Partial Discharge Pattern Recognition of Transformers Based on the Gray-Level Co-Occurrence Matrix of Optimal Parameters

SHENGYA SUN[1], **YUANYUAN SUN**[1], **(Member, IEEE), GONGDE XU**[1], **LINA ZHANG**[2]**, YIRU HU**[2]**, AND PING LIU**[3]

[1]School of Electrical Engineering, Shandong University, Jinan 250061, China
[2]China National Offshore Oil Corporation, Beijing 100027, China
[3]CNOOC Energy Development Equipment Technology Company Ltd., Tianjin 300452, China

Corresponding author: Yuanyuan Sun (sunyy@sdu.edu.cn)

ABSTRACT The partial discharge (PD) is the most common fault of transformers, which is the main factor affecting the stable operation of transformers. Therefore, the PD should be monitored and identified timely to improve the reliability of the transformers. In this paper, a transformer PD pattern recognition algorithm based on the gray-level co-occurrence matrix of optimal parameters and support vector machine (GLCMOP-SVM) is proposed. Firstly, the GLCM of optimal parameters (GLCMOP) is proposed to be determined by calculating the proportion of the off-diagonal elements (*PODE*) in GLCM. The GLCMOP has the advantage of avoiding the subjectivity of parameter selection and simplifying the calculation process. Then, the phase-resolved partial discharge (PRPD) maps are used as the PD samples and are converted into the GLCMOP to extract the PD features. Moreover, the feature space of the GLCMOP is dimensionally reduced by screening out the features with high distinguishability, which can improve the generalization ability and recognition speed of the classifier. Finally, the SVM classifier is trained to sort the PD samples and recognize the PD types, which include the tip discharge, surface discharge, and air discharge PD types. Lab tests are performed to verify the accuracy and validity of the proposed methodology. Compared with the traditional algorithms based on GLCM, XGBoost (eXtreme Gradient Boosting) and artificial neural network (ANN), the performance of GLCMOP-SVM is better. The GLCMOP-SVM has less memory consumption and faster recognition speed, so it is very suitable for the online and real-time monitoring of PD occurred in the transformers.

INDEX TERMS Partial discharge (PD), pattern recognition, gray-level co-occurrence matrix (GLCM), feature extraction, support vector machine (SVM), PRPD maps.

I. INTRODUCTION

The transformer PD monitoring can not only reflect the early insulation failure of the transformers but also evaluate the insulation state and severity by judging the type of PD [1], [2]. The accurate recognition of PD types is the prerequisite for analyzing the insulation faults and guiding the transformer's maintenance operation [3], [4]. Therefore, the PD pattern

recognition of the transformers is of great significance for the safe and stable operation of the power system [5]–[7].

The feature extraction and classifier design are the two most important stages in PD pattern recognition [8]. In recent years, varieties of feature extraction algorithms and pattern classifiers have been applied to the transformer PD pattern recognition. Based on the PRPD maps, the statistical data of PD amplitude, the number, and distribution characteristics of PD pulses can be obtained [9], [10]. The transformer PD pattern recognition can be realized by combining the PD feature extraction of PRPD maps and the PD pattern

---

The associate editor coordinating the review of this manuscript and approving it for publication was Wei-Yen Hsu.

classifier [11]. The PRPD maps are used as PD samples and are transformed into the PRPD grayscale images. The ANN classifier is used to extract the sample features and recognize the PD types [12]. However, practically, the PD samples of transformers are very limited, and the sample needs of the ANN classifier cannot be met, which leads to the underfitting of the ANN classifier. To reach a stable recognition performance of the classifier in practical applications, the simple clustering classifier, and statistical features are used in [13]. However, there are still some problems in this method, such as the poor distinguishability of PD statistical features and low recognition speed. In contrast, the SVM classifier has the advantages of fast recognition speed and less sample demand, therefore it gains wide application due to its good recognition performance [14].

The GLCM has been used to extract the PD features from the PRPD grayscale images effectively [15], [16], which has good PD feature distinguishability. However, the angle-offset parameter θ used to calculate the GLCM has several values and is usually determined artificially, which results in a bad self-adaptability [17]. If the value of the angle-offset parameter θ is not reasonably determined, the distinguishability of the corresponding GLCM features will be greatly reduced, and the features cannot reflect the essential characteristics of PD well [18]. To solve this problem, the gray-level co-occurrence matrices corresponding to all values of the angle-offset parameter θ are calculated. And then the average value of all gray-level co-occurrence matrices is calculated as the final GLCM to extract the features [19]. However, the method of calculating the average of all gray-level co-occurrence matrices requires mass computation and cannot retain the characteristics of all types of PD samples. In addition, the dimension of the feature space extracted by the existing PD pattern recognition algorithm is higher. The high dimension of feature space leads to a long recognition time and poor generalization ability, which cannot meet the real-time recognition speed and accuracy requirements.

Considering the above research gaps, a transformer PD pattern recognition algorithm based on the GLCM of optimal parameters and SVM (GLCMOP-SVM) is proposed. This algorithm can adaptively determine the optimal value of the angle-offset parameter θ according to the measured PRPD maps. We only need to calculate the GLCMOP corresponding to the optimal angle-offset parameter θ and extract the features of GLCMOP, the GLCM corresponding to other values of the angle-offset parameter θ does not need to be calculated. Therefore, the proposed method can not only reflect the PD features well, but also reduce the calculation burden. Moreover, the feature space of PD samples is dimensionally reduced by screening out the features with high distinguishability, which is based on the three-dimensional distribution maps and the intersection of the probability distribution. The feature space of PD samples which has low dimension and high distinguishability features reduces the training time of the classifier and improves the recognition accuracy. Furthermore, the GLCMOP-SVM is compared with other pattern
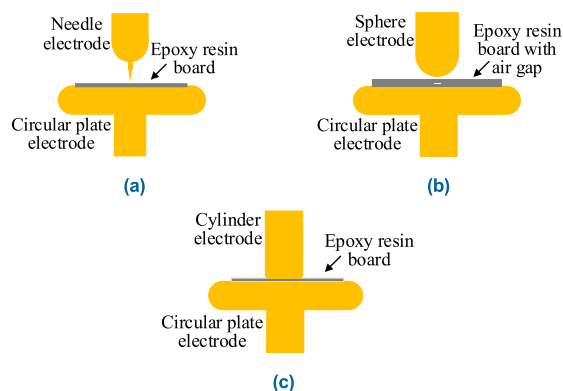
FIGURE 1. The insulation defect models (a), (b), (c) simulate the tip discharge, surface discharge, and air discharge respectively.

recognition methods. Comparison results show that the performance of GLCMOP-SVM is better than that of the other similar pattern recognition methods.

The remainder of this paper is organized as follows. In Section II, the PD test platform based on the integral insulation structure of the transformer and the generation processes of PRPD maps are illustrated, then the GLCM principle and how to use the GLCM to extract the features of PRPD maps are introduced. A description of the GLCMOP-based feature extraction method and the distinguishability screening of features are introduced in Section III. Section IV highlights the GLCMOP-SVM with a discussion about the experimental results and a comparison with a few previous pattern recognition methods. Finally, the conclusions are given in Section V.

