated, and the proportion is 80%.

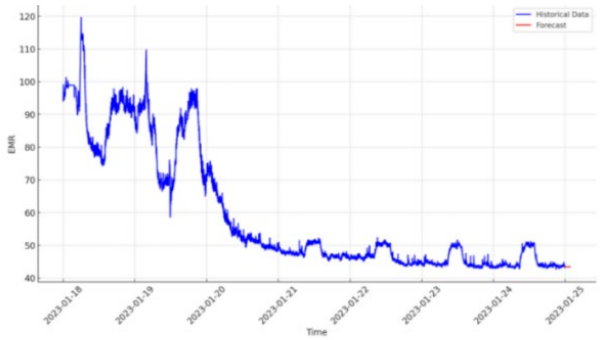


Fig.17 Trends in electromagnetic radiation

As shown in Fig.17, the trend of electromagnetic radiation has an upward trend, so the occurrence probability of the electromagnetic radiation precursor signal at the time shown in the ARIMA model is 78%, which is reasonable. For the probability of precursor signals appearing at other times, the optimal parameters can be searched by automatic model selection method, and the probability of precursor features appearing can be predicted by ARIMA model. The prediction results of the collected data are shown in Table 11 below:

Table 11 Probability of precursor features at the time of data collection

Electromagnetic radiation data	Probability of precursor features	Acoustic emission data	probability of precursor features
2023-1-24 23:58:36	80%	2023-1-24 23:58:36	73%
2023-2-11 23:59:20	51%	2023-2-11 23:59:20	77%
2023-2-26 23:59:27	67%	2023-2-26 23:59:27	69%
2023-3-10 23:58:14	89%	2023-3-10 23:58:14	88%
2023-3-30 23:58:13	75%	2023-3-30 23:58:13	58%

4 CONCLUSION

Comprehensiveness: The comparison of random forest, BP neural network and VCM model is established to improve the accuracy and reliability of prediction, which is crucial for improving the reliability of coal mine safety monitoring system. One of the advantages of the VCM model is that it can capture nonlinear relationships and interactions in the data, so it performs well when dealing with data with complex relationships. In addition, the Vcm model can provide an intuitive understanding of changes in the relationship between variables, can take into account the influence of geological conditions, mining processes, land cover and other factors on the ground pressure risk, and can flexibly describe how these factors change over time and space. It is helpful to analyze the mechanism and law of ground pressure danger.

Multi-feature engineering: For the prediction of rock burst risk in deep coal mining, wavelet analysis technology is used to obtain time-frequency domain information of electromagnetic radiation and acoustic emission signals, so as to capture local characteristics and abrupt changes of signals more accurately. In the process of model building, a variety of feature extraction techniques are used, including amplitude difference, noise level, duration and frequency features. These features are extracted from the actual signal and help capture complex patterns and changes in the signal. The ARIMA model is used for risk prediction. The ARIMA model is applicable to various types of time series data, including ground pressure monitoring data generated during deep mining of coal mines. The parameters of ARIMA model have clear statistical significance and can be used to explain the trend and dynamic changes of ground pressure data. This helps mine engineers and decision makers understand the mechanisms and laws behind ground pressure data, providing a deeper understanding of coal mining safety.

5 ACKNOWLEDGEMENTS

This work was supported by ministry of education industry-university cooperative education project (Grant No.: 231106441092432) and special research project on teaching reform(Grant No.: 30120300100-23-yb-jgkt03)

REFERENCES

[1] Wang, L. N., & Chen, L. (2017). Research on the application of support vector machine (SVM) combined with object-oriented method in information extraction of open pit mining area. *The 19th Academic Exchange Meeting of six Provinces and one City of Surveying and Mapping Society in East China and the 2017 Cross-Straits Surveying and mapping Technology Exchange and Academic Seminar*, 201-206.

[2] Hu, H. B., & Zhan, Y. L. (2018). Change characteristics of land use landscape pattern based on decision tree classification of remote sensing image. *Green Technology*, 24(072): 200-205.

[3] Wang, D. Y. (2021). Research on data reconstruction method of ground penetrating radar based on time series analysis. *Shandong Technology and Business University*.

[4] Wu, J. F. (2021). Research on anomaly detection and fault warning technology for in orbit satellites based on LSTM. *National University of Defense Technology*.

[5] Zhou,G. Y., Wan, S. P., & Chen.Y. N. (2022). Research on denoising algorithm for phase sensitive time domain reflectometer based on moving variance average algorithm. *Journal of Instruments and Meters*, 43(10): 233-240.

[6] Zhang, J. L., Guo, S. Y., & Ren, C. P. (2024). Research on personal credit rating card model based on logistic regression. *Modern Information Technology*, 8(05): 12-16.

[7] Du, B. Y., Gao, J. H., & Zhang, G. Z. (2024). Seismic prediction method and application of fracture density inversion for shale reservoir in-situ stress. *Petroleum Geophysical Exploration*, 59(2): 279-289.

[8] Zhou, Z. X., Zhen, X. J., & Liang, Y. G. (2024). Study on acoustic emission source location of ancient timber based on wavelet packet transform and cross-correlation. *Shanxi Architecture*, 50(9): 1-5.

[9] Yao, J. W., Li, Y., & Lv, Y. J. (2024). Research on the trend of water quality evolution in drinking water source areas based on water quality index method and M-K Test. *Environmental Science and Management*, 49(04): 43-48.

[10] Nadir, B., Azzeddine, D., & Ahmed, B. (2024). Exploring the effects of overvoltage unbalances on three phase induction motors: Insights from motor current spectral analysis and discrete wavelet transform energy assessment. *Computers and Electrical*

Engineering, 117109242.

[11] Zhang, X. G., & Luo R. (2024). Prediction and analysis of TCM constitution based on ARIMA time series model. *Asia-pacific traditional medicine*, 20(04): 156-16.