

Transformer Fault Detection Based on Partial Discharge Information Fusion and Machine Learning

Jian Wang*

State Grid Shandong Electric Power Research Institute, Jinan 250102, P. R. China
18511228492@163.com

Wen-Bing Zhu

State Grid Shandong Electric Power Research Institute, Jinan 250102, P. R. China
13526520680@163.com

Chao Gu

State Grid Shandong Electric Power Research Institute, Jinan 250102, P. R. China
talon521@163.com

Meng-Zhao Zhu

State Grid Shandong Electric Power Research Institute, Jinan 250102, P. R. China
min_ok@163.com

Zhao-Liang Gu

State Grid Shandong Electric Power Research Institute, Jinan 250102, P. R. China
sail-xsb@163.com

Tian-Feng Liu

College of Sciences.
Adamson University, Manila 1008, Philippines

*Corresponding author: Jian Wang

Received January 17, 2024, revised April 30, 2024, accepted July 15, 2024.

ABSTRACT. *Conventional transformer fault detection based on partial discharge signal detection suffers from the problem of failing to retain complete short-time spectral features. In addition, most of the machine learning-based detection models suffer from the disadvantage of lack of sparsity, which cannot be effectively applied to small-sample and unbalanced transformer partial discharge sound signals, leading to the problem of reduced classification accuracy easily. Therefore, a transformer fault identification method based on partial discharge information fusion and sparsifying machine learning techniques is proposed. Firstly, in order to retain more complete short-time spectral features, a Mel filter is used to obtain the Mel spectrum. Considering that the Mel filter cannot fully capture the high-frequency features of partial discharge sound signals, the inverse Mel spectrum is obtained by an inverse Mel filter. The method of feature fusion was adopted, and the Mel spectrum and the inverse Mel spectrum were feature fused to obtain the three-channel fused feature spectrum. Secondly, to address the lack of sparsity prevalent in Least Squares Support Vector Machine (LS-SVM), it is proposed to map the data samples into a high-dimensional kernel space and perform kernel space distance clustering on the mapped data by spectral clustering algorithm to achieve data preprocessing for LS-SVM, thus realising its sparsification. The experimental results show that for six types of transformer faults, the average diagnosis accuracy of sparsified LS-SVM is 88.3%. In addition, compared with the single spectral feature, the proposed fused spectral feature has richer acoustic features and can improve the performance of the classifier by nearly 5%.*

Keywords: Transformer faults; partial discharges; audio recognition; support vector machine; kernel function

1. **Introduction.** With the rapid development of the economy, power systems are widely used in various industries. Among them, the importance of high-voltage power transmission equipment cannot be ignored. However, switchgear, high-voltage transmission cables, and power transformers in high-voltage transmission equipment have been operated under high electric field strength and complex working conditions for a long time, and also affected by external factors such as external damage, natural disasters, and incorrect operation during installation and maintenance [1, 2], which lead to the degradation of the insulating performance of high-voltage transmission equipment. Partial discharge phenomenon occurs when the electric field strength of the insulator reaches the breakdown electric field [3, 4].

Partial discharges can cause galvanic corrosion and thermal corrosion of insulating materials, thus accelerating the occurrence of partial discharge phenomenon, which leads to accidents in the power system and causes serious power safety problems [5, 6]. A large number of research data show that the insulation system damage caused by partial discharge is the main reason for the failure of high-voltage transmission and substation equipment, occupying nearly 80% of the failure ratio [7]. In summary, the voltage electrical equipment monitoring system based on partial discharge signal detection is the main direction of the development of new high-voltage electrical equipment, and the intelligent monitoring platform built by using artificial intelligence and Internet of Things (IoT) technology is of great significance to improve the efficiency of the equipment and avoid the occurrence of electrical accidents.

Relative to identification methods such as vibration, sound detection has natural advantages such as non-contact, small size, low cost, easy to implement, and convenient installation of sensors, etc. However, the method of detecting by sound does not reflect the size of the discharge volume of a discharge, and it is difficult to assess the severity of fault damage [8, 9]. From the point of view of cost as well as monitoring and early warning, the continuous maturity of sound detection technology will make the acoustic detection method gradually become the mainstream method in partial discharge detection

methods. With the development of sound recognition technology and the generation of microphone array accurate sound positioning equipment, the use of sound signals in the partial discharge process to detect and locate partial discharges has become another new idea in this field.

1.1. Related Work. Traditional partial discharge identification methods often depend on experienced inspectors to identify whether the operating conditions of the equipment are abnormal or faulty by the sound of the equipment [10]. The use of audio recognition technology and the adoption of artificial intelligence algorithms to replace manual inspection is one of the goals of research in related fields.

Sadhu et al. [11] proposed a study on fault sound recognition of industrial equipment based on Hidden Markov model (HMM). Yaacob et al. [12] proposed a substation line detection based on discharge sound signals. Khan et al. [13] used Support Vector Machine (SVM) to identify audio signals of surface discharges on 10 kV solid insulated cabinets. Mnasri et al. [14] proposed a modelling system using staged speech recognition using HMM as an approximation of the sound density function for monitoring abnormal sounds in industrial equipment. The first stage of the system classifies the captured sounds into two categories, normal and abnormal sounds, and if the sound is determined to be abnormal, it proceeds to the next stage for further classification. The second stage is used to determine the type of abnormal sound, such as gunshots, explosions or screams. The final experimental results show that the system has a high detection rate for detecting abnormal sounds in noisy environments and the system is robust.

