Insulation Fault Diagnosis of Disconnecting Switches Based on Wavelet Packet Transform and PCA-IPSO-SVM of Electric Fields

TONGQIANG YI, YANZHAO XIE, HONGYE ZHANG, AND XU KONG

1State Key Laboratory of Electrical Insulation and Power, Xi’an Jiaotong University, Xi’an 710049, China
2School of Engineering, University of Edinburgh, Edinburgh EH9 3LJ, U.K.
3College of Electrical Engineering and Automation, Shandong University of Science and Technology, Tsingtao 266590, China

Corresponding author: Yanzhao Xie (yzxie@xjtu.edu.cn)

This work was supported by the National Natural Science Foundation of China under Grant 51677148.

ABSTRACT

Previous studies have shown that switching operations of gas insulated substations (GIS) can generate transient radiation fields outside the enclosure, namely switching transient electric fields (STEF). The waveform features of STEF can reflect the functioning performance of the switch. To monitor online the working states of disconnecting switches (DS), in this paper, we built an experimental platform to simulate their typical faulty types. Then, under different faulty status, a non-invasive three-dimensional (3D) electric field measurement system was applied to obtain STEF produced by DS. It is difficult for conventional methods to establish an accurate fault-diagnosis model, so we presented a novel method to identify the condition of DS. This innovative approach is based on feature extraction and machine learning and combined signal analysis to classify different defect types of DS. Measured STEF signals were analyzed by the wavelet packet transform (WPT) method in the time-frequency domain, which was transformed to the multi-dimensional feature matrix. The principal component analysis (PCA) algorithm was employed to reduce the dimensionality of the obtained feature matrix, which was also compared to other feature extraction algorithms. In addition, a support vector machine (SVM) with an improved particle swarm optimization (IPSO) algorithm was designed to achieve a PCA-IPSO-SVM model which can be used for signal recognition. The proposed IPSO technique can improve the convergence performance of the PSO through the dynamic adjustment of inertia weight and learning factors. Results show that the proposed fault diagnosis method based on WPT and PCA-IPSO-SVM can effectively identify the insulation faulty signals in STEF.

INDEX TERMS

Gas insulated substation (GIS); disconnecting switch (DS), switching transient electric field (STEF), wavelet packet transform (WPT), principal component analysis (PCA), support vector machine (SVM), improved particle swarm optimization (IPSO).

I. INTRODUCTION

Disconnecting switches (DS) are key components of gas insulated substations (GIS), which serve in isolating high voltage and ensuring the safety of high-voltage electrical equipment during maintenance. Given that DS has been widely used in the power grid, and they require high working reliability, the functioning performance of DS has a significant impact on the overall safe operation of GIS [1]. Therefore, to guarantee the normal operation of DS and GIS, it is necessary to monitor online the working states of DS and detect their potential insulation faults [2].

At present, the research on transient signals generated by GIS switching operations is mainly focused on the very fast transient overvoltage (VFTO) transmitted within GIS [3]. Many studies have been conducted regarding the mechanism, analytical analyses, and in-field measurement of VFTO, as well as some protective measures of GIS against VFTO [4]. When VFTO meets the discontinuous impedance points on the shell and outer lead interface of GIS, the wave refraction will lead to the occurrence of transient current in the shell and outer leads, and thus a strong
transient electromagnetic field can be stimulated in space [5], namely the switching transient electric field (STEF). STEF is closely related to the working states of the switch equipment. By analyzing STEF, the early-stage fault identification of GIS can be achieved [6], [7]. Many research institutions have conducted further research on the radiated electromagnetic fields. In [8], the transient electromagnetic field generated by switching operations inside GIS was modeled and simulated. Fei et al. [9], [10] compared the time-frequency features of the radiated electric fields generated by the circuit breaker (CB) and DS operations and analyzed their differences in terms of mechanism. Zhang et al. [11] proposed a non-invasive approach for high-voltage CB diagnosis, which is based on the time-frequency analysis of STEF outside the enclosure. The correlation between the different working status of CB and the time-frequency characteristics of their STEF is established.

Signal analysis is one of the critical parts in power system fault diagnosis. It can reduce and eliminate noise and interference signals, obtain useful signal information, and record and display it in a certain form, which has been a hotspot issue for many years. Signal analysis should consider the time-frequency feature to ensure the validity and completeness of feature extraction. In recent years, many researchers have performed the time-frequency analysis to extract features of the signal, such as short-time Fourier transform (STFT), wavelet transform (WT), and wavelet packet transform (WPT) [12]–[14]. STFT transforms the signal from time-domain to frequency-domain for feature extraction, and it can reflect the global features of the signal well. However, it has one single resolution, and the study of the local features is insufficient. Although WT can obtain the multi-resolution signal, its frequency resolution decreases as the wavelet scale increases. WPT can make up for the shortcomings of WT, and improve the wavelet transform to keep the excellent features of the orthogonal basis of wavelet. WPT can provide a more sophisticated analysis method for extracting signal features, so that different signal features have the adaptive ability and strong local analysis ability [15]. Guo et al. [16] proposed a novel method based on wavelet packet energy and modulation signal bispectrum analysis for planetary gearbox early fault diagnostics. Hao et al. [17] compared the differences in wavelet packet energy of the electric field in normal state and defect state of the CB, and then proposed a fault diagnosis method based on wavelet packet energy of STEF. In order to extract the feature parameters from a large amount of time-frequency domain data by WPT, it is necessary to highlight the characteristic ability of the STEF signal. PCA is a multivariate method based on the two order statistical characteristics, which converts the high dimensional space into the low ones by the idea of coordinate transformation. PCA is remarkable in information compression and elimination of data correlation [18]. Yang et al. [19] applied WPT combined with PCA to extract the feature to identify the health condition of wood utility poles. Using WPT to decompose the vibration signal, Sudhir et al. [20] developed bearing damage index (BDI) from the decomposed signal to select the useful signal from the originally recorded signal and PCA was employed to select significant features as the input of dendrogram support vector machine (DSVM) classifier to identify the faults of induction motor bearing.