II. THE PRPD MAPMS ACQUISITION AND GLCM-BASED FEATURE EXTRACTION
### A. THE PRPD MAPS ACQUISITION
According to the common PD types of insulation defects in the transformers, three typical insulation defects models are constructed, which include the tip discharge, surface discharge, and air discharge types [20] as shown in Fig.1. In this paper, the transformer with an air-epoxy resin insulation structure is used as an example to illustrate the insulation defects model. Moreover, the PD test insulation defect model is designed to occur in the closed air epoxy insulation to make the simulated PD reflect the actual transformer PD well. The enclosed air epoxy insulation model adopts the overall transformer structure as depicted in Fig.2.

The PD model is connected to the PD test circuit to form a test platform. The above three types of insulation defect models are used for the PD test data acquisition. The PD test platform circuit and the field test picture are depicted in Fig.3. In the PD test platform, T_1 is the tap-changing transformer, T_2 is the test transformer without PD, R_1 is the protective resistance, R_2 is the measuring resistance, C_1 is the coupling capacitor.

During the test, the voltage is increased step by step and the ultra-high frequency (UHF) sensors are used to collect
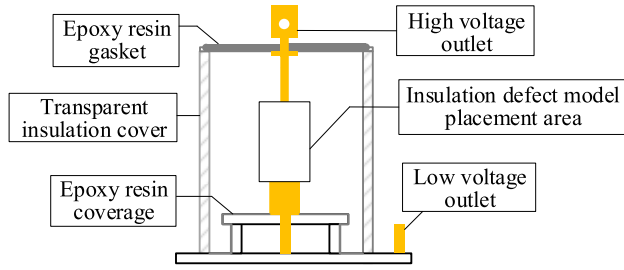
FIGURE 2. The PD test model has a closed air-epoxy resin insulation structure similar to the transformer.
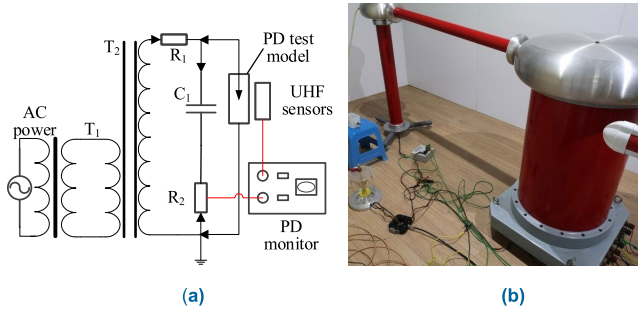

FIGURE 3. The PD test platform circuit diagram (a) and field test picture (b).

the PD data of the three PD models. The collected PD data is used to establish the sample database of PRPD maps for the subsequent feature extraction and pattern recognition. The specific test steps are listed as follows:

(1) A PD insulation defect model is selected and connected to the test circuit.

(2) The background noise of the laboratory is determined. Before the voltage is applied, the signal detected by the signal acquisition system is the background noise in the PD test environment.

(3) The inception voltage of PD is determined. The voltage of the tap-changing transformer is increased evenly and slowly step by step. When the PD signal is detected by the monitoring instrument, the corresponding voltage is recorded as the initial voltage U_1.

(4) The PD data are generated and collected with different voltage supplies. The voltage is continuously increased, when it is larger than 1.5 times U_1, a stable and obvious PD signal series can be seen in the monitoring system. And then the voltage is still increased to about twice U_1. The PD data generated during this process are collected by the UHF sensors.

(5) Different types of insulation defect models are replaced. After completing the PD test of the current PD insulation defect model, the voltage is slowly lowered and the power is turned off. The discharge rod is used to discharge the test circuit and the insulation defect model is replaced with another one. And then the (2) to (5) steps of the test are repeated until all the three types of PD data are collected.
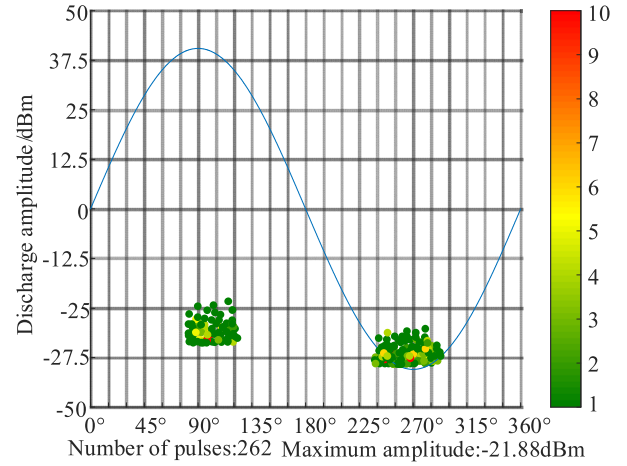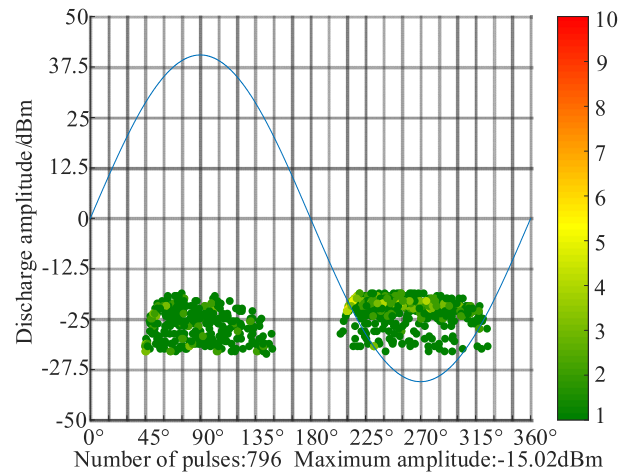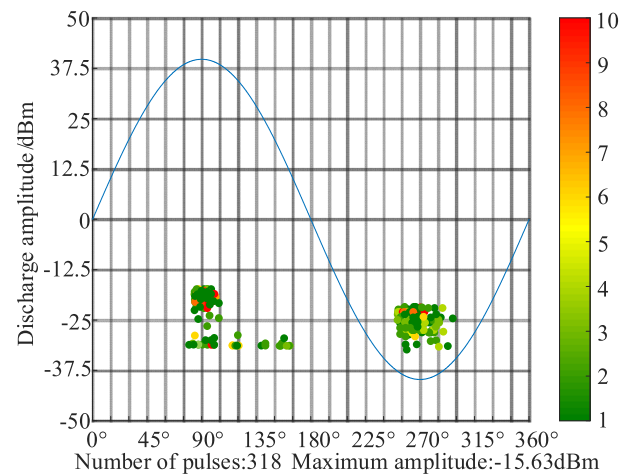
FIGURE 4. PRPD maps of the three different PD types. (a)Tip discharge, (b) Surface discharge, (c) Air discharge.