In recent years, with the wide application of machine learning technology, studies have shown that machine learning models are more effective than traditional classification models in dealing with complex sound classification problems. In the presence of a large amount of background noise interference, machine learning models can also effectively extract the features of the sound signal to achieve accurate classification and recognition. Currently, sound features based on human ear hearing have been widely used in feature extraction of machine learning sound models for speech recognition or classification. Demir et al. [15] used two Convolutional Neural Network (CNN) models with two different architectures to classify sounds with a large amount of ambient noise, one for classifying Mel's spectral features [16] and one for classifying the raw sound data, and finally achieved a classification accuracy of more than 85% on the dataset. Abdelhamid et al. [17] proposed an online speech emotion recognition system using CNN+Long Short-Term Memory (LSTM). Tippannavar et al. [18] proposed an online speech emotion recognition system based on CNN and Least Square Support Vector Machine (LS-SVM) network combination model using CNN to extract the best descriptive features of the speech signal and finally classify it with LS-SVM, which performs well on the dataset and obtains an average accuracy of 88%.

1.2. Motivation and contribution. When the transformer equipment works normally, the spark discharge and partial discharge have more energy overlap frequency bands, through the time-frequency analysis can distinguish between the normal operation of the equipment and the partial discharge phenomenon, but it can't distinguish the type of fault of the equipment. This is the common problem that exists in most of the current traditional transformer fault detection methods based on partial discharge signal detection. In addition, small and unbalanced samples are very common phenomena in most real-world data sources, and transformer partial discharge sound signals are no exception. However, the poor sparsity of LS-SVM leads to the problem of reduced classification accuracy when dealing with such samples.

In order to solve the above two problems, a partial discharge sound signal recognition method based on fusing spectral features with sparsified LS-SVM is proposed. The main innovations and contributions of this work include:

(1) Aiming at the problem that the traditional feature extraction method cannot retain the complete short-time spectral features, from the perspective of human hearing, the acoustic scale that is the most compatible with human hearing - Mel scale - is selected, and the Mel scale is used to get the short-time spectral features to retain the more complete Mel spectral features.

(2) Aiming at the poor sparsity of LS-SVM, it is proposed to map the data samples into a high-dimensional kernel space and perform kernel-space distance clustering on the mapped data by means of a spectral clustering algorithm in order to realise the data preprocessing of LS-SVM, thus realising its sparsification.

2. Time-frequency feature extraction based on partial discharge information fusion. In order to achieve speech separation and speech recognition of transformer partial discharge sound signals, it is necessary to extract meticulous sound features, such as pitch, volume and timbre. In this paper, we first study the theoretical basis of sound signal processing and preprocess the partial discharge sound signal.

2.1. Pre-processing of transformer partial discharge sound signals. Transformers can be divided into two categories: dry-type and oil-immersed transformers, the difference between them lies in the different cooling methods for the windings and the core and other major working parts. Dry-type transformers use air cooling and epoxy resin to wrap the windings as insulation and heat transfer material, while oil-immersed transformers use transformer oil to take away the heat generated by the work of the windings and core and to provide insulation protection. The object of study in this paper is the oil-immersed transformer, which has more heat dissipation capacity and more components, and the specific structure is shown in Figure 1 (hereinafter, the oil-immersed transformer will be referred to as the transformer).

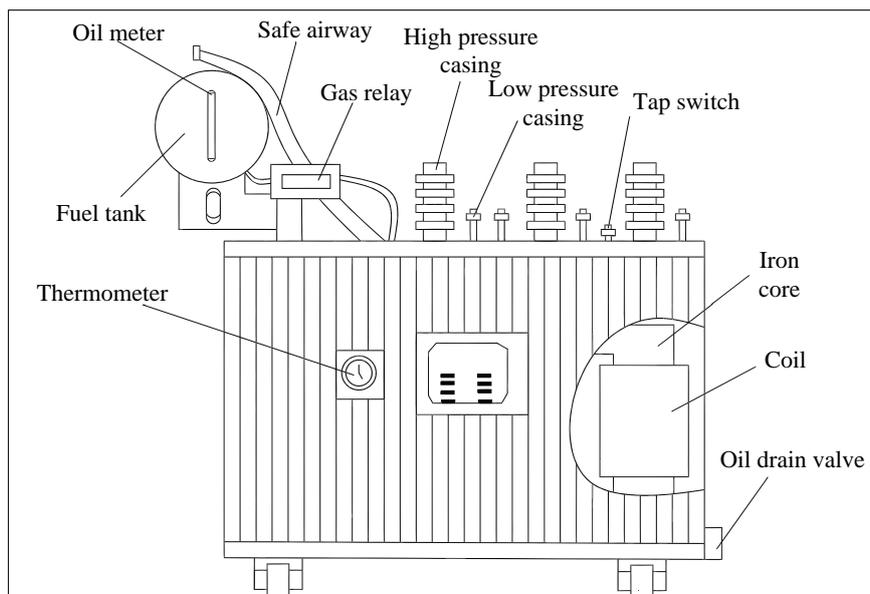


Figure 1. Transformer construction

The energy of mechanical noise is mainly distributed around 1KHz, while the high energy part of partial discharge sound is distributed around 5KHz-10KHz. There is an

obvious difference between the sound characteristics of the environment during normal operation of the transformer and those of the partial discharge, so it is necessary to pre-emphasise the high frequency part to highlight the high frequency characteristics of the partial discharge and to make up for the loss of its capacity. A first-order high-pass filter is usually used to realise the pre-emphasis of the sound signal, and its expression is as follows:

$$H(z) = 1 - \mu z^{-1} \quad (1)$$

where μ is the pre-emphasis factor, which usually takes the value of 0.96.

