ingegneria sismica
International Journal of Earthquake Engineering
Anno XLIII - Num. 2 - 2026

Research on multimodal fault recognition of main equipment of power grid supported by smart hub fused with acoustic features

Yong Qu[1], *Gewu Ren*[1], *Changbao Su*[1,*] *and Nan Yao*[1]

[1] Nanyang Power Supply Company, State Grid Henan Electric Power Company, Nanyang, Henan, 473000, China

SUMMARY: *The power distribution system in coal mining areas is not only one of the infrastructure supporting coal mining production and operation, but also an indispensable key link in the entire energy industry chain. And its stable operation is crucial for the production safety and efficiency of coal mining equipment. The operating conditions of the power distribution system of coal mining equipment are complex, and the current detection scheme of the power distribution system does not extract the voiceprint features of the power distribution system, so the accuracy of fault detection is low. A multimodal fault identification method based on voiceprint features has been proposed to address this issue. This method first designs a sound signal acquisition scheme to obtain raw data, and then uses an improved noise filtering method to preprocess the collected data and extract the main features. Finally, taking the main features of the voiceprint signal as input, a fault recognition model integrating voiceprint features was constructed by combining the improved Pelican optimization algorithm and convolutional neural network. The experimental results show that the accuracy of the training and testing sets of the recognition model constructed in the study is 95.63% and 96.45%, respectively. The above data indicates that the proposed method has significant effects on improving the accuracy and robustness of fault identification in energy mining equipment and power distribution systems. It also provides effective technical support for similar power distribution systems, which is of great significance for promoting the sustainable development of the energy industry.*

KEYWORDS: *Voiceprint features, Main equipment of the power grid, Multimodal faults, Noise filtering, Sound signal, Pelican optimization algorithm, Convolutional neural network.*

1 Introduction

As an important component of the energy industry, coal mining cannot improve its production efficiency and energy utilization efficiency without stable and reliable power support. The health status of the power distribution system is not only directly related to the stability and reliability of coal mining equipment, but also the cornerstone of the stable operation of the entire energy supply chain. However, due to the complex grid environment, diverse equipment types and uncertain working conditions, fault identification of grid main equipment faces huge challenges [1, 2]. First, the collected voiceprint signal has many impurities, which makes the recognition model unable to effectively identify the characteristic information of the fault voiceprint [3]. Secondly, in the energy industry, there are significant differences in energy mining equipment, power distribution systems, and operating environments among different coal mining areas, which makes it difficult to unify the parameters of traditional identification

*15737152799@163.com
https://doi.org/10.65102/is2026658
Published: 30 April, 2026

models, thereby affecting the applicability and robustness of the models [4]. In addition, the environment in which the equipment is located is also different, and the signal acquisition method is also a major problem [5]. At present, in order to solve the problem of poor voiceprint signal quality, signal processing technologies such as adaptive filtering and wavelet transform can be considered to pre-process the collected sound data, remove interference noise, and improve the clarity and accuracy of the signal [6]. In order to solve the problem of inconsistent model parameters, advanced technologies such as deep learning can be introduced to build an adaptive fault recognition model through big data training to improve the robustness and applicability of the model, while optimizing the model parameters and improving the operation efficiency [7]. According to the signal collection method in different environments, voiceprint collection schemes suitable for different working conditions can be selected, and verified and optimized in combination with actual scenarios to ensure that the collected sound data has high reliability and operability [8]. In view of this, a multimodal fault model for power distribution systems based on voiceprint features has been proposed. It is expected that this research can improve the accuracy and efficiency of fault diagnosis for energy mining equipment power distribution systems in coal mining areas, and provide strong support for the sustainable development of the entire energy industry. One of the innovations of the research is to use the whale optimization algorithm to optimize the parameters of the variational mode decomposition method and construct a novel voiceprint denoising method. One of the innovations of the research is to use the pelican optimization algorithm to optimize the convolutional neural network to improve the accuracy and robustness of the fault identification model.

2 Multi-modal fault identification of main power grid equipment by integrating voiceprint features

2.1 Voiceprint signal extraction and denoising method for main power grid equipment

The operational efficiency of the power distribution system directly affects the energy utilization efficiency of coal mining areas. Through sound signal analysis, equipment failures can be detected in a timely manner, reducing energy waste and improving energy utilization efficiency. The specific steps for fault detection of the power distribution system in coal mining areas are as follows: first, CVM-VP3 microphone is selected to collect the sound signal of the mining equipment power distribution system; Then use audio editing software to edit the collected sound signals of the power distribution system, and extract a stable 2-second sound signal; Then use MATLAB to organize and analyze this stable signal. The specific process is shown in Figure 1.