In parallel to the rise of machine learning techniques in industrial applications, scientists have become increasingly interested in applying machine learning to fundamental research. Machine learning and physics share some methods as well as goals, as both of them involve the process of gathering and analyzing data to design models that can predict the behavior of complex systems [21]. A support vector machine, developed by Vapnik [22], is a machine learning method based on the statistical learning theory, which solves the problem of over-fitting and low convergence rate, and it has been widely applied in the field of power system fault diagnosis [23]. Selection of kernel functions and kernel parameters is the key to SVM, and it directly affects the generalization ability of SVM. Due to the lack of theoretical guidance, traditional kernel parameters are mostly chosen through repeated experiments to get a satisfactory solution. However, it has some shortcomings, such as limited ability to find the global optimum solution and calculation complexity. Therefore, it has become significant and urgent to find a more efficient and reasonable parameter optimization method for SVM. Wang et al. [24] used a genetic algorithm to optimize the support vector machine (GA-SVM). This method can identify vehicle suspension shock absorber squeak and rattle noise. Sun et al. [25] proposed a fault diagnosis method for analog circuits based on PCA and PSO-SVM. PSO algorithm is used to optimize the penalty parameters and kernel parameters of SVM, which improves the recognition accuracy of fault diagnosis. Wang et al. [26] proposed a method for quality grading of resistance spot welding (RSW) by ultrasonic detection of signal characteristics and used PSO to optimize the parameters of SVM. The results have shown that the PSO-SVM classifier with all nine features has good classification accuracy.

In this paper, an intelligent defect diagnosis approach based on the measurement and analysis of STEF is proposed for DS, which combines WPT, PCA, IPSO, and SVM. Firstly, we built a DS test platform and used a 3D electric field measurement system to obtain STEF under different working conditions. Secondly, the pulse with the maximum amplitude was extracted as the research object, and the time-frequency analysis was carried out to obtain a set of union feature matrix. The PCA algorithm was used for feature extraction to get the fusion feature matrix as the input of the SVM model. To overcome the problem of premature convergence on PSO, this paper proposed an improved PSO algorithm for parameter optimization of the SVM. The IPSO algorithm improved the classification accuracy, convergence speed, and global optimization because of two strategies: dynamic inertia weight and dynamic learning factors. Finally, The effectiveness of the proposed algorithm was verified through comparison with different feature extraction algorithms.
The organization of the paper is as follows. Section 2 describes the related algorithms used to build the model, including the PCA algorithm and the IPSO-SVM algorithm. Section 3 introduces the experimental measurement of STEF, data processing and analysis. In Section 4, the prediction model is established and evaluated. The last section is the conclusion.

II. THE RELATED THEORIES

A. FEATURE EXTRACTION BASED ON PCA

PCA is a standard method applied to dimensionality reduction and feature extraction [18]. PCA is to study the correlation between variables and replace the original variables with a new set of less and unrelated variables, to retain as much information as possible. It is a data analysis method based on second-order statistics, which has a better ability for reducing dimension and noise filtration [27]. Mathematically, PCA relies on the eigen-decomposition or singular value decomposition of the covariance matrix [28]. The PCA algorithm processes are shown in the following below:

1) $Z = [z_1, z_2, \ldots, z_n]^T$ is the signal feature matrix and $z_i$ represents the element. The mean $M(z)$ of the samples can be calculated by

$$M(z) = \frac{1}{n} \sum_{i=1}^{n} z_i$$ (1)

2) According to the mean of the samples, the covariance matrix $C_x$ can be constructed as

$$C_x = \frac{1}{n} \sum_{i=1}^{n} (z_i - M)(z_i - M)^T$$ (2)

3) The eigen-values $\lambda$ and eigen-vectors $\xi$ of $C_x$ are written as

$$C_x \xi = \lambda_i \xi_i$$ (3)

4) The number of principal components is determined by the cumulative contribution rate $\eta$, which represents the proportion of the first $k$ variances in the total variance. At last, we can get the feature matrix $U$, as shown in (5).

$$\eta = \frac{\sum_{i=1}^{k} \lambda_i}{\sum_{i=1}^{n} \lambda_i}$$ (4)

$$U = [u_1, u_2, \ldots, u_k]$$ (5)

In (5), $u_k$ represents the $k$-th principal component of the feature matrix. When the cumulative contribution rate of the $k$ principal components is greater than 90%, $U$ contains most of the information of the original data.

