Pattern Recognition for Partial Discharge Using Multi-Feature Combination Adaptive Boost Classification Model

RUI YAO¹, JUN LI², MENG HUI¹, (Member, IEEE), LIN BAI¹, AND QISHENG WU¹

¹School of Electronic Control, Chang'an University, Xi'an 710064, China
²State Grid Shaanxi Electric Power Research Institute, Xi'an 710049, China

Corresponding author: Jun Li (18629635935@163.com)

This work was supported in part by the National Key Research and Development Program of China under Grant 2020YFC1512002, in part by the Natural Science Foundation of Shaanxi Province under Grant 2020JM-256, in part by the Technology Innovation Leading Program of Shaanxi under Grant 2020QFY03-01, in part by the Natural Science Basis Research Plan in Shaanxi Province of China under Grant 2020GY-060, and in part by the Xi'an Science and Technology Plan Project under Grant 2020KJRC0126.

ABSTRACT This paper proposes a multi-feature combination adaptive boost classification model, considering the difference and complementarity of the three different single feature sets for partial discharge pattern recognition. First, eight types of physical models are designed. Then, an UHF measurement system is used to collect partial discharge data. Second, three kinds of single feature sets extracted from the Phase Resolved Pulse Sequence (PRPS) data are combined with pairs and three to construct new feature sets. The final optimal feature set is selected from the single feature set and the combined feature set as the input of the classification model. Finally, using the boosting algorithm in combination learning to process the training data set, taking the support vector machine as the base classifier, and measuring the inconsistency between one base classifier and other base classifiers by using the “unpaired” diversity index based on information entropy. By this method, a series of various SVM-based classifiers with moderate accuracy are obtained, and finally an adaptive boost classification model based on the multi-feature combination method is obtained. For each defect, 25 samples were obtained at the same test voltage level, and a total of 150 samples were obtained at 6 voltage levels through multiple experiments. The proposed method was compared with the traditional methods using these data sets. The results revealed that the proposed method successfully identified the types of partial discharge insulation defects.

INDEX TERMS Partial discharge, pattern recognition, multi-features combination, ensemble learning.

I. INTRODUCTION

Partial discharge (PD) occurs to the area where the local electric field exceeds dielectric strength of the electrical insulating material in the insulation system [1], [2]. Detection of partial discharge can evaluate the insulation status of electrical equipment and provide an early indication of insulation material failure [3]. For partial discharge detection of electrical equipment, in addition to judging whether there is a partial discharge, it is necessary to further judge the type of insulation defect. Using pattern recognition methods to get the evaluation of the operation status of high-voltage electrical equipment includes two evaluations: insulation status and defect type characteristics. Based on the evaluation conclusions, valuable technical support can be provided for the maintenance of high-voltage electrical equipment.

The research on pattern recognition related issues has become a hot topic. In recent years, domestic and international scholar have conducted a lot of research on partial discharge data statistics and insulation defect type identification algorithms [4]–[7].

The methods commonly used for partial discharge classification are: neural network [8], [9], hidden Markov model [10], fuzzy logic classifier [11], self-organizing mapping network [12], inductive reasoning algorithm [13], Support vector machine [14] and sparse representation classifier [15]. Partial discharge recognition performances have been obtained by these methods. In recent years, the convolutional neural
network (CNN) for PD image recognition to distinguish various defect types is proposed in the literature [16]. The CNN is a type of deep learning method that time consumption is more than the other traditional methods. Moreover, the feature set automatically obtained by this algorithm cannot be used as the input of other algorithms that need to be compared. Pattern recognition of the detected PD signal is based on modified chaotic analysis of partial discharge (CAPD) method [17]. The multilayer perceptron technique of artificial neural network (ANN) is used to get the training data or the input vector from modified CAPD pattern data. This method considers the frequency domain analysis to improve the recognition rate. The classification accuracy decreases with higher noise level but Principal Component Analysis (PCA) features used in Support Vector Machine (SVM) and ANN showed the strongest tolerance against noise contamination [18]. The tolerance of the algorithms used in the PD pattern identification against noise contamination is analyzed.

However, it is found that the above classification techniques use a single classifier obtained from training data onto a certain characteristic extracted from a single partial discharge pattern to predict the class label of an unknown example. Different partial discharge patterns show the characteristics of partial discharge from different perspectives. The phase resolved pulse sequence analysis (PRPSA) pattern shows the time distribution of partial discharge pulses. Compared with the PRPSA pattern, the temporal characteristics of partial discharge with respect to time are lost in phase resolved partial discharge (PRPD) pattern, but it characterizes the phase characteristics of the partial discharge. Compare with the time resolved partial discharge pattern (TRPD), it only has the peak information of the pulse, and no PD pulsed waveform information. The TRPD pattern shows the change of the discharge pulse of time, including the waveform and time information. There is no phase information about pulse. Each partial discharge pattern only characterizes partial discharge from a certain angle. Therefore, if multiple partial discharge patterns can be selected and integrated, a more comprehensive analysis of partial discharge characteristics and a more accurate recognition rate can be obtained.

Additionally, some recent studies in the field of partial discharge recognition show that for certain selected partial discharge insulation defects, the recognition performance of integrated neural networks is better than that of single neural networks in [19]–[23]. However, the improvement in the classification performance of other algorithms by ensemble learning has not yet been effectively quantitatively evaluated, and because research shows that it has a high recognition ability when applied to data sets from other disciplines [24], [25]. The article considers using ensemble-boosting algorithm.