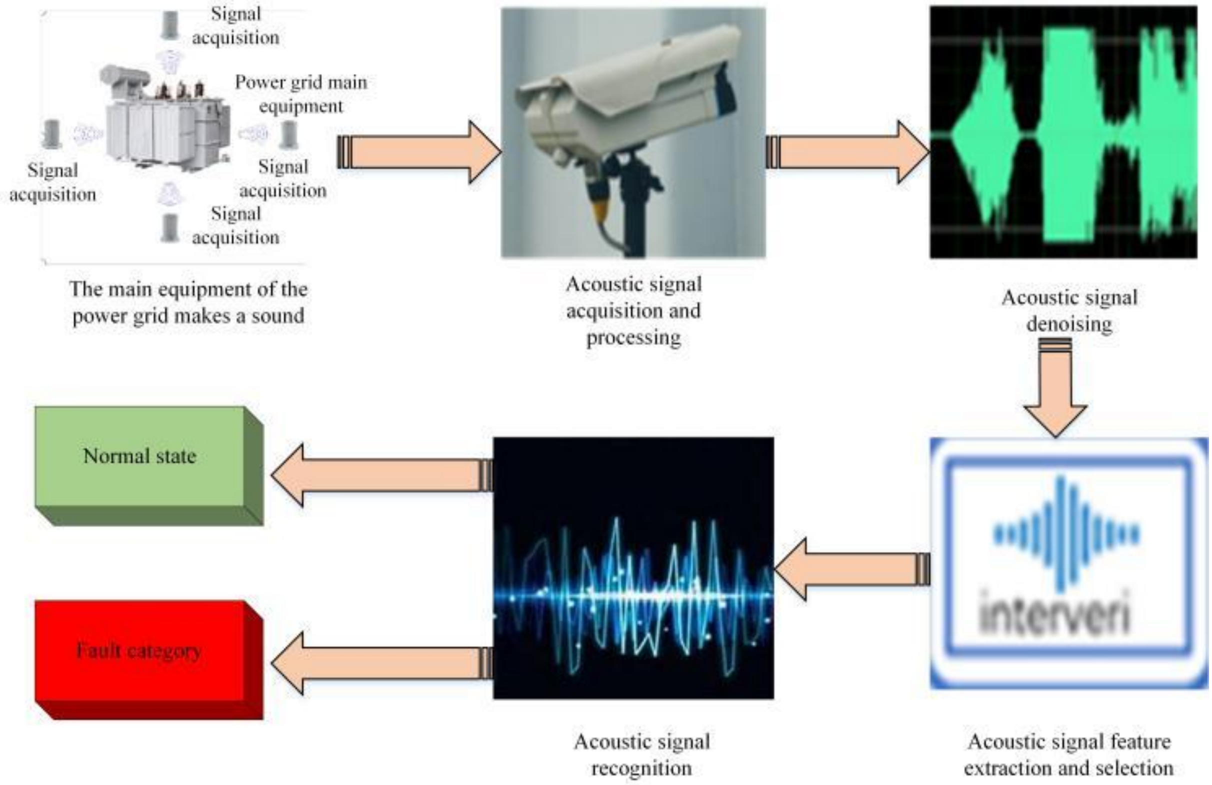


Figure 1: Grid main equipment signal acquisition and processing flow chart

As shown in Figure 1, acoustic signal acquisition devices are arranged around the main power grid equipment to be detected, with a distance of 40 cm, which are mainly responsible for collecting the acoustic signals of the main power grid equipment. The collected signals are then analyzed and processed in turn, and finally fault identification is performed. For acoustic signal denoising, the variational mode decomposition (VMD) method is used to decompose the original acoustic signal of the main power grid equipment into IMFs. The center frequency and frequency band range of each IMF are shown in formula (1) [9].

$$\begin{cases} \min\limits_{\{w_i\},\{\varphi_i\}} \sum\limits_i \left\| \varepsilon_t \left[\left(\delta(t) + j/\pi t\right) * w_t(t) e^{-j\varphi_i t} \right] \right\|_2^2 \\ s.t \sum\limits_{i=1}^{I} w_t = p \end{cases} \tag{1}$$

In formula (1), w_i is the modal component, φ_i is the center frequency, I is the number of modal components, t is the time, $*$ is the convolution operator, j is the imaginary unit, $\delta(t)$ is the impulse function, and p is the original waveform. Then, the Lagrangian multiplication operator is introduced to obtain the augmented Lagrangian expression as shown in formula (2) [10].

$$\begin{aligned} L(\{w_i\},\{\varphi_i\},\ \lambda) &= \alpha \sum_i \left\| \varepsilon_t \left[\left(\delta(t) + j/\pi t\right) * w_t(t) e^{-j\varphi_i t} \right] \right\|_2^2 \\ &+ \left\| p(t) - \sum_i w_i(t) \right\|_2^2 + \left[\lambda(t), p(t) - \sum_i w_i(t) \right] \end{aligned} \tag{2}$$

In formula (2), α is the penalty factor; $\lambda(t)$ is the Lagrange multiplier. Then, the alternating direction multiplier method and Pasaval theorem are used to solve each mode and center frequency. The result is shown in formula (3) [11].

$$\begin{cases} w_i^{n+1}(\omega) = \dfrac{p(\omega) - \sum_{j\neq i} w_i(\omega) + \lambda(\omega)/2}{1 + 2\alpha(\omega - \varphi_i)^2} \\ \varphi_i^{n+1} = \dfrac{\int_0^{\infty} \omega \left|\omega_i^{n+1}(\omega)^2\right| d\omega}{\int_0^{\infty} \left|\omega_i^{n+1}(\omega)^2\right| d\omega} \\ \lambda^{n+1}(\omega) = \lambda(\omega) + \vartheta\left(p(\omega) - \sum_i \omega_i^{n+1}(\omega)\right) \end{cases} \tag{3}$$