B. OPTIMIZE THE PARAMETERS OF SVM BY IMPROVED PSO

SVM is one of the supervised machine learning techniques based on statistical learning theory [29]. Figure 1 is a flowchart of the SVM, it is evident that the selection of kernel functions and kernel parameters is the key to SVM, and it directly affects the classification ability of SVM. Non-linear SVM can map input vector $x$ into a high-dimensional feature space to construct the optimal hyperplane. The kernel functions in SVM are generally composed of polynomial kernel function, radial basis kernel function (RBF), and sigmoid kernel function. For the selection of kernel function, there is no explicit standard [30]. RBF can be advantageous when there is not enough prior knowledge [31], thus this paper chooses RBF as the kernel function of SVM. The RBF function can be written as

$$K_{rbf} = \exp[-(x - x_i)^2/2\sigma^2]$$ (6)
By determining RBF as the kernel function, we need to consider the penalty coefficient C and kernel parameter σ. C impacts the generalization ability of the classifier, and σ affects the distribution of samples in the feature space. The determination of parameters has an essential influence on the performance of SVM. At present, there are many ways to optimize parameters, including GS, GA, PSO and so forth. In this paper, we use an improved PSO algorithm to optimize the SVM.

Kennedy and Eberhart firstly proposed the PSO algorithm in 1995 [32], and it is an evolutionary algorithm based on iterations. In PSO, a swarm of N particles is initialized, where each particle is assigned a random position in D-dimensional space. \( V_i = (v_{i,1}, v_{i,2}, \ldots, v_{i,D}) \) represents the speed of the i-th particle, \( S_i = (s_{i,1}, s_{i,2}, \ldots, s_{i,D}) \) is the position of the i-th particle. The optimal position of the i-th particle up to now is the individual extremum, which is denoted as \( p_i = (p_{i,1}, p_{i,2}, \ldots, p_{i,D}) \). The optimal position for the particle swarm is called the global extremum and expressed as \( p_g = (p_{g,1}, p_{g,2}, \ldots, p_{g,D}) \). Particles can be updated for velocity and position in terms of the following formulas:

\[
\begin{align*}
V_{i,M}^{(t+1)} &= W_i^{(t)} V_{i,d}^{(t)} + c_1 r_1 (p_{i,d}^{(t)} - s_{i,d}^{(t)}) + c_2 r_2 (p_{g,d}^{(t)} - s_{i,d}^{(t)}) \\
S_{i,d}^{(t+1)} &= s_{i,d}^{(t)} + V_{i,d}^{(t+1)}
\end{align*}
\]

where \( i = 1, 2, \ldots, N; d = 1, 2, \ldots, D; c_1 \) and \( c_2 \) are the positive constants in the range \([0, 2]\); \( r_1 \) and \( r_2 \) are the random numbers in the range \([0, 1]\); the inertial weight \( W_i \) is used to balance the capabilities of global exploration and local exploration; \( t \) is the number of iterations.

In PSO, \( W, c_1, \) and \( c_2 \) are essential factors for the convergence of the algorithm. The performance of PSO often suffers from the problems of slow convergence speed during the later period and being trapped in local optima [33]. In this paper, the proposed PSO is improved from two aspects: inertia weight \( W \) and learning factor \( c \) respectively. The detailed solution is given below.

1) DYNAMIC INERTIA WEIGHT STRATEGY
When \( W \) is small, the local search ability is strong, but the new area search effect is weak. As the particle reaches the lowest point, the rate of convergence will be extremely slow. If \( W \) is vast, the global search ability of the particle will be strong, but the local search ability will be inferior correspondingly. Therefore, the use of dynamic \( W \) can improve the efficiency of the parameter search. The design of adaptive \( W \) is mainly about three aspects: particle fitness, population size, and spatial dimension.

The fitness can reflect the merits and demerits of the current position of the particle, the local area where the particles with higher adaptability are located may be the global optimum. To quickly find the global optimum parameter, the local optimization ability can be enhanced by reducing \( W \). Conversely, for particles with low fitness, the current position is weak, and global search capabilities can be improved by increasing \( W \).

When the population size is large, the population diversity is relatively high, and the probability that the particle space covers the optimal solution is higher. Therefore, \( W \) should be reduced to improve the local search ability, so that it can quickly converge to the global best. When the population size is small, especially for some multi-peak functions, the particle space cannot cover the whole search space. In this case, \( W \) should be increased to improve the global search ability to avoid local optimum.

When the spatial dimension is oversizens, it causes premature convergence to the local optimum. For some complex problems in high-dimensional space, the global search ability should be improved to solve the question, as mentioned above. Therefore, it is indispensable to reduce \( W \). When the spatial dimension is too small, \( W \) should be increased to achieve fast convergence and to improve the search efficiency.

Finally, we can get the adjustment direction of \( W \) through analysis and studies as mentioned above,

\[
W_i = 1 \div \frac{1}{\alpha F_i (M \times \sum_{i=1}^{N} F_i) - \exp(-N/\beta) + 1}
\]

where \( F_i \) is the fitness of the i-th particle, \( N \) represents the number of particles, and \( \alpha \) and \( \beta \) are empirical parameters. After each iteration, \( W \) is calculated by (9) to realize the adaptive adjustment.

