Identification of ultra-high-frequency PD signals in gas-insulated switchgear based on moment features considering electromagnetic mode

Feng Bin\textsuperscript{1}, Feng Wang\textsuperscript{1,2}, Qiuqin Sun\textsuperscript{1,3}, She Chen\textsuperscript{1}, Jingmin Fan\textsuperscript{3}, Huisheng Ye\textsuperscript{4}

\textsuperscript{1}College of Electrical and Information Engineering, Hunan University, Changsha 410082, People’s Republic of China
\textsuperscript{2}College of Electronic and Information Engineering, Yili Normal University, Yili 835000, People’s Republic of China
\textsuperscript{3}School of Automation, Guangdong University of Technology, Guangzhou 510012, People’s Republic of China
\textsuperscript{4}Equipment Condition Assessment Center, Hunan Electric Power Research Institute, Changsha 410000, People’s Republic of China

E-mail: sunqq@hnu.edu.cn

Abstract: The feature extraction and pattern recognition techniques are of great importance to assess the insulation condition of gas-insulated switchgear. In this work, the ultra-high-frequency partial discharge (PD) signals generated from four types of typical insulation defects are analysed using S-transform, and the greyscale image in time-frequency representation is divided into five regions according to the cutoff frequencies of TE\(_{m1}\) modes. Then, the three low-order moments of every subregion are extracted and the feature selection is performed based on the J criterion. To confirm the effectiveness of selected moment features after considering the electromagnetic modes, the support vector machine, k-nearest neighbour and particle swarm-optimised extreme learning machine (ELM) are utilised to classify the type of PD, and they achieve the recognition accuracies of 92, 88.5 and 95\%, respectively. In addition, the results show that the ELM offers good generalisation performance at the fastest learning and testing speeds, thus more suitable for a real-time PD detection.

1 Introduction

Gas-insulated switchgear (GIS) is a compact metal encapsulated switchgear consisting of high-voltage components such as circuit-breakers and disconnectors. It is widely used in power system due to the high reliability, low maintenance and compact size \([1]\). A dielectric breakdown in GIS could result in serious outages and thereby cause enormous economic losses \([2]\). Partial discharge (PD) activity often happens before the insulation failure \([3]\). Thus, the PD detection is of great significance for diagnosing the incipient faults of GIS \([4]\).

Among the multifarious measurement methods, the ultra-high-frequency (UHF) method has attracted increasing attention because of its high sensitivity and strong anti-interference capability \([5]\). Different types of PDs give rise to harm with varying degrees. For example, damage caused by PD arising from the protrusion and void in the epoxy resin is more dangerous than that from the floating metal and bouncing particles \([6]\). Therefore, the PD source identification can provide a guideline for maintenance strategy.

Generally, the patterns to recognise PD in GIS are divided into two categories, i.e. phase-resolved PD (PRPD) \([1]\) and time-resolved partial discharge (TRPD) \([4, 6]\). Sometimes, it is difficult to obtain the phase information of PD in the field \([7]\). Besides, the PRPD mode requires massive memory space to record the data up to one thousand power frequency cycles \([8]\). In contrast, TRPD mode merely needs to acquire a single PD waveform. Since PD pulse typically belongs to the transient and non-stationary signal, separated time or frequency description cannot offer complete information. Comparatively, the time-frequency (TF) analysis is a more powerful tool for characterising PD signal \([9–12]\). The short-time Fourier transform (STFT) \([9]\), wavelet transform (WT) \([10]\), and S-transform (ST) which uniquely combines frequency-dependent resolution with absolutely reference phase \([11, 12]\), have been employed to extract the PD feature parameters. However, these features are applied to the recognition of PD source in power transformers or cables only. Due to the coaxial structure, the propagation properties of electromagnetic (EM) waves in GIS are distinctive. For instance, for the transverse EM (TEM) mode, the cutoff frequency \(f_c\) does not exist, whereas the high-order modes, including transverse electric (TE) mode and transverse magnetic (TM) mode, have a corresponding \(f_c\) \([13]\). In addition, when the UHF signals passed through the insulation spacer, the attenuation was mainly due to the superimposition of reduced TE mode and TM mode \([13]\), and in the case of the disconnecting part, the TEM mode component was reflected, whereas the higher frequency component over the \(f_c\) of TE\(_{11}\) could propagate \([14]\). Furthermore, it was found that the TEM mode component became the main component after the UHF signals passed through the L-shaped branch \([15]\). Therefore, the transmission characteristics of EM modes have a significant influence on the detected UHF signals. Nevertheless, the existing methods of extracting the features of UHF PD signals in GIS \([16–19]\) do not take into account the impact of EM mode on the recognition accuracy of the PD source. In our work, the method that using the EM mode divides the TF plane is investigated to improve the accuracy.