In formula (3), n is the number of iterations of the model; ϑ is the noise tolerance; $w_i^{n+1}(\omega)$, $\varphi_i^{n+1}(\omega)$, $p(\omega)$ and $\lambda^{n+1}(\omega)$ are the Fourier transforms of $w_i^{n+1}(t), \varphi_i^{n+1}(t)$, $p(t)$, and $\lambda^{n+1}(t)$, respectively. Then, the desired mode is inversely Fourier transformed to obtain the modal component IMF. The whale optimization algorithm (WOA) has a simple structure, few parameter settings, and strong optimization ability [12]. Therefore, the study uses the whale optimization algorithm to find the optimal solution for the parameters of VMD, and the optimization judgment index is the size of the envelope entropy value. A large envelope entropy value means that there are many types of noise in the IMF, and vice versa. The envelope entropy is shown in formula (4).

$$\begin{cases} E_p = -\sum_{i=1}^{S} A_i \lg A_i \\ A_i = B(i) / \sum_{i=1}^{S} B(i) \end{cases} \tag{4}$$

In formula (4), E_p is the envelope entropy; S is the number of samples; $B(i)$ is the signal after IMF demodulation; $A(i)$ is the probability distribution obtained after the standardization operation of $B(i)$, which is the envelope entropy value. The specific process of the whale optimization algorithm to optimize variational mode decomposition is shown in Figure 2 (a). The optimization process first initializes the parameters and population, and calculates the fitness of the individuals in the initial population. Subsequently, in each iteration, different strategies are selected to update the position of the whale individual according to the threshold of the current number of iterations until the preset maximum number of iterations is reached. During the update process, the optimal fitness and its corresponding parameter combination are retained. Finally, the whale individual with the optimal fitness and its corresponding fitness value are output, and this fitness value is the optimal parameter of VMD.

IThe study uses fast independent component analysis to convert the IMF obtained after VMD decomposition into signals. After the signal conversion, the fuzzy entropy value in information theory is used to identify the noise to be filtered. Suppose the main equipment of the power grid collects a noise signal of length M, and its time series is expressed as $f(i): 1 \leq i \leq M$. Sorting this noise signal to obtain a vector of dimension is shown in formula (5).

$$\begin{cases} X_i^n = \{ f(i), f(i+1), f(i+n-1) \} - f_0(i) \\ f_0(i) = \frac{1}{n} \sum_{j=0}^{n-1} x(i+j) \\ i = 1,2,3,\ldots,M-n+1 \end{cases} \tag{5}$$

Define the difference between two vectors X_i^n and X_j^n as x_{ij}^n, see equation (6).

$$x_{ij}^n = \left\| \left[f(i+k) - f_0(i) \right] \right\| - \left\| \left[f(j+k) - f_0(j) \right] \right\| \tag{6}$$

The definition of fuzzy membership function $A(x)$ is shown in formula (7).

$$A(x) \begin{cases} 1 \quad 0 \le x \le \lambda l \\ \exp\left[-(\ln 2)(x/l)^2 \right] \quad x > \lambda l \end{cases} \tag{7}$$

In formula (7), l is the similarity tolerance parameter; λ is the adjustment factor. Define the characteristic sound signal quality evaluation metric function $\varphi^n(l)$, see formula (8).

$$\varphi^n(l) = \frac{1}{M-n+1} * \sum_{i=1}^{M-n+1} A\left(\frac{1}{M-n} \right) \sum_{j=1,\, j \ne i}^{M-n+1} D_{ij}^n \tag{8}$$

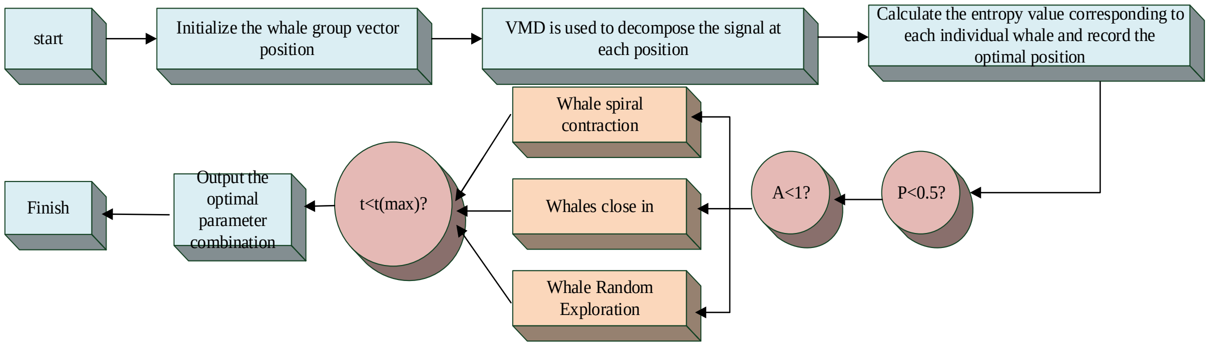
According to the above formula, the fuzzy entropy value $Fuzzy(n,l,M)$ can be obtained, see formula (9).