In the test, 90 groups of data are collected for each PD type and a whole of 270 groups of PD data are collected. Fig.4 shows the PRPD maps corresponding to the three PD

types obtained based on the field test data. The process of generating the PRPD maps shown in the paper are listed as follows:

(1) The UHF sensors monitor the PD pulse signal, and the monitored PD pulse signal carries time stamp and amplitude information.

(2) Based on the time stamp information and amplitude information of each pulse signal, the time stamp of the pulse signal is converted into the horizontal coordinate (phase) of the pulse point in the PRPD map, and the amplitude information of the pulse point is converted into the vertical coordinate of the pulse point in the PRPD map, and the vertical coordinate is the amplitude in dBm.

The PD pulse signal monitored by the UHF sensors is the power signal, and the amplitude unit of the PD pulse signal is mW. To be a more intuitive display of the pulse signal in the PRPD map, the amplitude of the pulse signal is converted from mW to dBm. The conversion formula corresponding to benchmark 0dBm is:

$$0\text{dBm} = 10\lg{(1\text{mW})} \tag{1}$$

(3) We draw the pulse points (50 cycles) in 1s into a PRPD map, i.e., each PRPD map contains information about the PD pulses collected by the sensors in 1s. Pulse signals with the same phase and amplitude are represented by a pulse point in the PRPD map, and the color of the pulse point represents the number of pulse signals with the same phase and amplitude.

In the PRPD map, by looking at the vertical and horizontal coordinates of a pulse point, we can directly obtain information about the phase and amplitude of the pulse. To visualize the phase corresponding to the moment of pulse generation, the waveform of one cycle of the sinusoidal function is added to the PRPD map. In addition, on the right side of the PRPD map is a color bar marked with values (from 1 to 10). Different colors on the color bar are marked with different values because the color of the pulse points represents the number of pulse points at that position in 50 cycles. For example, light green is marked with the value 4, and dark green is marked with the value 1 in the color bar.

As shown in Fig.4 (c), we can see that the color of the pulse points in the PRPD map is different. At a phase of 285° and an amplitude of −31.25 dBm, the pulse point is light green, the color of the pulse point at that position is similar to the color of the position marked by value 4 in the color bar, i.e., there are four pulse points at that position (i.e., there are four pulse points with the phase and amplitude corresponding to that position in 50 cycles). Similarly, at a phase of 300° and an amplitude of −25 dBm, the pulse point is light green, which means that there is only one pulse point at that position.

It can be seen that the PRPD maps can well reflect the number of PD pulses times, the amplitude, and spatial distribution of PD pulses, therefore, they can be used for the feature extraction. The extracted features can well reflect the characteristics of various types of PRPD maps to finally achieve pattern recognition of PD.
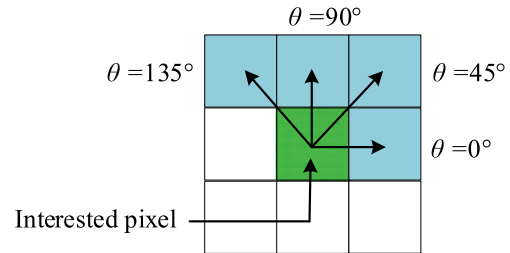

FIGURE 5. The different θ values of the interested pixel under $d = 1$.

B. THE FEATURE EXTRACTION OF PD BASED ON GLCM

The PRPD maps are converted into the PRPD grayscale images to determine the GLCM. Firstly, the number of PD pulses per unit area in the PRPD map is calculated and the maximum and the minimum number of PD pulses per unit area are found out, based on which the maximum and minimum of gray-level are determined. The higher the gray-level, the more PD pulses. After that, the gray-level of each pixel in the PRPD map can be determined, and then the PRPD grayscale image can be finally obtained [21]. The pulse phase, the pulse amplitude, and the number of PD pulses in the PRPD maps are reflected by the spatial structure, the spatial position, and the gray-level of the pixel pulses in PRPD grayscale images. Since the GLCM can statistically analyze the pixel pairs in the grayscale images, so it is very appropriate to be used to extract the features of PRPD grayscale images [22].

The GLCM is defined as the probability value of the appearance of a pair of pixels [23]. The gray-levels of the pixel pair are represented by i and j, which represent the gray-levels of the interested pixel and another pixel in the selected pixel pair, respectively. The mathematical expression of the GLCM with the angle-offset parameter θ and pixel distance d is shown in (1),

$$P(i,j;d,\theta) = \left\{ (x,y),(x+d,y+d) \,\middle|\, \begin{matrix} f(x,y) = i \\ f(x+d,y+d) = j \end{matrix} \right. \tag{2}$$

where (x,y) is the pixel coordinates, $x = 1, 2, \ldots, N_x$, $y = 1, 2, \ldots, N_y$, N_x and N_y represent the number of rows and columns of the grayscale images. θ is the angle-offset parameter between the pixel pair and the coordinate axis, whose values are 0 °, 45 °, 90 ° or 135 °. d is the distance between the interested pixel and another pixel in the pixel pair. To better explain the parameter θ and d of the pixel pair, the different θ values of the interested pixel under $d = 1$ is shown in Fig. 5.

The GLCM corresponding to the grayscale image with gray-level 8 when $\theta = 0°$ and $d = 1$ is shown in Fig.6 to better illustrate the generation process of GLCM. In this example, the left matrix is the gray-level matrix of the original image, the right matrix is the corresponding GLCM generated. The maximum value of gray-level in the left matrix determines the number of rows and columns of the right
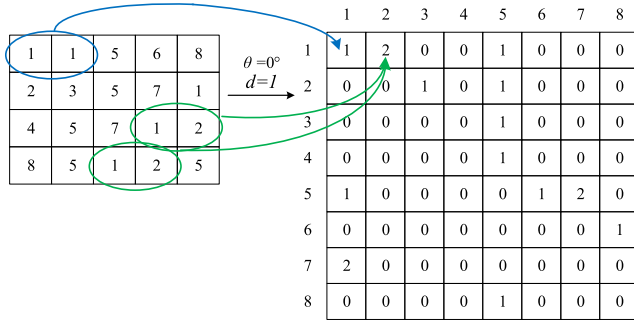
FIGURE 6. Schematic diagram of the GLCM generation.

matrix (GLCM). And the number of gray-level combinations in the left matrix corresponds to the value in the corresponding coordinate in the GLCM. For example, if $i = 1$ and $j = 1$ occurs once, GLCM $(1,1) = 1$, if $i = 1$ and $j = 2$ occurs twice, GLCM $(1,2) = 2$.

The GLCM can statistically analyze all the pixels in the three types of PRPD grayscale images, which can well reflect the comprehensive information of grayscale images in direction, variation amplitude, and local neighborhood distribution. After the GLCM corresponding to the PRPD grayscale image is obtained, we can extract the features from the GLCM for the subsequent pattern recognition.