2) DYNAMIC LEARNING FACTORS STRATEGY
\( c_1 \) and \( c_2 \) represent the weights of the statistical accelerators that push each particle to the individual extremum and global extremum. In PSO, usually take \( c_1 = c_2 = 2 \). That means it is not based on different stages of algorithm evolution to set different learning factors. In this way, it fails to make the population quickly search for the optimal value in the early evolution on the one hand, and does not converge quickly to the optimal solution in the later evolution on the other hand [34]. The regeneration pattern of \( c_1 \) and \( c_2 \) is related to the fitness of particle swarm. When the fitness is higher than the average value of population fitness, the speed of reaching the global optimum can be reduced by increasing \( c_1 \) and decreasing \( c_2 \). Contrarily, we should reduce \( c_1 \) and increase \( c_2 \) [35].

The fitness of the i-th particle in the t-th iteration is noted as \( F_i^t \), and the average fitness of the population is represented by \( \bar{F}^t \). For the t-th iteration, the adjustment formulas of \( c_1 \) and \( c_2 \) are calculated as follows

If \( F_i^t \geq \bar{F}^t \), then

\[
\begin{align*}
c_1^{t+1} &= 1 + \frac{F_i^t - \bar{F}^t}{\max{F_i}^t} \\
c_2^{t+1} &= 2 - c_1^{t+1}
\end{align*}
\]
Through the above two methods of improving the PSO algorithm, we can get the IPSO algorithm.

III. SIGNAL ACQUISITION AND FEATURE EXTRACTION
A. SIGNAL ACQUISITION
A DS fault simulation experimental platform was built in this paper. Figure 2 displays the equivalent circuit of the platform. C₁ and C₂ represent capacitive voltage dividers. C₁ is used to protect the power supply, and C₂ is the analog load. U represents an external high voltage power supply. DS₁ and DS₂ are the disconnecting switches of the test system. The research object in this paper is DS₂.

The 3D electric field measurement system designed is adopted in this paper to acquire the STEFs of the tested DS. It consists of a sensor, an optical transmitter, three optical fibers, an optical receiver, and an oscilloscope. The sensor contains an optical transmitter, and the output electrical signal is driven by a built-in coupling circuit to realize the conversion of an electrical signal to an optical signal. The optical signal is transmitted to an optical receiver through optical fibers to achieve photoelectric conversion. The oscilloscope finally receives the converted electrical signal. The adopted oscilloscope is of the type PicoScope 6404D, and its working bandwidth is 500 MHz, with a memory of 2 GS. The measuring range of the electric field sensor can be changed from 100 V/m to 50 kV/m. Its −3dB bandwidth is wide enough to cover 10 kHz−350 MHz.

Figure 4 shows the field test of the switching operations of the 110kV DS. The electric field sensor is placed at a horizontal distance of 1 m from the bus-bar. Figure 5 is the schematic diagram of the 3D electric field synthesis. The concrete calculating methods are as follows:

\[
\begin{align*}
E &= E_x \cdot \cos <E_x, E> + E_y \cdot \cos <E_y, E> + E_z \cdot \cos <E_z, E> \\
<E_x, E> &= \arccos(\sin \beta \cdot \cos \beta \cdot \cos \alpha) \\
<E_y, E> &= \arccos(\sin \beta \cdot \cos \beta \cdot \sin \alpha) \\
<E_z, E> &= \beta
\end{align*}
\] (12)

The DS experimental platform was established to simulate three types of insulation defects, including internal tip discharging, internal suspended metal particles, and external flashover. To simulate tip discharging, A 1 cm long metal tip is placed inside the bus tube, connected to the inner surface of the outer conductor, as shown in Figure 6 (a). Some metal particles have been pasted to an insulated tape connecting the inner and outer conductors to simulate suspended particles inside the bus tube. Figure 6 (b) shows the metal particles used in the experiment. A discharge tip is drawn from the high voltage terminal inside the bushing, and the external flashover is established between the discharge tip and the external shell.
as shown in Figure 6 (c). This article measured STEF of DS under the above-mentioned defect states.

In the experiment, the sampling frequency of the signal is 1.25 GHz. Hence, its Nyquist frequency is 625 MHz, which is the upper limit frequency for the wavelet packet decomposition. The signal sample label of the normal state, the internal tip discharging, suspended particles, as well as the external flashover are set as 0, 1, 2, and 3, respectively. TABLE 1 shows the sample distribution of the data set.

| Type label | Training set | Test set |
|------------|--------------|---------|
| 0          | 43           | 32      |
| 1          | 42           | 23      |
| 2          | 53           | 37      |
| 3          | 48           | 32      |
| Total      | 186          | 124     |

B. FEATURE SELECTION AND EXTRACTION

For a start, we take the STEF signals of DS switching under the normal state as an example to extract the feature parameters. Figure 7 shows the signal waveforms measured by the electric field sensor when DS is normally closed.

When pre-breakdown occurs in the initial stage, the contact distance is large and the arc instantaneous establishment
components. and has a shorter duration compared with the low-frequency Moreover, the high-frequency component attenuates rapidly

FIGURE 8. Time-frequency spectrum of $P$. 

FIGURE 7. The radiated electric field in space when DS is normally closed.

To further extract signal features, we used the WPT to obtain signal energy characteristics. The original signal is decomposed in terms of different frequency bands with certain energy. The normalized energy can be used as feature parameters to represent the operating status of the DS. Bior 5.5 wavelet is used as the wavelet base for WPT in this paper, which guarantees the wavelet base has excellent tight support in both the time-domain and the frequency-domain.