$$Fuzzy(n,l,M) = \ln \varphi^n(l) - \ln \varphi^{n+1}(l) \tag{9}$$

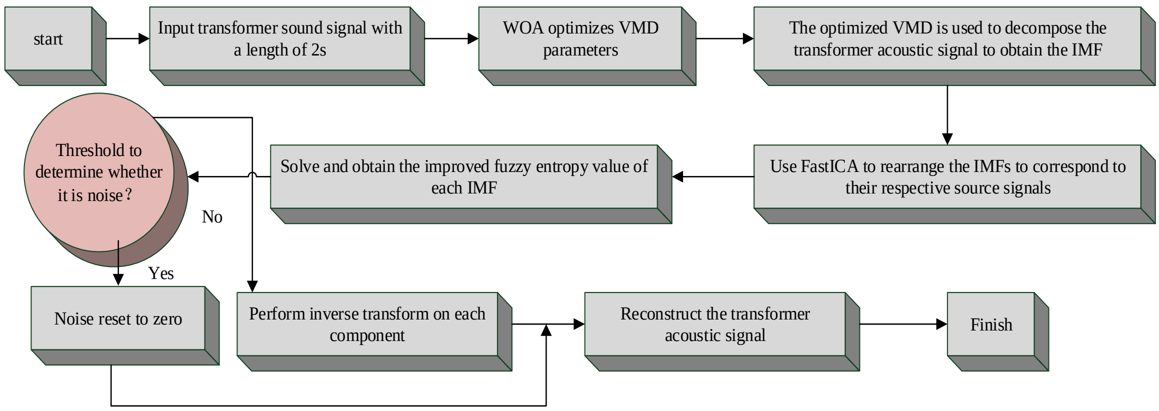
After calculating the fuzzy entropy values of each IMF, the fuzzy entropy values are discriminated using the discriminant formula to distinguish noise from normal power grid main equipment signals. The discriminant formula is shown in formula (10).

$$\begin{cases} 1 < h \le [M/2] \\ \xi(h+1) - \xi(h) < \xi(h) - \xi(h-1) \end{cases} \tag{10}$$

In formula (10), h is the variable; $\xi(h)$ is the fuzzy entropy value of the h th IMF. The IMF of the first fuzzy entropy value is defined as noise, and the other IMFs are reorganized by inverse transformation to obtain the main equipment signal of the power grid after noise removal. Based on the above analysis, the whole main equipment signal denoising process of the power grid is shown in Figure 2 (b).



(a) WOA optimization VMD process



(b) Acoustic signal noise filtering process

Figure 2: WOA optimization VMD process and acoustic signal noise filtering process

2.2 Voiceprint feature extraction and fault identification method for main power grid equipment

The study has extracted and denoised the original acoustic signal of the main equipment of the power grid, and then extracted the fault information features from the processed original acoustic signal as the input of the fault identification model of the main equipment of the power grid. The study uses a feature extraction algorithm based on Mel frequency cepstral coefficients to extract the voiceprint features of the original acoustic signal, and the results are shown in Figure 3.

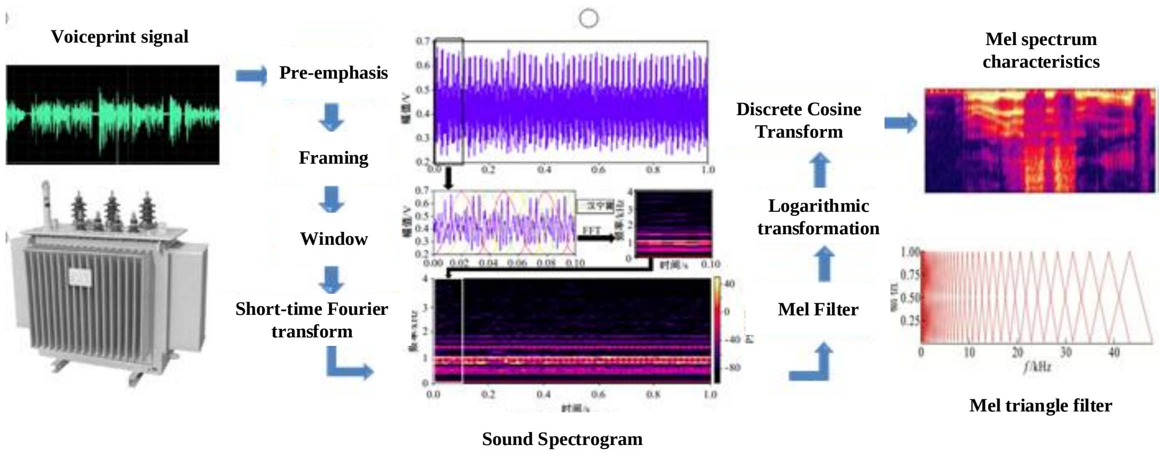


Figure 3: Voiceprint feature extraction based on Mel-frequency cepstral coefficients

