Substation equipment abnormal detection based on acoustic fingerprint learning

Dingmei Wang*, Li Liang, Shengting Liu, Dehong Jiang
Lanzhou Longneng Electric Power Science & Technology Co., Ltd., China

ABSTRACT

Predicting and evaluating the operation status of the equipment online can not only reflect the personalization of the equipment but also meet the actual working needs of intelligent substations. Through acoustic fingerprint learning, potential defects and hidden dangers can be identified, and fault handling and emergency repair time can be shortened. This article proposed a convolutional neural network to learn the acoustic fingerprint of substation equipment and discussed the feature selection and feature preprocessing of the proposed machine learning model. We then conducted a simulation experiment and analyzed the right parameters selection for the proposed model. The experiment results show that the proposed model can achieve a recognition accuracy of above 90% on all the different abnormal voiceprint test sets. The recognition results showed the effectiveness of the voiceprint recognition model and can thus provide a solid guarantee for the safe and stable operation of the power system.

Keywords: Machine learning, acoustic fingerprint, substation equipment, abnormal detection

1. INTRODUCTION

The internal production equipment in smart substations is subjected to electrical, magnetic, mechanical, and other stresses, which will be accompanied by vibration. The mechanical waves formed will be transmitted to the casing through the medium and can be captured by sensor devices. This signal contains a large amount of time-frequency domain characteristic information, like fingerprints. When there is an abnormality in production equipment, the acoustic fingerprint will change and can be used as the main characteristic parameter for diagnosing equipment defects and faults. Voiceprint has characteristics such as stability, measurability, and uniqueness, making it very suitable for monitoring intelligent substations.

In recent years, with the development of deep learning, the use of voice recognition has been widely applied in criminal investigation, finance, smart home, and other fields. At present, the inspection robots used in substations have the ability to recognize the appearance of the equipment, such as oil leakage, surface damage, pollution, etc. However, abnormal states and developmental faults inside the equipment, cannot be observed through imaging methods. The use of deep learning methods such as Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LTSM) Networks provides effective technical means for substation sound recognition. Therefore, conducting research on key technologies for voiceprint recognition and fault detection of key equipment in substations based on deep learning, realizing automatic analysis and judgment of the operating status of important primary equipment such as transformers and reactors, can detect equipment faults in substations in advance, which helps to accelerate the process of unmanned substations.

Predicting and evaluating the operation status of the equipment online at any time not only reflects the personalization of the equipment but also meets the actual working needs of intelligent substations. Therefore, the development of equipment status management tools based on machine hearing can provide data support for power system decision-making assistance. Through machine learning, potential defects and hidden dangers can be identified, and fault handling and emergency repair time can be shortened, providing a solid guarantee for the safe and stable operation of the power system.

*lory2366@163.com

Fifth International Conference on Green Energy, Environment, and Sustainable Development (GEESD 2024), edited by M. Aghaei, X. Zhang, H. Ren, Proc. of SPIE Vol. 13279, 132790K · © 2024 SPIE · 0277-786X
 · doi: 10.1117/12.3044598

2. RELATED WORKS

2.1 Acoustic fingerprint recognition

Acoustic Fingerprint Recognition technology was originally a biometric technology, mainly by converting sound signals into electrical signals, and then using computer technology for recognition. The development of voiceprint recognition technology can be mainly divided into the following three stages: In the 1930s, Kersta[1] from Bell Labs introduced the possibility of applying voiceprint recognition technology in speaker recognition, opening up research on the application of voiceprint recognition technology. The University of Washington Atal[2,3] proposed a linear prediction model and a linear cepstral coefficient LPCC by simulating human pronunciation, which greatly improves the accuracy of voiceprint recognition. Although LPCC can better describe the resonance peak characteristics of noise signals, its ability to describe consonants is insufficient. Davis and Mermelstein[4] proposed Mel Frequency Cepstra Coefficients (MFCCs), which has better accuracy and robustness in pattern recognition and has therefore developed into the most extensive and mature feature extraction method. Template matching technology is gradually being replaced by the Hidden Markov Model (HMM), Gaussian Mixture Model (GMM), and Artificial Neural Network (ANN) models. The increase in model parameters increases the difficulty of training, and the model puts forward higher requirements for the dimensionality and volume of data.