Wavelet packet decomposition layer $j$ determines the frequency features of the signal. With the increase of $j$, the spectrum window after decomposition will be fully subdivided, and the feature information of the signal in different frequency bands can be obtained. However, the increase of $j$ will lead to the complexity of calculation. Therefore, $j = 5$ has been chosen in this paper as a trade-off. Based on the above analysis, the bior 5.5 wavelet basis is used for the five-layer wavelet packet decomposition of the $P$. If the sampling rate is $2^j$, the $j$-layer wavelet packet of the signal can form $2^j$ equal-width bands, and the interval band is $f/2^j$. The $j$-layer wavelet packet coefficients $C_{j,m,k}$, $k = 0, 1, \ldots, 2^j - 1$, $m$ is the spatial position of the wavelet packet. After decomposition, TABLE 2 manifests the range of each frequency band and its corresponding node. According to (14)–(15), we can obtain the normalized energy of each frequency band.

$$E_{j,k} = \sum_{m} |C_{j,m,k}|^2$$  \hspace{1cm} (14)

$$\tilde{E}_{j,k} = \frac{E_{j,k}}{\sum_{2^j} E_{j,k}}$$  \hspace{1cm} (15)

It should be noted that, the upper cut-off frequency of the $-3$dB bandwidth of the employed electric field sensor is 350 MHz. Therefore, to ensure the accuracy of the extracted characteristic parameters, we have adopted the normalized energy values in the frequency range from 0 to 312.5 MHz. TABLE 3 shows that the normalized energy of the wavelet packet centralizes in the frequency band between 0 to 19.53 MHz, of which the energy value is much higher than those of other frequency bands. For the sake of convenience, $E_i (i = 0, 1 \ldots 15)$ is used to represent the normalized energy value of the $(i+1)$-th frequency band in the following.

Based on the above process, this paper selected the amplitude($A$), the center frequency($F$), and the wavelet packet normalized energy ($E_0, E_1, \ldots, E_{15}$) as the feature parameters to classify the type of a signal. In order to have an intuitive understanding of the feature distribution, we used a boxplot to analyze these features. In Figure 9 and Figure 10, it can be found that the amplitude is distributed below 10 kV/m. The STEF amplitude distribution under the normal state is the widest, and the difference between the internal tip discharging and the suspended particles is not obvious. The distribution range of the center frequency is below 200 MHz, but the distribution of the STEF signal under the suspended particle defect and that of the external flashover signal resemble each other.

For the energy distribution range after wavelet packet decomposition, Figure 11 shows that in different frequency bands, energy distribution under the internal tip discharging state is the widest, and the energy distribution difference between various types of signals is visible, thus the signal energy can be extracted as the feature parameters for signal classification. Through the above comprehensive analysis, we can find that the feature parameters extracted in this paper contain not only useful information about wave propagation or defect types, but also redundant information generated by relevant features, background noise, and measurement noise. Therefore, we need to process the extracted features further.
TABLE 2. Nodes and corresponding frequency ranges after five-layer wavelet packet decomposition.

| Node | Frequency Band(MHz) |
|------|---------------------|
| (5, 0) | 0–19.53 |
| (5, 1) | 19.53–39.06 |
| (5, 2) | 39.06–58.59 |
| (5, 3) | 58.59–78.13 |

| Node | Frequency Band(MHz) |
|------|---------------------|
| (5, 4) | 58.59–97.66 |
| (5, 5) | 97.66–117.19 |
| (5, 6) | 117.19–136.72 |
| (5, 7) | 136.72–156.25 |

| Node | Frequency Band(MHz) |
|------|---------------------|
| (5, 8) | 156.25–175.78 |
| (5, 9) | 175.78–195.31 |
| (5, 10) | 195.31–214.84 |
| (5, 11) | 214.84–234.38 |

| Node | Frequency Band(MHz) |
|------|---------------------|
| (5, 12) | 234.38–253.91 |
| (5, 13) | 253.91–273.44 |
| (5, 14) | 273.44–292.97 |
| (5, 15) | 292.97–312.5 |

| Node | Frequency Band(MHz) |
|------|---------------------|
| (5, 16) | 312.5–332.03 |
| (5, 17) | 332.03–351.56 |
| (5, 18) | 351.56–371.09 |
| (5, 19) | 371.09–390.63 |

| Node | Frequency Band(MHz) |
|------|---------------------|
| (5, 20) | 390.63–410.16 |
| (5, 21) | 410.16–429.69 |
| (5, 22) | 429.69–449.22 |
| (5, 23) | 449.22–468.75 |

| Node | Frequency Band(MHz) |
|------|---------------------|
| (5, 24) | 468.75–488.28 |
| (5, 25) | 488.28–507.81 |
| (5, 26) | 507.81–527.34 |
| (5, 27) | 527.34–546.88 |

| Node | Frequency Band(MHz) |
|------|---------------------|
| (5, 28) | 546.88–566.41 |
| (5, 29) | 566.41–585.94 |
| (5, 30) | 585.94–605.47 |
| (5, 31) | 605.47–625 |

TABLE 3. Nodes and corresponding normalized energy of P after five-layer wavelet packet decomposition.