Various classification techniques, like k-means clustering \([6]\), fuzzy c-means clustering \([20]\), probabilistic neural network \([21]\), support vector machine (SVM) \([22]\), and k-nearest neighbour (KNN) \([23]\), have been applied to the PD source identification of high-voltage equipment. Nevertheless, both the training and/or testing phases of these classifiers are time-consuming. Recently, a novel learning algorithm for single-hidden layer feedforward neural network (SLFN), the so-called extreme learning machine (ELM) \([24]\), is considered to be able to efficiently process large data set and have excellent classification capability. However, in \([24]\), the input weights and biases are randomly generated. Since these critical parameters have a great influence on the performance of ELM, they should be tuned.

The novelty and contribution of this paper are elaborated as follows:

(i) According to the cutoff frequencies of TE\(_{m1}\) modes in the range 0–2 GHz, the TF plane is divided into five regions. The three low-order moments of every subregion are extracted as the original feature space. For the ELM, SVM and KNN classifiers, the
recognition accuracies based on the selected six moments through $J$

criterion are raised up by 26.5, 29 and 37%, respectively, compared to those based on the three low-order moments of the whole
greyscale image. It demonstrates that our selected feature
parameters significantly improve the recognition accuracy of PD
source in GIS. Thus, the image segmentation method based on the
greyscale image. It demonstrates that our selected feature
parameters significantly improve the recognition accuracy of PD

(ii) In our work, the input weights and biases are tuned by particle
swarm optimisation (PSO) technique rather than generated randomly. When the ELM is applied to GIS PD pattern
recognition, it provides the comparative generalisation performance at the fastest testing speeds compared to SVM and
KNN, thereby more appropriate for the real-time PD online
monitoring. Moreover, the PSO-ELM algorithm is first elaborately
applied to GIS PD source identification.

This paper is organised as follows: Section 2 mainly deals with the
UHF PD measurement system. Then, the procedures of signal
preprocessing and feature extraction are elaborated in Section 3,
followed by the brief introduction of ELM optimised by PSO, as
presented in Section 4. Furthermore, based on the moment features,
the performance of ELM is investigated and compared with that
derived from SVM and KNN classifiers in Section 5. Finally,
Section 6 draws the conclusions.

2 Experiment’s setup

Fig. 1 depicts four types of artificial insulation defects. The
metallic protrusion defect consists of a pair of needle-to-plane
electrode; the floating metal defect is realised by suspending an
insulation ring to fasten a floating brass, the HV electrode and
floating brass use tapered tips in order to discharge easily and the
plane electrode employs rounded corners to avoid corona
discharge; to imitate the free metallic particle defect, fifteen
aluminium foils with various diameters are put in a bowl-shaped
electrode; the void defect is composed of three layers of epoxy
resin and the middle one has a small hole. Although the designed
artificial defect models are not exactly the same as the real
insulation defects in practice, their PD mechanisms [25–27] are
identical. Fig. 2 presents the experimental setup of UHF PD
detection. The artificial insulation defect is placed in a perspex
glass container to generate the PD signal. The PD source
composed of insulation defect and the container is located at the
left end of GIS. The central conductor diameter and enclosure
diameter of 220 kV GIS model are 90 and 320 mm, respectively. A
disc coupler is used as an UHF antenna. Due to the limitation of
measuring range of gigahertz TEM cell [28], its effective height for
characterising the sensitivity to the incident electric field is
measured only in the range of 0.2–2 GHz and the results are plotted
in Fig. 3. A Keysight DSO9404A digital oscilloscope, whose
analogue bandwidth is 4 GHz and the sampling rate is set to 10
GSa/s, is utilised to acquire the PD signals. To ensure the diversity
of samples from the same defect, two experimental parameters are
controlled. One is the test voltage. In our experiments, for each
class of defect, three test voltages are applied, and at each test
voltage, 100 samples are acquired. For example, the three test
voltages of the void defect are 18, 19 and 20 kV, respectively.
Changing the test voltage results in the different discharge quantity,
which reflects the severity of discharge. The other is the duration of
the applied voltage. The PD waveforms, such as the signal
amplitude, slightly vary with the lasting time of applied voltage.
In the experiments, 10 PD pulses are continuously acquired at
intervals of ten minutes. Thus, these samples are representative.
The PD inception voltage (PDIV), test voltage and the number of
samples are given in Table 1. The PDIV mainly depends on the
structure and dimension of defect models. Fig. 4 shows the typical
UHF signals after wavelet-based denoising [29] and Fig. 5 gives
the corresponding normalised frequency spectrums. For the
protrusion defect, the energy of UHF signals is mostly distributed
in the range of 0.4–1.9 GHz; for the floating metal defect, the
energy is within 0.3–1.4 GHz; for the particle defect, the energy is
within 0.5–1.8 GHz; for the void defect, the energy is within 0.5–
1.3 GHz and the lasting time is the shortest.