In the 21st century, statistical models represented by GMM have become the mainstream technology of voiceprint recognition methods. MIT Lincoln Laboratory[5] proposed a Gaussian Mixture Model Universal Background Model (GMM-UBM) to effectively solve the problem of excessive GMM parameters, making it possible for voiceprint recognition technology to move from experimentation to application. On this basis, Campbell et al.[6] from the same laboratory applied a Support Vector Machine (SVM) to GMM, effectively improving the expression ability of voiceprint recognition models. Kenny, Dehak, and others from the Montreal Computer Research Institute have proposed techniques such as Joint Factor Analysis[7] and i-vector[8], which further enhance the model's ability to compensate for channel variability and resist noise.

In recent years, with the improvement of computational performance, the application of artificial intelligence methods such as deep learning has driven the development of phonetics. In 2014, Google's Variani et al.[9] used Deep Neural Network (DNN) to automatically extract feature vectors from spectrograms and named the extracted vectors d-vector. In 2015, Chm et al.[10] from MIT applied convolutional neural networks to text-dependent speaker recognition and achieved good results. In 2017, Snyder et al.[11] from Johns Hopkins University proposed the famous x-vector by using a feedforward neural network (FNN) to extract embeddings instead of i-vectors, which aggregates the time pooling layers of the network on the input speech to capture the speaker's long-term features. This allows the network to be trained to distinguish speakers from speech segments of different lengths.

2.2 Equipment acoustic fault detection

With the development of voiceprint recognition technology, the use of acoustic signals for fault diagnosis of various mechanical equipment has attracted widespread attention from domestic and foreign scholars due to the characteristics of simple installation of acoustic sensors and no impact on equipment performance. In 1996, Gao et al.[12] proposed a method for diagnosing bearing faults by applying neural networks and spectral analysis techniques. Then some researchers proposed a blind source separation algorithm based on a genetic algorithm, which solved the problem of mutual information between audio diagnostic signals of bearing faults. In 2021, Pan et al.[13] from Northwestern Polytechnical University conducted a detailed study on the acoustic model of bearing faults. In the same year, Mao[14] simultaneously used acceleration sensors and voiceprint sensors to collect bearing operating status signals, and proposed a fusion diagnostic model based on Convolutional Neural Networks (CNNs), which achieved higher accuracy. In 2022, Tauheed et al.[15] established a professional acoustic laboratory to obtain low-noise bearing voiceprint signals, and analyzed the signals using feature extraction methods combined with SVM. The experimental results proved that using non-invasive voiceprint sensors for accurate fault diagnosis of bearings is reliable and effective. In 2023, Liu et al.[16] proposed a bearing fault diagnosis method that combines variational mode decomposition and Mel-CNN, achieving excellent fault diagnosis accuracy. Research has shown that for the fault diagnosis needs of rotating machinery in special environments, bearing fault diagnosis methods based on voiceprint recognition technology can effectively diagnose bearing faults through non-contact sensors, with unique advantages.

At present, research on fault diagnosis methods based on voiceprint recognition mainly focuses on feature extraction and template-matching techniques. This article will delve into related research and validate the effectiveness of the proposed method through experiments.

3. FEATURE SELECTION OF NOISE SIGNALS

The operation status of power transformers varies in various external manifestations, and noise signals are one of the important characteristics. However, it is difficult to quantitatively determine the changes in the noise signal of power transformers solely based on sensory perception. Therefore, selecting an appropriate feature parameter to characterize the operating condition of power transformers is an important task. Voiceprint contains the time-domain and frequency-domain characteristics of the speaker's speech signal. By analyzing the voiceprint of the speaker's speech signal, the speaker's identity can be effectively recognized. Similarly, the noise signal of a power transformer is also a type of sound signal, and its voiceprint characteristics can also reflect the working condition of the power transformer body.

The noise distribution data of power transformers under Undervoltage, rated voltage, overvoltage, and different DC bias coefficients have their characteristics. Based on these data, the extraction of voiceprint characteristic parameters can be carried out. Firstly, the energy and frequency characteristics of the noise signal will be analyzed; Secondly, we should preprocess the noise signal; Then, Mel frequency spectrogram from the preprocessed noise signal will be extracted; Finally, a comparative analysis and study will be conducted on the extracted Mel time-frequency spectra under different working conditions.

