Sound & Vibration 2025, 59(2), 2757.
https://doi.org/10.59400/sv2757

Article

Architecture design and implementation for sensing equipment faults from ultra-weak acoustic emission signals

Li'an Lu, Guangyu Liu, Hui Xiao, Xuefeng Li[*]

Tongji University, Shanghai 201800, China
* **Corresponding author:** Xuefeng Li, lixuefeng@tongji.edu.cn

CITATION

Lu L, Liu G, Xiao H, Li X. Architecture design and implementation for sensing equipment faults from ultra-weak acoustic emission signals. Sound & Vibration. 2025; 59(2): 2757. https://doi.org/10.59400/sv2757

ARTICLE INFO

Received: 9 February 2025
Accepted: 2 April 2025
Available online: 24 April 2025

Abstract: In acoustic emission (AE) detection, the weakness of the acoustic source signal, the interference from background noise, and the attenuation during signal propagation result in the sensor-received signal being completely submerged by noise, severely impacting downstream fault identification and anomaly analysis. This study focuses on the fault detection scenario of AE signals characterized by a low signal-to-noise ratio (SNR). It uses low-loss noise reduction and Mel-frequency spectrum transformation for preprocessing, then extracts features with an optimized stacked auto-encoder combined with CNN (OSAE-CNN) innovatively. These features are input into an SVM for fault classification. The proposed method significantly improves fault identification accuracy to 91.38% for signals with an SNR of −20 dB, a 30% increase over the previous method. The research findings can provide technical support for the fault monitoring and safe operation of electromechanical equipment and also offer empirical references for ultra-weak signal processing in various domains.

Keywords: acoustic emission detection; low signal-to-noise ratio; stacked auto-encoder; convolutional neural network; Mel time-frequency spectrum

1. Introduction

Acoustic emission (AE) technology, as a dynamic non-destructive testing method, is widely used to monitor the operating status of equipment in real time by detecting, recording, and analyzing the acoustic emission signals to infer the source of the acoustic emission. In the field of power electronics, acoustic emission technology is used to monitor the operating status of power semiconductor devices such as IGBTs (Insulated Gate Bipolar Transistors) and MOSFETs (Metal Oxide Semiconductor Field Effect Transistors). In addition, for large power equipment such as Gas-Insulated Switchgear (GIS), acoustic emission technology is also used to detect partial discharge (PD) faults within the equipment to prevent major accidents [1–3]. However, due to factors such as weak signal strength, dielectric absorption, scattering attenuation during transmission, and environmental noise interference, it is exceedingly challenging to extract effective information from signals that are completely submerged by noise for fault identification and anomaly analysis. Currently, relevant research efforts are concentrated on seeking breakthroughs in three key areas: reasonable noise reduction, feature extraction, and pattern recognition, with numerous significant technical advancements being achieved in these domains [4–8].

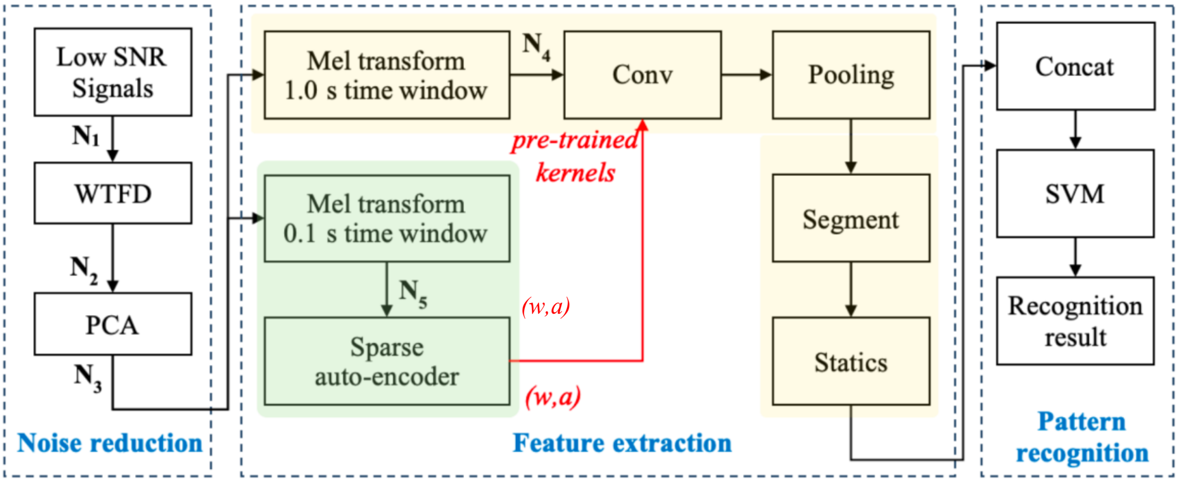
In the realm of low-loss noise reduction, Wang et al. [9] employed the S-transform method to filter noise from PD signals and trained a support vector machine (SVM) based on the downscaled statistical features, with a classification accuracy exceeding 90%. Chaudhuri et al. [10] proposed the use of Total Variation Denoising

(TVD) filters and Denoising Autoencoders to significantly enhance the processing of PD signals corrupted by white Gaussian noise. We [11] introduced a novel Wavelet Threshold Function Denoising (WTFD) method, which, under −10 dB signal-to-noise ratio conditions, enabled the SVM model to attain a recognition accuracy of 91.58%. Regarding feature extraction and pattern recognition, Chen et al. [12] presented a fractional Fourier transform-based PD feature extraction approach for the defect pattern recognition of GIS. Even with ± 10% noise interference, the correlation degree between the tested pattern and the matter-element models maintained high accuracy and tolerance. Zheng et al. [13] utilized one-dimensional time-domain feature sequences of PD pulses and corresponding wavelet time-frequency maps as inputs, combined with an Improved Convolutional Neural Network (IFCNN), achieving a PD pattern recognition accuracy of 95.8%. Sun et al. [14] constructed hybrid features from time-domain, frequency-domain, and entropy characteristics, employed t-SNE for dimensionality reduction, and proposed an Improved Northern Giant Eagle Optimization Algorithm (SCNGO) to automatically optimize the parameters of Variational Mode Decomposition (VMD) and SVM for GIS PD fault identification. We also validated the effectiveness of Mel frequency cepstrum coefficients (MFCC) in the low-frequency domain used as feature vectors, which can improve the accuracy and stability of the pattern recognition algorithm [15]. Furthermore, based on FBank features, we optimized the time-frequency spectral density distribution method and utilized image classification networks such as ResNet50 to classify PD faults, achieving an accuracy of 98% [16]. Liu et al. [17] proposed a convolutional neural network and long-short-term memory (CNN-LSTM) model, effectively extracting and utilizing spatiotemporal features of the PD input signals, with an average pattern recognition rate of 97.9%. Jing et al. [18] proposed a transfer learning method based on Stacked Autoencoders (SAE), which mitigated the issues of incomplete category labels and limited data samples in GIS fault data, achieving a PD recognition rate of 87.13%. Tian et al. [19] presented a knowledge-driven algorithm consisting of a feature space and a knowledge space, leveraging Deep Residual Networks (ResNet) and Graph Convolutional Networks (GCN) to automatically extract PD features. This method demonstrated improved performance on noisy, insufficient, and imbalanced datasets, achieving an accuracy of 99.58% on the experimental dataset and 88.79% on the noisy dataset.