3 Signal preprocessing and feature extraction

In this work, the PD source identification is split into seven steps,
and the flowchart is shown in Fig. 6.
3.1 S-transform

In ST, the widths of time windows and frequency exhibit contrary change characteristics, resulting in a higher frequency resolution at low frequencies and a higher temporal resolution at high frequencies. Thus, ST integrates the merits of STFT and WT [11, 12]. The ST of signal $x(t)$ is defined as

$$S(\tau, f) = \int_{-\infty}^{+\infty} x(t)w(t - \tau)e^{-j2\pi ft}dt$$

where $f$ represents the frequency of the UHF signal and $\tau$ is a parameter that controls the location in the timeline. Through ST, modulo operation and min–max normalisation, the TF matrix is projected into a greyscale image, as shown in Fig. 7. The deeper the red colour, the larger the magnitude of instantaneous frequency of the UHF signal. On the contrary, the deeper the blue colour, the smaller the magnitude.

3.2 TF plane division

GIS, which consists of a central conductor and an enclosure, is modelled as a cylindrical coaxial waveguide. Theoretically, the cutoff frequency of the TEM mode, $f_c$, is 0 Hz and all frequencies of the TEM mode from direct current upwards can exist in GIS. However, for the high-order modes, only the frequencies above $f_c$ of corresponding modes can propagate. The formula to calculate $f_c$ of TE mode can be found in [30],

$$G_n(ka)Y_n(kb) - Y_n'(ka)G_n'(kb) = 0$$

where $f$ represents the frequency of the UHF signal and $\tau$ is a parameter that controls the location in the timeline. Through ST, modulo operation and min–max normalisation, the TF matrix is projected into a greyscale image, as shown in Fig. 7. The deeper the red colour, the larger the magnitude of instantaneous frequency of the UHF signal. On the contrary, the deeper the blue colour, the smaller the magnitude.

### Table 1

| Defect type | Inception voltage, kV | Test voltage, kV | Number of samples |
|-------------|-----------------------|-----------------|-------------------|
| protrusion  | 8.7                   | 10              | 100               |
|             | 11                    | 100             |                   |
|             | 12                    | 100             |                   |
| floating    | 5.3                   | 7               | 100               |
|             | 8                     | 100             |                   |
|             | 9                     | 100             |                   |
| particle    | 9.2                   | 11              | 100               |
|             | 12                    | 100             |                   |
|             | 13                    | 100             |                   |
| void        | 16.4                  | 18              | 100               |
|             | 19                    | 100             |                   |
|             | 20                    | 100             |                   |

### Fig. 4

**UHF PD signals in the time domain**

(a) Metallic protrusion, (b) Floating metal, (c) Free metallic particles, (d) Void in the epoxy resin

### Fig. 5

Normalised frequency spectrums of UHF signals

(a) Metallic protrusion, (b) Floating metal, (c) Free metallic particles, (d) Void in the epoxy resin

### Fig. 6

Flowchart of PD pattern recognition

1. Input UHF signals after denoising
2. TF representation by ST
3. TF plane division based on propagation modes
4. Moment features extraction
5. Feature selection using $J$ criterion
6. ELM optimized by PSO
7. Type of PD

$$S(\tau, f) = \int_{-\infty}^{+\infty} x(t)w(t - \tau)e^{-j2\pi ft}dt$$

$$w(t - \tau, f) = \frac{1}{\sqrt{2\pi}}e^{-\left(t^2 - \tau^2\right)/2\sigma^2}$$

(1)

---

This is an open access article published by the IET and CEPRI under the Creative Commons Attribution License (http://creativecommons.org/licenses/by/3.0/)

---

High Volt., 2020, Vol. 5 Iss. 6, pp. 688-696
first derivatives of the \( n \)th-order Bessel function of the first and second kind, respectively. Table 2 lists the cutoff frequencies of \( \text{TE}_{m1} \) modes \((m=1, 2, 3, 4\) or \(5)\) modes <2 GHz. The higher the order of \( \text{TE} \) mode, the greater the cutoff frequency.

In addition, \( \text{TEM} \) component in GIS propagates at the light speed, but the high-order modes display the velocity dispersion characteristics and their velocities depend on the frequencies. The propagation velocity of the \( \text{TE} \) mode is given by

\[
v_{\text{TE}} = c \sqrt{1 - \left(\frac{f}{f_c}\right)^2}
\]

where \( c \) is the light speed in the free space. Fig. 8 presents the velocities of \( \text{TEM} \) and \( \text{TE}_{m1} \) modes. It is observed that all frequency components of \( \text{TE} \) mode can propagate, whereas each \( \text{TE}_{m1} \) mode has a corresponding cutoff frequency, which is consistent with Table 2. Moreover, the higher the order of \( \text{TE}_{m1} \) mode is, the slower is the propagation velocity at the same frequency.

Considering the different propagation velocities and attenuation characteristics of EM modes in GIS, the TF plane is divided into five rectangular regions according to the cutoff frequency of \( \text{TE}_{m1} \) mode. The split lines of these regions along the vertical axis are at frequencies 0 Hz, 479, 1249, 1586 MHz, and 2 GHz, respectively, as illustrated in Fig. 9.