In dealing with complex pattern classification problems, Support Vector Machine (SVM) is a commonly used supervised learning method with high classification performance. Literature [14] applied SVM to successfully distinguish three kinds of artificial insulation defects, but showed excessive recognition of internal discharge. SVM is used to determine the defects of dirty particles of transformer oil [26], [27], and it can also classify the local discharge sources of oil-impregnated paper insulated cables [28]. Literature [29] verifies that Fuzzy Support Vector Machine (FSVM) is effective in dealing with data sets flooded by noise. Literature [30] first proposed a new method of transformer partial discharge pattern recognition based on multi-core multi-class correlation vector machine, which has higher diagnostic accuracy and better practicability than single-core classifiers. Support vector machine is proposed based on the concept of Vapnik-Chervonenkis (VC) dimension and the principle of structural risk minimization. It has a strong theoretical basis and has been proved to be effective and feasible in partial discharge pattern recognition. For support vector machines, parameter selection is a critical problem in model training, because the quality of model parameters will directly affect the generalization ability of the algorithm. For the support vector machine with Gaussian kernel, its performance is affected by the penalty parameter and the Gaussian kernel parameter. How to find an efficient search method to determine the model parameter values and obtain the minimum generalization error of the support vector machine is a key issue in the current research work in this direction [31]–[35].

Therefore, in this paper, an ensemble-boosting method based on support vector machines is used to establish a classification model for insulation defects identification. This method improves the partial discharge recognition rate for two aspects. One is to use multiple partial discharge patterns, and the other is the use of ensemble learning method to improve the learning of single SVM.

The research idea is as follows:

1. First, eight types of physical models are designed. Then, an UHF measurement system is used to collect partial discharge data.
2. Second, three kinds of single feature sets obtained by Phase Resolved Pulse Sequence (PRPS) data are combined with pairs and three combined to construct a new feature set.
3. Then, the final optimal feature set is selected from the single feature set and the combined feature set as the input of the classification model.
4. Finally, using the boosting algorithm in combination learning to process the training data set, taking the support vector machine as the base classifier, and measuring the inconsistency between one base classifier and other base classifiers by using the “unpaired” diversity index based on information entropy. By this method, a series of various SVM-based classifiers with moderate accuracy are obtained, and finally an adaptive boost classification model based on the multi-feature combination method is obtained.

It should be noted that this paper studies the classification model on the basis of feature extraction in the literature [36]. Literature [36] studied effective feature parameter extraction methods and feature selection algorithms. It proposed random forest sequential forward selection method based on variety analysis (RF-VA) for the optimal subset selection. The Part III of this paper is its corresponding content. And the
task of this paper is: based on the partial discharge feature set selection results in [36], through exploring the maximum use of the rich insulation state information and combining the difference and complementarity of different single feature sets for identifying insulation defects, this paper proposes the multi-feature combination adaptive boost classification model, instead of the way of using feature sets extracted by a certain partial discharge pattern.

The sample data used in this paper is: 6 voltage level experiments for each defect, 25 sample data for each voltage level, 150 sample data in total. The proposed method is validated for validity and correctness on the partial discharge pulse sequence data. The results obtained with the proposed method were compared with the traditional methods. The results revealed that the proposed method successfully identified the types of partial discharge insulation defects.

It is known that the classification performance of SVM with Gaussian kernel is affected by the penalty parameter C and the kernel parameter σ. Therefore, this paper obtains a series of SVM base classifiers with moderate accuracy by the adjustment of σ value. The adjustment is by means of the diversity index, which measuring the inconsistency between the base classifiers. Combining with other related algorithms and bagging approaches, the proposed adaptive boost classification model with diverse SVM base classifiers can achieve better classification performance. Furthermore, compared with individual SVM, diverse AdaBoost SVM can achieve much better classification performance on PD identification.

In our proposed method, the boosting approach focuses on the misclassified samples, and the classification accuracy of each base SVM can be controlled by adjusting σ value. It is observed that, the proposed method can solve the accuracy/diversity problem. By parameter adjusting method, a balance between accuracy and diversity can be get and better generalization performance than other related algorithms, bagging approaches and AdaBoost SVM.

II. DATA SETS ACQUISITION SYSTEM

A. PHYSICAL MODELS OF THE INSULATION DEFECTS

Eight types of physical models are designed. They include the following: suspending electrode (Label 1), metal protrusion on the enclosure (Label 2) and conductor (Label 3), metal particles contamination on the spacer (Label 4), metal contamination on the spacer (II and II, Label 5, 6), a void in the insulator (Label 7), free metal particles on the enclosure (Label 8).

B. EXPERIMENTAL PROCEDURE

In this paper, an UHF measurement system is used to collect partial discharge data. The frequency band range of UHF signal is 0–2 GHz. The sensor had a mean effective height over the frequency range of 500 MHz to 1500 MHz. The UHF PD signal was obtained by the internal UHF sensor. A Tektronix oscilloscope DPO7354 was used for data acquisition. The parameters for the oscilloscope were a 3.5 GHz bandwidth and a 40 GS/s max. sample rate. After placing the insulation defect model, using the voltage rise and fall method to determine the initial discharged voltage and breakdown voltage of various insulation defects. The degree of partial discharge is correlated with the voltage level. When collecting partial discharge data, in order to obtain complete discharge data in a short time, this article uses the following methods for data collection.