Feature extraction is a key step in PD pattern recognition. The features of PRPD grayscale images extracted in this paper are divided into two categories. One is the moment features of GLCM, such as Energy, Entropy, Contrast, Homogeneity, and Correlation [24], [25]. The other is the fractal features of the two-dimensional image, such as the Solidity, Euler number of the binary image (*Euler*), mean of matrix elements (*Mean2*), and two-dimension correlation coefficient (*Corr2*). The combination use of the GLCM features and the fractal features of the two-dimensional image can reflect the characteristics of PRPD grayscale images more comprehensively. For the n-level GLCM, the mathematical formulas of the above the moment features of GLCM are listed as follows:

(1) Energy:

$$ENE = \sum_{i=1}^{n} \sum_{j=1}^{n} [P(i, j; d, \theta)]^2 \tag{3}$$

ENE is the sum of squares of the elements of GLCM, which reflects the uniformity of the gray-level distribution and texture thickness.

(2) Entropy:

$$ENT = -\sum_{i=1}^{n} \sum_{j=1}^{n} P(i, j; d, \theta) \log P(i, j; d, \theta) \tag{4}$$

ENT represents the amount of information in the image, which reflects the complexity of the texture. When the element distribution in the co-occurrence matrix is scattered and random, the gray-level of the texture is complex and the entropy will have a large value.

(3) Contrast:

$$CON = \sum_{i=1}^{n} \sum_{j=1}^{n} \left[(i-j)^2 P(i, j; d, \theta) \right]^2 \tag{5}$$

CON is also known as the moment of inertia, reflecting the clarity of the image and the depth of the texture grooves. The deeper the texture groove, the greater the contrast, the clearer the visual effect.

(4) Homogeneity:

$$HOM = \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{P(i, j; d, \theta)}{1 + |i - j|} \tag{6}$$

HOM measures how close the elements in GLCM are to the diagonal elements of GLCM in the same row.

(5) Correlation:

$$COR = \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} (ij) P(i, j; d, \theta) - \mu_x \mu_y}{\sigma_x \sigma_y} \tag{7}$$

COR measures the similarity of GLCM elements in each row or each column, which reflects the linear correlation of the image gray-level. If the gray-level is continuous in a certain direction, the correlation value is high. In (7), μ_x, μ_y and σ_x, σ_y are defined as follows:

$$\mu_x = \sum_{i=1}^{n} \sum_{j=1}^{n} x \times P(i, j; d, \theta) \tag{8}$$

$$\mu_y = \sum_{i=1}^{n} \sum_{j=1}^{n} y \times P(i, j; d, \theta) \tag{9}$$

$$\sigma_x = \sum_{i=1}^{n} \sum_{j=1}^{n} (i - \mu_x)^2 \times P(i, j; d, \theta) \tag{10}$$

$$\sigma_y = \sum_{i=1}^{n} \sum_{j=1}^{n} (j - \mu_y)^2 \times P(i, j; d, \theta) \tag{11}$$

It can be seen that the features of GLCM described above can well reflect the characteristics of the PRPD grayscale images, including the uniformity and thickness of the gray-level distribution, the complexity and non-uniformity of the texture, the clarity and depth of the texture, and the similarity of the local gray-level of the images [26], [27]. Therefore, the GLCM is very suitable for extracting the features of the PRPD grayscale images.

III. FEATURE EXTRACTION OF PRPD MAPS BASED ON GLCMOP

A. DETERMINE THE GLCMOP

From Section II, we know that the values of angle-offset parameter θ and distance d need to be determined before the calculation of GLCM. The value of θ includes 0 °, 45 °, 90 ° or 135 °. Each angle-offset parameter θ corresponds to a different GLCM. The value of θ has a greater impact on the characteristics of GLCM than the value of d.

The feature space of PRPD maps for the pattern recognition comes from GLCM, that is to say, the distinguishability of features depends on the ability of GLCM to reflect the characteristics of PRPD grayscale images. Therefore, it is very important to determine the optimal value of the angle-offset parameter θ. The feature space of PD samples corresponding to the GLCMOP is the one with the highest distinguishability.

The pixel pairs in GLCM corresponding to the PRPD grayscale images are analyzed. The value of the elements in GLCM can reflect the number of the three types of pixel pairs. GLCM (1, 1) is the number of background pixel pairs in the PRPD grayscale images. Secondly, the gray-level of the two pixels in the pixel pairs are different, which are the unequal pixel pairs. And the number of the unequal pixel pairs in the PRPD grayscale images is reflected by the off-diagonal element of GLCM (i, j). In addition, the number of the third type of pixel pairs, which have the same gray-level images called the equal pixels, is reflected by the element of GLCM (i, j), where $i = j \neq 1$. Among the three types of pixel pairs, the unequal pixel pairs can best reflect the characteristics of PRPD grayscale images. Because the unequal pixel pairs contain the boundary information between the PD pulses and the background of PRPD grayscale images, also contain the density changing information of the PD pulses aggregation area.

Therefore, the GLCMOP with the optimal angle-offset parameter θ should meet the requirement that the unequal pixel pairs have the highest proportion in all pixel pairs. Among all gray-level co-occurrence matrices, the ratio of the number of unequal pixel pairs to the number of equal pixel pairs in GLCMOP is the largest. Therefore, the proportion of the off-diagonal elements in GLCM (*PODE*) is proposed to determine GLCMOP. The formula of *PODE* is shown in (12).

$$PODE = \frac{ODE_S}{DE_S} \tag{12}$$

where ODE_S is the sum of the values of off-diagonal elements, and DE_S is the sum of the values of diagonal elements after removing the background pixels, and they can be calculated according to (13) and (14),

$$ODE_S = \sum_{i=1}^{n}\sum_{j=1}^{n} GLCM(i,j) - \sum_{i=j=1}^{n} GLCM(i,j) \tag{13}$$

$$DE_S = \sum_{i=1}^{n}\sum_{j=1}^{n} GLCM(i,j) - GLCM(1,1) \tag{14}$$

where n is the maximum order of GLCM, which is also the maximum gray-level.

After all the *PODE* of gray-level co-occurrence matrices corresponding to all angle-offset parameters θ are calculated, the angle-offset parameter θ corresponding to the maximum value of *PODE* is determined to be the optimal angle-offset parameter θ. The GLCM with the optimal angle-offset parameter θ is called the GLCMOP, whose features
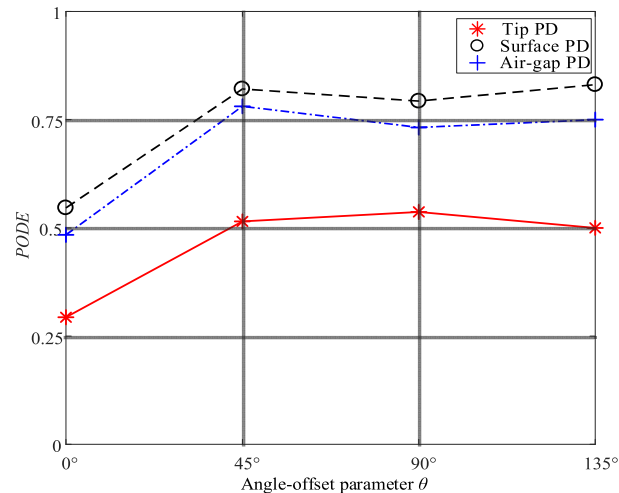

FIGURE 7. The values of *PODE* with different angle-offset θ.