3.1 Energy characteristics

In the analysis of sound signals in power transformers, the vertical axis amplitude in the time-domain waveform of sound pressure corresponds to the instantaneous energy magnitude of sound pressure. Therefore, the larger the amplitude of sound pressure, the greater its energy. According to the national standard, the noise level of a 110kV transformer is below 80 decibels at a distance of 3 meters from the transformer. In order to make the extracted noise voiceprint features more prominent, this article refers to the requirements of GB/T 1094.10-2003 "Power Transformers—Part 10: Sound Level Measurement". The noise signal of this point is measured at the corresponding position on the envelope surface of the transformer oil tank, with a distance of 20 cm between the measurement points, and a distance of 100cm between the specified contour line of the measurement points and the transformer oil tank. One of the measurement points was selected as the observation point, which is 100cm away from the transformer and has a height of 1/2 of the transformer's height. The distance between the two edges of the transformer is equal. Measure the time-domain waveform of noise data at observation points of power transformers during no-load operation under rated voltage, undervoltage, overvoltage, and different DC bias coefficients, that is, the change in total sound pressure amplitude at different times.

The amplitude fluctuation of the total sound pressure field at the starting time in the time-domain waveform of the power transformer under different working conditions is significant. This is because the time-domain waveform of the total sound pressure field is obtained through simulation experiments, and multi physics field simulation software requires a certain transition time to achieve stable results in the model calculation. By comparing and analyzing the time-domain waveform of the total sound pressure of power transformers under different working states, it can be found that there is a significant difference in the total sound pressure amplitude between the DC bias working state and the normal working state, that is, there is a significant difference in energy characteristics. However, the total sound pressure amplitude between the undervoltage and overvoltage working states and the normal working state is basically the same, and the difference in energy characteristics is small. Therefore, the energy characteristics cannot distinguish which working state the power transformer is in. Therefore, further analysis using other methods is needed.

3.2 Frequency characteristics

In order to further analyze the vibration characteristics of the sound waveform of power transformers, the sound signals of power transformers under different working conditions can be decomposed into a series of trigonometric functions with different amplitudes and periods. Using the amplitude and frequency of a series of trigonometric functions to express the characteristics of this sound segment, their correspondence is only related to frequency. Therefore, the correspondence with the time-domain spectrum is called the frequency-domain spectrum. The Fourier transform can be used to transform from time domain to frequency domain.

The sound spectrum of power transformers is mainly composed of 100 Hz and its harmonics, which are mostly distributed in the frequency range of 0-800 Hz. However, in the DC bias working state, the sound spectrum will still exhibit 50 Hz and its odd harmonic components. With the continuous increase of excitation voltage, the main frequency shifts from 100 Hz to 200 Hz, and higher harmonics gradually increase. As the DC bias coefficient continues to increase, the proportion of the main frequency gradually decreases, and higher and odd harmonics continue to increase.

Based on the above analysis, due to the extremely complex information contained in the noise signals generated by power transformers under different working conditions, conventional energy and frequency characteristics can only reflect some differences in characteristics between time and frequency domains, but it is difficult to accurately and clearly determine the operating state of the power transformer. Therefore, further research is needed on noise signals to select more effective voiceprint feature vectors in order to achieve accurate recognition of different working states of power transformers.

4. FEATURE PREPROCESSING

Sound signals are non-stationary signals that are difficult to process directly. Therefore, in order to accurately extract their feature vectors, it is necessary to preprocess the sound signals. Due to the short-term stability of sound signals, processing short-term sound signals can also achieve high accuracy. Therefore, longer sound signals can be divided into several short-term sound signals and then processed. Similarly, as a type of sound signal, the noise signal of power transformers is also a non-stationary signal with the characteristic of short-term stability. Therefore, it is possible to preprocess the noise signal of power transformers. However, due to the different characteristics of power transformer noise signals and speech signals, the preprocessing methods used for power transformer noise signals will also be different.

4.1 Pre-emphasis

In the spectrum of power transformer noise signals, the energy of the noise signal is usually distributed in the low-frequency part, and the energy in the low-frequency part is much higher than that in the high-frequency part. In addition, as the frequency of the power transformer noise signal continues to increase, the power spectrum will gradually decrease, which will lead to a significant decrease in the signal-to-noise ratio of the high-frequency part of the noise signal. In order to avoid numerical problems in subsequent Fourier transform operations, it is necessary to enhance high-frequency information, which is very useful for balancing the spectrum and can also improve the overall signal of noise signals and the accuracy of noise signal recognition.