As shown in Figure 3, the voiceprint signal is first pre-emphasized, and the voiceprint signal $b(n)$ obtained after the processing is shown in formula (11).

$$b(n)=a(n)-\gamma(n-1) \tag{11}$$

In formula (11), $a(n)$ is the voiceprint signal of the current frame; $a(n-1)$ is the voiceprint signal of the previous frame; γ is the system. Then the complete voiceprint signal is divided into countless frame sub-signals, and each segment of the signal is processed separately, that is, framing. Then, let the frame signal and the window signal be $q(n)$ and $u(n)$ respectively, and after the signal is windowed, the obtained signal $p(n)$ is shown in formula (12).

$$\begin{cases} p(n)=q(n)u(n), 0\le n\le N-1 \\ u(n)=\begin{cases} 0.54-0.46\cos\left(\dfrac{2n}{L-1}\right) & 0\le n\le L-1 \\ 0 & Others \end{cases} \end{cases} \tag{12}$$

Then, the acoustic signal of the main equipment of the power grid in each frame is subjected to short-time Fourier transform, and the spectral line energy $T(i,k)$ is shown in formula (13)

$$\begin{cases} Q(i,k)=FFT\left[q_i(n)\right] \\ T(i,k)=\left|Q(i,k)\right|^2 \end{cases} \tag{13}$$

The frequency of each frame is obtained, and then the frequency is filtered. The filtering method is shown in formula (14).

$$G(i,m)=\sum_{k=0}^{N-1} T(i,k)I_m(k), 0\le m\le M \tag{14}$$

In formula (14), m is the number of filters; $I_m(k)$ is the corresponding value of the filter in the frequency domain. When frequency filtering, take the logarithm of the filter energy, and then use discrete cosine transform to transform the frequency spectrum factor to the time domain, and then you can get the spectrum characteristics of the voiceprint signal $MFCC(i)$, see formula (15).

$$MFCC(i)=\sqrt{\frac{2}{m}}\sum_{M-1}^{m-1}\log\left[G(i,M)\right]\cos\left[\frac{\pi n(2M-1)}{2m}\right] \tag{15}$$

In formula (15), M is the M th filter. The calculated optimal fusion voiceprint feature is used as the input of the convolutional neural network model. CNN performs convolution operation on the voiceprint feature through the convolution kernel. Then the pooling layer downsamples the multiple feature maps obtained in the convolution layer. Finally, the fully connected layer integrates the fault information features of the first two layers for classification and output. However, the convolution kernel in the convolution layer of this neural network is a traditional convolution kernel, and its computational efficiency will become slow as the parameters increase. To this end, the study introduces the Pelican optimization algorithm to optimize the convolution kernel, and uses the Pelican optimization algorithm to continuously find the optimal convolution kernel, thereby improving the network performance. The fault voiceprint recognition model based on POA-CNN is shown in Figure 4.

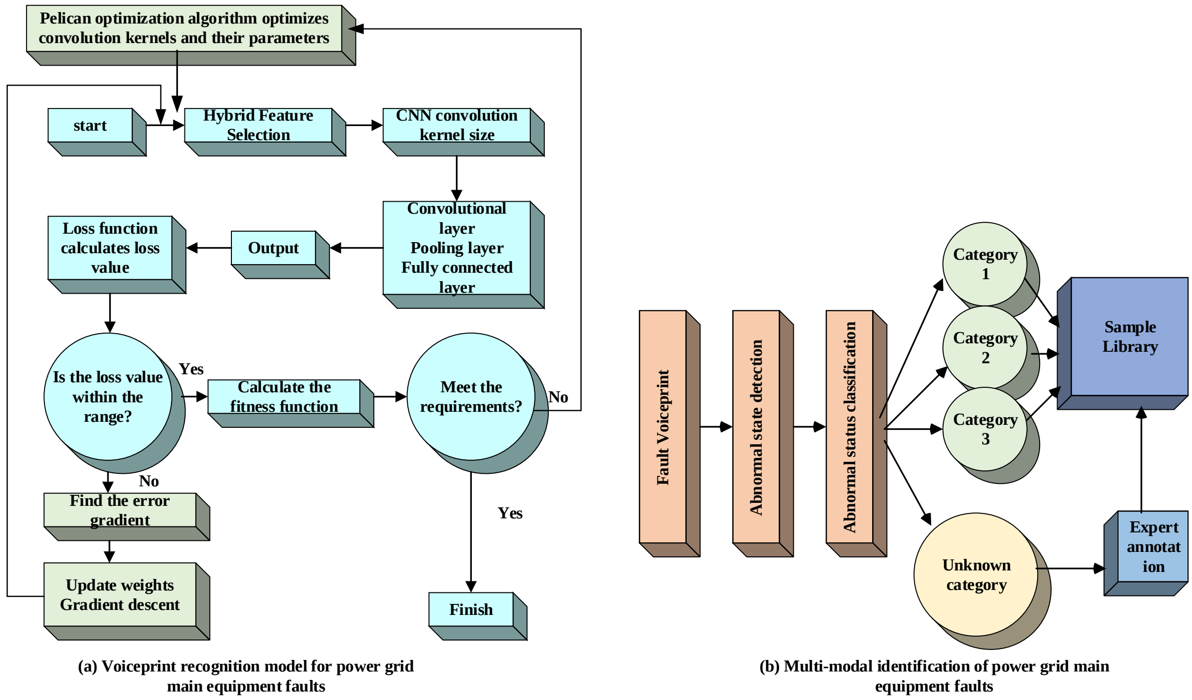


Figure 4: Fault soundprint and multi-modal fault recognition of main power grid equipment