The specific implementation steps are: uniformly and slowly increase the test voltage, observe the oscilloscope, if the oscilloscope has a discharge pulse, record the test voltage, which is the initial discharge voltage of the defective discharge model. Continue to increase the voltage until the defect occurs flashover or breakdown, this voltage is the breakdown voltage.

Then, from initial discharge to breakdown, a total of 6 voltage levels of sample data were recorded, and a total of 150 sample data of each defect were obtained at 6 voltage levels through multiple experiments.

The data recording format is a PD pulse sequence \( q_s(t_s, u(t_s)) \). It is the PD activity of the PD pulse, which has the discharge amplitude of the PD pulse \( q_s \), the appearance time of the PD pulse \( t_s \) and the applied test voltage \( u(t_s) \) within a certain measurement time \( t_m \) [37], [38].

C. NOISE LEVEL MEASUREMENT

As shown in Figure 1, it can be seen that there is a lower level of background noise. UHF method is adopted to obtain partial discharge pulse data to ensure that the discharge pulse is not submerged by noise. Figure 2 shows the background signal data obtained by the acoustic method. The periodic peak value changes between 0.5mV and 2.3mV, and the signals of the other two components are zero.

D. PD PATTERNS ACQUISITION

In this section, three PD patterns were got to carry out pattern recognition of partial discharge. There are phase resolved pulse sequence analysis (PRPSA) pattern [39], [40], phase resolved partial discharge (PRPD) pattern and polar coordinate phase resolved analysis (PCPRA) pattern [40]. Figure 3 is an example of the metal particles contamination on spacer defect at 120kV.

III. FEATURE PARAMETERS EXTRACTION

In order to identify PD defects, the section is to extract the characteristic parameters from the acquired PD patterns.
These feature parameters include: 15 basic feature parameters extracted from the PRPSA pattern constituted the input feature vector \( \mathbf{F}_1 \) (1200 x 15) to identify the PD types shown in Table 1.

PRPD pattern determines the maximum pulse height distribution \( H_{\text{max}}(\varphi) \), the pulse height distribution \( H_{\text{p}}(\varphi) \), the number of discharges vs. discharge magnitude distribution can be determined by the five discharge quantities. Then the selected statistical operators are calculated [42]–[44]. 33 statistical operators constituted the input feature vector \( \mathbf{F}_2 \) (1200 x 33) as shown in Table 2.

Additionally, the feature parameters extracted from PCPRA pattern include centroid \( (c_{\varphi}, c_{q}) \), discharge width \( W_{p} \), number of discharges \( N_c \), median phase \( Q_{02} \), median amplitude \( Q_{02} \), phase quartile \( Q_{01}, Q_{03} \), and amplitude quartile \( Q_{a1}, Q_{a3} \) of each discharge cluster, quadrant-based parameters \( D_i (i=1,2,3,4) \), cosine similarity of the centroid vector \( \text{Cossim}(c_1, c_2) \), ratios of amplitude quartiles \( A_{\text{ratio1}}, A_{\text{ratio1}} \). The centroid, discharge width and phase quartile, amplitude quartile are shown in Figure 4. A total of 34 parameters which constituted the input feature vector \( \mathbf{F}_3 \) (1200 x 34) to identify the PD types [41].

### IV. ADAPTATIVE BOOST CLASSIFICATION MODEL BASED ON MULTI-FEATURE COMBINATION

#### A. THE COMBINATION METHOD

The combination method, also known as ensemble method, builds a classification model by gathering multiple classifiers. It constructs a set of base classifiers from the training set, and then performs classification by voting on the prediction of each base classifier. Let \( \varepsilon \) is the error rate of the base classifier, and \( \varepsilon_{\text{ensemble}} \) is the error rate of the combined classifier. The application of the combination is meaningless When \( \varepsilon_{\text{ensemble}} = \varepsilon \), that is to say all the base classifiers are exactly the same. When the classification errors of all base classifiers are independent of each other, for a combination of \( T \) binary classifiers, the error rate of the combined classifier is:

\[
\varepsilon_{\text{ensemble}} = \sum_{i=\lceil T/2 \rceil}^{T} C_i^T \varepsilon^i (1 - \varepsilon)^{T-i} \quad (1)
\]
where $T$ is the number of the base classifiers, $C^T_i$ represents \( \left( \begin{array}{c} T \\ i \end{array} \right) \).

According to Hoeffding’s inequality, the error rate of the combined classifier is:

\[
e_{\text{ensemble}} = \sum_{i=\lfloor T/2 \rfloor}^{T} C^T_i e^i (1 - \varepsilon)^{T-i} \leq \exp \left(-\frac{1}{2}T (1 - 2\varepsilon)^2 \right) \tag{2}\]

It can be seen that the error rate of the combined classifier will decrease exponentially as $T$ increases. The key to the obvious effectiveness of the combined method is to construct a base classifier with an error rate $\varepsilon < 0.5$ and the classification error is not correlated. It can also be said that the base classifier should be accurate and diverse.

### B. ADABOOST

The AdaBoost algorithm, called the adaptive boosting algorithm, uses the boosting method to construct a combined classifier. Let \( \{ (x_j, y_j) | j = 1, 2, \ldots, N \} \) denote a set of $N$ training samples. In this algorithm, the importance of the base classifier depends on its error rate $\varepsilon$. The error rate is defined as:

\[
\varepsilon_i = \frac{1}{N} \sum_{j=1}^{N} w_j I \left( C_i(x_j) \neq y_j \right) \tag{3}
\]

where $I$ represents predicate logic. If the predicate $C_i(x_j) \neq y_j$ is true, then $I (C_i(x_j) \neq y_j) = 1$, otherwise it is 0.