### 3.3 Feature extraction

In our work, the grey distribution of the image is characterised by its three low-order moments, i.e. the mean, the standard deviation and the third root of skewness. The extracted moments from the whole image are lack of spatial information. Therefore, the greyscale image is first divided into five regions according to the EM modes, and then the three low-order moments of every subregion are computed as the original feature space.

\[
\begin{align*}
u_i &= \frac{1}{N_i} \sum_{j=1}^{N_i} p_{i,j} \\
\sigma_i &= \left( \frac{1}{N_i} \sum_{j=1}^{N_i} (p_{i,j} - u_i)^2 \right)^{1/2} \\
s_i &= \left( \frac{1}{N_i} \sum_{j=1}^{N_i} (p_{i,j} - u_i)^3 \right)^{1/3}
\end{align*}
\]

where \( u_i \) is the mean of the \( i \)th region, \( \sigma_i \) is the standard deviation of the \( i \)th region, \( s_i \) is the third root of skewness of the \( i \)th region, \( N_i \) is the total number of pixels in the \( i \)th region and \( p_{i,j} \) is the grey level of the \( j \)th pixel in the \( i \)th region. The mean can represent the average grey level of the image. The larger the mean, the greater the energy of UHF PD signals. The standard deviation can measure how far a set of the pixel value is spread out from their mean. The greater the standard deviation is, the more dispersed the data is.

### Table 2

| Mode | \( \text{TE}_{11} \) | \( \text{TE}_{21} \) | \( \text{TE}_{31} \) | \( \text{TE}_{41} \) | \( \text{TE}_{51} \) |
|------|----------------|----------------|----------------|----------------|----------------|
| \( f_c, \text{ MHz} \) | 479 | 891 | 1249 | 1586 | 1914 |
around the mean. The skewness can describe the symmetrical degree of pixel value around their mean. The larger the skewness, the less symmetric the distribution of data. This is the reason why three low-order moments are chosen as features.

3.4 Feature selection

In order to address the issue caused by dimensionality, the $J$ criterion is utilised to reduce the dimensionality of the original feature space [31]. The $J$-value is equal to the ratio of the between-class scatter value $S_b$ to the within-class scatter value $S_w$. For $L$-class problems, the $J$-value is calculated as follows

$$ J = \frac{S_b}{S_w} = \frac{\sum_{c=1}^{L} (N_c/N)(m_c - m_0)^2}{\sum_{c=1}^{L} (N_c/N)\sigma_c^2} \tag{5} $$

where $N_c$ is the number of samples belonging to class $c$, $N$ is the total number of samples, $m_c$ is the mean of selected feature for class $c$, $m_0$ is the mean of selected feature for all samples, and $\sigma_c$ is the standard deviation of selected feature for class $c$. The $J$-values of all the 15 moment features are shown in Fig. 10. A greater $J$-value denotes that the feature parameter has good capability to separate the different classes. Applying this criterion, six features with the $J$-value $>0.2$, namely, $s_2$, $s_4$ and $s_5$, $s_2$, $s_4$ and $s_5$ (a) $s_2$, $s_4$ and $s_5$, (b) $s_2$, $s_4$ and $s_5$, are chosen as the input of classifiers and they are plotted in Fig. 11.

4 Principle and parameter optimisation of ELM

4.1 Principle of ELM

ELM is a highly efficient learning algorithm [24]. Given $N$ observations $(x_i, y_i)$, where $x_i = [x_{i1}, x_{i2}, \ldots, x_{in}]^T \in \mathbb{R}^n$, $y_i = [y_{i1}, y_{i2}, \ldots, y_{il}]^T \in \mathbb{R}^l$, the output of SLFN with $M$ neurons in the middle layer is expressed as

$$ y_j = \sum_{i=1}^{M} \beta_i g(w_i \cdot x_j + b_i), \quad 1 \leq j \leq N \tag{6} $$

which can also be written in the form of the matrix $H\beta = Y$, with

$$ H = \begin{bmatrix} g(w_1 \cdot x_1 + b_1) & \cdots & g(w_M \cdot x_1 + b_M) \\ \vdots & \ddots & \vdots \\ g(w_1 \cdot x_N + b_1) & \cdots & g(w_M \cdot x_N + b_M) \end{bmatrix}_{N \times M} \tag{7} $$

$$ \beta = \begin{bmatrix} \beta_1^T \\ \vdots \\ \beta_M^T \end{bmatrix}_{M \times 1} \quad \text{and} \quad Y = \begin{bmatrix} y_1^T \\ \vdots \\ y_N^T \end{bmatrix}_{N \times 1} \tag{8} $$

where $g(x)$ is the activation function, $w_i = [w_{i1}, w_{i2}, \ldots, w_{in}]^T$ is the input weight vector linking the $i$th hidden node and the input nodes, $b_i$ is the bias of the $i$th hidden node, $\beta_i = [\beta_{i1}, \beta_{i2}, \ldots, \beta_{il}]^T$ is the output weight vector linking the $i$th hidden node and the output nodes, and $H$ is called the hidden-layer output matrix. In the ELM

---

This is an open access article published by the IET and CEPRI under the Creative Commons Attribution License (http://creativecommons.org/licenses/by/3.0/)
Input: $p = 80$, $c_2 = 2$, $c_2 = 0.2$, $c_{max} = 0.9$, $h_{max} = 0.4$, $h_{max} = 600$, $V_{min}$, $V_{max}$, $X_{min}$, $X_{max}$.