After the first-order high-pass filtering and pre-emphasis of the original partial discharge signal, part of the low-frequency signal is filtered out, and the high-frequency signal is preserved to be used for high-frequency feature extraction for the subsequent analysis. After the first-order high-pass filtering and pre-emphasis, the discharge sound signal is a typical non-smooth signal, and its signal is in a smooth state only in a short period of time. If the Fourier Transform (FFT) [19] is performed directly, it will make the discharge sound signal transformed in the whole duration, which will not be able to get the timing information of the change of the discharge sound signal, and at the same time, the localised features of the non-smooth discharge sound signal will be preserved. Local features of the non-smooth discharge sound signal cannot be focused. Therefore, in order to solve this problem, the STFT is adopted, which can make the FFT effectively extract the local features and judge the sequence of signal appearance, so that both local and temporal information can be preserved. The core idea of the STFT is to add a window to the original signal by frames, and then do a Discrete Fourier Transform (DFT) [20] on the signal after adding the window. Observed from the overall point of view of the discharge sound signal, the signal is not smooth as a whole, but it is in a smooth state in the time range of 10-30 ms.

STFT is defined as follows:

$$x(n, w) = \sum_{m=-\infty}^{\infty} x(m)w(n - m)e^{-jwm} \quad (2)$$

where $w(m)$ is the window function used when adding windows. Generally, frame shifting is required after overlapping segments when performing frame splitting operations, as a way to avoid signal loss when analysing sound.

STFT uses window functions to achieve signal framing, so the choice of window functions is crucial to the quality of the signal after framing. Different window functions have different effects on the signals, because different window functions produce signals with different leakage sizes and different frequency resolution capabilities. From the above analysis of the partial discharge sound signal, it can be concluded that the energy of the partial discharge sound signal is mainly distributed in the high-frequency region and the high-energy frequency has a certain span range, so it is necessary to choose a window function suitable for multi-frequency components and complex spectral analysis.

The Hamming window with more complex weighting coefficients has a smaller side-lobe [21], so its low-pass is better than other window functions (the signal spectrum after adding the window can be closer to the original signal spectrum), so this paper chooses the Hamming window as the window function of the STFT to preprocess the partial discharge sound signals, and its window function is defined as follows:

$$w(n) = \begin{cases} 0.54 - 0.46 \cos\left(\frac{2\pi n}{M-1}\right), & 0 \leq n \leq M-1 \\ 0, & \text{otherwise} \end{cases} \quad (3)$$

where $n = 1, 2, \dots, N - 1$, N denotes the total length of the window function, and M is the effective length of the window function.

2.2. Feature extraction. When the transformer equipment works normally, the sound signal energy is mainly concentrated in the low frequency band of 2K 5KHz. The energy of partial discharge is mainly concentrated in the middle and high frequency band of 5K 10K, while the energy of spark discharge is more concentrated in the high frequency band of 7K 15K. After the time-frequency analysis of partial discharge sound signals, this work improves the traditional sound feature extraction technique. Two types of filters, the Mel filter [22] and the inverse Mel filter [23], are used for feature extraction, in which the Mel filter can effectively capture the low and medium frequency features of the partial discharge sound signal, while the inverse Mel filter can effectively capture the high frequency features of the partial discharge sound signal.

(1) Mel spectral feature extraction.

Research has shown that humans do not perceive frequencies linearly and are more sensitive to low-frequency signals than high-frequency signals. For example, humans can easily detect the difference between 500Hz and 1000Hz, but find it difficult to detect the difference between 10000Hz and 12000Hz. The auditory mechanism of the human ear can be more closely approximated by using the Mel scale, which is a nonlinear transformation based on the spectrum, and its conversion equation is defined as follows [24]:

$$mel = 2595 \cdot \log_{10} \left(1 + \frac{f}{700} \right) \quad (4)$$

where f is the frequency of the signal.

Select M frequency points at equal intervals throughout the frequency domain, take the corresponding values of these frequency points in the Mel scale as the centre frequency point $f(m)$, and use the Hamming window to add a window to the signal in order to construct a Mel filter bank with a frequency corresponding to $H_m(k)$, whose frequency domain response $H_m(k)$ is defined as follows:

$$H_m(k) = \begin{cases} 0 & k < f(m-1) \\ \frac{2(k-f(m-1))}{(f(m+1)-f(m-1))(f(m)-f(m-1))} & f(m-1) \leq k \leq f(m) \\ \frac{2(f(m+1)-k)}{(f(m+1)-f(m-1))(f(m+1)-f(m))} & f(m) \leq k \leq f(m+1) \\ 0 & k > f(m+1) \end{cases} \quad (5)$$

The number of Mel filters selected in this paper is 188. For the partial discharge sound signal dataset, the window functions are mainly concentrated in the low-frequency band 0-5KHz, but in the high-frequency band above 12KHz, the window functions are less distributed. The main range of the Mel filter bank is in the low frequency band, which can effectively capture the low frequency characteristics of the partial discharge sound signal. However, the energy of the discharge sound signal is widely distributed in the frequency range, and also has high energy in the high frequency band, so the Mel spectrum cannot represent all the characteristics of the partial discharge sound signal.

(2) Inverse Meier spectrum feature extraction

It can be seen from the above that the Mel filter has some limitations. The Mel spectrum can characterise the low frequency part of the partial discharge sound, but it cannot fully characterise the high frequency information. Therefore, it is necessary to optimise the characteristics of the Mel spectrum. In order to compensate for the limitations of the Mel filter, the inverse Mel filter, which is complementary to the Mel filter, is used, and the inverse Mel scale of the inverse Mel filter is defined as follows:

$$imel = 2195.286 - 2595 \log_{10} \left(1 + \frac{4031.25 - f_{\text{Hz}}}{700} \right) \quad (6)$$

Similarly, the number of inverse Mel filters chosen in this paper is 188. Based on the partial discharge sound signal dataset in this paper, the inverse Mel filter mainly acts on the high frequency part of the signal, which can be complemented with the Mel filter.