| Node | Normalized energy | Node | Normalized energy |
|------|-------------------|------|-------------------|
| (5, 0) | 0.3560 | (5, 8) | 0.0220 |
| (5, 1) | 0.0712 | (5, 9) | 0.0200 |
| (5, 2) | 0.0400 | (5, 10) | 0.0360 |
| (5, 3) | 0.0739 | (5, 11) | 0.0289 |
| (5, 4) | 0.0282 | (5, 12) | 0.0406 |
| (5, 5) | 0.0454 | (5, 13) | 0.0473 |
| (5, 6) | 0.0405 | (5, 14) | 0.0457 |
| (5, 7) | 0.0433 | (5, 15) | 0.0610 |

We extracted the feature parameters of 310 signals by the time-frequency analysis, each signal contains 18 characteristic parameters. The union feature matrix $E (310 \times 18)$ can be obtained. Before PCA, the features should be normalized to eliminate the influence of different order of magnitudes and units. In addition, normalization will accelerate the convergence speed of the model. The normalized formula can be represented as follows:

$$X^* = \frac{X - X_{\text{min}}}{X_{\text{max}} - X_{\text{min}}} \quad (16)$$

where $X, X^* \in \mathbb{R}^n$, the purpose of normalization is that the original data is normalized to the range $[0, 1]$, and then we used the PCA to extract the comprehensive characteristics of the signal. According to (1)–(5), we can obtain the correlation coefficient matrix, the eigenvalues, the corresponding eigenvectors, the contribution rate, as well as the cumulative contribution rate. TABLE 4 demonstrates that the first three main components contribute 93.08% to classifying signals. Thus we can replace the original 18 feature variables with three new feature variables $(u_1, u_2, u_3)$ and use them as the input of the SVM. The linear combinations of $u_1, u_2$, and $u_3$ are:

$$u_1 = -0.1122X_1 + 0.8272X_2 - 0.3054X_3 - 0.1168X_4 - 0.3326X_5 - 0.2898X_6 - 0.1562X_7 + 0.2194X_8 + 0.1940X_9 + 0.2873X_{10} - 0.2046X_{11} - 0.3019X_{12}$$

FIGURE 9. Amplitude box plot of different types of signals.
T. Yi et al.: Insulation Fault Diagnosis ofDisconnecting Switches

FIGURE 10. Center frequency box plot of different types of signals.

\[ u_2 = 0.2295X_1 - 0.3349X_2 - 0.0813X_3 - 0.4088X_4 \]
\[ - 0.2692X_5 - 0.4401X_6 + 0.0486X_7 + 0.1802X_8 \]
\[ - 0.1370X_9 - 0.2583X_{10} + 0.2296X_{11} + 0.1859X_{12} \]
\[ + 0.1823X_{13} + 0.2705X_{14} - 0.2546X_{15} + 0.2549X_{16} \]
\[ + 0.2745X_{17} + 0.2514X_{18} \]  \hspace{1cm} (17)

\[ u_3 = -0.0823X_1 - 0.5302X_2 + 0.2738X_3 - 0.2527X_4 \]
\[ - 0.2814X_5 - 0.2308X_6 - 0.2211X_7 + 0.3151X_8 \]
\[ + 0.2618X_9 + 0.2185X_{10} + 0.2744X_{11} - 0.2287X_{12} \]
\[ - 0.0387X_{13} - 0.0970X_{14} + 0.3087X_{15} - 0.0358X_{16} \]
\[ - 0.0332X_{17} - 0.0789X_{18} \]  \hspace{1cm} (19)

Therefore, the first three principal components are selected to get a fusion feature matrix \( U(310 \times 3) \). To test the validity

TABLE 4. The contribution rate of the first five principal components after PCA dimensionality reduction.

| Principal component | Eigenvalue | Contribution rate/% | Cumulative contribution rate/% |
|---------------------|------------|---------------------|-------------------------------|
| 1                   | 17.91      | 81.41               | 81.41                         |
| 2                   | 1.71       | 7.76                | 89.17                         |
| 3                   | 0.86       | 3.91                | 93.08                         |
| 4                   | 0.63       | 2.85                | 95.93                         |
| 5                   | 0.33       | 1.48                | 97.41                         |

FIGURE 11. Box plot of wavelet packet energy in different frequency bands.
of the PCA, we compared the clustering effects of several feature extraction algorithms. Figure 12 shows the visualization result after dimension reduction by LLE, MDS, LDA, and PCA.

In Figure 12, PCA has the best clustering result, as it can distinguish different types of signals, and only a few samples are overlapped. LLE is the worst-performing algorithm among the algorithms compared. The reason behind is that LLE has specific restrictive requirements on the original distribution of data sets to some extent, and the distribution of STEF in high-dimensional space may not meet closed manifolds. The clustering effect of the LDA algorithm is better than LLE and MDS, it can classify the signals of various categories. However, the STEF signals under the internal tip discharging are mixed with other types of signals, thus it is worse than the PCA clustering effect.

**IV. FORECAST RESULTS AND ANALYSIS**

When using the SVM for model training, the size of parameters needs to be determined, which directly influences the quality of the final model. In this paper, RBF is selected as the kernel function of SVM, and IPSO is used to optimize the penalty coefficient \( C \) and RBF kernel parameter \( \sigma \). Based on the above analysis, the model training has been carried out by using the software Matlab2016a and the SVM toolbox-Libsvm. Figure 13 describes the realization process of the proposed method.