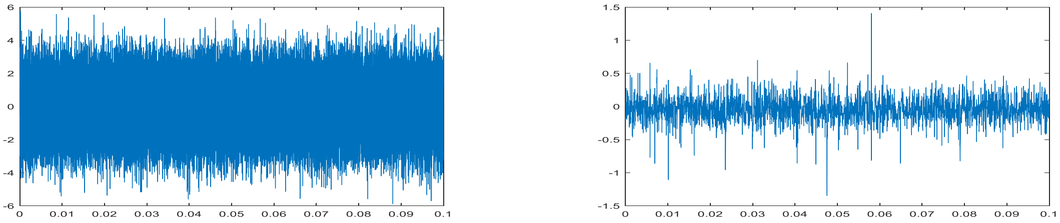
While the field of PD-type identification within GIS has witnessed numerous outstanding research accomplishments and engineering applications, research targeting the analysis of weak fault signals in strong noise environments remains insufficient. Considering the substantial background noise levels in real-world operating conditions, exploring fault detection and corresponding strategies for extremely low SNRs is both practical and forward-thinking. This study, grounded in low-loss noise reduction techniques [11], focuses on innovative feature extraction methods. We propose a methodology that harnesses a sparse autoencoder to extract detailed signal features for training the convolutional kernels of a CNN, capturing both the global and intricate aspects of the time window. By employing the Optimized Stacked Autoencoder with Convolutional Neural Network (OSAE-CNN) to extract image features from the Mel time-frequency spectrum as input, we achieve a significant enhancement in the accuracy of downstream recognition models. This advancement offers technical support for the monitoring and analysis of PDs in GIS, contributing to enhanced system reliability and safety.

2. Method

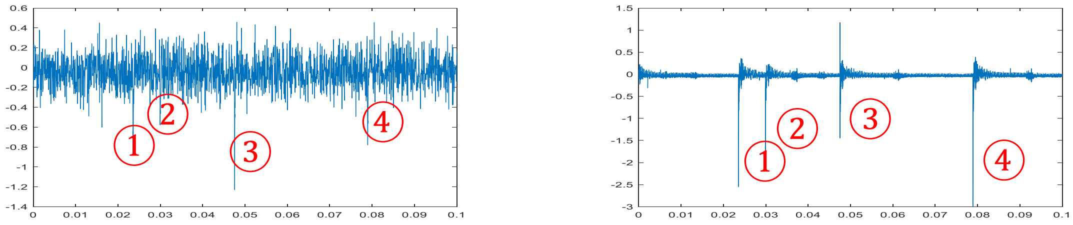
The innovations and model flow proposed in this study are shown in **Figure 1a**, which contains three modules: noise reduction, feature extraction, and pattern recognition.



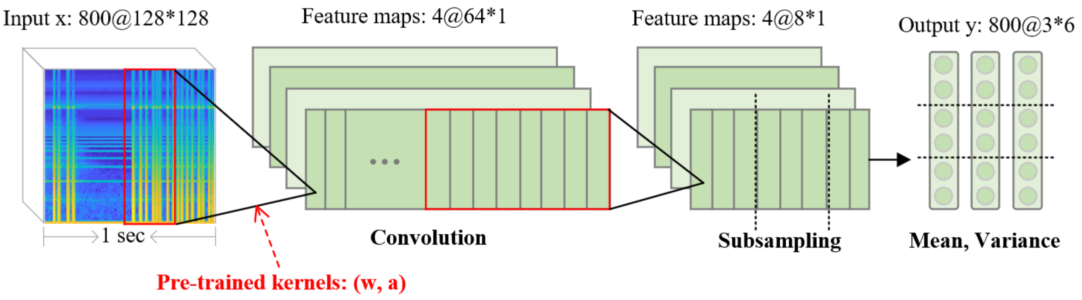
(a) Model flowchart.



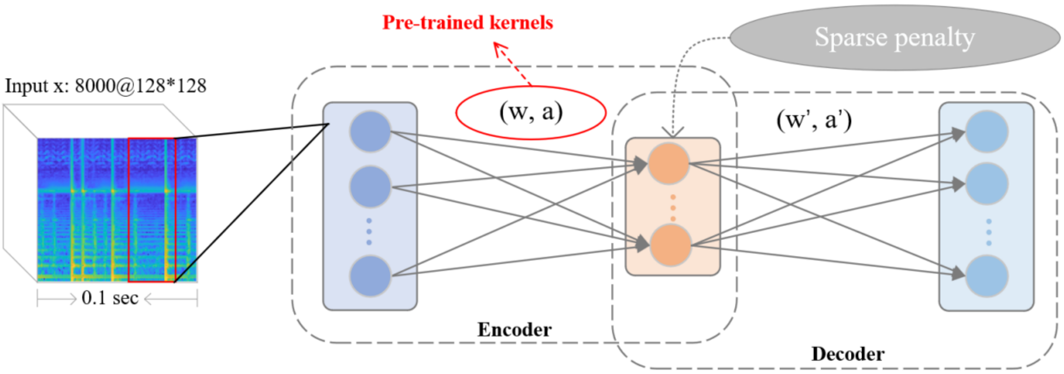
(b) N1: Typical faulty signal dyed noise −20 dB. **(c)** N2: N1 after WTFD.



(d) N3: N2 after PCA. **(e)** Raw signal.



(f) N4: Feature mapping.



(g) N5: Kernel learning

Figure 1. **(a)** Model flowchart; **(b)** N1: Typical faulty signal dyed noise −20 dB; **(c)** N2: N1 after WTFD; **(d)** N3: N2 after PCA; **(e)** Raw signal; **(f)** N4: Feature mapping; **(g)** N5: Kernel learning.

The flowchart of this study is shown in **Figure 1a**. To simulate weak AE signals in the real industrial scene, strong noise of varying degrees was first added to the AE signals collected in the laboratory. The effect in the time domain is presented in **Figure 1b**. Compared to the original signal in **Figure 1e**, a −20 dB signal was added, with the maximum amplitude being approximately 4 times that of the original signal's maximum amplitude, rendering the signal completely obscured by noise. To restore the discharge peaks in the signal, the new threshold wavelet denoising method proposed in the literature [11] was initially applied to reduce noise in the time domain, as depicted in **Figure 1c**. At this point, the signal's maximum amplitude had been reduced to a range similar to the original signal, but the discharge features were still not sufficiently pronounced. PCA reconstruction was subsequently combined to further denoise the signal by discarding secondary components, resulting in an AE signal with more distinct discharge features, as shown in **Figure 1d**. Next, considering the feasibility of converting an AE signal into a Mel spectrogram for fault detection as demonstrated in literature [14], this experiment selectively expanded the frequency band containing important information in the AE signal by restricting the Mel spectrogram bands, also outputting Mel-frequency spectra with a pixel size of 128 × 128. Finally, leveraging the main innovation of this study, the OSAE-CNN network, the detailed and global features of the Mel-frequency spectra were fused and recognized, and SVM was used to classify the feature vectors for fault diagnosis, which was then compared with previous research conducted in our laboratory.