The importance of the base classifier $C_i$ is given by:

\[
\alpha_i = \frac{1}{2} \ln \left( \frac{1 - \varepsilon_i}{\varepsilon_i} \right) \tag{4}
\]

The parameter $\alpha_i$ is also used to update the weights of the training samples. Assume that $w_j^{(i)}$ represents the weight assigned to example $(x_j, y_j)$ during the $j$-th boosting round. The weight update mechanism of the adaptive boosting algorithm is to increase the weight of incorrectly classified examples and decrease the weight of that have been correctly classified, and is given by the following equation:

\[
w_j^{(i+1)} = \frac{w_j^{(i)}}{Z_i} \begin{cases} e^{-\alpha_i} & \text{if } C_i(x_j) = y_j \\ e^{\alpha_i} & \text{if } C_i(x_j) \neq y_j \end{cases} \tag{5}
\]

where $Z_i$ is a normalizing factor, which is used to ensure $\sum_{j=1}^{N} w_j^{(i+1)} = 1$.

The adaptive boosting algorithm penalizes models with poor accuracy by weighting the predicted value of each classifier $C_i$ according to $\alpha_i$. If $\varepsilon > 0.5$, the weight will be restored to the initial consistent value $w_j = 1/N$ and sampling will be performed again [45].

The entire adaptive boosting algorithm iterative algorithm has three steps as follows:

1. Initialize the weight distribution of the training data. If there are $N$ samples, each training sample is given the same weight at the beginning: $w = \{ w_j = 1/N | j = 1, 2, \ldots, N \}$

2. Training the base classifiers. In the specific training process, if a sample point has been accurately classified, its weight is improved. Then, the sample set with updated weights is used to train the next classifier, and the entire training process proceeds iteratively in this way.

3. Combine the weak classifiers obtained from each training into a strong classifier. After training each weak classifier, increasing the weight of the weak classifier with a small classification error rate to make it play a greater decisive role in the final classification function, and reduce the weak classifier with a large classification error rate. The weight makes it play a smaller decisive role in the final classification function. In other words, a weak classifier with a low error rate has a larger weight in the final classifier, otherwise it is smaller.

### C. ALGORITHM DESCRIPTION

The partial discharge can be identified correctly to a certain extent when using a single feature set as the input of the classification model. In order to improve the recognition rate, this paper combines the difference and complementarity of different single feature sets for various defects, and proposes a multi-feature combination adaptive boost classification model.

Combination method is to combine the three single feature sets obtained from the PRPSA pattern, $F_3 (1200 \times 33)$ extracted from the PRPD pattern, and $F_2 (1200 \times 34)$ extracted from the PCPRA pattern. The feature sets are in pairs or three combinations to construct a new feature set.
The multi-feature combination adaptive boost classification model uses SVM as the base classifier, the kernel function adopts the Gaussian kernel function, and the corresponding kernel parameter is the Gaussian pulse width $\sigma$. The combined classification model is obtained from the training data set to carry out the partial discharge pattern recognition. How to set the kernel parameters $\sigma$ of the Gaussian kernel function in the AdaBoost iteration process is the problem to be solved. If the value of $\sigma$ is too large, the base classifier SVM is weaker, and the classification accuracy rate is less than 50%, which does not meet the requirements of the AdaBoost base classifier. On the other hand, the value of $\sigma$ is too small will make the SVM very strong. Combining strong SVMs is ineffective because the classification errors of these base classifiers are highly correlated. In addition, too small $\sigma$ will cause the SVM overfit. Therefore, when using AdaBoost, it is very important to find a suitable value of $\sigma$ for the SVM-based classifier.

Therefore, this paper obtains a series of SVM base classifiers with moderate accuracy by the adjustment of $\sigma$ value. The adjustment is by means of the diversity index, which measuring the inconsistency between the base classifiers. Diversity measures include “paired” and “unpaired” indicators [46]. This paper uses the “unpaired” diversity index based on information entropy to measure the inconsistency between base classifiers. If $C_i(x_j)$ is the predicted result of sample $x_j$ by the $i$ - th classifier, defined $f_i(x_j)$ as predicting result consistency measurement of the sample $x_j$ by the base classifier $i$ and the base classifier $i - 1$. Then the diversity of the base classifier $i$ to predicting the label of sample $x_j$ is:

$$D_i(x) = \frac{1}{(N - \lceil N/2 \rceil)} \min \left\{ \sum_{j=1}^{N} f_i(x_j), N - \sum_{j=1}^{N} f_i(x_j) \right\}$$

$$f_i(x_j) = \begin{cases} 1: & \text{if } C_i(x_j) = C_{i-1}(x_j) \\ 0: & \text{others} \end{cases}$$

where $N$ is the number of sample data sets. If all base classifiers have the same output, the diversity will reach the minimum value of 0.

The multi-feature combination adaptive boost classification model is described as follows and its flowchart is illustrated in Figure 5:

Input: a set of labeled training samples $\{(x_1, y_1), \ldots, (x_N, y_N)\}$.