2.3. Fused spectrum feature extraction. The fused spectral feature extraction process proposed in this paper is shown in Figure 2.

Firstly, the Mel spectrum and the inverse Mel spectrum of the partial discharge sound signal are extracted according to the above method and converted into feature matrices as $X \in \mathbb{R}^{188 \times 188 \times 1}$ and $Y \in \mathbb{R}^{188 \times 188 \times 1}$, respectively.

The features are then fused in the channel dimension, with the Mel Spectrum feature matrix X constituting the first channel data of the fusion matrix. The inverse Mel Spectral Feature Matrix Y constitutes the second channel of the fused features. For the third channel of the fused features, the first 94 dimensions of the Mel Spectral Feature Matrix and the last 94 dimensions of the Inverse Mel Spectral Feature Matrix are used to form a 188-dimensional matrix as the third channel of the fused features.

Finally, based on the partial discharge sound signal, the acoustic fusion feature spectrograms of three channels $Z \in \mathbb{R}^{188 \times 188 \times 3}$ are constructed as the feature inputs for the subsequent machine learning recognition model.

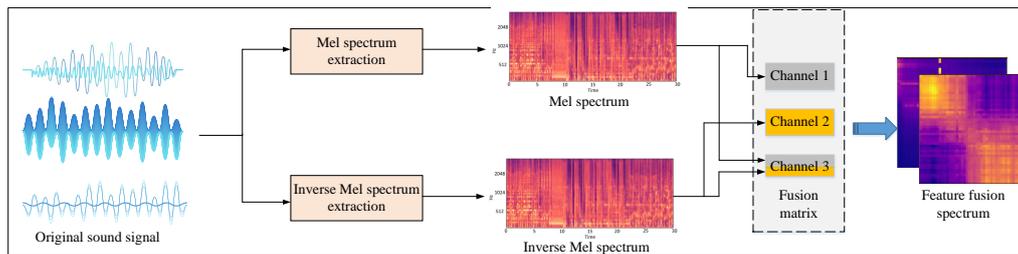


Figure 2. Flow Chart of Fusion Feature Extraction

3. Machine Learning Based Partial Discharge Acoustic Signal Recognition.

3.1. Problems with LS-SVM. SVM is a commonly used discriminative method in the field of machine learning, which can usually be used for pattern recognition, classification and regression analysis. LS-SVM is widely used because of its simplicity and speed in solving small-sample, nonlinear and high-dimensional pattern recognition. Compared with SVM, LS-SVM algorithm has good generalisation ability and is widely used because the overall algorithm has been optimised to have a simpler structure, and the computational efficiency has been improved [25, 26]. However, compared with the SVM model, the LS-SVM model has the disadvantage of lack of sparsity. The Lagrange multiplier of LS-SVM is $\alpha_i = C\xi_i$. The comparison of the two models is based on the distribution of values of α_i , as shown in Figure 3.

It can be clearly seen that the curve of the SVM model increases with i and after a certain point the Lagrange multiplier part will be equal to zero [27]. However, the LS-SVM curve does not keep infinitely converging to 0. In contrast to SVM, most of the values of α_i of the LS-SVM model are not 0, which results in none of the Lagrange multipliers being 0. This shows that the LS-SVM model is not sparse. Small, unbalanced samples are a very common phenomenon in most real-world data sources, and transformer partial

discharge sound signals are no exception. However, the poor sparsity of LS-SVM leads to the problem of reduced classification accuracy when dealing with such samples.

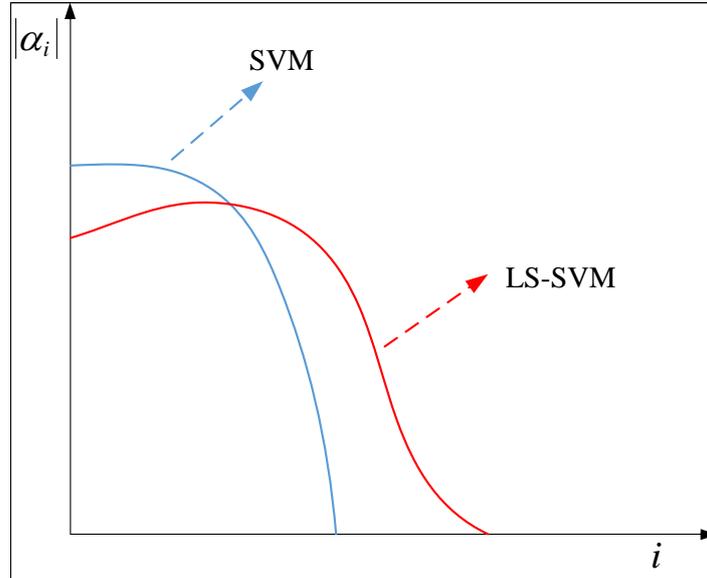


Figure 3. Schematic comparison of SVM and LS-SVM

3.2. Sparsified LS-SVM based on kernel space clustering. Kernel Space Distance Clustering (KSDC) is a data preprocessing method used to map data samples to a high-dimensional kernel space, which allows for the downscaling of the data and feature extraction.