4.2 Framing

Performing Fourier transform on the noise signal of the power transformer transforms the noise signal from time domain to the frequency domain. However, if the Fourier transform is applied to the entire segment of noise signal, temporal information will be lost. Due to the short-term stability of power transformer noise signals. Therefore, assuming that the frequency information remains unchanged for a short period of time t and performing a Fourier transform on a frame of length t, the frequency and time domain information of the noise signal can be appropriately represented. So it is necessary to perform frame division processing on the original noise signal, dividing it into N segments of fixed-size noise signals, where each segment of noise signal is called a frame. Compared with voice signals, power transformer noise signals are more stable and can be appropriately increased in frame length to achieve higher accuracy. However, excessively long frame lengths can have a serious impact on recognition speed. After comprehensive consideration, this article chooses a frame length of 40ms and a frameshift of 10ms, which means that adjacent frames have a 25% overlap rate. This ensures the temporal invariance of the noise signal while ensuring the spectral resolution of the power transformer noise signal.

4.3 Adding windows

After framing the noise signal of the power transformer, the noise signal is cut off at the boundary, causing discontinuity, which can cause significant distortion in Fourier analysis. Therefore, it is necessary to add a window function to every frame of the original noise signal and set the value outside the window to 0 to eliminate possible signal discontinuity. The commonly used window functions include rectangular windows and Hamming windows. The main lobe width of the Hamming window is larger than that of the rectangular window, and the side lobe width is smaller than that of the rectangular window. This not only effectively reduces the loss of effective information, but also makes the low-pass characteristics smoother, which can better reflect the frequency characteristics of short-term signals. Therefore, this article uses Hamming windows for windowing processing.

5. DESIGN OF DEEP LEARNING MODEL

5.1 Model selection

Deep learning is an important branch of machine learning, originally designed to enable computers to automatically learn

existing data like the human brain. Deep learning can extract and compute low-level features, thereby obtaining more abstract high-level features and effectively mining distributed features of data. Compared with other learning methods, deep learning has a stronger learning ability. Therefore, applying deep learning to the voiceprint feature recognition of power transformers under different working conditions has strong feasibility and practical value.

Convolutional neural networks, as one of the widely used network frameworks in deep learning, have excellent performance in image processing, video analysis, natural language processing, and other fields. At present, some scholars have studied how to apply convolutional neural networks to the state detection of power transformers. The difference between it and traditional deep neural networks is that on the one hand, the neurons in the network are interconnected in different ways. The connection method of convolutional neural networks is incomplete connection, which can effectively reduce the complexity and number of parameters of the network; On the other hand, convolutional neural networks use the same connection weights for operation, which reduces the number of weight values.

Convolutional neural networks take the original image as input and can obtain output by training the network, thus avoiding the difficulties of traditional methods that require feature selection and finding classifiers. Convolutional neural networks have the following advantages in processing image signals:

(1) The input image can always maintain its original structural features throughout the entire network; (2) The two independent processes of pattern classification and feature extraction in pattern recognition can be combined into one structure;(3) Using the same connection weights reduces the number of parameters that need to be trained in the network, making parameter training easier and the network more adaptable (Figure 1).

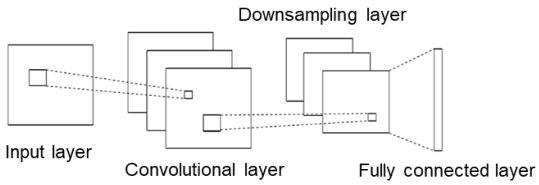

Figure 1. Convolutional neural network structure.

5.2 Training process

The detailed training process of convolutional neural network parameters is as follows:

1) The training set of voiceprint samples is randomly selected;

2) Parameters are initialized;

3) A set of labeled data is input into a convolutional neural network;

4) The output vector of the intermediate layer of the convolutional neural network and the actual output vector is calculated;

5) The actual output vector of the convolutional neural network is compared with the label values of the data, and the corresponding error is calculated by using the corresponding loss function;

6) The adjustment amount of the threshold and the adjustment amount of each weight value in sequence are calculated;

7) Weights and thresholds are adjusted;

8) We perform forward calculation on the updated parameters to determine if the loss function is below the threshold. If it is lower, it indicates that the loss is small and the actual output is close to the label value. Then it continues to the next step. If greater than, it indicates a significant difference between the actual output and the label value. It should return to step (3) and perform the iterative calculation again;

9) Training is over.