Output: $g_{Best}$ under the different activation functions and number of hidden nodes.

For $af = 1:5$

1. set $act\_func[af]$ as the activation function;
2. for neuron $num = 10:20$
3. set $num_{max}$ as the number of hidden nodes;
4. randomly initialize all the particles;
5. for $h = 1: h_{max}$
6. for each particle $k = 1: p$
7. compute $\hat{\beta}$ based on Eq. (9);
8. solve $Y$ according to $Y = \beta H$;
9. calculate fitness value according to Eq. (11);
10. update velocity and position based on Eq. (10), and constrain $V^{n+1}(k)\in (V_{min}, V_{max})$ and $X^{n+1}(k)\in (X_{min}, X_{max})$;
11. update the $p_{Best}(k)$;
12. end for
13. record the $g_{Best}$ under the current type of activation function and number of hidden nodes;
14. end for
15. end for
16. end for

Fig. 12 Pseudocode of PSO-ELM algorithm

Fig. 13 Number of hidden nodes versus training accuracy under different activation functions

Table 3 Key parameters of ELM and PSO algorithms

| Algorithm | Parameter | Value |
|-----------|-----------|-------|
| ELM [24] | number of hidden nodes | 16 |
| | activation function | sig |
| | input weights $w$ | $(-1, 1)$ |
| | input biases $b$ | $(0, 1)$ |
| PSO [32] | swarm size $p$ | 60 |
| | inertia weight $\epsilon$ | 0.9–0.4 |
| | acceleration const $c_1$ and $c_2$ | 2 |
| | maximum number of iterations $h_{max}$ | 600 |

algorithm, the $w$ and $b$ are randomly generated, and the output weight matrix $\beta$ is estimated as

$$\hat{\beta} = \hat{H}Y$$

where $\hat{H}$ is the Moore–Penrose generalised inverse of $H$.

4.2 Parameter optimisation of ELM

To obtain better generalisation performance, the parameters of the hidden layer of ELM need optimisation. As an evolutionary algorithm, the PSO possesses a strong global search ability to avoid local optimum. In our work, the PSO is utilised to obtain the optimal input weights $w = [w_1, w_2, \ldots, w_{16}, w_{17}, \ldots, w_{600}](-1, 1)$ and input biases $b = [b_1, b_2, \ldots, b_20](0, 1)$. Each particle $[w, b]$ is a potential solution in $(n \times M + M)$-dimensional space. At each iteration, it is manipulated using [32]

$$V^{h+1}(k) = \epsilon V^{h}(k) + c_1 r_1(p_{Best}(k) - X^{h}(k)) + c_2 r_2(g_{Best} - X^{h}(k))$$

$$X^{h+1}(k) = X^{h}(k) + V^{h+1}(k)$$

where $h$ is the current number of iterations; both $V^{h}(k)$ and $X^{h}(k)\in \mathbb{R}^{n \times M + M}$, and they refer to the velocity and position of particle $k$ at the $h$th epoch, respectively; $c_1$ and $c_2$ represent the acceleration const$; r_1$ and $r_2$ denote two random numbers within $(0, 1)$; $p_{Best}(k)$ is the best position of particle $k$; $g_{Best}$ is the global best position of swarm; $\epsilon$ is the current inertia weight; $e_{min}$ and $e_{max}$ are the initial and final inertia weights, respectively; $h_{max}$ is the maximum number of iterations. In addition, the velocity and position are limited within a certain range, i.e. $V^{h+1}(k)\in (V_{min}, V_{max})$ and $X^{h+1}(k)\in (X_{min}, X_{max})$. Additionally, for each particle, the root means standard error (RMSE) [24] is calculated as its fitness value

$$RMSE = \sqrt{\frac{1}{T \times N} \sum_{i=1}^{N} \| y_i - \hat{y}_i \|}$$

where $y_i$ is the expected output of ELM and $l$ is the length of the vector $y_i$. The pseudocode about the optimisation procedure is given in Fig. 12.

5 Results and discussion

The simulations are carried out in MATLAB 2016b environment running in an Intel core i5–6500 CPU with clock speed 3.2 GHz. Each class of measured PD signals is randomly divided into two groups, one of which contains 200 samples as training data set and the other contains 100 samples as testing dataset. Hence, the numbers of total training and testing samples are 800 and 400, respectively.