Therefore, this paper proposes to adopt the KSDC method to preprocess the data of LS-SVM in order to achieve the sparsification of the LS-SVM model, so as to improve the efficiency and generalisation performance of the model. In addition, different from the traditional KSDC based on division clustering, this paper proposes to use the spectral clustering algorithm based on graph theory [28] to cluster the mapped data, and the specific steps of implementation are as follows:

Firstly, the kernel function commonly used in LS-SVM is similar to the traditional SVM, which achieves nonlinear modelling capability by mapping the input samples to a high-dimensional feature space. In this work, the polynomial kernel function is chosen to implement data preprocessing for kernel space distance clustering.

$$K(x, x') = (\gamma x^T x' + r)^d \quad (7)$$

where γ is the scale parameter, r is the constant term, and d is the number of polynomials.

A polynomial kernel function is used to map the input data samples to a high dimensional kernel space, specifically using an implicit mapping method and using a Kernel Trick to reduce the computational complexity by avoiding the explicit computation of the feature vectors in the high dimensional space. The implicit mapping is computed as follows:

$$K'(x, y) = \phi(x) \cdot \phi(y) \quad (8)$$

where $K'(x, y)$ denotes the value of the kernel function computed in the low-dimensional space; $\phi(x)$ and $\phi(y)$ denote the implicit mapping of the input data samples x and y , respectively, in the high-dimensional kernel space.

Using the Euclidean distance metric in a high dimensional kernel space, the distance between samples is computed in the kernel space which can be shown as follows:

$$D(\phi(x_i), \phi(x_j)) = \sqrt{\|\phi(x_i)\|^2 + \|\phi(x_j)\|^2 - 2\langle\phi(x_i), \phi(x_j)\rangle} \quad (9)$$

where x_i and x_j are the features of the two samples in the input space respectively; $\phi(\cdot)$ denotes the kernel mapping function; $\|\cdot\|$ denotes the norm of the vector; $\langle\cdot, \cdot\rangle$ denotes the inner product of the vector.

Instead of the inner product $\langle\phi(x_i), \phi(x_j)\rangle$, we can use the kernel function $K(x_i, x_j)$ in the computation, thus avoiding explicitly computing vector representations in the high-dimensional kernel space.

$$D(\phi(x_i), \phi(x_j)) = \sqrt{K(x_i, x_i) + K(x_j, x_j) - 2K(x_i, x_j)} \quad (10)$$

In this way, we can obtain the Euclidean distance between samples in the high-dimensional kernel space by calculating the kernel function in the input space, and thus use the Euclidean distance metric in the kernel space-based sparsified LS-SVM classifier for sample clustering.

Secondly, spectral clustering is applied to the mapped samples in the kernel space. Spectral clustering achieves cluster category division by graph partitioning method, and the number of partition subsets is equal to the cluster categories, and spectral clustering needs to solve the vertex similarity and the eigenvalues of the partition subsets.

Let the graph $G = (V, E)$ contain a total of V vertices and the set of edge relations formed by n vertices is $E = \{e_{ij} = \langle v_i, v_j \rangle | v_i, v_j \in V\}$, where the degree of similarity between vertices v_i and v_j is w_{ij} , $w_{ij} \geq 0$.

$$w_{ij} = \begin{cases} \exp\left(-\frac{d(v_i, v_j)^2}{\sigma^2}\right), & e_{ij} \in E \\ 0, & \text{otherwise} \end{cases} \quad (11)$$

where σ is a constant and $d(v_i, v_j)$ denotes the distance between two vertices.

Let the normalised cut set classify the graph G into k categories (A_1, A_2, \dots, A_k) , and the k categories are represented by the set $h_j = \{h_{1j}, h_{2j}, \dots, h_{nj}\}$ for computation.

$$h_{ij} = \begin{cases} \frac{1}{\sqrt{\text{vol}(A_j)}}, & v_i \in A_j \\ 0, & \text{otherwise} \end{cases} \quad (12)$$

$$\text{vol}(A_i) = \sum_{v_i \in A_i} d_i \quad (13)$$

$$d_i = \sum_{j=1}^n w_{ij} \quad (14)$$

The set of categories is constructed as a sub-set $H = \{h_1, h_2, \dots, h_k\}$. For k categories, the normalised cut set can be optimised to the features corresponding to the k eigenvalues of L_s to obtain the features corresponding to L_r . The spectral clustering algorithm groups the samples into different categories to form the clustering result.

Finally, based on the clustering results, the representative samples in each category are screened out, e.g., the clustering centre in each category is selected as the representative sample. The screened representative samples are used as the training set to train the LS-SVM model. Due to the preprocessing of kernel spatial distance clustering, the number of samples in the required training set is less, which achieves the sparsification of the LS-SVM classifier.

For example, assuming that N is the set of training sample data, M_1 and M_2 need to each apply a spectral clustering algorithm to compute the centre of each sample cluster as a representative sample for the class A samples M_1 and class B samples M_2 in the training set.

$$d_i = \sum_{j=1}^k d(x_i, y_j) \quad (15)$$

$$d_j = \sum_{i=1}^k d(y_j, x_i) \quad (16)$$

where d_1 is the nuclear distance from each class in M_1 to M_2 and d_2 is the nuclear distance from each class in M_2 to M_1 .

The screened representative samples are recorded as the new training set N' , and modelled using N' to obtain the sparsified LS-SVM classifier based on kernel space clustering. The sparsification principle of kernel space clustering is shown in Figure 4.