1) Firstly, we perform the time-frequency analysis on the pulse with the maximum amplitude extracted from the experimental data and obtain the union feature matrix \( E(310 \times 18) \). PCA is used to reduce the number of features to three and obtain the fusion feature matrix \( U(310 \times 3) \) for model training of SVM.

2) Secondly, the parameters \( C \) and \( \sigma \) of the SVM are optimized by IPSO. IPSO needs parameter initialization: the particle spatial dimension is 2, the population number is set to 50, and the iteration number \( T = 200 \). The fitness function can evaluate the performance of the PCA-IPSO-SVM model. In (20), the fitness function \( F \) is the average accuracy of six-fold cross-validation.

\[
F = \frac{1}{k} \sum_{i=1}^{k} \left( \frac{P_{II}}{P_{I}} \times 100\% \right) \quad (20)
\]
where $P_{TI}$ is the number of correct classification of the verification set, and $P_I$ is the number of samples in the verification set.

3) Finally, the PCA-PSO-SVM model is obtained by the parameter optimization of the PSO algorithm, and the optimal parameters are $C = 3.94$ and $\sigma = 0.4$. The final test accuracy attains $97.58\%$. The iteration process of fitness obtained using the IPSO-SVM and PSO-SVM are presented in Figure 14 and Figure 15. After several iterations, the algorithm converges to the best values. The fitness curves have been gradually improved as iteration progress until some iteration that they have not embodied significant improvements, which means that it has obtained the optimal parameters. In Figure 14 and Figure 15, the best fitness of IPSO-SVM converges to the optimal values quickly, while PSO-SVM needs about 100 iterations to reach the best results. Meanwhile, the best fitness and average fitness of the IPSO...
algorithm are relatively higher. Based on the above analysis, we can conclude that IPSO has faster convergence speed and higher fitness compared with the PSO algorithm.

To further demonstrate the superiority of the PCA algorithm and the IPSO algorithm, we used the LLE algorithm, the MDS algorithm, the LDA algorithm, and the PCA algorithm for feature extraction. We applied the IPSO-SVM (the parameters of SVM are optimized by IPSO) and the PSO-SVM (the parameters of SVM are optimized by PSO) algorithm for model training.

Figure 16 shows the classification results of the IPSO-SVM model with different feature extraction algorithms. IPSO-SVM using the LDA algorithm has a good classification effect, with only six samples misclassified, and the recognition accuracy reaches 95.16%. IPSO-SVM using the LLE algorithm has the worst classification effect, and the recognition accuracy is only 73.39%. The IPSO-SVM model using the PCA has the best classification effect, with only three samples misclassified. The reason may be that the time-frequency features of the STEF contain many redundant characteristics, and the PCA algorithm can remove the redundancy information and maximize the fault information contained in the signal, and effectively improve the recognition accuracy of the signal.

In Figure 17, similar to the IPSO-SVM model, the PSO-SVM model obtained by the PCA algorithm has the best recognition effect, and the classification effect of the LLE algorithm is still weak. From Figure 16 and Figure 17, compared with PSO-SVM, IPSO-SVM has a higher recognition rate under the same feature extraction algorithm. IPSO overcomes the shortcomings of the PSO, such as premature convergence, being easy to fall into local optimal, and low search accuracy, thus it can achieve a better effect on parameter optimization.

To verify the effectiveness of the IPSO algorithm proposed in this paper compared with other parameter optimization algorithms. The fusion matrix $U$ after feature extraction by the PCA algorithm is employed as the input of SVM, and different algorithms (IPSO, GS, and GA) are used to optimize the SVM parameters and compare the classification results. Figure 18 shows that the SVM classifier optimized by the above-mentioned algorithms has a very high classification accuracy, all of which are over 90%. The order from the best to the worst of the recognition effect is PCA-IPSO-SVM, PCA-GA-SVM, and PCA-GS-SVM, and the recognition
accuracies are 97.58%, 93.55%, and 91.94%, respectively. PCA-IPSO-SVM has the best classification effect, because compared with GA, PSO has neither crossover nor mutation operations, and it relies on particle speed to search. In the iterative process, only the optimal particle transmits the information to other particles, so the search efficiency is high, and
the search scope and often fails to find the optimal solution. Compared with PSO, IPSO improves search efficiency to some extent.

TABLE 5 shows the optimal parameters of different models, as well as the recognition accuracy in the training set and test set. Compared with other methods, PCA-IPSO-SVM has the highest recognition accuracy for the training set and test set, and the classification accuracies are 99.46% and 97.58%, respectively. Based on the above analysis, we can conclude that the PCA algorithm and the IPSO algorithm proposed in this paper have the best effect in feature extraction and optimizing the parameters of SVM, as far as the analysis of STEF is concerned.

V. CONCLUSION

In this paper, a PCA-IPSO-SVM algorithm has been proposed to identify STEF and finally achieved the insulation fault diagnosis for DS. The results are summarized as follows:

1) An experiment platform was built to simulate the four working states (normal status and three defects) of DS. By the use of a 3D electric field sensor, the STEFs in the space could be measured. Based on a large amount of experimental data, time-frequency analysis, including WPT, was applied for extracting the original signal features.