Fig. 13 plots the number of hidden nodes versus training accuracy under different activation functions. To avoid overfitting, the optimal number of hidden neurons is first estimated according to the empirical formula in [33]. The training accuracy is raised up with increasing the number of hidden nodes, but when the number of hidden nodes is >16, the training accuracy grows at a snail’s pace on matter the type of activation functions. Moreover, the ‘sig’ function performs the best, whereas the ‘hardlim’ function does the worst. Ultimately, the ‘sig’ function is selected and the number of hidden nodes is set to 16. The key parameters of the two algorithms are listed in Table 3. The fitness in the optimisation process is illustrated in Fig. 14. The part of the output results of ELM is listed in Table 4. For the second case, the correct class should be protrusion defect, whereas it is misclassified as a particle defect. Similarly, for the sixth case, the correct class should be particle defect, whereas it is misclassified as a floating metal defect. Accordingly, the classification results from ELM are shown in Table 5, where six samples are misclassified as protrusion defect, and one sample is misclassified as floating defect and four samples are misclassified as gap defect in terms of particle defect. It can be seen that the main classification errors are concentrated on the type of particle defect. This is attributed to the great dispersity of UHF signals induced by the free metallic particles. For the ELM classifier, the overall accuracy is 95%.

The multiclass SVM classifier model [34] is established to directly achieve multiclass classification by using LIBSVM 3.22 toolbox. Its kernel function is set to radial basis function, and the
penalty factor $C$ and kernel parameter $\gamma$ are selected with ten-fold cross-validation and grid-search method in the range from $2^{-8}$ to $2^8$. From Fig. 15, the highest training accuracy can reach 98.75% when $C = 2^4$ and $\gamma = 2^6$. Regarding KNN, its highest testing accuracy can reach 88.5% when the type of distance is set to Euclidean distance and the number of nearest neighbours is equal to 5 through considerable trials.

Table 6 lists the centroid of training dataset for six moment features. After building the SVM and KNN classifier models, the remaining samples are tested. Table 7 summarises the overall accuracy and time cost of three classifiers. It can be observed that although the training accuracy of SVM is slightly superior to the ELM, the ELM performs the best in the testing accuracy. Since the $K$-fold cross-validation initially randomly divides the data into $K$ subsets, there is still a possibility of a non-representative data split being used for training and validation [35], resulting in that the testing accuracy of SVM is 6.75% lower than the training accuracy. From the perspective of time cost, the ELM only requires 5.25% of the training time taken by the SVM, and the testing speed of ELM is 21.33 and 9.67 times faster than those of SVM and KNN, respectively. Tables 8 and 9 list the precision and recall, respectively. For each class of defect, three classifiers have high precision and recall at the same time. It demonstrates that the overfitting does not occur.

Table 10 lists the computational complexity. Since the KNN is a type of lazy learning [37], no explicit training step is required, and for the training phase, both the time cost and computational complexity are 0. The parameters in Table 10 are illustrated as follows: $D_i = 6$ (the dimension of input features), $D_o = 4$ (the dimension of the output result of ELM), $N_h = 16$ (the number of hidden neurons of ELM), $N_t = 200$ (the number of training samples) and $N_s$ is the number of support vectors of SVM. Thereby, for the training and testing phases of ELM, the complexities are $\tau(32,000)$ and $\tau(160)$, respectively; for the testing phase of KNN, the complexity is $\tau(1200)$. Besides, the random allocation of training and testing samples leads to the varied number of support vectors at each run [36], and the time in

![Fig. 14](image1.png)  
*Fitness versus epoch*

![Fig. 15](image2.png)  
*Training accuracy of SVM based on the grid-search method*

2. From Fig. 15, the highest training accuracy can reach 98.75% when $C = 2^5$ and $\gamma = 2^5$. Regarding KNN, its highest testing accuracy can reach 88.5% when the type of distance is set to Euclidean distance and the number of nearest neighbours is equal to 5 through considerable trials. Table 6 lists the centroid of training dataset for six moment features. After building the SVM and KNN classifier models, the remaining samples are tested. Table 7 summarises the overall accuracy and time cost of three classifiers. It can be observed that although the training accuracy of SVM is slightly superior to the ELM, the ELM performs the best in the testing accuracy. Since the $K$-fold cross-validation initially randomly divides the data into $K$ subsets, there is still a possibility of a non-representative data split being used for training and validation [35], resulting in that the testing accuracy of SVM is 6.75% lower than the training accuracy. From the perspective of time cost, the ELM only requires 5.25% of the training time taken by the SVM, and the testing speed of ELM is 21.33 and 9.67 times faster than those of SVM and KNN, respectively. Tables 8 and 9 list the precision and recall, respectively. For each class of defect, three classifiers have high precision and recall at the same time. It demonstrates that the overfitting does not occur.