3. OSAE-CNN network

In the OSAE-CNN network, which addresses the common challenge of sample scarcity in electromechanical fault diagnosis, the network integrates the logical structure of CNN for feature mapping with SAE to enhance the detailed feature extraction capabilities of convolutional kernels. This integration is illustrated in **Figure 1f,g**. Previous research has shown that improving the features extracted from one-dimensional time domain signals is not sufficient to support fault type identification in strong noise backgrounds. To increase the effectiveness of subsequent feature extraction and pattern recognition modules, this study implemented improvements in both global and local feature extraction methods. First, 800 one-dimensional time-domain signals were transformed into two-dimensional Mel-frequency spectra. These spectra were then processed by CNN, which is adept at handling two-dimensional signals, resulting in the extraction of richer global features. Secondly, SAE was used to pre-train the convolutional kernels of the CNN, giving them the ability to capture local feature information from the signals. The auxiliary data set was generated by segmenting each sample of the original data set into 10 parts along the time dimension, resulting in a total of 8000 shuffled samples. In summary, the main structure of the OSAE-CNN network consists of two parts, N4 and N5, which are detailed below.

3.1. N5: Kernel learning

Autoencoders, as unsupervised learning models, have the capability to automatically extract representative expressions and deep features from limited sample data. They consist of two primary components, namely the encoder and the decoder, and rely on the back-propagation algorithm along with optimization techniques to achieve data abstraction and reconstruction. Specifically, the encoder utilizes the input $x \in D_x$ as a supervisor to direct the neural network in learning a mapping function h to the feature domain, thereby obtaining the most informative feature vector $h\left(x\right) \in D_{x'}$. The decoder then transforms the hidden variables of the hidden layer back to their original dimensions by minimizing the reconstruction error, ensuring that the output can approximately reconstruct the initial input. Initially, to enhance the adaptability of the convolutional kernel to the samples, during each iteration of kernel training, we randomly select samples from the auxiliary dataset and extract time-frequency spectrogram segments at arbitrary locations as training instances, as illustrated in **Figure 1g**. The dimensions of the extracted spectrograms match those of the trained convolutional kernel, with their heights being consistent with the height of the original time-frequency spectrograms. In this case, the encoder function h maps the 0.1 s-long Mel spectral fragment x to the potential representation $x' \in D_{x'}$. The encoder function can be written in the following form:

$$x' = h\left(x\right) = s\left(Wx + \alpha\right) \tag{1}$$

where, x is the selected Mel spectrum segment, x' is the feature vector of the hidden layer, s is the nonlinear activation function. In this paper, we use the logistic sigmoid function $sigmoid(z) = 1/(1+e^{-z})$, W is the weight matrix, and α is the bias vector. Then, the decoder function is reconstructed based on the feature vector, denoted as

$$g\left(x'\right) = g\left(h\left(x\right)\right) = s'\left(W^T x' + \beta\right) \tag{2}$$

where s' is the activation function of the decoder, again a logistic sigmoid function $\left(s = s'\right)$. W^T is the weight matrix shared with the encoder and β is the bias vector. The autoencoder trains the network by tuning the parameters $\theta = \left(W, \alpha, \beta\right)$ on the auxiliary dataset corresponding to the training set, reconstructing the input with the aim of minimizing the total reconstruction error.

$$\min_\theta \sum_{x\in D_x} L\left(x, g\left(h\left(x\right)\right)\right) \tag{3}$$

where L is the reconstruction error calculated using the squared error $L\left(x,x'\right)=\left\|x-x'\right\|^2$. The sparse self-encoder, a variant of the traditional self-encoder, introduces sparsity constraints on the coding layer. This limits the information capacity of the hidden layer, enabling the self-encoder to concentrate on abstracting and compressing input information during training, rather than merely replicating the input with excessive detail. A common way to implement a sparse autocoder is to add an additional sparsity penalty to the loss function.

$$\min_\theta \sum_{x\in D_x} L\left(x, g\left(h\left(x\right)\right)\right)+\lambda_1\sum_{k=1}^{s_1} KL\left(\rho\|\hat{\rho}\right) \tag{4}$$

where $x\in D_x$ is the randomly taken out temporal spectral picture segment, $L\left(\cdot\right)$ denotes the reconstruction error, which is used here as the squared error $L\left(x,x'\right)=\left\|x-x'\right\|^2$, $g\left(\cdot\right)$ and $h\left(\cdot\right)$ denote the encoder and the decoder, respectively, λ_1 is the sparsity penalty term weight, and ρ is the sparsity parameter, which is set to 0.05 and 1, respectively, in this experiment. Furthermore, s_1 denotes the active unit of the encoding layer, $\hat{\rho}_j = 1/\rho\sum_{t=1}^{p} h_j\left(x_t\right)$ is the average excitation of the encoding layer, p denotes the number of training samples, and $KL\left(\cdot\right)$ is the KL dispersion (Kullback-Leibler Divergence), which is used as a measure of the distance between two probability distributions.

$$KL\left(P\|Q\right)=-\sum P\left(i\right)\ln\left(Q\left(i\right)/P\left(i\right)\right) \tag{5}$$

In this study, we pre-train the kernels using sparse self-encoders instead of generating them randomly. To capture the input Mel spectral structure at various scales and enrich the feature details, we consider using different numbers of kernels with multiple sizes. The sparse self-encoder is trained on small, randomly extracted segments from different locations, with each segment's size matching the learned convolutional kernel. After conducting numerous trials, we select a single kernel size of 64 × 128 and use Equation (6) to independently pre-train the sparse self-encoder. This process ultimately results in four efficient and lightweight $\left(W^i,\alpha^i\right)$ groups, which serve as convolutional kernel spares for subsequent feature vector extraction.

3.2. Characteristic mapping

The feature vector extraction methodology presented in this paper consists of four main steps: convolution, pooling, segmentation for statistical analysis, and splicing. A visual representation of this process is shown in **Figure 1f**. When the original Mel spectrum is fed into the convolutional layer, which has been trained on an auxiliary dataset, it extracts more representative features from the temporal spectrogram. This is achieved because each neuron in the output layer receives input from a limited range, capturing detailed local features embedded in the convolutional kernel. The extracted features are then processed through a pooling layer, which mitigates the effect of noise on feature quality and increases the relevance of the output features for fault classification purposes. Global features are then derived by computing statistical values that capture the relationships between local features, rather than focusing on their specific locations. This approach facilitates the extraction of higher-level features.