Step 1: Initialize the weights of $N$ samples, $w = \{w_j = 1/N | j = 1, 2, \ldots, N\}$, assign initial values $\sigma_{\text{ini}}$ to $\sigma$, set the minimum value $\sigma_{\text{min}}$, set $\sigma_{\text{step}}$ as the step size of $\sigma$, and the diversity threshold DIV.

Step 2: According to $w$, the training set $D_i$ is generated by sampling the original training set (with replacement).

Step 3: Training the base classifier $C_i$. Using $C_i$ to classify all samples in the original training set and calculate the classification error:

$$\varepsilon_i = \frac{1}{N} \sum_{j=1}^{N} w_j \delta (C_i(x_j) \neq y_j)$$

Step 4: Calculate the diversity, according to Eq. (6) and Eq. (7).

Step 5: Determine whether the current classification error satisfies $\varepsilon_i > 0.5$ and diversity $D_i < \text{DIV}$. If it is satisfied, then adjust the kernel parameters $\sigma = \sigma - \sigma_{\text{step}}$ and go to step 3, if not, then continue to step 6.

Step 6: Set the weight of the weak classifier $\alpha_i = \frac{1}{2} \ln \left( \frac{1 - \varepsilon_i}{\varepsilon_i} \right)$ and update the weight of each sample according to the following formula.

$$w_j^{(i+1)} = \frac{w_j^{(i)}}{Z_i} \begin{cases} e^{-\alpha_i} & \text{if } C_i(x_j) = y_j \\ e^{\alpha_i} & \text{if } C_i(x_j) \neq y_j \end{cases}$$

where $Z_i$ is a normalizing factor, which is used to ensure $\sum_{j=1}^{N} w_j^{(i+1)} = 1$.

Step 7: Determine whether the current kernel parameter is satisfied $\sigma > \sigma_{\text{min}}$ or the number of boosting rounds $i \leq 10$. If satisfied, then return to step 3, if not satisfied, then end the iteration.

Output: $C^+(x) = \arg \max_y \sum_{j=1}^{k} \alpha_j \delta (C_i(x) = y)$

### D. EXPERIMENTAL TESTS DESIGN

The experimental test steps are shown in Figure 6.

Step 1: Using multiple feature combination methods to form multiple feature sets.
TABLE 3. Kernel parameters and searching space description.

| name         | Kernel function       | Kernel parameter | Step  \( s \) | Min | Max |
|--------------|-----------------------|------------------|--------------|-----|-----|
| Linear       | \( K(x, x') = x^T x' \)| -                | -            | -   | -   |
| Polynomial   | \( K(x, x') = (x^T x' + 1)^d \) | \( d \) | 1            | 1   | 8   |
| Gaussian     | \( \exp\left[\frac{-||x-x'||^2}{2\sigma^2}\right] \) | \( \sigma \) | \( 1/2\sigma^2 \) | 2\( d \) | 2\( d \) | 2\( d \) | 2\( d \) |
| Sigmoid      | \( \tanh((x^T x' + 1)) \) | \( g \) | 2\( d \) | 2\( d \) | 2\( d \) | 2\( d \) |
| Semi Local   | \( (x^T x' + 1) \cdot \exp\left[\frac{-||x-x'||^2}{2\sigma^2}\right] \) | \( \sigma \) | \( 1/2\sigma^2 \) | 2\( d \) | 2\( d \) | 2\( d \) | 2\( d \) |

In the experiment, firstly, combined \( F_1 \) (1200 × 15), \( F_2 \) (1200 × 33) and \( F_3 \) (1200 × 34) in pairs or three together to form Multi-feature sets, \( (F_1, F_2) \), \( (F_1, F_3) \), \( (F_2, F_3) \) and \( (F_1, F_2, F_3) \). Using the principal component analysis (PCA) perform feature dimensionality reduction on the single feature sets and the combined feature sets.

Step 2: Select the feature set and establish SVM model.

Then the single feature set and the combined feature set are trained by the SVM algorithm composed of different kernel functions. Before that, the feature set is linearly normalized first, and then the original data set is divided according to the ratio of training set and test set samples of 4:1. Use 10-fold cross-validation for model training to obtain the final classification result. For each kernel function test, the parameter combination with the highest classification accuracy will be used to build the final SVM classification model. Finally, the generalization error of the SVM classification model composed of different kernel functions is estimated on the test set.

The optimization range of each parameter and the optimization step length is \( [100,801] \), and the selection range of the maximum number of examples required for subdivision of CART’s internal nodes \( S_{min} \) is \([100,801]\), and the selection range of the maximum depth of the tree \( d \) is \([1,10]\).

Step 4: Using the support vector machine selected in step 2 as the base classifier, using the selected optimal feature set as the input of the base classifier, and combining the performance of multi-feature combination adaptive boost classification model with a single support vector machine and other combinations methods for comparative analysis. The framework of combined learning proposed to this paper is shown in Figure 7. Other combination methods use SVM as the base classifier and use the bagging algorithm to form the combined classification model- Bagging SVM, use the existing AdaBoost method to form the AdaBoost SVM.

V. EXPERIMENTAL RESULTS AND DISCUSSION

A. FEATURE SET SELECTION

Performing feature dimensionality reduction on single feature sets and combined feature sets. The dimensionality reduction method used is the PCA. The dimensionality reduction results are divided into training set and test set by the fence method grouping strategy, and the SVM classification model composed of different kernel functions are established. The statistical classification results are shown in Table 4.