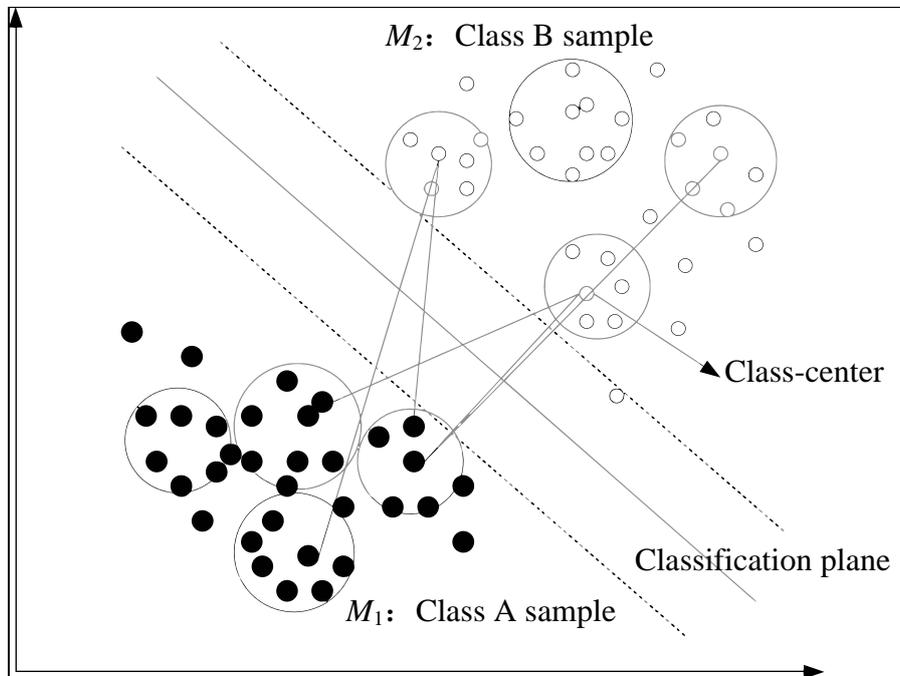


Figure 4. Sparse principle of kernel space clustering

3.3. Steps for fault detection. The specific steps for fault detection of the transformer using sparsified LS-SVM are as follows:

Step 1: Based on the spectrogram dataset extracted by fusing spectral features, construct a partial discharge energy level binary classification dataset (including only partial discharge and spark discharge), load the spectrogram to convert it into a feature vector;

Step 2: Normalise the feature matrix in Step 1. In order to reduce the amount of computation, the feature size is set to 32×32 as the input of the machine learning model. Create a sparse LS-SVM classifier and set the parameters, including kernel function, kernel function coefficients, maximum number of iterations, and so on;

Step 3: Use the proposed sparsified LS-SVM classifier to train the model on the data in the training set and solve the parameters of the classifier by minimising the objective function;

Step 4: Train the sparsified LS-SVM classifier until the error requirement is met or the maximum number of iterations is reached, and save the classifier parameters as a .pkl file for later testing;

Step 5: Identify the new samples using the trained sparsified LS-SVM classifier and calculate the accuracy, F1-score.

4. Experimental results and analyses.

4.1. Experimental environment design. Pycharm software was used to design the experimental procedure, and the hardware and software parameters used for the experiment are shown in Table 1. When using deep learning for model experimental training, the GPU module is used to increase the training speed to accelerate the training of the network. The Pytorch deep learning framework is used to build the model in this experiment. The python environment suitable for this experiment was chosen to design the programme and the experiment was developed in Windows.

Table 1. The software and hardware environment used in this experiment

Appliances	Descriptions
Cpu	Inter(R) core (TM) i7-12700 3.9GHz
Ram	32.00GB
Display card	NVIDIA RTX3090T
Operating system	Windows 11
Software environment	Python 3.8, Pytorch 1.13.1, CUDA 11.8

The experimental dataset was obtained from the transformer public database IEC TC10. 40 sets of sample data were selected for each category, totalling 150 sets. Five groups are selected as representative from each class of samples. The kernel parameter of the sparsified LS-SVM classifier is 0.6 and the penalty factor is 0.9. 80% of the sound sample data is selected as the training set for the experiment and 20% of the sound sample data is selected as the test set.

4.2. Validation of the effect of sparsification. In order to validate the effect of sparsification, the sparsified LS-SVM and LS-SVM are tested as a control group. Using the fused spectral features as inputs, the sparsified and un-sparse classifiers are trained and compared by applying the same data.

Table 2. Comparison of the accuracy of the two classifiers

Types of classifiers	Training data	Accuracy
Sparsification LS-SVM	Normalcy	0.9714
	Introduce a fault	0.8727
LS-SVM	Normalcy	0.9416
	Introduce a fault	0

It can be seen that the classification accuracy of the sparsified LS-SVM training data under the same conditions is 0.9714, while the classification accuracy of the typical LS-SVM is 0.9416, and the comparison shows that the sparsified classifiers have a certain effect on the classification accuracy. When fault data is added to the training data 250 to 310, the comparison of the accuracy of the two classifiers is shown in Table 2. It can be seen that when the sparsifying classifier has fault data in the training data, it has little effect on the classification accuracy. Therefore, compared with the typical LS-SVM, the sparsified LS-SVM then has better robustness and higher classification accuracy.

4.3. Fault detection effect analysis. Next, Mel-Frequency Cepstral Coefficients (MFCC) features, Inverse Mel-Frequency Cepstral Coefficients (IMFCC) features, and fused spectral features, respectively, are input to the two classifiers (LS-SVM and sparsified LS-SVM) for experimental comparison. The performance of the two models in partial discharge sound recognition was compared in order to verify whether the fused features could improve the accuracy of partial discharge sound detection. Using accuracy as well as F-score as evaluation metrics, the experimental results of the two classifiers at different features are shown in Table 3.