Upon acquiring the kernel $\left(W^i, \alpha^i\right)$, it is applied segment-by-segment to the input Mel spectrum with a fixed number of rows. Given that the number of rows in the trained convolutional kernel matches the number of rows in the Mel spectrum, the corresponding feature map is generated after processing with the *i*-th convolutional kernel in the layer:

$$f^i\left(x\right) = s\left(conv\left(W^i, x\right) + \alpha^i\right) \tag{6}$$

where $conv\left(\cdot\right)$ represents the convolution operation and $f^i\left(x\right)$ signifies the sequence resulting from the convolution of the Mel spectrum with the *i*-th kernel. Next, the feature map is down-sampled using a mean operation with a window size of *m* × *n*: $F_{t,j}^i = mean_{k \in win_j}\left(f_k^i\left(x\right)\right)$, for k within the *j*-th window win_j. The pooled features from all patches associated with the *i*-th kernel are then concatenated to form a feature vector, whose length is determined by the number of windows:

$$F_{t,j}^i = \left[F_{(t,1)}^i\left(x\right), \ldots, F_{\left(t,N_w^i\right)}^i\left(x\right)\right] \tag{7}$$

where N_w^i denotes the number of windows associated with the *i*-th kernel. Due to variations in the number of patches (segments) within each discharge signal, we compute the mean and variance of all patches for a given signal $F_{t,j}^i$. The result is a feature vector $F^i(x)$ with a fixed length of $2 \times N_w^i$, defined as:

$$F^i(x) = \left[mean\left(F_t^i(x)\right)\Big|_t, var\left(F_t^i(x)\right)\Big|_t\right] \tag{8}$$

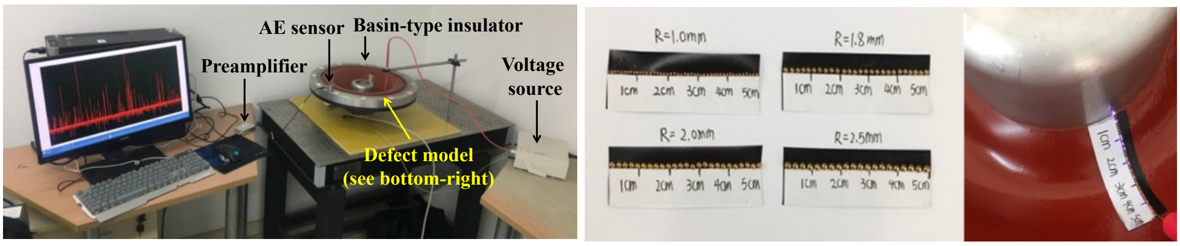
Finally, the feature vectors from all kernels are concatenated to produce the ultimate feature vector y:

$$y = \left[F^1(x), \ldots, F^q(x)\right] \tag{9}$$

4. Data preprocessing

4.1. Data samples

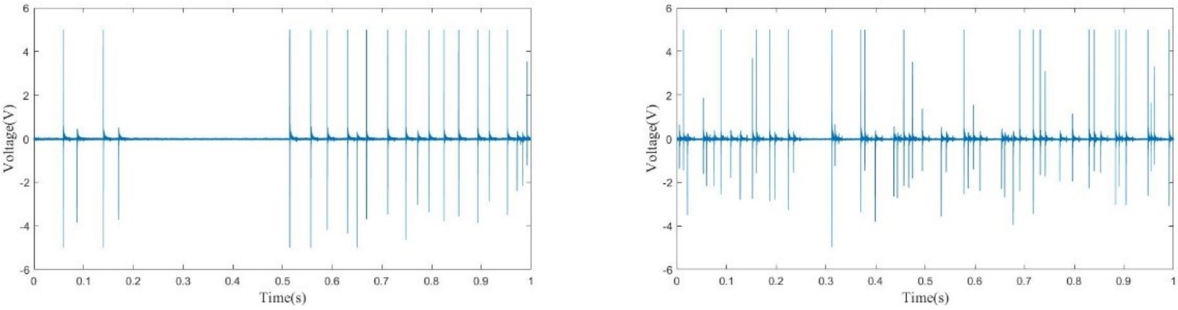
About signal collection, a sort of defect model is designed to simulate the accumulation of free metal particles on the surface of the GIS internal insulation structure in the last work [11]. The detection system used in this study is shown in **Figure 2a**, which includes a 15 kV high voltage source to supply power to the defect model; a piezoelectric AE sensor with a resonance frequency of 30 kHz and a bandwidth of 10 dB (20–110 kHz) for signal acquisition; a preamplifier with a gain of 40 dB and bandwidth of 0.01–2.0 MHz for amplification of the output signal; and a processing terminal PC with a high-speed data acquisition card. Different degrees of PD manifestations were distinguished by four metal particle diameters ranging from 1.0 to 2.5 mm which are shown in **Figure 2b**.



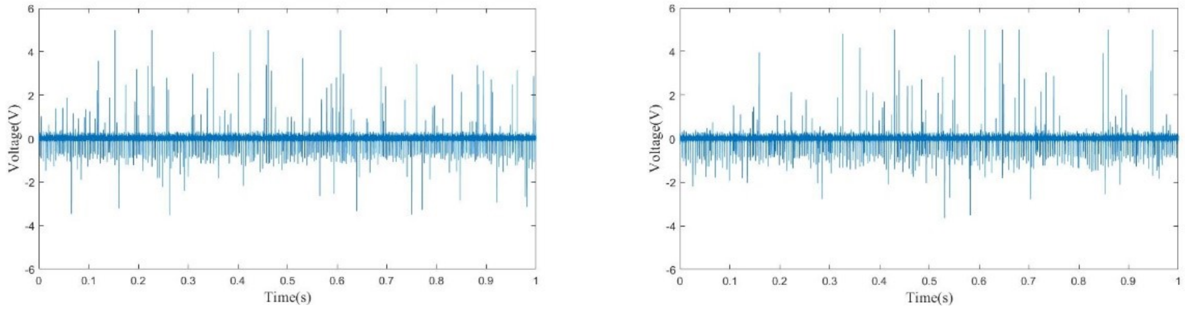
(a) **(b)**

Figure 2. The experiment of PD-induced AE signals detection: **(a)** experimental platform; **(b)** defect models.

PD-induced AE signals were collected over 200.0 s at a sampling frequency of 1.0 MHz and segmented into 1.0-second units for time-domain analysis, as depicted in **Figure 3**. This resulted in a limited dataset of 200 (unit) × 4 (group). Due to the common challenge of limited sample availability in electromechanical fault diagnosis, this study utilized the aforementioned small dataset. The SVM model, which is well-suited for small sample sizes and nonlinear relationships, was employed, also enabling direct comparison with prior work.