As shown in Figure 4, this study uses the mixed acoustic signal characteristics of the main power grid equipment as the input of the convolutional neural network, and then uses the Pelican optimization algorithm to optimize the convolution kernel to determine the optimal convolution step size. The convolution kernel is then used to train the neural network to extract the acoustic signal characteristics of the main power grid equipment. Subsequently, the loss function is updated using gradient descent, and multiple iterations are performed to obtain the best fault voiceprint recognition model for abnormal voiceprint samples. Finally, a multimodal fault category recognition model is constructed to output the specific fault type. When the model determines that it is an abnormality of an unknown category, experts will identify it to update the sample library, and finally continuously optimize the fault multimodal recognition model.

3 Results Analysis

3.1 Analysis of denoising effect based on WOA-VMD-FastICA

In order to verify the effectiveness of WOA-VMD-FastICA denoising, WOA was used to iteratively optimize VMD, and the optimal number of parameter modes was found to be 5, and the optimal penalty factor coefficient was 230. The transformer of the main power equipment in a coal mine area was taken as the object of acoustic signal collection, and a group of collected acoustic signals were analyzed as samples. The group of acoustic signals was decomposed into 3 IMFs by WOA-VMD, and each IMF was transformed by FastICA to obtain the FastICA time-frequency domain diagram, as shown in Figure 5.

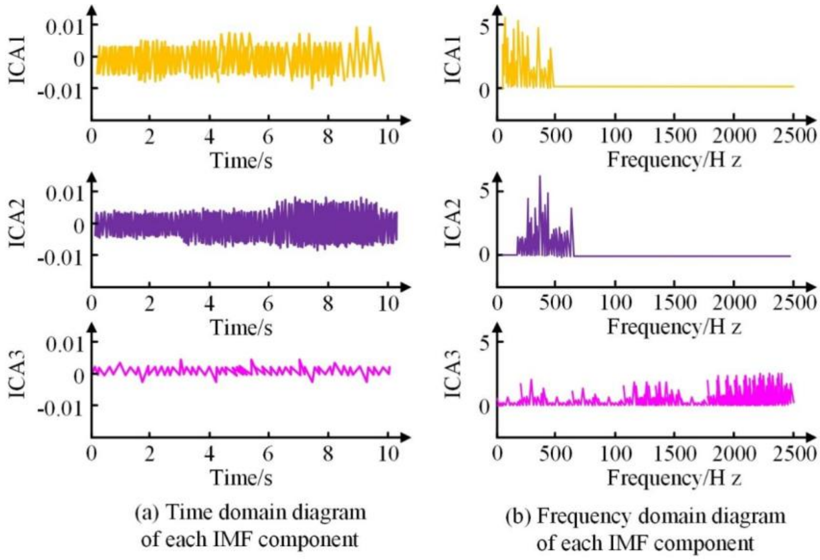


Figure 5: Time-frequency diagram of three-segment IMF

The frequency distribution of the transformer noise signal and the main acoustic signal can be seen from the spectrum analysis in Figure 5. The noise frequency of IFM1 is mainly distributed in 0-500HZ, the noise frequency of IFM2 is mainly distributed in 200-650HZ, and the noise frequency of IFM2 is distributed in 0-2500HZ, but mainly distributed in 1800-2500HZ. The three components are discriminated by the improved fuzzy entropy threshold discrimination method, and the entropy values of ICA1, ICA2, and ICA3 are 0.0065, 0.0325, and 0.0056, respectively. It can be seen that ICA1 and ICA3 are noise signals, which are directly removed and returned to 0, and ICA2 is converted into a transformer acoustic signal using inverse Fourier. The time-frequency comparison results of the transformer acoustic signal before and after denoising are shown in Figure 6.