The noise signal of a power transformer can reflect its own operating status information. Under different working conditions, the noise signal will have significant changes in the time and frequency domains. However, the changes in this operating status information are very complex and difficult to distinguish directly. By using a Mel-CNN-based power transformer voiceprint recognition model, the noise signals of power transformers under different operating states

can be preprocessed to obtain the Mel time-frequency spectrum. Then, the Mel time-frequency spectrum can be used as the original input image of the CNN network for learning, thereby achieving the extraction of power transformer voiceprint features and recognition of different operating state modes (Figure 2).

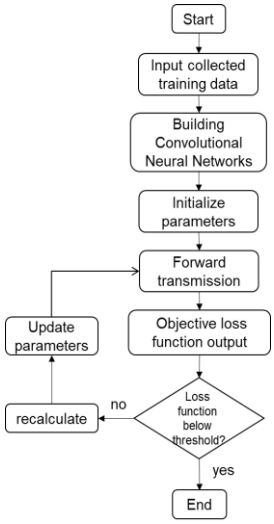

Figure 2. Flow chat of neural network parameter training.

6. PERFORMANCE ANALYSIS

The noise signal of a power transformer can reflect its own operating status information. Under different working conditions, the noise signal will have significant changes in the time and frequency domains. However, the changes in this operating status information are very complex and difficult to distinguish directly. By using a Mel-CNN-based power transformer voiceprint recognition model, the noise signals of power transformers under different operating states can be preprocessed to obtain the Mel time-frequency spectrum. Then, the Mel time-frequency spectrum can be used as the original input image of the CNN network for learning, thereby achieving the extraction of power transformer voiceprint features and recognition of different operating state modes.

6.1 Simulation experiment design

This article obtains noise signal data of power transformers under different working states through simulation experiments. Due to the influence of computer calculation speed, it is not possible to measure a long period of noise signal like in actual measurement. The simulation time in this article is 1 second. However, when training the model, it is necessary to learn a large amount of voiceprint sample data. In order to overcome this difficulty, this paper extracts sound pressure time-domain data from multiple measurement points in the sound field simulation results to compensate for the lack of data. The noise signal at the corresponding position on the envelope surface of the power transformer oil tank is measured. The specific selection rules are as follows: measurement points are selected on the front, back, left, and right sides of the power transformer, with the distance between the envelope surface of the oil tank and the power transformer being 50 cm, the distance between the two adjacent measurement points on the left and right being 10cm, and the distance between the two adjacent measurement points on the top and bottom is 50 cm. Using this method, a total of 96 measurement points were selected for the front and rear surfaces of the fuel tank, and 54 measurement points were selected for the left and right surfaces of the fuel tank. A total of 150 measurement points were selected for one simulation. In order to prevent overfitting and improve the effectiveness of model training, it is necessary to randomly shuffle the extraction path of processed voiceprint sample data, and input the voiceprint samples into the model in random order for training and recognition, greatly ensuring the effectiveness of training.

A convolutional neural network structure was designed using a deep network designer. Firstly, image features were extracted from the input image through convolution, and the image features were standardized; Secondly, activate nodes were through activation functions; Then, it is pulled into a one-dimensional vector through a fully connected layer for classification; Finally, the recognition accuracy of the test set is obtained through the classifier, and the output label is determined using class output.

6.2 Sample distribution

To ensure the effectiveness of deep learning, it is necessary to complete the dataset partitioning of batch preprocessed sample data during model training. As can be seen from the previous research, under the excitation of different voltage levels and DC bias coefficients, the voiceprint characteristics of power transformers have a certain trend of variation. In order to reduce the time required for model training, this paper classifies the voiceprint data of power transformers under the same operating state into one category and labels them uniformly. To verify the generalization ability of the model, 90% of the voiceprint sample data under the same label was randomly selected as the training set, and the remaining 10% was used as the testing set. The training set is used for algorithm training, model selection, and parameter adjustment, while the test set is used to evaluate the recognition results of the algorithm. The sample distribution of the voiceprint database is shown in Table 1.

Table 1. Sample distribution of voiceprint database.

Operating conditions	Label	Number of training sets	Number of testing sets
Normal condition	1	1000	100
Undervoltage condition	2	500	50
Overvoltage condition	3	500	50
DC bias	4	500	50
Other abnormal condition	5	500	50
Total	\	3000	300

6.3 Learning rate and recognition accuracy

The convergence speed of convolutional neural networks is determined by the size of the learning rate, which has a significant impact on the weights in the network structure, thereby affecting the final classification accuracy. It is very important to set an appropriate learning rate size. The learning rate cannot be set too large or too small. If it is too small, it will result in very small changes in weight values during each iteration, slow convergence speed of the system, and easy to cause local minima; When it is too large, the change in weight values during each iteration becomes very significant, making it difficult to find the optimal parameters for the system. This article applies empirical methods to the selection of learning rates. Figure 3 shows the detailed training results under different learning rates.