Table 3. Experimental results of the two classifiers at different features

Methodologies	Diagnostic property	Accuracy/%	F-score
LS-SVM	MFCC	53.4981	0.65
	IMFCC	52.5912	0.64
	Fusion of spectral features	57.9167	0.7
Sparsification LS-SVM	MFCC	61.8495	0.74
	IMFCC	60.6386	0.73
	Fusion of spectral features	66.5557	0.78

It can be seen that the sparsified LS-SVM has a better performance in the sound feature dataset compared to LS-SVM. The accuracy of sparsified LS-SVM is higher than that of LS-SVM on MFCC, IMFCC and fused spectral features. Meanwhile, both classifiers have better performance when using fused spectral features, and the classification accuracies are much higher than that of MFCC and IMFCC, with a performance enhancement of nearly 5%, which indicates that the fused spectral features have richer acoustic features.

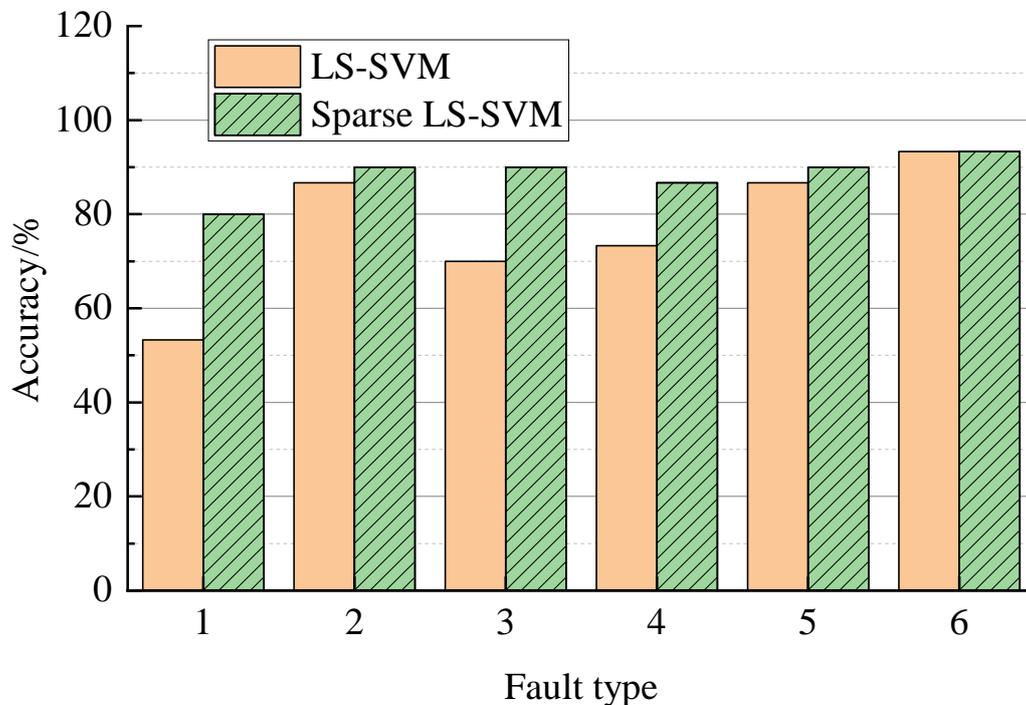


Figure 5. Detection correctness of the two classifiers

For six transformer fault types (1: low and medium temperature overheating, 2: high temperature overheating, 3: low energy discharge, 4: high energy discharge, 5: partial

discharge, and 6: normal), and all of them using fused spectral features as inputs, the detection correctness of the two classifiers is shown in Figure 5.

It can be seen that the sparsified LS-SVM achieves 80.0%, 90.0%, 90.0%, 86.7%, 90.0%, and 93.3% correct diagnosis for the six states. Compared with LS-SVM, the sparsified LS-SVM has the highest diagnosis accuracy for all faults. The average accuracy of sparsified LS-SVM for all faults reached 88.3%, while the average accuracy of LS-SVM algorithm was 75.6%. This indicates that sparsified LS-SVM has better classification than LS-SVM and has higher accuracy for fault detection in transformers.

5. Conclusions. In this work, a partial discharge sound signal recognition method based on fused spectral features and sparsified LS-SVM is proposed. Firstly, for the problem that traditional feature extraction methods have the inability to retain complete short-time spectral features, from the perspective of human hearing, the acoustic scale that is most compatible with human hearing-Mel scale is selected, and the Mel scale is used to obtain short-time spectral features to retain more complete Mel spectral features. Secondly, to address the problem of poor sparsity of LS-SVM, it is proposed to map the data samples to the high-dimensional kernel space and perform kernel space distance clustering on the mapped data by spectral clustering algorithm to achieve data preprocessing of LS-SVM, so as to achieve its sparsification. The experimental results show that the sparsified LS-SVM achieves an average accuracy of 88.3% for all faults, while the average accuracy of the LS-SVM algorithm is 75.6%. This indicates that the classification effect of sparsified LS-SVM is better than LS-SVM, and it has higher accuracy for transformer fault detection.

Acknowledgment. This work supported by the Science and Technology Project of State Grid Corporation (No. 5500-202116119A-0-0-00), and the National Natural Science Foundation of China (No. 62003191).