It can be seen from the classification accuracy results on the test set that the classification accuracy rate corresponding to the combined feature set \( (F_1, F_2, F_3) \) is the highest, and the classification accuracy rate corresponding to the combined
B. COMPARISON OF CLASSIFICATION PERFORMANCE OF PROPOSED METHOD, SVM AND OTHER COMBINATION METHODS

The bagging algorithm is used to form the combined classification model-Bagging SVM, the existing AdaBoost method is used to form the combined classification model-AdaBoost SVM. This paper compares the Multi-feature combination adaptive boost classification model with the classification results of single SVM, Bagging SVM and AdaBoost SVM.

As shown in Table 6, the experimental results on the selected feature set \( (F_1, F_2, F_3) \) show that the Adaboost SVM is better than Bagging SVM, which is 6.7% higher than the classification accuracy of Bagging SVM, but it is worse than a single support vector machine, which is 1.63% lower than a single support vector. Compared with a single support vector machine, the classification accuracy of the proposed adaptive boosting algorithm based on multi-feature combination is 1.67% higher than a single support vector machine. Because of that the proposed algorithm uses a diversity index to adjust the parameters of the Gaussian kernel function to obtain the SVM base classifier. The parameters solve the problem of SVM kernel parameter selection and obtain a balance between classification accuracy and diversity.

VI. DISCUSSION

A. SENSITIVE TO THE PENALTY PARAMETER AND THE KERNEL PARAMETER

1) SENSITIVE TO \( C \) AND \( \sigma_{ini} \)

Changing the \( C \) value from 1 to 120, and performing experiments on the feature set \( (F_1, F_2, F_3) \) to obtain the average generalization level. Figure 8 shows the result of the comparison. The generalization performance of the algorithm shows that the \( C \) value change has little effect on the result (less than 2%). In the experiment, as the number of SVM base classifiers increases (as shown on the horizontal axis in the Figure 8), the kernel parameter \( \sigma \) of the Gaussian kernel function decrease from the initial value \( \sigma_{ini} \) to the
minimum value $\sigma_{\text{min}}$. The small platform at the top left of the figure means that the error rate of the test data set does not decrease until the value of the $\sigma$ decreases to a certain value. Then, the test error rate quickly dropped to the lowest value and remained stable. This experimental result shows that $\sigma_{\text{min}}$ has little effect on the performance of the proposed algorithm.

2) SENSITIVE TO $\sigma_{\text{step}}$

Experiments on the feature set ($F_1, F_2, F_3$) are performed to study the influence of the $\sigma_{\text{step}}$ value on the performance of the algorithm. Different step sizes of the kernel parameters are set in the experiment. Figure 9 shows the results. It can be seen from the figure that although the number of learning cycles for multi-feature combination adaptive boost changes as the $\sigma_{\text{step}}$ value changes, the final test error rate is stable.

### B. ALGORITHM ANALYSIS

As shown in Figure 7, firstly the boosting method trains an SVM base learner $C_1$ from the initial training set $D_1$, and then adjusts the training samples according to the performance of the base learner. The training samples that the previous base learner did wrong will receive more attention in the follow-up. Secondly, the sample distribution $D_2$ (after adjusted) is used to train the next base learner $C_2$. The process is repeated until the number of base learners reaches the specified value or the parameters meet the requirements. Finally, the $T$ base learners are combined with weight.

The adaptive boost classification model requires the base classifier SVM being able to learn with specified distributions. This is accomplished by re-weighting, that is, weighting training examples in each round according to the sample distribution. The model proposed in this paper checks whether the currently generated base learner meets the basic conditions in each round of training ($\epsilon_i > 0.5$ and $D_i < \text{DIV}$). From the perspective of bias-variance decomposition, the model focuses on reducing bias. The difference with Adaboost SVM is that it obtains a series of SVM base classifiers with moderate accuracy by the adjustment of $\sigma$ value. The adjustment is by means of the diversity index, which measuring the inconsistency between the base classifiers. The classification accuracy of each base SVM can be controlled by adjusting $\sigma$ value. It is observed that, the proposed method can solve the accuracy/diversity problem. By parameter adjusting method, a balance between accuracy and diversity can be get and better generalization performance than other related algorithms, bagging approaches and AdaBoost SVM.

Compared with the Bagging method, the ensemble learning model proposed is a sequential ensemble method while Bagging is a parallel ensemble method. The basic motivation of parallel ensemble methods is to exploit the independence between the base learners, since the error can be reduced dramatically by combining independent base learners. Another advantage of parallel ensemble methods is that the use of multi-core computing processors or parallel computers can easily speed up training. This is attractive because multi-core processors are commonly used today.

The Bagging uses bootstrap sampling to obtain the data set for training the base classifier. It trains a base learner based on each sampling set, and then combines these learners. When combining the classification results, use simple voting method. From the perspective of deviation-variance decomposition, the Bagging mainly focuses on reducing variance.