TABLE 1. PODE of GLCM corresponding to different angle-offset parameters θ.

PD types	*PODE*			
	0°	**45°**	**90°**	**135°**
Tip discharge	0.294	0.515	0.537	0.500
Surface discharge	0.546	0.821	0.793	0.831
Air discharge	0.484	0.781	0.732	0.750
Average	0.441	0.706	0.687	0.694

have the highest distinguishability. Moreover, the subjectivity in determining the value of angle migration parameter θ is avoided and the calculation burden of feature extraction is reduced by determining the GLCMOP. The *PODE* of gray-level co-occurrence matrices corresponding to three types of PRPD grayscale images under different angle-offset parameters θ are shown in Fig.7.

It can be seen from Fig.7 that when the angle-offset parameter θ are 45°, 90°and 135°, the *PODE* of gray-level co-occurrence matrices corresponding to the PRPD grayscale images is high. In other words, with these values of angle-offset parameter θ, the calculated GLCM can better reflect the characteristics of PRPD grayscale images. The optimal angle-offset parameter θ is determined according to the average of all *PODE* of the gray-level co-occurrence matrices for the three types of PD samples under different θ values. The *PODE* results are shown in Table 1.

It can be seen that when the angle-offset parameter θ is 45°, the average of *PODE* for the three types of PD samples is the highest, which has a value of 0.706. Therefore, the optimal angle-offset θ is determined to be 45° and its corresponding GLCM is called the GLCMOP of PRPD grayscale images, which is used to extract the features of PRPD grayscale images.

When the measured PRPD data changes, the GLCMOP is updated according to the *PODE* index, which improves the adaptability of the algorithm.

B. DISTINGUISHABILITY SCREENING AND DIMENSIONALITY REDUCTION OF THE PD FEATURES

After the optimal angle-offset θ is determined, nine features are extracted from the GLCMOP corresponding to each PRPD grayscale image. These features can be represented by a nine-dimensional vector which can be used for the PD classifiers. However, with the increase of the number of PRPD samples, the data amount of the feature space of PD samples composed of this nine-dimensional vector is large. As a result, the training time of the classifier increases, and the convergence of the classifier decreases. Therefore, in actual applications, to reduce the memory occupation and improve the recognition speed, the dimension of the feature space should be reduced as much as possible on the premise of ensuring recognition accuracy. In this paper, the probability distribution parameter and three-dimensional distribution maps of PD samples are used to evaluate the feature distinguishability quantitatively and qualitatively. Then the features having the highest distinguishability are selected as the final features for the pattern recognition.

The intersection of probability distribution J_b of features in various types of PD is used as the distinguishability evaluation index of features [28], which can comprehensively evaluate the distinguishability from the quantitative perspective. Its mathematical expression is shown in (15).

$$J_b = \int [p(x|\lambda_1) p(x|\lambda_2)]^2 dx \qquad (15)$$

In (15), $p(x|\lambda_1)$ and $p(x|\lambda_2)$ represent the probability density of the feature distribution of x in two types of PD. The larger the intersection of probability distribution J_b of feature x in the two PD types λ_1 and λ_2, the worse the ability of feature x to distinguish the PD types λ_1 and λ_2. That is to say, feature x has a lower distinguishability. In the worst case, $p(x|\lambda_1)$ and $p(x|\lambda_2)$ are equal and the value of the intersection of probability distribution J_b reaches the maximum of 1. Feature x is independent of PD types λ_1 and λ_2, which means that feature x cannot distinguish PD types λ_1 and λ_2 at all. When the two data distributions have no intersection at all, the intersection of probability distribution J_b is equal to 0, which means that feature x can completely distinguish the PD types λ_1 and λ_2. The values of the intersection of probability distribution J_b for the nine types of features mentioned in the three types of PD are shown in Table 2.

It can be seen from Table 2 that among the nine selected features, the intersection of the probability distributions of feature 7 (*HOM*) in the combination of tip discharge and surface discharge, and that of feature 9 (*CORR2*) in the combination of surface discharge and air discharge is of the smallest values, which indicates that among all the characteristic parameters, feature 7 (*HOM*) and feature 9 (*CORR2*) are the best ones to distinguish the tip discharge and surface

**TABLE 2.** The intersection of probability distribution (J_b) of features in three types of PD.

Features	Tip-Surface	Surface - Air	Tip-Air	Average value
1(*ENE*)	0.1867	0.0815	0.0074	0.0919
2(*ENT*)	0.0154	0.0758	0.0028	0.0313
3(*CON*)	0.1181	0.0725	0.0351	0.0752
4(*SOL*)	0.0987	0.1022	0.0784	0.0931
5(*EUL*)	0.1428	0.1364	0.1859	0.1550
6(*MEAN2*)	0.1556	0.1874	0.1123	0.1517
7(*HOM*)	0	0.1424	0.1312	0.0912
8(*COR*)	0.0401	0.0011	0.1825	0.0745
9(*CORR2*)	0.0142	0.0001	0.1931	0.0691

discharge, surface discharge and air discharge, respectively. Similarly, feature 8 (*COR*) and feature 9 (CORR2) have the best discriminating ability to distinguish the surface discharge and air discharge, and the J_b of them is approximately to be 0. Feature 1 (*ENE*) and feature 2 (*ENT*) are the best features to distinguish the tip discharge and air discharge.

By analyzing the intersection of the probability distribution of features in three types of PD, it is easy to know that the J_b of features for the combination of tip discharge and air gap discharge is larger than those in the other two discharge combinations. That is to say, the combination of tip discharge and air discharge is the most difficult one to be distinguished, while the combination of tip discharge and surface discharge, and the combination of surface discharge and air discharge are easier to be distinguished. From the average value of the intersection of the probability distribution of features, it can be seen that feature 2 (*ENT*) is the optimal feature of PRPD grayscale images. In addition, among the three kinds of combinations of discharges, the moment features have the smallest intersection of the probability distribution, so the moment features of PRPD grayscale images are better than the fractal features in distinguishing discharge types.

In addition, the distribution of PD samples in three-dimensional feature space can complementarily verify the quantitative evaluation results. The three-dimensional distribution map of PD samples can directly reflect the overlapping degree and distribution of PD samples in a certain coordinate axis direction, i.e., the feature direction, and then qualitatively judge the distinguishability of features as shown in Fig.8.