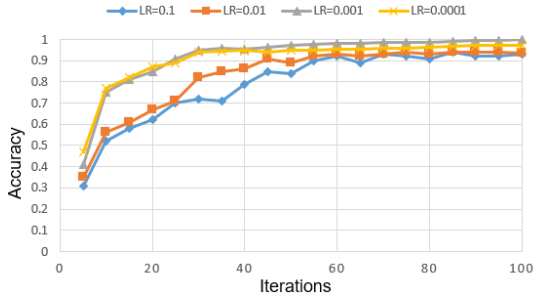

Figure 3. Accuracy of voiceprint feature recognition under different learning rates.

Table 2 shows the specific recognition accuracy corresponding to different learning rates. By comparison, it can be seen that when the learning rate is 0.001, the model has the highest recognition accuracy and the training speed is also fast. Therefore, a learning rate of 0.001 was selected as the optimal learning rate for the Mel-CNN model.

Table 2. Accuracy of voiceprint feature testing at different learning rates.

Learning rates	0.1	0.01	0.001	0.0001
Accuracy	91.2%	93.7%	98.6%	95.0%

6.4 Optimizers and recognition accuracy

Hyperparameters are important factors that affect the training speed and accuracy of convolutional neural networks. The optimizer is the most important hyperparameter that can affect the optimization process of reducing loss values. Therefore, it is necessary to choose a suitable optimizer to ensure the reduction of loss values and the improvement of accuracy in deep learning networks. Common optimizers include stochastic gradient descent (SGD) and adaptive moment estimation (Adam). In order to study the impact of different optimizers on recognition accuracy, this paper selected SGD and Adam optimizers for Mel-CNN recognition model tuning, with 100 iterations. During the training process, the dataset, network model, and other hyperparameters remained consistent. The detailed recognition performance of the two optimizers is shown in Figure 4.

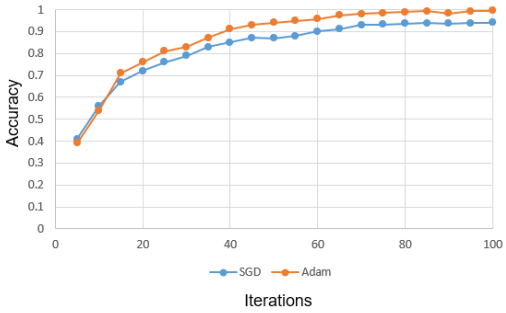

Figure 4. Accuracy of voiceprint feature recognition under different optimizers.

The Adam optimizer can calculate different parameters and adjust the adaptive learning rate. Its advantage is that it determines the range of learning rate during the iteration process, ensuring that the parameters remain relatively stable throughout the entire training process. From Table 3, it can be seen that using the Adam optimizer can achieve a recognition accuracy of 97.53%, which shows better performance compared to the SGD optimizer.

Table 3. Accuracy of voiceprint feature testing under different optimizers.

Optimizer	SGD	Adam
Accuracy	91.7%	97.53%

6.5 Accuracy of voiceprint recognition

Based on the above research analysis, this article selects Adam as the optimizer of the Mel-CNN voiceprint recognition model and sets the learning rate to 0.001. At this time, the recognition accuracy of the voiceprint recognition model for three operating conditions is shown in Table 4.

Table 4. The accuracy of the voiceprint recognition model for voiceprint feature test under different working conditions.

Operating conditions	Normal	Undervoltage	Overvoltage	DC bias	Other abnormal
Accuracy	98.1%	97.3%	96.5%	99.2%	95.2%

The experimental results show that the proposed voiceprint recognition model has achieved high recognition accuracy for the voiceprint features of power transformers under different working conditions, and has excellent performance.

7. CONCLUSION AND DISCUSSION

This article first analyzes the sound characteristics in the rated voltage operating state, undervoltage operating state, overvoltage operating state, and DC bias operating state through simulation software. The noise signals obtained through simulation experiments under different operating states are extracted with voiceprint features. The training and recognition of voiceprint features were completed using deep learning methods, achieving accurate recognition of different working states of power transformers.