Assuming that the feature dimension of the training data set is $m$ and the number of training samples is $n$, the time complexity of the base classifier (SVM) is $O(m^2 n^2)$, then the time complexity of the bagging SVM is roughly $T (O(m^2 n^2) + O(s)).$ Considering that the time complexity of sampling and voting/averaging process is very small, in this paper, $T$ is 130. Therefore, training a Bagging

### TABLE 6. Experimental results of combined SVM (%).

| Defect Label | SVM-Gaussian | Bagging SVM | Adaboost SVM | The Proposed Model |
|-------------|-------------|-------------|-------------|-------------------|
| 1           | 86.67       | 83.33       | 90.00       | 86.67             |
| 2           | 93.33       | 90.00       | 83.33       | 93.33             |
| 3           | 90.00       | 76.67       | 80.00       | 86.67             |
| 4           | 80.00       | 83.33       | 90.00       | 93.33             |
| 5           | 86.67       | 73.33       | 76.67       | 90.00             |
| 6           | 93.33       | 76.67       | 93.33       | 93.33             |
| 7           | 90.00       | 93.33       | 96.67       | 93.33             |
| 8           | 86.67       | 63.33       | 83.33       | 83.33             |
| Total       | **88.33**   | **80.00**   | **86.70**   | **90.00**         |

**FIGURE 8.** The classification performance of the proposed method with different $C$.

**FIGURE 9.** The classification performance of the proposed method with different $\sigma_{\text{step}}$. The small platform at the top left of the figure means that the error rate of the test data set does not decrease until the value of the $\sigma$ decreases to a certain value. Then, the test error rate quickly dropped to the lowest value and remained stable. This experimental result shows that $\sigma_{\text{min}}$ has little effect on the performance of the proposed algorithm.
ensemble has the same level of complexity as training a learner using the base learning algorithm directly. The time complexity of the adaptive boost classification model is \( T \left( O(mn^2) + O(n) + O(s) \right) \). Therefore, the time complexity of the model assembled in this paper is the same as that of directly training a learner using the base learning algorithm.

VII. CONCLUSION

This paper proposes a multi-feature combination adaptive boost classification model by combining the differences and complementarities of different single feature sets for insulation defect recognition. The types of partial discharge insulation defects are successfully identified by the algorithm, and it solved the problem of setting the kernel function parameters of the support vector machine-based classifier by using the diversity index.

In order to explore the maximum use of the rich insulation state information obtained by the UHF sensor and combine the difference and complementarity of different single feature sets for identifying insulation defects. The model uses the multi-feature combination method to obtain the combined feature set as the input of the base classifier.

Through comparison and analysis with commonly used classification algorithms, the model in this paper takes the support vector machine as the base classifier, and the kernel function adopts the Gaussian kernel function. Aiming at the problem of how to set the kernel function parameters of the support vector machine base classifier, the method of adjusting the Gaussian kernel function kernel parameters by measuring the diversity index of the inconsistency between the base classifiers is used to obtain the support vector machine base classifier with a moderate classification accuracy rate. So that the adaptive boosting algorithm has better classification performance.

In summary, the multi-feature combination adaptive boost classification model proposed in this paper not only makes full use of the original data information, but also optimizes the model parameters. The experimental results on partial discharge date sets demonstrate that the combined feature set obtained by the combined use of the three single feature sets is more resolving than the traditional single feature set and other combined feature sets. The author also compares the support vector machine models using different kernel functions and finds that the Gaussian kernel function has better classification performance than other kernel functions. The experimental results on the selected feature set show that the classification model proposed in this paper successfully identified the types of partial discharge insulation defects.

ACKNOWLEDGMENT

The authors would like to thank the contributions of the staff of Xidian Group High Voltage Switch Company Ltd. Wang Xianle and Jiang Xingquan for their work on the formulation of the experimental plan and the measurement and recording of experimental data of this document.