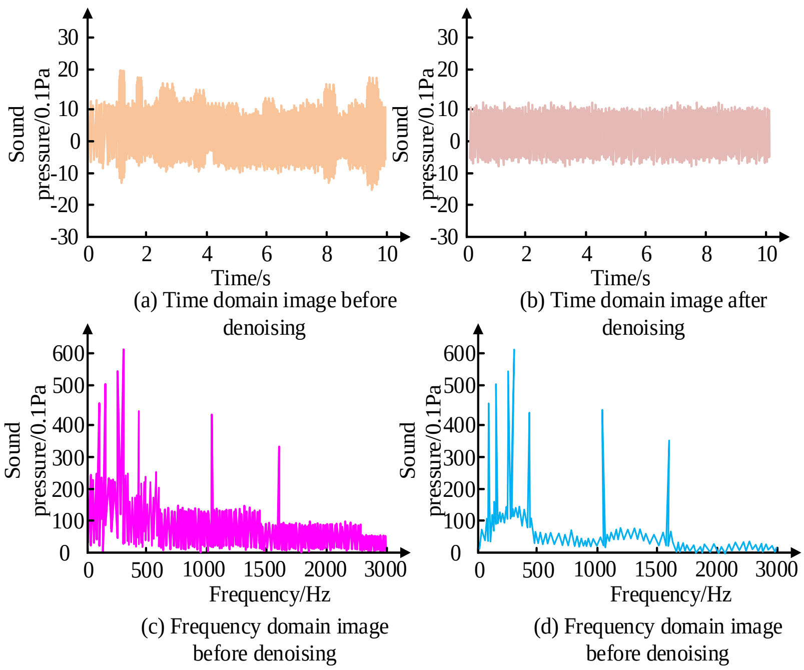


Figure 6: Time-frequency spectrum comparison of transformer acoustic signal before and after denoising

Comparing Figure 6 (a) and (b), it can be clearly seen that the time domain curve before denoising fluctuates greatly, while the overall waveform of the sound signal becomes smooth after denoising. Comparing Figure 6 (c) and (d), it can be seen that the frequency of the sound signal after denoising is clear and many interferences are eliminated. This proves the effectiveness of the WOA-VMD-FastICA denoising method.

3.2 Effect analysis of fault recognition model based on POA-CNN

In order to verify the effectiveness of the fault identification model constructed by the institute, the experimental environment was built using Windows 10 system, Intel Core I7-6500U processor, and 16G memory. The basic frequency was set to 4.0GHz, the learning rate was 0.1, and the number of iterations was 200. 500 sets of acoustic signal data of the main power transformer in the coal mine area were collected and divided into test set and training set according to the ratio of 1:4. The study used CNN model, CNN-LSTM model [13], CNN-CWT model [14], CNN-BiGRU model [15] and the research model for training and testing. The results are shown in Table 1.

Table 1: Test results of five algorithms

Network Model	Accuracy (%)		Loss value	
	Training set	Test Set	Training set	Test Set
CNN Model	84.65	85.54	0.026	0.028
CNN-LSTM Model	91.52	92.65	0.018	0.019
CNN-CWT Model	92.63	93.62	0.020	0.021
CNN-BiGRU Model	93.36	94.12	0.019	0.020
POA-CNN Model	95.63	96.45	0.012	0.010

As can be seen from Table 1, the accuracy of the training set and test set used to construct the POA-CNN model are 95.63% and 96.45% respectively. The results are much higher than the CNN model, which clearly shows that POA plays a significant optimization role on the CNN model. At the same time, it is slightly higher than the CNN-LSTM model, CNN-BiGRU model, and CNN-CWT model, indicating that the POA-CNN model has certain advantages among similar optimized fault identification models. This is because POA can improve the convergence speed of the model during the model training process, allowing the model to reach a stable state faster, thereby improving training efficiency. And POA can better capture the correlation between features and enhance the generalization ability of the CNN model. The training set loss value and test set loss value of the POA-CNN model are 0.012 and 0.010 respectively, which are lower than the other four models, which also demonstrates the effectiveness of the POA-CNN model. The final study uses a confusion matrix to measure the recognition results of the POA-CNN model on transformer acoustic signal samples. The results are shown in Figure 7.

True label

	1	2	3	4	5	6	7	8	9	10	11	12	13
1	1.00												
2		0.95	0.05										
3		0.05	0.95										
4				1.00									
5					1.00								
6						1.00							
7							1.00						
8							0.05	0.95					
9									1.00				
10										1.00			
11											1.00	0.05	
12												0.95	
13													1.00

Figure 7: Confusion matrix of voiceprint recognition of transformer fault in main power grid equipment

In Figure 7, label "1" represents the voiceprint signal of the transformer of the main power grid equipment under normal working conditions; labels "2", "3" and "4" represent the signals collected by the three signal collection points of upper, middle and lower when the transformer of the main power grid equipment is short-circuited; labels "5", "6" and "7" represent the signals collected by the three signal collection points of upper, middle and lower when the transformer is overloaded; labels "8", "9" and "10" represent the signals collected by the three signal collection points of upper, middle and lower when the transformer is mechanically damaged; labels "11", "12" and "13" represent the signals collected by the three signal collection points of upper, middle and lower when the transformer is unknown. From the results of the entire confusion matrix, the POA-CNN model has an efficiency of 98% in identifying various types of faults of the main power grid equipment transformer.