REFERENCES

- [1] J. Zhao, and X. Yue, "Condition monitoring of power transmission and transformation equipment based on industrial internet of things technology," *Computer Communications*, vol. 157, pp. 204-212, 2020.
- [2] C. Wu, P. Jia, X. Yu, J. Guan, J. Deng, and H. Cheng, "Function orientation and typical application scenarios of the Internet of Things construction for power transmission and transformation equipment," *Energy Reports*, vol. 8, pp. 109-116, 2022.
- [3] H. Yang, Z. Zhang, X. Yin, J. Han, Y. Wang, and G. Chen, "A novel short-term maintenance strategy for power transmission and transformation equipment based on risk-cost-analysis," *Energies*, vol. 10, no. 11, 1865, 2017.
- [4] J. Liu, "The analysis of innovative design and evaluation of energy storage system based on Internet of Things," *The Journal of Supercomputing*, vol. 78, no. 2, pp. 1624-1641, 2022.
- [5] W. J. K. Raymond, H. A. Ilias, and H. Mokhlis, "Partial discharge classifications: Review of recent progress," *Measurement*, vol. 68, pp. 164-181, 2015.
- [6] E. Lemke, "A critical review of partial-discharge models," *IEEE Electrical Insulation Magazine*, vol. 28, no. 6, pp. 11-16, 2012.
- [7] C. Hudon, and M. Belec, "Partial discharge signal interpretation for generator diagnostics," *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 12, no. 2, pp. 297-319, 2005.
- [8] N. Sahoo, M. Salama, and R. Bartnikas, "Trends in partial discharge pattern classification: a survey," *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 12, no. 2, pp. 248-264, 2005.
- [9] R. Bartnikas, "Partial discharges. Their mechanism, detection and measurement," *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 9, no. 5, pp. 763-808, 2002.
- [10] G. C. Stone, "Partial discharge diagnostics and electrical equipment insulation condition assessment," *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 12, no. 5, pp. 891-904, 2005.
- [11] A. Sadhu, G. Prakash, and S. Narasimhan, "A hybrid hidden Markov model towards fault detection of rotating components," *Journal of Vibration and Control*, vol. 23, no. 19, pp. 3175-3195, 2017.

- [12] M. Yaacob, M. Alsaedi, A. Al Gizi, and N. Zareen, "Partial discharge signal detection using ultra high frequency method in high voltage power equipments: A review," *International Journal of Scientific and Engineering Research*, vol. 4, no. 1, pp. 1-6, 2013.
- [13] Q. Khan, S. S. Refaat, H. Abu-Rub, and H. A. Toliyat, "Partial discharge detection and diagnosis in gas insulated switchgear: State of the art," *IEEE Electrical Insulation Magazine*, vol. 35, no. 4, pp. 16-33, 2019.
- [14] Z. Mnasri, S. Rovetta, and F. Masulli, "Anomalous sound event detection: A survey of machine learning based methods and applications," *Multimedia Tools and Applications*, pp. 1-50, 2022.
- [15] F. Demir, M. Turkoglu, M. Aslan, and A. Sengur, "A new pyramidal concatenated CNN approach for environmental sound classification," *Applied Acoustics*, vol. 170, 107520, 2020.
- [16] H. Pulakka, and P. Alku, "Bandwidth extension of telephone speech using a neural network and a filter bank implementation for highband mel spectrum," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 19, no. 7, pp. 2170-2183, 2011.
- [17] A. A. Abdelhamid, E.-S. M. El-Kenawy, B. Alotaibi, G. M. Amer, M. Y. Abdelkader, A. Ibrahim, and M. M. Eid, "Robust speech emotion recognition using CNN+ LSTM based on stochastic fractal search optimization algorithm," *IEEE Access*, vol. 10, pp. 49265-49284, 2022.
- [18] S. S. Tippannavar, S. Yashwanth, K. Puneeth, M. S. MP, and E. A. Madappa, "Advances and Challenges in Human Emotion Recognition Systems: A Comprehensive Review," *Journal of Trends in Computer Science and Smart Technology*, vol. 5, no. 4, pp. 367-387, 2023.
- [19] S. Winograd, "On computing the discrete Fourier transform," *Mathematics of Computation*, vol. 32, no. 141, pp. 175-199, 1978.
- [20] S. Weinstein, and P. Ebert, "Data transmission by frequency-division multiplexing using the discrete Fourier transform," *IEEE Transactions on Communication Technology*, vol. 19, no. 5, pp. 628-634, 1971.
- [21] P. Podder, T. Z. Khan, M. H. Khan, and M. M. Rahman, "Comparative performance analysis of hamming, hanning and blackman window," *International Journal of Computer Applications*, vol. 96, no. 18, pp. 1-7, 2014.
- [22] O. Farooq, and S. Datta, "Mel filter-like admissible wavelet packet structure for speech recognition," *IEEE Signal Processing Letters*, vol. 8, no. 7, pp. 196-198, 2001.
- [23] H. K. Kathania, S. Shahnawazuddin, W. Ahmad, and N. Adiga, "Role of linear, mel and inverse-mel filterbanks in automatic recognition of speech from high-pitched speakers," *Circuits, Systems, and Signal Processing*, vol. 38, pp. 4667-4682, 2019.
- [24] D. Liu, X. Wang, J. Zhang, and X. Huang, "Feature extraction using Mel frequency cepstral coefficients for hyperspectral image classification," *Applied Optics*, vol. 49, no. 14, pp. 2670-2675, 2010.
- [25] T.-Y. Wu, Q. Meng, Y.-C. Chen, S. Kumari, and C.-M. Chen, "Toward a Secure Smart-Home IoT Access Control Scheme Based on Home Registration Approach," *Mathematics*, vol. 11, no. 9, pp. 2123, 2023.
- [26] T.-Y. Wu, A. Shao, and J.-S. Pan, "CTOA: Toward a Chaotic-Based Tumbleweed Optimization Algorithm," *Mathematics*, vol. 11, no. 10, pp. 2339, 2023.
- [27] K. Wang, C.-M. Chen, Z. Liang, M. M. Hassan, G. M. L. Sarné, L. Fotia, and G. Fortino, "A trusted consensus fusion scheme for decentralized collaborated learning in massive IoT domain," *Information Fusion*, vol. 72, pp. 100-109, 2021.
- [28] K. K. Sharma, and A. Seal, "Multi-view spectral clustering for uncertain objects," *Information Sciences*, vol. 547, pp. 723-745, 2021.