2) The PCA algorithm was used to further feature extraction to reduce redundant features and obtain the 3D fusion matrix $U$, and it was taken as the input of SVM. Given the slow convergence rate and being easy to fall into local optimal of traditional PSO, the IPSO algorithm (from the weight coefficient and learning factor to improve the PSO algorithm) was used to optimize SVM parameters, and it effectively improves the convergence speed and signals recognition accuracy. Finally, we get a PCA-IPSO-SVM model with 97.58% classification accuracy.

3) The PCA-IPSO-SVM model proposed in this paper was compared with a variety of other models. The results show that the SVM classifier using the PCA algorithm has better generalization performance than other classifiers using LLE, MDS, and LDA. In parameter optimization, the SVM classifier optimized by IPSO is better than the classifier based on PSO, GA, and GS. Therefore, the PCA-IPSO-SVM has a better recognition accuracy and classification effect than other models.

This paper proposes an intelligent fault diagnosis method based on machine learning. Compared with the traditional fault diagnosis method, it is a method based on data training, which can significantly save labor costs. Specifically, due to the fact that the proposed methodology outperforms in both high recognition accuracy and fast recognition speed with application to STEFs, it has become the research direction of intelligent fault diagnosis in power systems. To further refine the method and diagnose different types of defects, we will focus on analyzing the relationship between different types of faults and STEF in space to establish a fault diagnosis model and apply to an intelligent fault diagnosis system of GIS to realize automated signal analysis and fault diagnosis.

REFERENCES

[1] C. X. Zhang, W. Z. Sun, Y. P. Tu, and Y. H. Lu, “Analysis of cable sheath overvoltage due to not completely closing of GIS disconnector,” High Voltage Eng., vol. 37, no. 10, pp. 2498–2505, 2011.

[2] L. Cheng, H. Y. Zhang, T. Q. Yi, Y. Z. Xie, X. Liu, and C. Zhang, “Insulation defect detection method for disconnecting switches based on time-frequency analysis of 3D switching electric fields,” High Voltage Technol., vol. 46, no. 4, pp. 1417–1423, 2016.

[3] G. Lupo, C. Petrarca, M. Vitelli, and V. Tucci, “Multiconductor transmission line analysis of steep-front surges in machine windings,” IEEE Trans. Dielectr. Electr. Insul., vol. 9, no. 3, pp. 467–478, Jun. 2002.

[4] Y. Liu, X. Y. Zhao, Y. J. Xie, and D. W. Wu, “Research on collaborative simulation of VFTO radiation interference in GIS by field-circuit-wave,” J. Air Force Eng. Univ. (Natural Sci. Ed.), vol. 17, no. 4, pp. 59–63, 2016.

[5] X. Kong, F. Guo, and T. Liang, “Measurement and analysis of the transient radiated electric field excited by the breaker when closing the unloaded transformer in GIS,” Proc. CSEE, vol. 36, no. 18, pp. 5087–5093, 2016.

[6] Y. Zhang, S. H. Liu, X. F. Hu, and W. D. Liu, “Properties of transient electromagnetic signals radiation field from various radiation sources,” Chin. J. Radio Science., vol. 30, no. 2, pp. 316–322, 2015.

[7] W. Zhang, P. Chen, W. Chen, X. Cui, L. Wang, H. Huang, R. Hu, and C. Wu, “Measurement and simulation of disturbance voltage generated by VFTO in UHV GIS substation on the secondary cables,” in Proc. Chin. Soc. Electr., vol. 33, 2013, pp. 187–196.

[8] H. Y. Zhang, Y. Z. Xie, Y. Zhou, and Y. X. Qiu, “Development and application of simulator for transient radiation field of GIS switching operation,” High Voltage App., vol. 56, no. 4, pp. 172–178 and 183, 2020.

[9] F. Guo, X. Kong, X. T. Wu, J. Guo, Q. Liu, A. F. Ainizio, and Y. Z. Xie, “Measurement and analysis of the radiation electric field excited by the switching operation of 330 kV GIS,” High Voltage Apparatus., vol. 50, no. 9, pp. 119–125, 2014.

[10] B. X. Liu, Z. Z. Wang, C. R. Li, L. J. Ding, W. Wang, and J. C. Wang, “Measurements and research of switching operation transient electric field in 500KV substations,” Proc. CSEE., vol. 24, no. 4, pp. 133–138, 2004.

[11] H.-Y. Zhang, Y.-Z. Xie, T.-Q. Yi, X. Kong, L. Cheng, and H.-J. Liu, “Fault detection for high-voltage circuit breakers based on time–frequency analysis of switching transient E-fields,” IEEE Trans. Instrum. Meas., vol. 69, no. 4, pp. 1620–1631, Apr. 2020.

[12] H.-B. Xie, Y.-P. Zheng, and J.-Y. Guo, “Classification of the mechanomyogram signal using a wavelet packet transform and singular value decomposition for multifunction prosthesis control,” Physiol. Meas., vol. 30, no. 5, pp. 441–457, May 2009.

[13] J. U. Chu, I. Moon, Y. J. Lee, and S. K. Kim, “A supervised feature-projection-based real-time EMG pattern recognition for multifunction myoelectric hand control,” IEEE Trans. Mechatronics, vol. 12, pp. 282–290, Jun. 2007.