4 Conclusion

As the forefront of energy extraction, the power security and stability of coal mining areas are directly related to the implementation of national energy strategies and the balance of supply and demand in the energy market. Therefore, strengthening the power safety in coal mining areas, improving the reliability and fault identification ability of the power distribution system, is of great significance for ensuring energy supply and promoting economic and social development. To enhance the power safety in coal mining areas, the study first designed a scheme for extracting voiceprint signals from the power distribution system of mining equipment, and then the feature extraction algorithm based on Mel frequency cepstrum coefficients was used to extract the voiceprint features of the acoustic signal. Finally, the voiceprint features were used as the input of the convolutional neural network model to construct a fault voiceprint recognition model for the main power grid equipment based on POA-CNN. The results show that WOA-VMD-FastICA can effectively denoise the collected acoustic signals, and the time-frequency diagram of the denoised acoustic signal is clear. The training set and test set accuracies of the POA-CNN model constructed in the study are 95.63% and 96.45% respectively, and the loss values of the training set and test set are 0.012 and 0.010 respectively. The accuracy of the model is higher than that of the other four test models, and the loss values are lower than those of the other four test models. The effectiveness and superiority of the recognition model constructed in the study are proved. The POA-CNN model was used to perform fault identification tests. According to the results of the confusion matrix, the identification model constructed in the study has an efficiency of 98% in identifying various types of faults in the transformer of the main power grid equipment. The above data show that the method used in the study can solve the problem of multi-modal fault identification of the main power grid equipment in the smart power grid in the coal mine area. However, there are many types of main power grid equipment in the coal mine area, and the fault types are not the same. The equipment fault database will be further expanded in the future.

Funding

This work was supported by Research and industrialization of key technologies of digital twin based on source - grid - load - storage integration (3502Z20231006) and Science and Technology Project of State Grid Henan Electric Power Company (Project No. 521780230002).

References

[1] He S, Zhang Y, Zhu R, et al. Electric signature detection and analysis for power equipment failure monitoring in smart grid[J]. IEEE Transactions on Industrial Informatics, 2020, 17(6): 3739-3750.

[2] Ghasemi H, Farahani E S, Fotuhi-Firuzabad M, et al. Equipment failure rate in electric power distribution networks: An overview of concepts, estimation, and modeling methods[J]. Engineering Failure Analysis, 2023, 145: 107034.

[3] Wang S, Zhou Y, Ma Z. Research on fault identification of high-voltage circuit breakers with characteristics of voiceprint information[J]. Scientific Reports, 2024, 14(1): 9340.

[4] Jiang Y, Sun H, Hong L, et al. Research on Voiceprint Recognition of Electrical Faults With Lower False Alarm Rate[C]//2021 Power System and Green Energy Conference (PSGEC). IEEE, 2021: 599-603.

[5] Chen L, Wang R, Hu F, et al. Research on Voice Print Recognition of Electrical Faults Based on Attention-MFCC Algorithm[C]//2021 Power System and Green Energy Conference (PSGEC). IEEE, 2021: 748-751.

[6] Xu, C. Xu, H. Zhang, L. Huang, Y. Liu, Y. Nojima, et al., "A multi-population multi-objective evolutionary algorithm based on the contribution of decision variables to objectives for large-scale multi/many-objective optimization," IEEE Transactions on Cybernetics, vol. 53, no. 11, pp. 6998-7007, Nov. 2022. DOI: 10.1109/TCYB.2022.

3180214.

[7] Dang X J, Wang F H, Ma W J. Fault diagnosis of power transformer by acoustic signals with deep learning[C]//2020 IEEE International Conference on High Voltage Engineering and Application (ICHVE). IEEE, 2020: 1-4..

[8] Cui Y, Huang X, Zhang X. Deep neural network based acoustic pattern recognition system for fault localization application[J]. Applied Mathematics and Nonlinear Sciences, 2023.

[9] Zhou H, Lu L, Shen M, et al. An Efficient Noise Reduction Method for Power Transformer Voiceprint Detection Based on Poly-Phase Filtering and Complex Variational Modal Decomposition[J]. Electronics, 2024, 13(2): 338.

[10] Lu L, Zhang X, Ma H, et al. Transformer fault acoustic identification model based on acoustic denoising and DBO-SVM[J]. Journal of Electrical Engineering & Technology, 2024: 1-13.

[11] You X, Wu H, Li J, et al. Fault Diagnosis of Driving Gear in Battery Swapping System based on Audio Features and SRC-Adaboost[J]. Measurement Science and Technology, 2024..

[12] Li B, Wang G. Research and Application of Acoustic Sensor Array and WOA-VMD-FastICA Based Voiceprint Separation Method for 750kV Reactor[C]//2024 6th International Conference on Energy Systems and Electrical Power (ICESEP). IEEE, 2024: 1462-1465.

[13] Lv Z, Yang D, Wei B. Technical analysis of transformer vowel print fault diagnosis based on CNN-LSTM network[J]. International Journal of COMADEM, 2024, 27(1).

[14] Isyanto H, Arifin A S, Suryanegara M. Voice biometrics for Indonesian language users using algorithm of deep learning CNN residual and hybrid of DWT-MFCC extraction features[J]. International Journal of Advanced Computer Science and Applications, 2022, 13(5).

[15] Zhou X, Wu M, Wang H. Leveraging Artificial Intelligence and Data Mining to Identify Chinese Dialects: A Comparative Analysis of Tone and Segment Importance[C]//2023 6th International Conference on Artificial Intelligence and Big Data (ICAIBD). IEEE, 2023: 127-132.