It can be seen from the three-dimensional distribution map of PD samples that in the feature space of 1 (*ENE*), 2 (*ENT*), 3 (*CON*) and feature space of 7 (*HOM*), 8 (*COR*), 9 (*CORR2*), all types of PD samples have strong aggregation and low sample overlap degrees. This is consistent with the calculation result of the intersection of the probability distribution. Therefore, six features (1 (*ENE*), 2 (*ENT*), 3 (*CON*), 7 (*HOM*), 8 (*COR*), and 9 (*CORR2*)) with high distinguishability are selected as the final features of pattern recognition
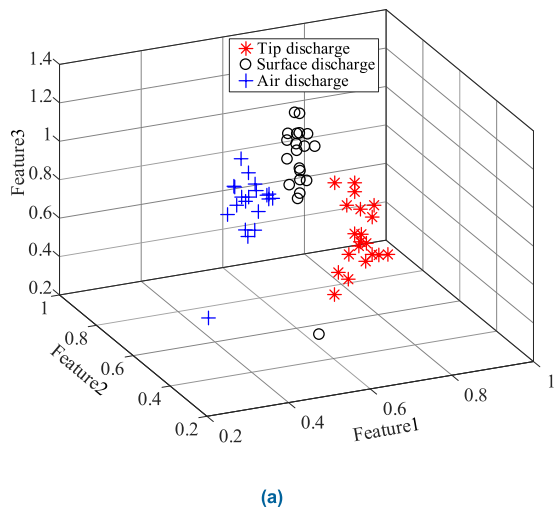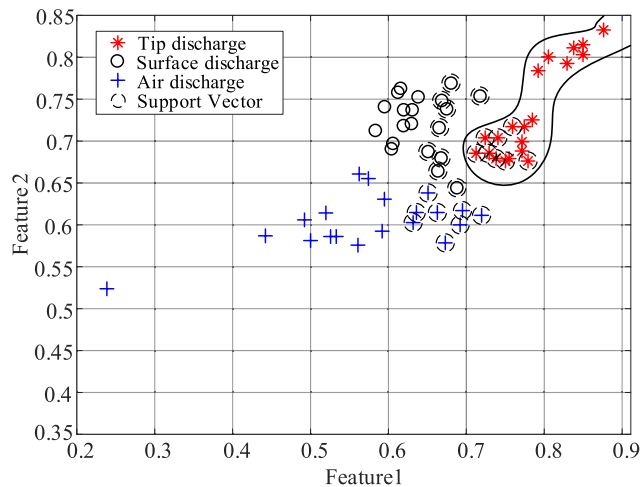
(a)
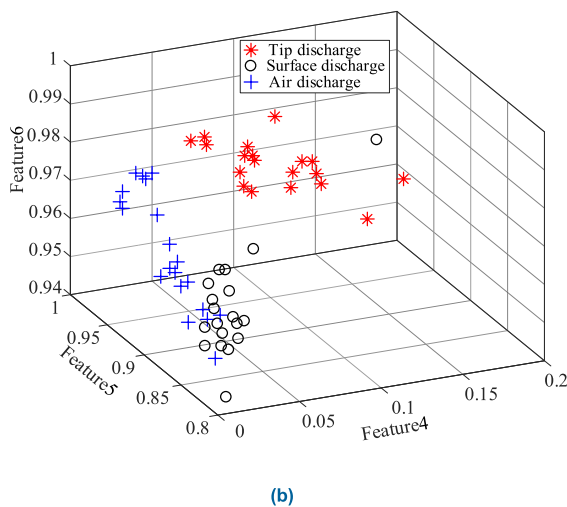
(b)
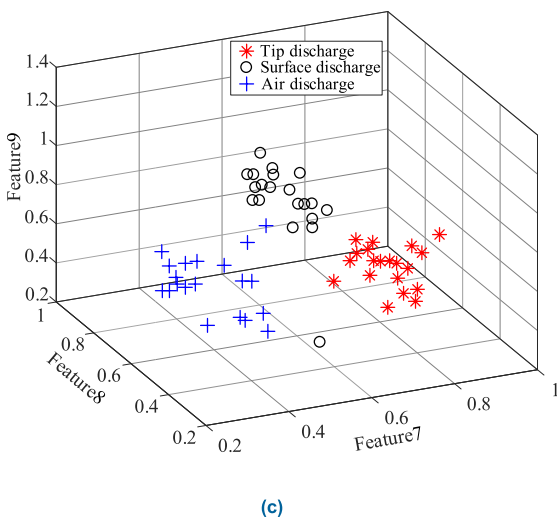
(c)

FIGURE 8. Three-dimensional distribution maps of PD samples in different feature spaces. (a)Space of feature 1 (*ENE*), 2 (*ENT*) and 3 (*CON*), (b) Space of feature 4(*SOL*), 5(*EUL*) and 6(*MEAN2*), (c) Space of feature 7(*HOM*), 8(*COR*) and 9(*CORR2*).

FIGURE 9. OVR-SVM finds the estimation boundary between the positive and negative classes, the support vectors are observations that occur on or beyond their estimated class boundaries.

reduce the training time, improve the classifier convergence and further improve the recognition accuracy.

C. CLASSIFIER TRAINING AND RECOGNITION PROCESS

Another key step of pattern recognition is to build a classifier. The classifier adopted in this paper is the one-to-rest SVM (OVR-SVM). The main goal of SVM is to find the hyperplane which can separate the two types of PD samples accurately and has the largest interval of classification in the PD sample space. The SVM classifier can obtain good statistical laws with a small number of PD samples and has a high classification accuracy in the test sample set, and also has good adaptability to the sample set.

Three OVR-SVM classifiers are constructed to realize the multi-class classification. Each classifier sets one of the three types of PD samples as positive class and the other two as negative class. Taking tip discharge in two-dimensional feature space as an example to illustrate the OVR-SVM. The tip discharge is set to be in a positive class, and the surface discharge and air discharge are set to be in the negative class.

As shown in Fig.9, when the training samples are input into the OVR-SVM classifier, the SVM classifier will find the class boundaries between the positive and negative classes. When the test samples are input to the classifier, the samples will be identified by three classifiers in turn and the discriminant function of each classifier will calculate a function value. If only one classifier outputs a positive value, the result can be directly judged as the positive class in the corresponding classifier. Otherwise, the class corresponding to the maximum value (the highest confidence) of the discriminant function is selected as the final type of the test samples.

According to the analysis in Sections II and III, the optimal angle-offset parameter θ is determined and the GLCMOP corresponding to the PRPD grayscale images are calculated. On this basis, a six-dimensional feature space of PD samples is constructed after the distinguishability screening and
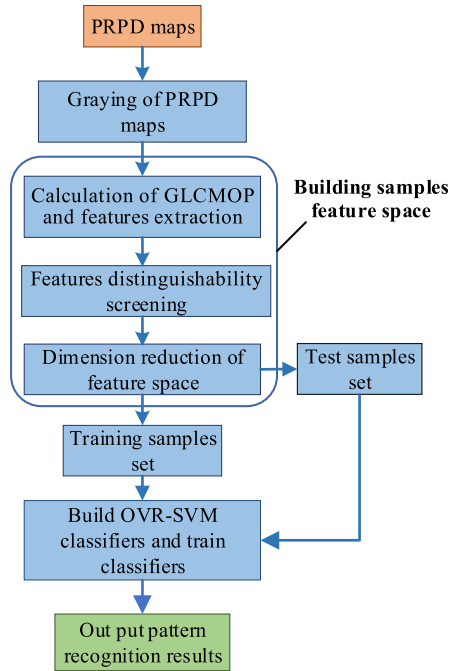
to construct a low dimension and high distinguishability feature space for the PD samples. The derived feature space can

FIGURE 10. The flow chart of GLCMOP-SVM pattern recognition.