In addition, Table 10 lists the computational complexity. Since the KNN is a type of lazy learning [37], no explicit training step is required, and for the training phase, both the time cost and computational complexity are 0. The parameters in Table 10 are illustrated as follows: $D_i = 6$ (the dimension of input features), $D_o = 4$ (the dimension of the output result of ELM), $N_h = 16$ (the number of hidden neurons of ELM), $N_t = 200$ (the number of training samples) and $N_s$ is the number of support vectors of SVM. Thereby, for the training and testing phases of ELM, the complexities are $O(32,000)$ and $O(160)$, respectively; for the testing phase of KNN, the complexity is $O(1200)$. Besides, the random allocation of training and testing samples leads to the varied number of support vectors at each run [36], and the time in
Table 10  Computational complexity

| Classifier | Complexity | Training | Testing |
|------------|------------|----------|---------|
| ELM        | \(O(N_{\text{h}}(N_{\text{h}}+D))\) | \(O(N_{\text{H}}(N_{\text{H}}+D))\) |
| SVM        | \(O(N_{\text{h}} + N_{\text{H}} + N_{\text{H}}D)\) [36] | \(O(DN_{\text{H}})\) [36] |
| KNN        | \(O(0)\) | \(O(DN_{\text{H}})\) |

Table 11  Training and testing accuracies based on the three low-order moments of the whole greyscale image

| Classifier | Training accuracy, % | Testing accuracy, % |
|------------|----------------------|---------------------|
| ELM        | 70                   | 68.5                |
| SVM        | 70.5                 | 63                  |
| KNN        | —                    | 51.5                |

Table 12  15 moment features based on the descending order of \(J\)-value

| Order | Feature | \(J\)-value | Order | Feature | \(J\)-value |
|-------|---------|-------------|-------|---------|-------------|
| 1     | \(s_4\) | 0.705       | 9     | \(u_2\) | 0.086       |
| 2     | \(a_4\) | 0.585       | 10    | \(\sigma_1\) | 0.060       |
| 3     | \(s_5\) | 0.517       | 11    | \(\sigma_1\) | 0.059       |
| 4     | \(a_2\) | 0.379       | 12    | \(\mu_1\) | 0.058       |
| 5     | \(\sigma_5\) | 0.362       | 13    | \(\sigma_3\) | 0.043       |
| 6     | \(\sigma_2\) | 0.347       | 14    | \(\sigma_3\) | 0.034       |
| 7     | \(u_4\) | 0.198       | 15    | \(u_3\) | 0.033       |
| 8     | \(u_5\) | 0.140       | —     | —       | —           |

This work is supported by the National Natural Science Foundation of China (No. 51677061, 51507058).

8 References

[1] Zeng, F., Dong, Y., Tang, J.: ‘Feature extraction and severity assessment of partial discharge under protrusion defect based on fuzzy comprehensive evaluation’, IET Gener. Transm. Distrib., 2015, 9, (16), pp. 2493–2500
[2] Ma, Z., Yang, Y., Cowan, K., et al.: ‘Fractal-based autonomous partial discharge pattern recognition method for MV motors’, *High Voltage*, 2018, 3, (2), pp. 103–114.

[3] Behzad, K., Hamed, J.: ‘Towards automated statistical partial discharge source classification using pattern recognition techniques’, *High Voltage*, 2018, 3, (3), pp. 162–169.

[4] Luo, L., Han, B., Chen, J., et al.: ‘Partial discharge detection and recognition in random matrix theory paradigm’, *IEEE Access*, 2017, 5, pp. 8025–8031.

[5] Li, Z., Luo, L., Sheng, G., et al.: ‘UHF partial discharge localisation method in substation based on dimension-reduced RSSI fingerprint’, *IET Gener. Transm. Distrib.*, 2018, 12, (2), pp. 398–405.

[6] Gao, W., Ding, D., Liu, W.: ‘Research on the typical partial discharge using the UHF detection method for GIS’, *IEEE Trans. Power Deliv.*, 2011, 26, (4), pp. 2621–2629.

[7] Tang, J., Wang, D., Fan, L., et al.: ‘Feature parameters extraction of GIS partial discharge signal with multifractal detrended fluctuation analysis’, *IEEE Trans. Dielectr. Electr. Insul.*, 2015, 22, (5), pp. 3037–3045.

[8] Li, J., Jiang, T., Harrison, R., et al.: ‘Recognition of ultra high frequency partial discharge signals using multi-scale features’, *IEEE Trans. Dielectr. Electr. Insul.*, 2012, 19, (4), pp. 1412–1420.

[9] Luo, L., Han, B., Chen, J., et al.: ‘A hybrid algorithm based on S transform and affinity propagation clustering for separation of two simultaneously artificial partial discharge sources’, *IEEE Trans. Dielectr. Electr. Insul.*, 2015, 22, (2), pp. 1042–1060.

[10] Hikita, M., Ohtsuka, S., Ueta, G., et al.: ‘Influence of insulating spacer type on propagation properties of PD-induced electromagnetic wave in GIS’, *IEEE Trans. Dielectr. Electr. Insul.*, 2010, 17, (5), pp. 1642–1648.

[11] Hikita, M., Ohtsuka, S., Ueta, G., et al.: ‘Influence of disconnecting part on propagation properties of PD-induced electromagnetic wave in model GIS’, *IEEE Trans. Dielectr. Electr. Insul.*, 2010, 17, (6), pp. 1731–1737.

[12] Hikita, M., Ohtsuka, S., Wada, J., et al.: ‘Propagation properties of PD-induced electromagnetic wave in 66 kV GIS model tank with I branch structure’, *IEEE Trans. Dielectr. Electr. Insul.*, 2011, 18, (5), pp. 1678–1685.