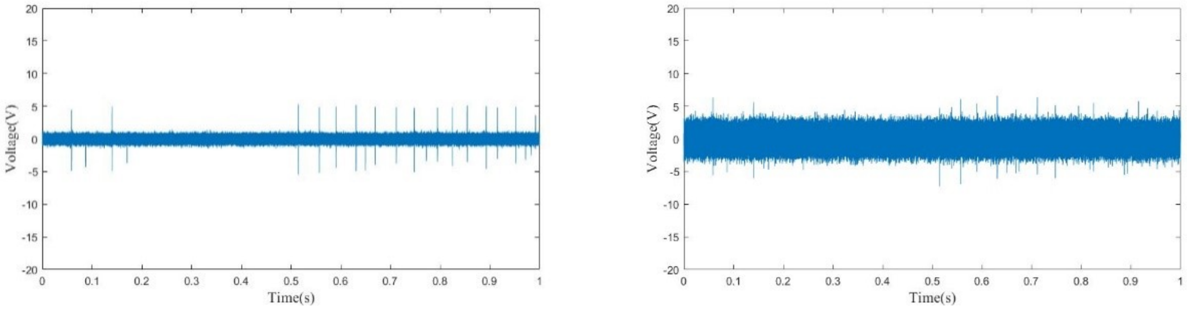
(a) **(b)**



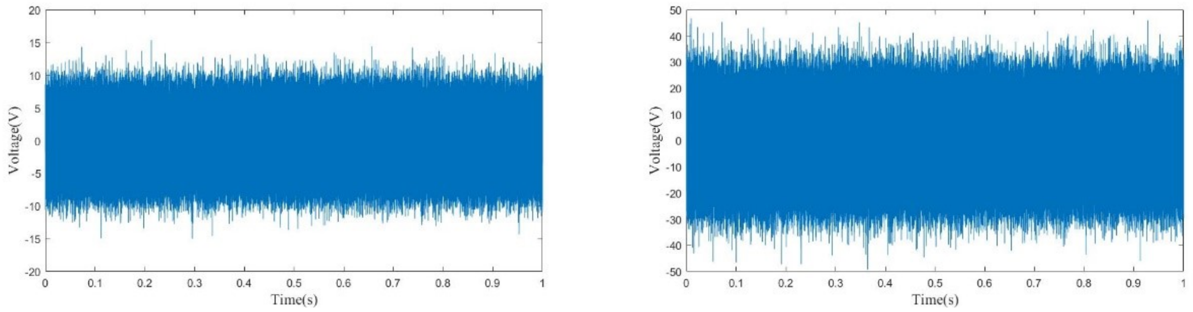
(c) **(d)**

Figure 3. Raw signal 1 s units: **(a)** 1.0 mm; **(b)** 1.8 mm; **(c)** 2.0 mm; **(d)** 2.5 mm.

Figure 3 illustrates that as the diameter of the metal particles increases, the period of PD-induced AE signals progressively shortens, while the frequency of small-amplitude pulses rises. The variation in signal characteristics can be attributed to significant differences in arc resistance, stemming from the varying cross-sectional areas of the discharge channels (maintained at a constant spacing), which convey essential fault source information. During the signal-to-noise ratio analysis, the experiment involved increasing the voltage to 14 kV to amplify PD intensity and utilizing an optical damping platform to mitigate background noise interference. These measures led to the acquisition of AE signals with a notably high signal-to-noise ratio. Therefore, in the subsequent experiments involving noise addition and reduction, the sensor signals were treated as pure for comparative analysis. Additionally, datasets with varying signal-to-noise ratios were created by introducing white noise, with a particular focus on analyzing low signal-to-noise conditions where signals were entirely obscured by background noise and assessing their recognition efficacy. Take the original signal in **Figure 3a** as an example; the amplitude of the original signal is about 5 V. The signal-to-noise ratio shown in **Figure 4a** is −10 dB, and the noise increases obviously; the signal-to-noise ratio shown in **Figure 4b** is −20 dB, and the signal pulse is almost drowned by the noise; and the signal-to-noise ratio shown in **Figure 4c** is −30 dB, and the intensity of the noise is more than 10 V, which completely drowns the signal pulse. The S/N ratio shown in **Figure 4d** is −40 dB, and the noise intensity further increases to more than 20 V.



(a) **(b)**

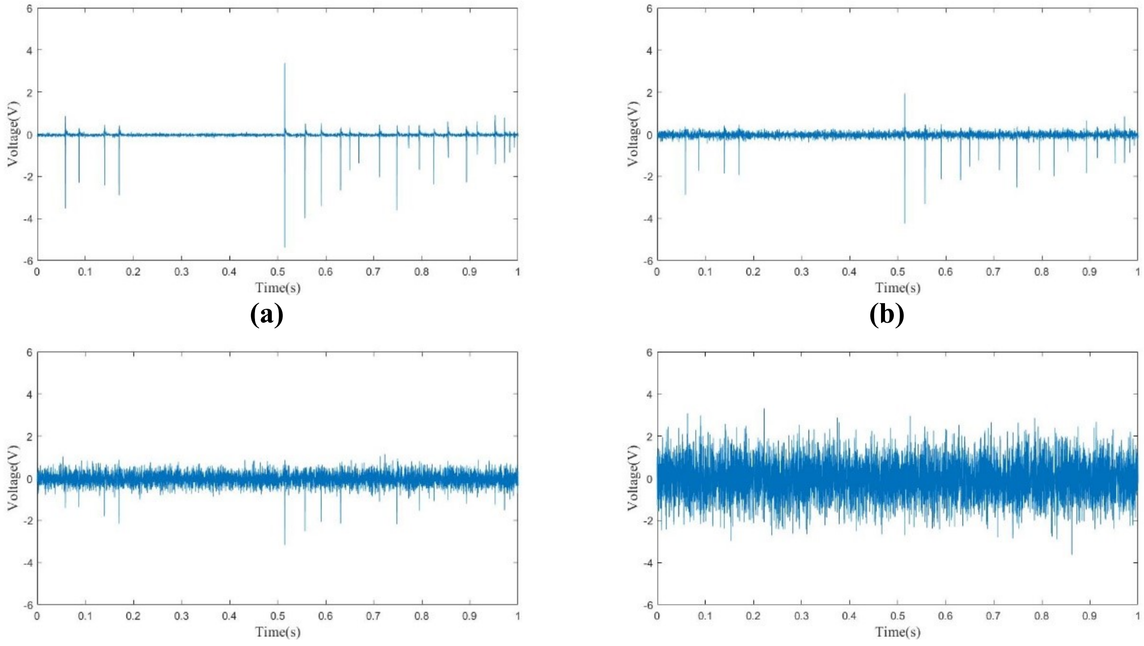


(c) **(d)**

Figure 4. Adding different levels of noise to **Figure 1a**: **(a)** −10 dB; **(b)** −20 dB; **(c)** −30 dB; **(d)** −40 dB.