TABLE 3. Comparison of training time and number of SVs.

Methods	Training time/s	Number of SVs
GLCM-SVM	10	36
GLCMOP-SVM	2	64

dimensionality reduction of the features. The feature space of PD samples is divided into the training sample set and the test sample set. The training sample set is used for the classifier training and the test sample set is input into the SVM classifier for the PD pattern recognition. Finally, the classification result is output. The flow chart of GLCMOP-SVM pattern recognition is shown in Fig.10.

IV. EXPERIMENT RESULTS AND DISCUSSION

Based on the test platform illustrated in Section II, 270 PD samples are generated for the three types of PD. 190 PD samples are used as the training sample set for the PD pattern recognition, and the rest 80 PD samples are used as the test samples set. To evaluate quantitatively the influence of the GLCMOP algorithm on the classifier, the training time, and the number of support vectors (SVs) of the SVM classifier in GLCM-SVM and GLCMOP-SVM are calculated and compared in Table 3. At the same time, the performance of GLCMOP-SVM with other pattern recognition methods are also compared in the aspects of recognition time, memory consumption, and recognition accuracy as shown in Table 4 and Table 5.

The results in Table 3 show that the training time of GLCMOP-SVM is shortened. Because the feature distinguishability screening process can derive a low dimension

TABLE 4. Comparison of recognition time and memory consumption.

Methods	Recognition Time (s)	Memory Consumption (%)
ANN	5.1	7.9%
XGBoost	4.0	4.0%
GLCM-SVM	3.8	4.5%
GLCMOP-SVM	2.2	2.1%

TABLE 5. Comparison of recognition results.

PD types	Recognition rate/%			
	ANN	GLCM-XGBoost	GLCM-SVM	GLCMOP-SVM
Tip discharge	85.0%	81.3%	87.5%	91.3%
Surface discharge	88.8%	85.0%	90.0%	97.5%
Air discharge	87.5%	83.8%	88.8%	95.0%
Total	87.1%	83.3%	88.8%	94.6%

and high distinguishability feature space, which speeds up the training process of the classifier. The processes of determining the optimal angle-offset parameter θ and calculating the GLCMOP corresponding to the PRPD greyscale images can increase the number of support vectors (SVs) in the PD samples. And thus, the class boundaries of different types of PD samples are much clearer than before, which is beneficial to improve the generalization ability and recognition accuracy of the classifier.

It can be seen from Table 4 that the GLCMOP-SVM outperforms the ANN, GLCM-XGBoost, and GLCM-SVM in terms of recognition time and memory consumption. This is because after calculating the GLCMOP corresponding to the PRPD maps, we do not need to calculate the GLCM corresponding to other values of the angle-offset parameter θ. In addition, the recognition speed can be improved and the memory consumption of the classifier can be reduced due to the distinguishability screening and dimensionality reduction of PD feature space.

According to Table 5, the recognition accuracies of GLCMOP-SVM for the tip discharge, surface discharge, and air discharge are 91.3%, 97.5%, 95.0% respectively, and the overall average recognition rate is 94.6%. And GLCMOP-SVM has higher recognition accuracy for each type of PD compared with the ANN and GLCM-SVM, and the total average recognition rate is about 7% higher. In summary, compared with the traditional classification methods, such as GLCM-SVM, XGBoost, and ANN, the GLCMOP-SVM has a faster recognition speed and higher accuracies.

V. CONCLUSION

A transformer PD pattern recognition algorithm based on the GLCM of optimal parameters and SVM (GLCMOP-SVM) is proposed in this paper. The main contributions of this paper are summarized as follows.

(1) The algorithm can adaptively determine the corresponding GLCMOP according to the measured PRPD data, and the characteristics of PRPD maps can be extracted and reflected by the GLCMOP as much as possible.

(2) The processes of determining the GLCMOP and feature distinguishability screening lower the dimension and improve the distinguishability of the feature space.

(3) The total recognition rate of the algorithm is as high as 94.6%, which is about 7% higher than the traditional pattern recognition algorithm. Moreover, the training and recognition time of the algorithm can be greatly reduced.

Therefore, the algorithm is suitable for the on-line monitoring of transformers PD due to its high recognition speed and accuracy, and the proposed method can be used in small mobile devices or integrated systems.

REFERENCES

[1] C. C. Wang, Q. F. Li, and Y. S. Gao, *On Line Monitoring and Fault Diagnosis of Power Equipment*. Beijing, China: Tsinghua Univ. Press, 2006, pp. 30–50.

[2] C. R. Lou and G. W. Li, "Development of on-line monitoring system for partial discharge within power transformer," *Power Syst. Technol.*, vol. 28, no. 16, pp. 56–59, 2004.

[3] S. Haikun, Y. Jinsha, W. Yu, and J. Song, "Feature extraction for partial discharge based on cross-wavelet transform and correlation coefficient matrix," *Trans. China Electrotech. Soc.*, vol. 29, no. 4, pp. 274–281, 2014.

[4] J. A. Ardila-Rey, J. M. Martinez-Tarifa, G. Robles, and M. V. Rojas-Moreno, "Partial discharge and noise separation by means of spectral-power clustering techniques," *IEEE Trans. Dielectr. Elect. Insul.*, vol. 20, no. 4, pp. 1436–1443, Aug. 2013.

[5] G. Z. Tang, Z. M. Tang, Z. J. Li, Y. J. Wang, C. Wu, and K. Wang, "Review on partial discharge pattern recognition of electrical equipment," *High Voltage Eng.*, vol. 8, no. 1, pp. 491–495, 2013.

[6] B. Liu and J. Zheng, "Partial discharge pattern recognition in power transformers based on convolutional neural networks," *High Voltage App.*, vol. 53, no. 5, pp. 70–74, 2017.

[7] J. Guo, G. N. Wu, X. Q. Zhang, and W. Shu, "The actuality and perspective of partial discharge detection techniques," *Trans. China Electrotech. Soc.*, vol. 20, no. 2, pp. 29–35, Feb. 2005.

[8] H.-G. Kranz, "Diagnosis of partial discharge signals using neural networks and minimum distance classification," *IEEE Trans. Electr. Insul.*, vol. 28, no. 6, pp. 1016–1024, Dec. 1993.