[13] Zhu, M., Xue, J., Zhang, J., et al.: ‘Classification and separation of partial discharge ultrahigh-frequency signals in a 252 kV gas insulated substation by using cumulative energy technique’, *IET Sci. Meas. Technol.*, 2016, 10, (4), pp. 316–326.

[14] Zou, F., Chang, C.S., Chang, C., et al.: ‘Classification of partial discharge events in gas-insulated substations using wavelet packet transform and neural network approaches’, *IEEE Proc. – Sci. Meas. Technol.*, 2006, 153, (2), pp. 55–63.

[15] Dai, D., Wang, X., Long, J., et al.: ‘Feature extraction of GIS partial discharge signal based on S-transform and singular value decomposition’, *IET Sci. Meas. Technol.*, 2017, 11, (2), pp. 186–193.

[16] Li, X., Wang, X., Xie, D., et al.: ‘Time-frequency analysis of PD-induced UHF signal in GIS and feature extraction using invariant moments’, *IET Sci. Meas. Technol.*, 2018, 12, (2), pp. 169–175.

[17] Wang, K., Li, J., Zhang, S., et al.: ‘A new image-oriented feature extraction method for partial discharges’, *IEEE Trans. Dielectr. Electr. Insul.*, 2015, 22, (2), pp. 1015–1024.

[18] Evagorou, D., Kyprianou, A., Lewin, P.L., et al.: ‘Feature extraction of partial discharge signals using the wavelet packet transform and classification with a probabilistic neural network’, *IET Sci. Meas. Technol.*, 2010, 4, (3), pp. 177–192.

[19] Zhong, K., Si, G., Diao, L., et al.: ‘Partial discharge recognition based on SF6 decomposition products and support vector machine’, *IET Sci. Meas. Technol.*, 2012, 6, (4), pp. 198–204.

[20] Harbaj, M.M.O., Zahed, A.H., Habbour, S.A., et al.: ‘Design of hilbert fractal antenna for partial discharge classification in oil-paper insulated system’, *IEEE Sensors J.*, 2017, 17, (4), pp. 1037–1045.

[21] Huang, G.B., Zhu, Q.Y., Siew, C.K.: ‘Extreme learning machine: theory and applications’, *Neurocomputing*, 2006, 70, (1), pp. 489–501.

[22] Boggs, S.A.: ‘Partial discharge: overview and signal generation’, *IEEE Electr. Insul. Mag.*, 1999, 6, (4), pp. 33–39.

[23] Hara, M., Negara, Y., Setoguchi, M., et al.: ‘Particle-triggered pre-breakdown phenomena in atmospheric air gap under AC voltage’, *IEEE Trans. Dielectr. Electr. Insul.*, 2005, 12, (5), pp. 1071–1081.

[24] Boggs, S.A.: ‘Partial discharge III: cavity-induced PD in solid dielectrics’, *IEEE Electr. Insul. Mag.*, 1996, 6, (6), pp. 11–16.

[25] Judd, M.D., Farish, O.: ‘A pulsed GTEM system for UHF sensor calibration’, *IEEE Trans. Instrum. Meas.*, 1998, 47, (4), pp. 875–880.

[26] Florkowski, M., Florkowska, B.: ‘Wavelet-based partial discharge image denoising’, *IET Gener. Transm. Distrib.*, 2007, 1, (2), pp. 340–347.

[27] Eberhart, R., Kennedy, J.: ‘A new optimizer using particle swarm theory’. Proc. Sixth Int. Symp. on Micro Machine and Human Science, Nagoya, Japan, 1995, pp. 39–43.

[28] Sinaga, H.H., Phung, B.T., Blackburn, T.R.: ‘Recognition of single and multiple partial discharge sources in transformers based on ultra-high frequency signals’, *IEEE Trans. Dielectr. Electr. Insul.*, 1996, 3, (2), pp. 213–222.

[29] Huang, G.B., Zhu, Q.Y., Siew, C.K.: ‘Extreme learning machine: theory and applications’, *Neurocomputing*, 2006, 70, (1), pp. 489–501.

[30] Mathur, A., Foody, G.M.: ‘Multiclass and binary SVM classification: implications for training and classification users’, *IEEE Geosci. Remote Sens. Lett.*, 2008, 5, (2), pp. 241–245.

[31] Clark, A.R.J., Everson, R.M.: ‘Multi-objective learning of relevance vector machine classifiers with multi-resolution kernels’, *Pattern Recogn.*, 2012, 45, (9), pp. 3533–3543.

[32] Burgs, C.J.: ‘A tutorial on support vector machines for pattern recognition’, *Data Min. Knowl. Discov.*, 1998, 2, (2), pp. 121–167.

[33] Garcia, E.K., Feldman, S., Gupta, M.R., et al.: ‘Completely lazy learning’, *IEEE Trans. Knowl. Data Eng.*, 2010, 22, (9), pp. 1274–1285.