4.2. Noise reduction

The objective of noise reduction is to eliminate complex exogenous interference while maximizing the preservation of the original source information. In this section, a combination of wavelet threshold denoising and the PCA method is used for noise reduction of AE signals with a low signal-to-noise ratio. Wavelet thresholding denoising finely decomposes the signal into subbands of different frequencies by means of wavelet transform, a time-frequency analysis method in which the high-frequency subbands often contain a large number of noise components. The method utilizes a thresholding mechanism to filter the wavelet coefficients in the high-frequency subbands and sets coefficients smaller than a set threshold to zero, thus effectively suppressing the high-frequency noise while retaining the important features in the signal. This processing benefits from the multi-resolution analysis capability of the wavelet transform, which makes it excellent in dealing with non-smooth signals and detail noise. In the low SNR case, the conventional hard and soft threshold functions may not be effective in removing the noise and retaining the useful information of the signal due to the high noise level. In contrast, the novel threshold wavelet denoising method previously proposed by our team [11] is reconstructed based on the expressions of the hard and soft threshold functions, which utilizes the convexity of the exponential function to ensure the continuity of the function and reduce the constant deviation between the reconstructed signal and the original signal. The new threshold function is able to better balance the denoising effect and signal retention ability by adaptively adjusting the threshold size. The new threshold function is able to remove noise while retaining as much useful information of the signal as possible. This is due to its continuous functional form and adjustable parameter m, which enables the algorithm to maintain a good signal retention ability even with low SNR. The tuning coefficient m is a positive integer that is used to adjust the shape of the new threshold function and the denoising effect. For different SNR signals, m can be adjusted to obtain the optimal threshold function. For different types of noise, wavelet threshold denoising may not be effective in dealing with certain types of noise, especially when the noise overlaps the signal spectrum or has complex correlations. PCA, on the other hand, achieves data reduction and reconstruction by projecting the original variables into a new space consisting of principal components. In this process, PCA reconstruction removes more types of noise, including noise that overlaps with the signal spectrum, by retaining the principal components. In terms of working principle, wavelet thresholding denoising relies on thresholding of local features, while PCA reconstruction removes noise by projection of global feature vectors. This makes PCA reconstruction more robust when dealing with complex noise. The denoising ability of PCA reconstruction is due to its ability to extract the main features of the signal by maximizing the variance, as well as the mechanism of removing noise and redundant information by dimensionality reduction. **Figure 5** illustrates the noise reduction outcomes for the four groups of dyed-noise signals presented in **Figure 4**. Using **Figure 5a** as an illustration, it is evident that our proposed noise reduction scheme can restore the discharge characteristics of the original signals with a dyed noise of −10 dB or −20 dB. In particular, for signals with a dyed noise of −30 dB, the weak discharge characteristics can still be retained by the noise reduction scheme, while for signals with noise levels of −30 dB and below, the discharge characteristics are not discernible in the time-domain signals, and it has been shown in a previous study [15] that when the signal noise level is below −30 dB, it is impossible to identify the type of partial discharge faults by simply extracting the time-frequency-domain features or the MFCC features from the time-domain signals.



(a) (b)

(c) (d)

Figure 5. Denoising the signal in **Figure 2**: **(a)** −10 dB; **(b)** −20 dB; **(c)** −30 dB; **(d)** −40 dB.

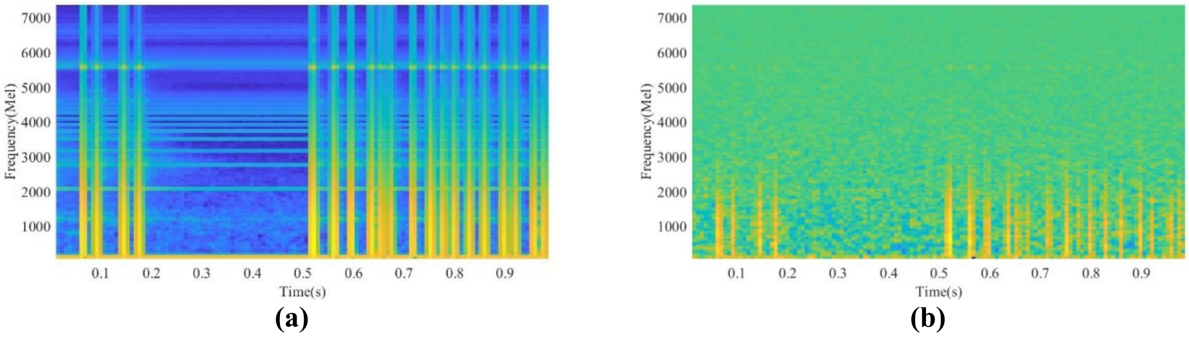
To delve deeper into features that are inadequately captured in time-domain signals, this study focuses on transforming one-dimensional time-domain signals into speech spectrograms and integrating them with deep learning models for feature extraction. Spectrograms integrate information from both time and frequency dimensions, allowing deep learning models, such as CNNs, to learn more intricate and comprehensive feature representations, thereby enhancing the accuracy of fault identification. Furthermore, previous research has relied on manually designed feature extractors, which depend heavily on experts' a priori knowledge. Conversely, spectrograms paired with deep learning models can automatically learn features from signals, diminishing the reliance on manual design and enhancing feature extraction efficiency. Additionally, manually designed feature extractors are intimately tied to the signal itself and struggle to adapt to new data while deep learning models can grasp the intrinsic laws and patterns of the data during training, exhibiting robust generalization capabilities.

4.3. Mel time-frequency spectrum

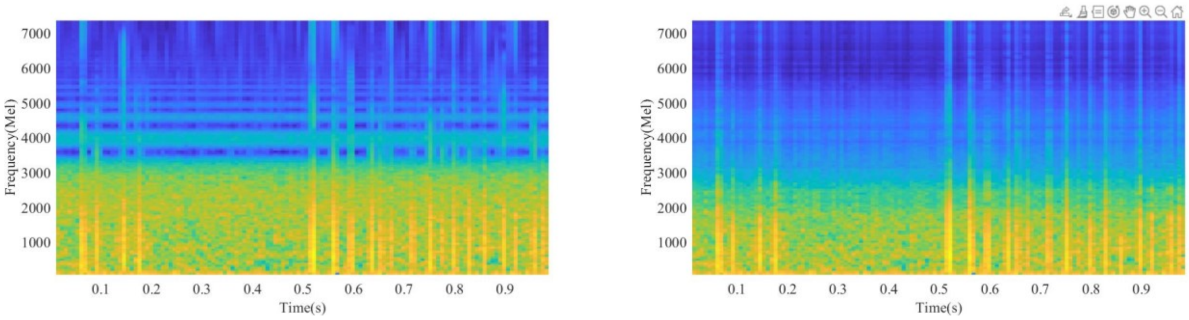
Motivated by human sensitivity to signal variations, researchers have introduced nonlinear transformations on the frequency axis of the short-time Fourier transform, utilizing units known as Mel scales. In speech synthesis, the Mel spectrum functions as a crucial intermediate representation of speech features, effectively bridging the gap between the vocoder and the acoustic model. By analyzing the bandwidth of the frequency band containing the signal's primary energy, an appropriate range is selected for the bandwidth of the delta filter bank. A mathematical mapping exists between the Nyquist frequency and the Mel frequency, with the conversion function defined as follows:

$$Mel\left(f\right)=1127\times\ln\left(1+f/700\right) \tag{10}$$

The Mel spectrum represents a spectrogram in the Mel scale, and the corresponding features are called FBank features. Furthermore, MFCC features can be derived by applying the Discrete Cosine Transform (DCT) to the FBank. The specific approach of this project is shown as follows, which is roughly divided into three steps: A Mel-scale triangular filter bank is established according to the spectrum of the AE signal, and the spectral energy of each frequency band is calculated. To improve the efficiency of the subsequent feature extraction tasks, the Mel spectrum after frequency reduction will be obtained in this step. Experiment has determined that the optimal size for the Mel spectrum is 128×128. Taking 1s samples as an example, each frame is initially set to contain 20480 samples, with an overlap region of 12800 samples between adjacent frames to mitigate excessive variations. Subsequently, each frame is multiplied by a Hamming window to enhance continuity between the frame's endpoints. Finally, a fast Fourier transform (FFT) is applied to obtain the frequency and power spectra. The Mel filter banks were applied to the power spectrum, and the logarithm of the energy of all filter banks was computed to obtain the Mel spectrum for the four discharge types. MATLAB is utilized to generate visualization images of the Mel spectrum for signals at each stage. As an example, −20 dB white noise is added to the initial segment of the 1.8mm fault model. Comparison graphs depicting the signals before and after denoising are presented in **Figure 6**.