[9] H. Ma, J. C. Chan, T. K. Saha, and C. Ekanayake, "Pattern recognition techniques and their applications for automatic classification of artificial partial discharge sources," *IEEE Trans. Dielectr. Electr. Insul.*, vol. 20, no. 2, pp. 468–478, Apr. 2013.

[10] R. Bartnikas, "Partial discharges. Their mechanism, detection and measurement," *IEEE Trans. Dielectrics Electr. Insul.*, vol. 9, no. 5, pp. 763–808, Oct. 2002.

[11] E. Gulski and A. Krivda, "Neural networks as a tool for recognition of partial discharges," *IEEE Trans. Elect. Insul.*, vol. 28, no. 6, pp. 984–1001, Dec. 1993.

[12] Y. Khan, "Partial discharge pattern analysis using PCA and back-propagation artificial neural network for the estimation of size and position of metallic particle adhering to spacer in GIS," *Electr. Eng.*, vol. 98, no. 1, pp. 29–42, Mar. 2016.

[13] C. T. Lee and S. C. Horng, "Abnormality detection of cast-resin transformers using the fuzzy logic clustering decision tree," *Energies*, vol. 13, no. 10, pp. 2546–2565, 2020.

[14] H. F. Ye, Y. Qian, L. Q. G. H. Si Sheng, and X. C. Jiang, "UHF partial discharge pattern recognition for GIS with support vector machine," *Int. Rev. Electr. Eng.*, vol. 8, no. 1, pp. 491–496, 2013.

[15] S. M. Song, Y. Qian, H. Wang, Y. M. Zang, G. H. Sheng, and X. C. Jiang, "Partial discharge pattern recognition based on 3D graphs of phase resolved pulse sequence," *Energies*, vol. 16, no. 13, pp. 4103–4119, 2020.

[16] M. H. Zhao, B. P. Tang, and Q. Tan, "Fault diagnosis of rolling element bearing based on S transform and gray level co-occurrence matrix," *Meas. Sci. Technol.*, vol. 8, no. 26, pp. 5008–5019, 2015.

[17] S. Barburiceanu, R. Terebes, and S. Meza, "3D texture feature extraction and classification using GLCM and LBP-based descriptors," *Appl. Sci.*, vol. 11, no. 5, pp. 2332–2357, 2021.

[18] M. Benco, R. Hudec, P. Kamencay, M. Zachariasova, and S. Matuska, "An advanced approach to extraction of colour texture features based on GLCM," *Int. J. Adv. Robotic Syst.*, vol. 11, no. 7, p. 104, Jul. 2014.

[19] P. Jiao, Y. Guo, L. Liu, and X. Wei, "Implementation of gray level co-occurrence matrix texture feature extraction using MATLAB," *Comput. Technol. Develop.*, vol. 22, no. 11, pp. 169–171, 2012.

[20] S. Haikun, Y. Jinsha, W. Yu, and Z. Liwei, "Partial discharge pattern recognition in power transformer based on multi-kernel multi-class relevance vector machine," *Trans. China Electrotech. Soc.*, vol. 29, no. 11, pp. 221–228, 2014.

[21] Y. Liu, F. Lu, and C. Li, "Study on pattern spectrum of patial discharge grayscale image," *Proc. CSEE*, vol. 24, no. 5, pp. 183–187, 2004.

[22] H. Bo, F. L. Ma, and L. C. Jiao, "Research on computation of GLCM of image texture," *Acta Electronica Sinica*, vol. 34, no. 1, pp. 134–155, 2006.

[23] G. D. Wang, P. L. Zhang, G. Q. Ren, and X. Kou, "Texture feature extraction method fused with LBP and GLCM," *Comput. Eng.*, vol. 38, no. 11, pp. 199–201, 2012.

[24] L. Zhao, Y. L. Zhu, and Y. F. Jia, "Feature extraction for partial discharge grayscale image based on gray level co-occurrence matrix and local binary pattern," *Elect. Meas. Instrum.*, vol. 54, no. 1, pp. 77–82, 2017.

[25] Y. Z. Min, T. D. Cheng, and C. Yin, "Texture image segmentation based on GLCM and self-adaptive Gabor filter bank," *Comput. Eng.*, vol. 43, no. 1, pp. 1280–1286, 2017.

[26] M. Benčo and R. Hudec, "Novel method for color textures features extraction based on GLCM," *Radioengineering*, vol. 16, no. 4, pp. 64–67, 2007.

[27] R. Prabu and R. Harikumar, "Electrical impedance tomography (EIT) image classification using GLCM based feature extraction in artificial neural network (ANN)," *Int. J. Appl. Eng. Res.*, vol. 10, no. 38, pp. 28828–28833, 2015.

[28] K. X. Huo, "Study on partial discharge pattern recognition in transformers based on grayscale image and affinity propagation algorithm," M.S. thesis, Dept. Electron. Eng., Shandong Univ, Shandong, China, 2016.
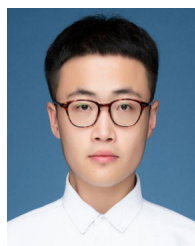
SHENGYA SUN was born in Puyang, Henan, China, in 1996. He received the B.S. degree in electrical engineering from the School of Electrical Engineering, Shandong University, Jinan, China, in 2019, where he is currently pursuing the M.S. degree in electrical engineering. His research interests include PD pattern recognition, partial discharge signal denoising, and flexible DC transmission.

YUANYUAN SUN (Member, IEEE) was born in Zibo, Shandong, China, in 1981. She received the B.S. and Ph.D. degrees in electrical engineering from the School of Electrical Engineering, Shandong University, Jinan, China, in 2003 and 2009, respectively. She is currently a Professor with the School of Electrical Engineering, Shandong University; a Ph.D. Supervisor; and a Candidate for Future Plans of Young Scholars of Shandong University. Her current research interests include harmonic modeling, on-line monitoring of transformers, harmonic contribution determination, and active distribution system analysis.

GONGDE XU was born in Zhucheng, Shandong, China, in 1996. He received the B.S. degree in electrical engineering from the School of Electrical Engineering, Shandong University, Jinan, China, in 2019, where he is currently pursuing the M.S. degree in electrical engineering. His research interests include hot spot temperature prediction of the transformer and energy router.

YIRU HU received the B.S. degree in control science and engineering from the College of Information Science and Engineering, China University of Petroleum, Beijing. She is currently an Electrical Engineer with the China National Offshore Oil Corporation. Her current research interests include the electrical design of offshore oil and gas power grid, smart grid, and electrical equipment condition-based maintenance.

LINA ZHANG is currently an Electrical Engineer with the China National Offshore Oil Corporation. Her current research interests include the electrical design of offshore oil and gas power grid, smart grid, and electrical equipment condition-based maintenance.

PING LIU received the B.S. degree in electrical engineering from the School of Electrical Engineering, Shandong University, Jinan, China. She is currently an Electrical Engineer with CNOOC Energy Development Equipment Technology Company Ltd. Her main research interests include power system automation and so on.

• • •