REFERENCES

[1] S. A. Boggs, “Partial discharge: Overview and signal generation,” IEEE Electr. Insul. Mag., vol. 6, no. 4, pp. 33–39, Jul. 1990.
[2] M. G. Danikas, “The definitions used for partial discharge phenomena,” IEEE Trans. Electr. Insul., vol. 28, no. 6, pp. 1075–1081, Dec. 1993.
[3] R. Baumgartner, B. Fruth, W. Lanz, and K. Pettersson, “Partial discharge. IX. PD in gas-insulated substations—fundamental considerations,” IEEE Electr. Insul. Mag., vol. 7, no. 6, pp. 5–13, Nov./Dec. 1991.
[4] M. Yan et al., “UHF signal characteristics study of partial discharge generated in three-phase-in-one-tank type GIS,” Insulating Mater., vol. 3, pp. 79–84, 2014.
[5] Z. Ren, M. Dong, L. Y. Wu, M. Ren, and W. R. Si, “Analysis of partial discharge in GIS under impulse voltage by using quadratic short-time Fourier transform,” Electr. Power, vol. 46, no. 4, pp. 48–53, 2013.
[6] W. Gao, D. Ding, and W. Liu, “Research on the typical partial discharge using the UHF detection method for GIS,” IEEE Trans. Power Del., vol. 26, no. 4, pp. 2621–2629, Oct. 2011.
[7] X. Wang, T. Li, D. Ding, and M. Rong, “The influence of L-shaped structure on partial discharge radiated electromagnetic wave propagation in GIS,” IEEE Trans. Plasma Sci., vol. 42, no. 10, pp. 2536–2537, Oct. 2014.
[8] C. M. Bishop, Neural Networks for Pattern Recognition. Birmingham, U.K.: Oxford Univ. Press, 1995, pp. 164–191.
[9] A. Belayadi, L. Ait-Gougam, and F. Mekideche-Chafa, Pattern Recognition and Neural Networks. Cambridge, U.K.: Cambridge Univ. Press, 1996, pp. 123–126.
[10] T. K. Abdel-Galil, Y. G. Hegazy, M. M. A. Salama, and R. Bartnikas, “Partial discharge pulse pattern recognition using hidden Markov models,” IEEE Trans. Dielectrics Electr. Insul., vol. 11, no. 4, pp. 715–723, Aug. 2004.
[11] M. M. A. Salama and R. Bartnikas, “Fuzzy logic applied to PD pattern classification,” IEEE Trans. Dielectr. Electr. Insul., vol. 7, no. 1, pp. 118–123, Feb. 2000.
[12] Y. Han and Y. H. Song, “Using improved self-organizing map for partial discharge diagnosis of large turbogenerators,” IEEE Trans. Energy Convers., vol. 18, no. 3, pp. 392–399, Sep. 2003.
[13] T. K. Abdel-Galil, R. M. Sharakawy, M. M. A. Salama, and R. Bartnikas, “Partial discharge pulse pattern recognition using an inductive inference algorithm,” IEEE Trans. Dielectrics Electr. Insul., vol. 12, no. 2, pp. 320–327, Apr. 2005.
[14] L. Hao and P. Lewin, “Partial discharge source discrimination using a support vector machine,” IEEE Trans. Dielectr. Electr. Insul., vol. 17, no. 1, pp. 189–197, Feb. 2010.
[15] R. Ambikairajah, B. T. Phung, J. Ravishankar, and T. Blackburn, “Spectral features for the classification of partial discharge signals from selected insulation defect models,” IET Sci., Meas. Technol., vol. 7, no. 2, pp. 104–111, Mar. 2013.
[16] F.-C. Gu, “Identification of partial discharge defects in gas-insulated switchgears by using a deep learning method,” IEEE Access, vol. 8, pp. 163894–163902, 2020.
[17] M. Y. A. Khan and J.-Y. Koo, “Neural network based diagnosis of partial discharge defects patterns at XLPE cable under DC stress,” Electr. Eng., vol. 99, no. 1, pp. 119–132, Mar. 2017.
[18] W. J. K. Raymond, H. A. Illias, and A. H. Abu Bakar, “Classification of partial discharge measured under different levels of noise contamination,” PLOS ONE, vol. 12, no. 1, 2017, Art. no. e0170111.
[19] A. A. Mas‘ud, J. A. Ardila-Rey, R. Albarracín, and F. Muhammad-Sukki, “An ensemble-boosting algorithm for classifying partial discharge defects in electrical assets,” Machines, vol. 5, no. 3, p. 18, 2017.
[20] A. A. Mas‘ud, B. G. Stewart, and S. G. Mcemekin, “Application of an ensemble neural network for classifying partial discharge patterns,” Electr. Power Syst. Res., vol. 110, pp. 154–162, May 2014.
[21] A. A. Mas‘ud, B. G. Stewart, S. G. Mcemekin, and A. Nesbitt, “Recognition of partial discharges using an ensemble of neural networks,” in Proc. Annu. Rep. Conf. Electr. Insul. Dielectric Phenomena, Cancun, Mexico, Oct. 2011, pp. 497–500.
[22] A. A. Mas‘ud, B. G. Stewart, S. G. Mcemekin, and A. Nesbitt, “An ensemble Neural Network for recognizing PD patterns,” in Proc. 45th Int. Universities Power Eng. Conf., Cardiff, U.K., Aug./Sep. 2010, pp. 1–6.
[23] A. A. Mas‘ud, B. G. Stewart, S. G. Mcemekin, and A. Nesbitt, “Partial discharge pattern classification for an oil-pressboard interface,” in Proc. IEEE Int. Symp. Electr. Insul., San Juan, PR, USA, Jan. 2012, pp. 122–126.
RUI YAO received the B.S. and M.S. degrees in control theory and control engineering from North China Electric Power University, in 2008 and 2011, respectively, and the Ph.D. degree in electrical engineering from Xi’an Jiaotong University, in 2017. She is currently a Lecturer with the School of Electronic and Control Engineering, Chang’ an University. Her research interests include the areas of partial discharge of high-voltage electrical equipment and fault diagnosis of electrical equipment based on machine learning.

JUN LI received the B.S. and master’s degrees in computer science and technology from North China Electric Power University, in 2008 and 2011, respectively. He is currently working with the State Grid Shaanxi Electric Power Research Institute. He is also a Senior Engineer with the Power Grid Research Institute. His research interests include the areas of partial discharge of high-voltage electrical equipment and fault diagnosis of electrical equipment based on machine learning.

MENG HUI (Member, IEEE) received the bachelor’s and master’s degrees in mechanical engineering and the Ph.D. degree in electrical engineering from Xi’an Jiaotong University, in 2004, 2007, and 2011, respectively.

He is currently an Associate Professor with the School of Electronic and Control Engineering, Chang’ an University. His research interests include the areas of fault diagnosis of electrical equipment based on machine learning and nonlinear dynamics in complex networks.

LIN BAI received the bachelor’s degree in electronic information science and technology from Northwest University, in 2003, the master’s degree in electronic science and technology, in 2006, and the Ph.D. degree in signal and information processing from Northwestern Polytechnical University, in 2010.

He is currently an Associate Professor with the School of Electronic and Control Engineering, Chang’ an University. His research interests include the areas of fault diagnosis of electrical equipment based on deep learning and information processing.

QISHENG WU received the B.S. degree from the Dalian University of Technology, in 1984, and the master’s degree from Chang’an University, in 1996. His current research interests include artificial intelligence and automobile control.