(a) **(b)**



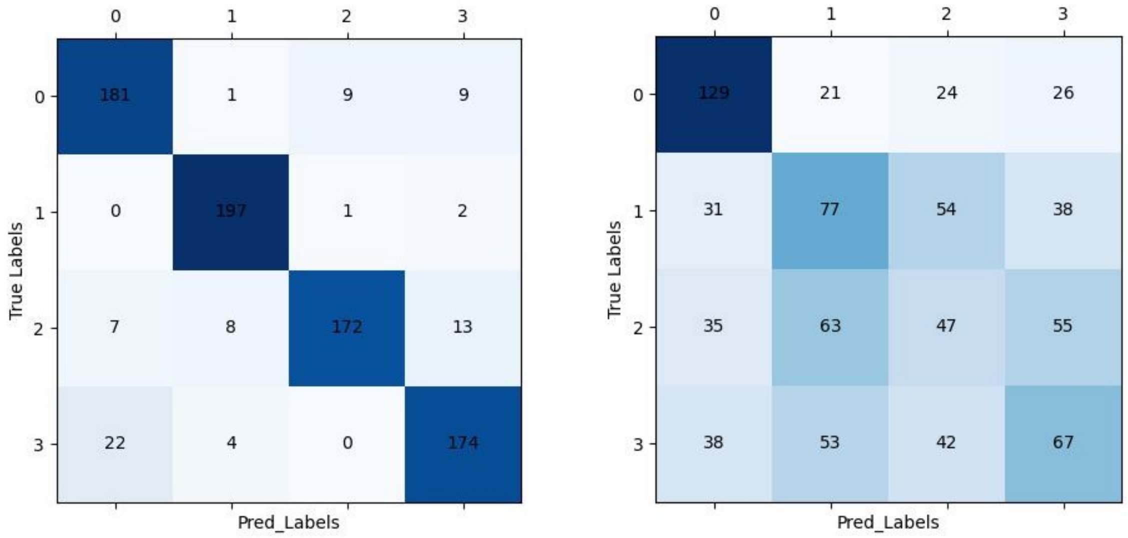
(c) **(d)**

Figure 6. Mel spectrogram: **(a)** original signal; **(b)** additive noise −20 dB for **(a)**; **(c)** wavelet new threshold denoising for **(b)**; **(d)** PCA denoising for **(c)**.

Figure 6a shows the Mel spectrum of the unprocessed signal, reflecting auditory characteristics with a non-linear frequency-Mel scale mapping. The Mel spectrogram of the original signal shows clear, orderly vertical lines representing frequency components over time. When −20 dB noise is added (**Figure 6b**), the spectrogram becomes complex with randomly distributed noise components obscuring the original spectral characteristics. After denoising using the wavelet transform (**Figure 6c**) and a new thresholding method (**Figure 6d**), significant improvements are observed. Wavelet denoising reduces noise and clarifies vertical lines, but horizontal noise lines remain. PCA denoising is more robust, further reducing redundant horizontal lines and enhancing the spectral characteristics of the original signal, resulting in a more concentrated energy distribution and high signal quality.

5. Results and analyses

In contrast with previous studies, this study continues to employ SVM with a Gaussian kernel function for feature classification and fault type diagnosis. Through extensive research, SVM has proven to be adept at handling small-sample, nonlinear, and high-dimensional pattern recognition challenges, emerging as the optimal classifier-based method for detecting PD fault types. For each PD model type, 200 sets of 1 s signals were gathered and subsequently segmented into 2000 sets of 0.1 s signals as auxiliary datasets by wavelet threshold denoising and PCA reconstruction. Mel spectra were extracted from the signal data, and locally consistent feature vectors were obtained using the proposed feature extraction method. These feature vectors were then normalized, with 90% serving as training samples and 10% as test samples for SVM-based fault classification. Meanwhile, this study uses the grid search method to optimize the two hyperparameters of the RBF kernel function, namely the parameter g and the penalty factor C. A 10-fold cross-validation was conducted to demonstrate the generalization and stability of the proposed approach. In order to validate the recognition accuracy of the model, several conventional classification methods were selected for the pattern recognition of four types of GIS local discharge defects. These included a previous study in our lab [11], a classification scheme using MFCC with SVM, and 2DCNN [20] and CFFO-CNN [21], which were compared with the scheme proposed in this study. Taking the feature extraction method as a benchmark, **Table 1** presents the 10-fold average accuracy achieved by the proposed scheme under varying signal-to-noise ratios. **Figure 7** gives the confusion matrix of the recognition results for typical signal-to-noise ratios (−20dB and −30dB).



(a) **(b)**

Figure 7. Classical confusion matrix: (**a**) −20 dB; (**b**) −30 dB.

Table 1. Comparison of average accuracy for different signal-to-noise ratio cases.

Feature Extraction Methods	SNR before noise reduction (dB)						
	−10	−15	−20	−25	−30	−35	−40
	Average recognition accuracy (%)						
MFCC + SVM [11]	91.25	79.13	60.88	(31.50)	(25.00)	-	-
2DCNN [20]	92.00	85.50	70.75	36.25	30.00	26.25	-
CFFO-CNN [21]	93.45	88.62	79.93	40.75	32.25	25.34	-
OSAE-CNN	92.38	91.38	90.88	68.62	45.25	31.13	25.63
	Standard deviation						
MFCC + SVM [11]	0.030	0.051	0.029	(0.043)	(0.056)	-	-
2DCNN [20]	0.064	0.140	0.210	0.045	0.009	0.002	-
CFFO-CNN [21]	0.012	0.210	0.341	0.092	0.010	0.001	-
OSAE-CNN	0.001	0.001	0.000	0.002	0.001	0.002	0.001

Table 1 reveals that, with the SVM model held constant, the recognition accuracy significantly improves when using the OSAE-CNN proposed in this paper for feature extraction and fault classification, outperforming traditional statistical feature combinations. Regarding accuracy upper limits, the average recognition accuracy exceeds 90% at signal-to-noise ratios of −10 dB, −15 dB, and −20 dB, highlighting the accuracy and effectiveness of the proposed fault recognition system. In terms of accuracy improvement, fault type recognition accuracies are 30.00%, 37.00%, and 20.00% higher than previous methods at signal-to-noise ratios of −20 dB, −25 dB, and −30 dB, respectively. Furthermore, as the signal-to-noise ratio decreases, the standard deviation (SD) remains stable at approximately 1.0×10^{-3}, which is an order of magnitude lower than in previous studies, demonstrating the stability and robustness

of the extracted features. Besides, it can be seen that the OSAE-CNN proposed in this paper achieves the best results under each noise intensity. In contrast, although CFFO-CNN and 2DCNN are able to outperform the method of simply extracting MFCC for SVM classification in terms of accuracy, their accuracy fluctuates a lot in terms of the 10-fold standard deviation, which suggests that the performances of these two schemes for small-sample and strongly noisy datasets are questionable. This experiment also verifies that the low-frequency-domain Mel spectrum, based on the frequency-shift phenomenon, contains valuable fault source information. Utilizing it as feature vectors dramatically enhances the accuracy and stability of the pattern recognition algorithm.