In order to accurately identify the different working states of power transformers, preprocessing was performed on the noise signals under different working states. The preprocessed noise signals were then extracted with voiceprint features,

and Mel time-frequency spectrograms were used to replace the original noise signals. It was found that the voiceprint feature extraction method effectively reduced the dimensionality of the noise sample data while retaining the time-domain and frequency-domain characteristics of the original noise signal, improving the diagnostic speed and accuracy of subsequent recognition.

Finally, the structure of convolutional neural networks and how to train network parameters were described in detail. The construction of the voiceprint dataset and the design of the network structure were introduced. The learning rate and optimizer of the voiceprint model were selected. By comparing the impact of four learning rates on recognition accuracy, choose to set the learning rate to 0.001; By comparing the impact of two optimizers on recognition accuracy, adaptive moment estimation is selected as the optimizer for the voiceprint recognition model. The processed voiceprint training set was trained using this voiceprint recognition model, and the trained model achieved a recognition accuracy of above 90% on the voiceprint test set. The recognition results showed the effectiveness of the voiceprint recognition model.

However, this article only obtains the noise signals of power transformers under different working conditions through simulation experiments. In the future when conditions permit, the noise signals of power transformers under normal working conditions, under voltage working conditions, over voltage working conditions, and DC bias working conditions can be measured on-site to further verify the effectiveness of the voiceprint recognition model.

REFERENCES

[1] Kersta, L. G., "Voiceprint-identification infallibility," The Journal of the Acoustical Society of America, 34(12), 1978-1978 (1962).

[2] Atal, B. S., "Automatic recognition of speakers from their voices," Proceedings of the IEEE, 64(4), 460-475 (1976).

[3] Atal, B. S., "Effectiveness of linear prediction characteristics of the speech wave for automatic speaker identification and verification," Journal of the Acoustical Society of America, 55(6), 1304-1322 (1974).

[4] Davis, S. and Mermelstein, P., "Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences," Acoustics Speech and Signal Processing IEEE Transactions, 28(4), 357-366 (1980).

[5] Reynolds, D. A., et al., "Speaker verification using adapted Gaussian mixture models," Digital Signal Processing, 10(1-3), 19-41 (2000).

[6] Campbell, W. M., Sturim, D. E. and Reynolds, D. A., "Support vector machines using GMM supervectors for speaker verification," IEEE Signal Processing Letters, 13(5), 308-311 (2006).

[7] Kenny, P., Boulianne, G., Ouellet, P., et al., "Joint factor analysis versus eigenchannels in speaker recognition," IEEE Transactions on Audio, Speech, and Language Processing, 15(4), 1435-1447 (2007).

[8] Dehak, N., Kenny, P. J., Dehak, R., et al., "Front-end factor analysis for speaker verification," IEEE Transactions on Audio Speech and Language Processing, 19(4), 788-798 (2011).

[9] Variani, E., Xin, L., Mcdermott, E., et al., "Deep neural networks for small footprint text-dependent speaker verification," IEEE International Conference on Acoustics, (2014).

[10] Chen, Y. H., Lopez-Moreno, I., et al., "Locally-connected and convolutional neural networks for small footprint speaker recognition," 16th Annual Conference of the International Speech Communication Association, (2015).

[11] Snyder, D., Garcia-Romero, D., Povey, D., et al., "Deep neural network embeddings for text-independent speaker verification," Interspeech 2017, (2017).

[12] Huang, X. D. and Huang, S. S., "Application of neural network and acoustic spectrum analysis technology in bearing fault diagnosis," Bearing, (8), 2-5 (1996).

[13] Liu, J., Xu, Y. J. and Pan, G., "A combined acoustic and dynamic model of a defective ball bearing," Journal of Sound and Vibration, 501(9), 116029 (2021).

[14] Wang, X., Mao, D. X. and Li, X. D., "Bearing fault diagnosis based on vibro-acoustic data fusion and ID-CNN network," Measurement, 173, 108518 (2021).

[15] Tauheed, M., Anurag, C. and Shahab, F., "An efficient diagnosis approach for bearing faults using sound quality metrics," Applied Acoustics, 195L108839 (2022).

[16] Shan, S., Liu, J., Wu, S., Shao, Y. and Li, H., "A motor bearing fault voiceprint recognition method based on Mel-CNN model," Measurement, 207, 112408 (2023).