6. Conclusion

With the wide application of deep learning to the intelligent diagnosis of defective faults in electrical equipment, the selection of feature quantity is no longer limited to traditional schemes. Among them, to improve the accuracy of the small sample fault classification task, making full use of its detailed information to enrich the global features has proven effective in feature extraction. In this study, an OSAE-CNN-based method is proposed to complete feature extraction on the Mel spectrum of the discharged signal. The method employs sparse self-encoder mining to extract detailed information from PD-induced AE signals for training the CNN convolution kernel. Experimental results indicate that the optimized CNN exhibits strong abstraction capabilities for the Mel spectrum and effectively integrates detailed and global information. The results indicate that, even with noisy signals having signal-to-noise ratios of −20 dB or lower, the proposed AE signal fault detection systems excel in fault classification tasks and attain superior performance limits compared to previous studies. The recognition accuracy under low-quality signals with signal-to-noise ratios below −20 dB is nearly 30% higher than that of all previous studies, and particularly, the recognition accuracy can still reach more than 90% at −20 dB. However, this study still uses SVM for signal classification, leaving considerable scope for enhancing the pattern recognition method. For the small samples and low-quality sample set, the subsequent research plans to abandon SVM and complete fault identification through CNN combined with GAN or an attention mechanism. Furthermore, the universality of the proposed feature extraction method and its adaptability to other pattern recognition approaches require further validation in future work.

Author contributions: Conceptualization, LL, GL, HX and XL; methodology, LL and XL; software, LL; validation, LL, GL, HX and XL; formal analysis, LL, GL, HX and XL; investigation, LL and GL; resources, LL and XL; data curation, LL and GL; writing—original draft preparation, LL; writing—review and editing, LL and XL; visualization, LL; supervision, HX and XL; project administration, HX and XL; funding acquisition, HX and XL. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Natural Science Foundation of China under Grant 62175187; the Shanghai Municipal Science and Technology Major Project under Grant 2021SHZDZX0100; and the Shanghai Municipal Commission of

Science and Technology Project under Grant 19511132101.

Institutional review board statement: Not applicable

Informed consent statement: Not applicable

Conflict of interest: The authors declare no conflict of interest.

References

1. Faizol Z, Zubir F, Saman NM, et al. Detection method of partial discharge on transformer and gas-insulated switchgear: A review. Applied Sciences. 2023; 13(17): 9605.
2. Prajna K, Mukhopadhyay CK. Fractional Fourier transform based adaptive filtering techniques for acoustic emission signal enhancement. Journal of Nondestructive Evaluation. 2020; 39(1): 14.
3. Gu FC. Identification of partial discharge defects in gas-insulated switchgears by using a deep learning method. IEEE Access. 2020; 8: 163894–163902.
4. Hussain GA, Hassan W, Mahmood F, et al. Review on partial discharge diagnostic techniques for high voltage equipment in power systems. IEEE Access. 2023; 11: 51382–51394.
5. Sun W, Ma H, Wang S. A Novel Fault Diagnosis of GIS Partial Discharge Based on Improved Whale Optimization Algorithm. IEEE Access. 2024.
6. Ji H, Liu H, Wang J, et al. Mechanical fault diagnosis of gas-insulated switchgear based on saliency feature of auditory brainstem response under noise background. Measurement Science and Technology. 2023; 35(1).
7. Li D, Wang Y, Yan WJ, Ren WX. Acoustic emission wave classification for rail crack monitoring based on synchrosqueezed wavelet transform and multi-branch convolutional neural network. Structural Health Monitoring. 2021; 20(4): 1563–1582.
8. Lv F, Liu G, Wang Q, et al. Pattern recognition of partial discharge in power transformer based on infogan and cnn. Journal of Electrical Engineering & Technology. 2023; 18(2): 829–841.
9. Wang Z, Zhao Y, Guo J, et al. Partial discharge pattern recognition in gis based on s transform denoising. In: Proceedings of the 22nd International Symposium on High Voltage Engineering (ISH 2021); 21–26 November 2021; Xi'an, China. pp. 897–901.
10. Chaudhuri S, Ghosh S, Dey D, et al. Denoising of partial discharge signal using a hybrid framework of total variation denoising-autoencoder. Measurement. 2023; 223: 113674.
11. Lin Q, Lyu F, Yu S, et al. Optimized denoising method for weak acoustic emission signal in partial discharge detection. IEEE Transactions on Dielectrics and Electrical Insulation. 2022; 29(4): 1409–1416.
12. Gu FC, Chen HC, Chen BY. A fractional Fourier transform-based approach for gas-insulated switchgear partial discharge recognition. Journal of Electrical Engineering & Technology. 2019; 14: 2073–2084.
13. Zheng J, Chen Z, Wang Q, et al. GIS partial discharge pattern recognition based on time-frequency features and improved convolutional neural network. Energies. 2022; 15(19): 7372.
14. Sun W, Ma H, Wang S. Application of SCNGO-VMD-SVM in Identification of Gas Insulated Switchgear Partial Discharge. IEEE Access. 2024; 12: 43838–43848.
15. Lyu F, Yang Z, Wang L, et al. Behavior anomaly detection fused with features of Mel frequency cepstrum coefficients. In: Proceedings of the 2020 International Conference on Sensing, Measurement & Data Analytics in the era of Artificial Intelligence (ICSMD); 15–17 October 2020; Xi'an, China. pp. 89–93.
16. Lu L, Tao W, Liu G, et al. Effectiveness Analysis of the Mel spectrum features of AE signals in the detection of partial discharge faults. In: Proceedings of the 2022 International Conference on Sensing, Measurement & Data Analytics in the era of Artificial Intelligence (ICSMD); 30 November–2 December 2022; Harbin, China. pp. 1–5.
17. Liu T, Yan J, Wang Y, et al. GIS partial discharge pattern recognition based on a novel convolutional neural networks and long short-term memory. Entropy. 2011; 23(6): 774.
18. Jing Q, Yan J, He R, et al. Intelligent diagnosis of GIS partial discharge via stacked autoencoder and transfer learning. In: Proceedings of the 2022 6th International Conference on Electric Power Equipment-Switching Technology (ICEPE-ST); 15–18 March 2022; Seoul, Korea.
19. Tian J, Song H, Sheng G, Jiang X. Knowledge-driven recognition methodology of partial discharge patterns in GIS. IEEE Transactions on Power Delivery. 2021; 37(4): 3335–3344.

20. Rauscher A, Kaiser J, Devaraju M, Endisch C. Deep learning and data augmentation for partial discharge detection in electrical machines. Engineering Applications of Artificial Intelligence. 2024; 133: 108074.
21. Gao A, Zhu Y, Cai W, Zhang Y. Pattern recognition of partial discharge based on VMD-CWD spectrum and optimized CNN with cross-layer feature fusion. IEEE Access. 2020; 8: 151296–151306.